Abstract

This is the reference documentation for the expl3 programming environment. The expl3 modules set up an experimental naming scheme for \LaTeX\ commands, which allow the \LaTeX\ programmer to systematically name functions and variables, and specify the argument types of functions.

The \TeX\ and \epsilon-\TeX\ primitives are all given a new name according to these conventions. However, in the main direct use of the primitives is not required or encouraged: the expl3 modules define an independent low-level \LaTeX3\ programming language.

The expl3 modules are designed to be loaded on top of \LaTeX2. With an up-to-date \LaTeX2 kernel, this material is loaded as part of the format. The fundamental programming code can also be loaded with other \TeX\ formats, subject to restrictions on the full range of functionality.
Contents

I Introduction 1

1 Introduction to expl3 and this document 2
  1.1 Naming functions and variables ..................................... 2
  1.1.1 Scratch variables ............................................... 5
  1.1.2 Terminological inexactitude ..................................... 5
  1.2 Documentation conventions .......................................... 5
  1.3 Formal language conventions which apply generally .............. 7
  1.4 \TeX\ concepts not supported by \LaTeX{} ......................... 7

II Bootstrapping 8

2 The l3bootstrap package: Bootstrap code 9
  2.1 Using the \LaTeX{} modules ........................................ 9

3 The l3names package: Namespace for primitives 11
  3.1 Setting up the \LaTeX{} programming language .................... 11

III Programming Flow 12

4 The l3basics package: Basic definitions 13
  4.1 No operation functions ........................................... 13
  4.2 Grouping material ................................................ 13
  4.3 Control sequences and functions .................................. 14
    4.3.1 Defining functions ........................................... 14
    4.3.2 Defining new functions using parameter text ............... 15
    4.3.3 Defining new functions using the signature ............... 17
    4.3.4 Copying control sequences .................................. 19
    4.3.5 Deleting control sequences ................................ 20
    4.3.6 Showing control sequences .................................. 20
    4.3.7 Converting to and from control sequences ............... 20
  4.4 Analysing control sequences ..................................... 22
  4.5 Using or removing tokens and arguments .......................... 23
    4.5.1 Selecting tokens from delimited arguments ............... 25
  4.6 Predicates and conditionals ..................................... 25
    4.6.1 Tests on control sequences ................................ 27
    4.6.2 Primitive conditionals ..................................... 27
  4.7 Starting a paragraph .............................................. 28
  4.8 Debugging support ................................................ 29
5 The \texttt{\textbf{l3\textbackslash expand}} package: Argument expansion \hfill 30
  5.1 Defining new variants \hfill 30
  5.2 Methods for defining variants \hfill 31
  5.3 Introducing the variants \hfill 32
  5.4 Manipulating the first argument \hfill 34
  5.5 Manipulating two arguments \hfill 36
  5.6 Manipulating three arguments \hfill 36
  5.7 Unbraced expansion \hfill 37
  5.8 Preventing expansion \hfill 38
  5.9 Controlled expansion \hfill 39
  5.10 Internal functions \hfill 42

6 The \texttt{\textbf{l3sort}} package: Sorting functions \hfill 43
  6.1 Controlling sorting \hfill 43

7 The \texttt{\textbf{l3\textbackslash tl-analysis}} package: Analysing token lists \hfill 45

8 The \texttt{\textbf{l3\textbackslash regex}} package: Regular expressions in \LaTeX{} \hfill 46
  8.1 Syntax of regular expressions \hfill 47
    8.1.1 Regular expression examples \hfill 47
    8.1.2 Characters in regular expressions \hfill 48
    8.1.3 Characters classes \hfill 48
    8.1.4 Structure: alternatives, groups, repetitions \hfill 49
    8.1.5 Matching exact tokens \hfill 50
    8.1.6 Miscellaneous \hfill 52
  8.2 Syntax of the replacement text \hfill 52
  8.3 Pre-compiling regular expressions \hfill 54
  8.4 Matching \hfill 55
  8.5 Submatch extraction \hfill 56
  8.6 Replacement \hfill 57
  8.7 Scratch regular expressions \hfill 59
  8.8 Bugs, misfeatures, future work, and other possibilities \hfill 59

9 The \texttt{\textbf{l3\textbackslash prg}} package: Control structures \hfill 62
  9.1 Defining a set of conditional functions \hfill 62
  9.2 The boolean data type \hfill 64
    9.2.1 Scratch booleans \hfill 66
  9.3 Boolean expressions \hfill 66
  9.4 Logical loops \hfill 68
  9.5 Producing multiple copies \hfill 69
  9.6 Detecting \LaTeX{}'s mode \hfill 70
  9.7 Primitive conditionals \hfill 70
  9.8 Nestable recursions and mappings \hfill 70
    9.8.1 Simple mappings \hfill 71
  9.9 Internal programming functions \hfill 71
10 The \texttt{latex} package: System/runtime functions 72
\hspace{1em} 10.1 The name of the job 72
\hspace{1em} 10.2 Date and time 72
\hspace{1em} 10.3 Engine 73
\hspace{1em} 10.4 Output format 73
\hspace{1em} 10.5 Platform 74
\hspace{1em} 10.6 Random numbers 74
\hspace{1em} 10.7 Access to the shell 74
\hspace{1em} 10.8 Loading configuration data 75
\hspace{2em} 10.8.1 Final settings 76

11 The \texttt{msg} package: Messages 77
\hspace{1em} 11.1 Creating new messages 77
\hspace{1em} 11.2 Customizable information for message modules 78
\hspace{1em} 11.3 Contextual information for messages 78
\hspace{1em} 11.4 Issuing messages 80
\hspace{2em} 11.4.1 Messages for showing material 83
\hspace{2em} 11.4.2 Expandable error messages 83
\hspace{1em} 11.5 Redirecting messages 83

12 The \texttt{file} package: File and I/O operations 85
\hspace{1em} 12.1 Input–output stream management 85
\hspace{2em} 12.1.1 Reading from files 87
\hspace{2em} 12.1.2 Writing to files 90
\hspace{2em} 12.1.3 Wrapping lines in output 92
\hspace{2em} 12.1.4 Constant input–output streams, and variables 93
\hspace{2em} 12.1.5 Primitive conditionals 93
\hspace{1em} 12.2 File operation functions 93

13 The \texttt{latex} package: Lua\TeX{}-specific functions 98
\hspace{1em} 13.1 Breaking out to \LaTeX{} 98
\hspace{1em} 13.2 Lua interfaces 99

14 The \texttt{legacy} package: Interfaces to legacy concepts 100

IV Data types 101

15 The \texttt{tl} package: Token lists 102
\hspace{1em} 15.1 Creating and initialising token list variables 102
\hspace{1em} 15.2 Adding data to token list variables 103
\hspace{1em} 15.3 Token list conditionals 104
\hspace{2em} 15.3.1 Testing the first token 106
\hspace{1em} 15.4 Working with token lists as a whole 107
\hspace{2em} 15.4.1 Using token lists 107
\hspace{2em} 15.4.2 Counting and reversing token lists 108
\hspace{2em} 15.4.3 Viewing token lists 109
\hspace{1em} 15.5 Manipulating items in token lists 110
\hspace{2em} 15.5.1 Mapping over token lists 110
\hspace{2em} 15.5.2 Head and tail of token lists 112

iv
15.5.3 Items and ranges in token lists ........................................ 113
15.5.4 Sorting token lists .................................................. 116
15.6 Manipulating tokens in token lists ...................................... 116
15.6.1 Replacing tokens ..................................................... 116
15.6.2 Reassigning category codes ......................................... 117
15.7 Constant token lists ..................................................... 118
15.8 Scratch token lists ...................................................... 118

16 The l3str package: Strings ................................................. 120
16.1 Creating and initialising string variables ............................... 121
16.2 Adding data to string variables ........................................ 122
16.3 String conditionals ....................................................... 122
16.4 Mapping over strings ..................................................... 124
16.5 Working with the content of strings ................................... 126
16.6 Modifying string variables ............................................. 129
16.7 String manipulation ..................................................... 130
16.8 Viewing strings ......................................................... 131
16.9 Constant strings ......................................................... 132
16.10 Scratch strings ........................................................... 132

17 The l3str-convert package: string encoding conversions ............ 133
17.1 Encoding and escaping schemes ....................................... 133
17.2 Conversion functions .................................................... 135
17.3 Conversion by expansion (for PDF contexts) ......................... 135
17.4 Possibilities, and things to do ........................................ 135

18 The l3quark package: Quarks ............................................. 137
18.1 Quarks .................................................................... 137
18.2 Defining quarks .......................................................... 138
18.3 Quark tests ............................................................... 138
18.4 Recursion .................................................................. 139
18.4.1 An example of recursion with quarks ......................... 140
18.5 Scan marks ............................................................... 141

19 The l3seq package: Sequences and stacks ............................. 142
19.1 Creating and initialising sequences .................................... 142
19.2 Appending data to sequences .......................................... 144
19.3 Recovering items from sequences ...................................... 144
19.4 Recovering values from sequences with branching ................ 145
19.5 Modifying sequences .................................................... 147
19.6 Sequence conditionals ................................................... 147
19.7 Mapping over sequences ................................................. 148
19.8 Using the content of sequences directly ............................... 150
19.9 Sequences as stacks ...................................................... 151
19.10 Sequences as sets ....................................................... 152
19.11 Constant and scratch sequences ...................................... 153
19.12 Viewing sequences ..................................................... 154
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td><strong>The l3prop package: Property lists</strong></td>
<td></td>
</tr>
<tr>
<td>24.1</td>
<td>Creating and initialising property lists</td>
<td>198</td>
</tr>
<tr>
<td>24.2</td>
<td>Adding and updating property list entries</td>
<td>199</td>
</tr>
<tr>
<td>24.3</td>
<td>Recovering values from property lists</td>
<td>200</td>
</tr>
<tr>
<td>24.4</td>
<td>Modifying property lists</td>
<td>201</td>
</tr>
<tr>
<td>24.5</td>
<td>Property list conditionals</td>
<td>201</td>
</tr>
<tr>
<td>24.6</td>
<td>Recovering values from property lists with branching</td>
<td>202</td>
</tr>
<tr>
<td>24.7</td>
<td>Mapping over property lists</td>
<td>203</td>
</tr>
<tr>
<td>24.8</td>
<td>Viewing property lists</td>
<td>204</td>
</tr>
<tr>
<td>24.9</td>
<td>Scratch property lists</td>
<td>205</td>
</tr>
<tr>
<td>24.10</td>
<td>Constants</td>
<td>205</td>
</tr>
<tr>
<td>25</td>
<td><strong>The l3skip package: Dimensions and skips</strong></td>
<td>206</td>
</tr>
<tr>
<td>25.1</td>
<td>Creating and initialising dim variables</td>
<td>206</td>
</tr>
<tr>
<td>25.2</td>
<td>Setting dim variables</td>
<td>207</td>
</tr>
<tr>
<td>25.3</td>
<td>Utilities for dimension calculations</td>
<td>207</td>
</tr>
<tr>
<td>25.4</td>
<td>Dimension expression conditionals</td>
<td>208</td>
</tr>
<tr>
<td>25.5</td>
<td>Dimension expression loops</td>
<td>210</td>
</tr>
<tr>
<td>25.6</td>
<td>Dimension step functions</td>
<td>211</td>
</tr>
<tr>
<td>25.7</td>
<td>Using dim expressions and variables</td>
<td>212</td>
</tr>
<tr>
<td>25.8</td>
<td>Viewing dim variables</td>
<td>213</td>
</tr>
<tr>
<td>25.9</td>
<td>Constant dimensions</td>
<td>214</td>
</tr>
<tr>
<td>25.10</td>
<td>Scratch dimensions</td>
<td>214</td>
</tr>
<tr>
<td>25.11</td>
<td>Creating and initialising skip variables</td>
<td>214</td>
</tr>
<tr>
<td>25.12</td>
<td>Setting skip variables</td>
<td>215</td>
</tr>
<tr>
<td>25.13</td>
<td>Skip expression conditionals</td>
<td>216</td>
</tr>
<tr>
<td>25.14</td>
<td>Using skip expressions and variables</td>
<td>216</td>
</tr>
<tr>
<td>25.15</td>
<td>Viewing skip variables</td>
<td>216</td>
</tr>
<tr>
<td>25.16</td>
<td>Constant skips</td>
<td>217</td>
</tr>
<tr>
<td>25.17</td>
<td>Scratch skips</td>
<td>217</td>
</tr>
<tr>
<td>25.18</td>
<td>Inserting skips into the output</td>
<td>217</td>
</tr>
<tr>
<td>25.19</td>
<td>Creating and initialising muskip variables</td>
<td>218</td>
</tr>
<tr>
<td>25.20</td>
<td>Setting muskip variables</td>
<td>218</td>
</tr>
<tr>
<td>25.21</td>
<td>Using muskip expressions and variables</td>
<td>219</td>
</tr>
<tr>
<td>25.22</td>
<td>Viewing muskip variables</td>
<td>219</td>
</tr>
<tr>
<td>25.23</td>
<td>Constant muskips</td>
<td>220</td>
</tr>
<tr>
<td>25.24</td>
<td>Scratch muskips</td>
<td>220</td>
</tr>
<tr>
<td>25.25</td>
<td>Primitive conditional</td>
<td>220</td>
</tr>
<tr>
<td>26</td>
<td><strong>The l3keys package: Key–value interfaces</strong></td>
<td>221</td>
</tr>
<tr>
<td>26.1</td>
<td>Creating keys</td>
<td>222</td>
</tr>
<tr>
<td>26.2</td>
<td>Sub-dividing keys</td>
<td>227</td>
</tr>
<tr>
<td>26.3</td>
<td>Choice and multiple choice keys</td>
<td>227</td>
</tr>
<tr>
<td>26.4</td>
<td>Key usage scope</td>
<td>230</td>
</tr>
<tr>
<td>26.5</td>
<td>Setting keys</td>
<td>230</td>
</tr>
<tr>
<td>26.6</td>
<td>Handling of unknown keys</td>
<td>231</td>
</tr>
<tr>
<td>26.7</td>
<td>Selective key setting</td>
<td>231</td>
</tr>
<tr>
<td>26.8</td>
<td>Digesting keys</td>
<td>232</td>
</tr>
<tr>
<td>26.9</td>
<td>Utility functions for keys</td>
<td>233</td>
</tr>
<tr>
<td>26.10</td>
<td>Low-level interface for parsing key–val lists</td>
<td>233</td>
</tr>
</tbody>
</table>
27 The \texttt{l3intarray} package: fast global integer arrays \hfill 236
  27.1 \texttt{l3intarray} documentation \hfill 236
  \hspace*{1em} \texttt{l3intarray} Implementation notes \hfill 237

28 The \texttt{l3fp} package: Floating points \hfill 238
  28.1 Creating and initialising floating point variables \hfill 240
  28.2 Setting floating point variables \hfill 240
  28.3 Using floating points \hfill 241
  28.4 Floating point conditionals \hfill 242
  28.5 Floating point expression loops \hfill 244
  28.6 Some useful constants, and scratch variables \hfill 246
  28.7 Floating point exceptions \hfill 247
  28.8 Viewing floating points \hfill 248
  28.9 Floating point expressions \hfill 249
  \hspace*{1em} 28.9.1 Input of floating point numbers \hfill 249
  \hspace*{1em} 28.9.2 Precedence of operators \hfill 250
  \hspace*{1em} 28.9.3 Operations \hfill 250
  28.10 Disclaimer and roadmap \hfill 257

29 The \texttt{l3fparray} package: fast global floating point arrays \hfill 260
  29.1 \texttt{l3fparray} documentation \hfill 260

30 The \texttt{l3ctab} package: Category code tables \hfill 261
  30.1 Creating and initialising category code tables \hfill 261
  30.2 Using category code tables \hfill 261
  30.3 Category code table conditionals \hfill 262
  30.4 Constant category code tables \hfill 262

V Text manipulation \hfill 263

31 The \texttt{l3unicode} package: Unicode support functions \hfill 264

32 The \texttt{l3text} package: text processing \hfill 265
  32.1 Expanding text \hfill 265
  32.2 Case changing \hfill 266
  32.3 Removing formatting from text \hfill 267
  32.4 Control variables \hfill 268

VI Typesetting \hfill 269
33 The \texttt{l3box} package: Boxes

33.1 Creating and initialising boxes \hfill 270
33.2 Using boxes \hfill 271
33.3 Measuring and setting box dimensions \hfill 272
33.4 Box conditionals \hfill 273
33.5 The last box inserted \hfill 273
33.6 Constant boxes \hfill 273
33.7 Scratch boxes \hfill 273
33.8 Viewing box contents \hfill 274
33.9 Boxes and color \hfill 274
33.10 Horizontal mode boxes \hfill 274
33.11 Vertical mode boxes \hfill 275
33.12 Using boxes efficiently \hfill 277
33.13 Affine transformations \hfill 278
33.14 Primitive box conditionals \hfill 281

34 The \texttt{l3coffins} package: Coffin code layer \hfill 282

34.1 Creating and initialising coffins \hfill 282
34.2 Setting coffin content and poles \hfill 283
34.3 Coffin affine transformations \hfill 284
34.4 Joining and using coffins \hfill 284
34.5 Measuring coffins \hfill 285
34.6 Coffin diagnostics \hfill 285
34.7 Constants and variables \hfill 286

35 The \texttt{l3color} package: Color support \hfill 288

35.1 Color in boxes \hfill 288
35.2 Color models \hfill 288
35.3 Color expressions \hfill 289
35.4 Named colors \hfill 290
35.5 Selecting colors \hfill 291
35.6 Colors for fills and strokes \hfill 291
35.6.1 Coloring math mode material \hfill 292
35.7 Multiple color models \hfill 292
35.8 Exporting color specifications \hfill 293
35.9 Creating new color models \hfill 293
35.9.1 Color profiles \hfill 294

36 The \texttt{l3pdf} package: Core PDF support \hfill 295

36.1 Objects \hfill 295
36.2 Version \hfill 296
36.3 Compression \hfill 297
36.4 Destinations \hfill 297

VII Additions and removals \hfill 299
### 37 The \l3candidates package: Experimental additions to \l3kernel

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>37.1</td>
<td>300</td>
</tr>
<tr>
<td>37.2</td>
<td>300</td>
</tr>
<tr>
<td>37.3</td>
<td>301</td>
</tr>
<tr>
<td>37.4</td>
<td>301</td>
</tr>
<tr>
<td>37.5</td>
<td>301</td>
</tr>
<tr>
<td>37.6</td>
<td>302</td>
</tr>
<tr>
<td>37.7</td>
<td>302</td>
</tr>
<tr>
<td>37.8</td>
<td>302</td>
</tr>
<tr>
<td>37.9</td>
<td>303</td>
</tr>
<tr>
<td>37.10</td>
<td>303</td>
</tr>
<tr>
<td>37.11</td>
<td>304</td>
</tr>
<tr>
<td>37.12</td>
<td>305</td>
</tr>
<tr>
<td>37.13</td>
<td>306</td>
</tr>
<tr>
<td>37.14</td>
<td>307</td>
</tr>
</tbody>
</table>

### VIII Implementation

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>38 \l3bootstrap implementation</td>
<td>311</td>
</tr>
<tr>
<td>38.1 Lua\TeX-specific code</td>
<td>311</td>
</tr>
<tr>
<td>38.2 The \texttt{pdfstrcmp} primitive in Xe\TeX</td>
<td>311</td>
</tr>
<tr>
<td>38.3 Loading support Lua code</td>
<td>312</td>
</tr>
<tr>
<td>38.4 Engine requirements</td>
<td>312</td>
</tr>
<tr>
<td>38.5 Extending allocators</td>
<td>313</td>
</tr>
<tr>
<td>38.6 The \texttt{\LaTeX}3 code environment</td>
<td>314</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>39 \l3names implementation</td>
<td>317</td>
</tr>
</tbody>
</table>

### 40 \l3kernel-functions: kernel-reserved functions

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>40.1 Internal kernel functions</td>
<td>344</td>
</tr>
<tr>
<td>40.2 Kernel backend functions</td>
<td>350</td>
</tr>
</tbody>
</table>

### 41 \l3basics implementation

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>41.1 Renaming some \TeX primitives (again)</td>
<td>352</td>
</tr>
<tr>
<td>41.2 Defining some constants</td>
<td>354</td>
</tr>
<tr>
<td>41.3 Defining functions</td>
<td>354</td>
</tr>
<tr>
<td>41.4 Selecting tokens</td>
<td>355</td>
</tr>
<tr>
<td>41.5 Gobbling tokens from input</td>
<td>357</td>
</tr>
<tr>
<td>41.6 Debugging and patching later definitions</td>
<td>357</td>
</tr>
<tr>
<td>41.7 Conditional processing and definitions</td>
<td>358</td>
</tr>
<tr>
<td>41.8 Dissecting a control sequence</td>
<td>364</td>
</tr>
<tr>
<td>41.9 Exist or free</td>
<td>366</td>
</tr>
<tr>
<td>41.10 Preliminaries for new functions</td>
<td>368</td>
</tr>
<tr>
<td>41.11 Defining new functions</td>
<td>369</td>
</tr>
<tr>
<td>41.12 Copying definitions</td>
<td>371</td>
</tr>
<tr>
<td>41.13 Undefining functions</td>
<td>371</td>
</tr>
<tr>
<td>41.14 Generating parameter text from argument count</td>
<td>372</td>
</tr>
<tr>
<td>41.15 Defining functions from a given number of arguments</td>
<td>373</td>
</tr>
<tr>
<td>41.16 Using the signature to define functions</td>
<td>374</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>45.3.4</td>
<td>Framework</td>
</tr>
<tr>
<td>45.3.5</td>
<td>Quantifiers</td>
</tr>
<tr>
<td>45.3.6</td>
<td>Raw characters</td>
</tr>
<tr>
<td>45.3.7</td>
<td>Character properties</td>
</tr>
<tr>
<td>45.3.8</td>
<td>Anchoring and simple assertions</td>
</tr>
<tr>
<td>45.3.9</td>
<td>Character classes</td>
</tr>
<tr>
<td>45.3.10</td>
<td>Groups and alternations</td>
</tr>
<tr>
<td>45.3.11</td>
<td>Catcodes and csnames</td>
</tr>
<tr>
<td>45.3.12</td>
<td>Raw token lists with \u</td>
</tr>
<tr>
<td>45.3.13</td>
<td>Other</td>
</tr>
<tr>
<td>45.3.14</td>
<td>Showing regexes</td>
</tr>
<tr>
<td>45.4</td>
<td>Building</td>
</tr>
<tr>
<td>45.4.1</td>
<td>Variables used while building</td>
</tr>
<tr>
<td>45.4.2</td>
<td>Framework</td>
</tr>
<tr>
<td>45.4.3</td>
<td>Helpers for building an NFA</td>
</tr>
<tr>
<td>45.4.4</td>
<td>Building classes</td>
</tr>
<tr>
<td>45.4.5</td>
<td>Building groups</td>
</tr>
<tr>
<td>45.4.6</td>
<td>Others</td>
</tr>
<tr>
<td>45.5</td>
<td>Matching</td>
</tr>
<tr>
<td>45.5.1</td>
<td>Variables used when matching</td>
</tr>
<tr>
<td>45.5.2</td>
<td>Matching: framework</td>
</tr>
<tr>
<td>45.5.3</td>
<td>Using states of the NFA</td>
</tr>
<tr>
<td>45.5.4</td>
<td>Actions when matching</td>
</tr>
<tr>
<td>45.6</td>
<td>Replacement</td>
</tr>
<tr>
<td>45.6.1</td>
<td>Variables and helpers used in replacement</td>
</tr>
<tr>
<td>45.6.2</td>
<td>Query and brace balance</td>
</tr>
<tr>
<td>45.6.3</td>
<td>Framework</td>
</tr>
<tr>
<td>45.6.4</td>
<td>Submatches</td>
</tr>
<tr>
<td>45.6.5</td>
<td>Csnames in replacement</td>
</tr>
<tr>
<td>45.6.6</td>
<td>Characters in replacement</td>
</tr>
<tr>
<td>45.6.7</td>
<td>An error</td>
</tr>
<tr>
<td>45.7</td>
<td>User functions</td>
</tr>
<tr>
<td>45.7.1</td>
<td>Variables and helpers for user functions</td>
</tr>
<tr>
<td>45.7.2</td>
<td>Matching</td>
</tr>
<tr>
<td>45.7.3</td>
<td>Extracting submatches</td>
</tr>
<tr>
<td>45.7.4</td>
<td>Replacement</td>
</tr>
<tr>
<td>45.7.5</td>
<td>Peeking ahead</td>
</tr>
<tr>
<td>45.8</td>
<td>Messages</td>
</tr>
<tr>
<td>45.9</td>
<td>Code for tracing</td>
</tr>
</tbody>
</table>

### 46 \texttt{t3prg} implementation

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>46.1</td>
<td>Primitive conditionals</td>
<td>563</td>
</tr>
<tr>
<td>46.2</td>
<td>Defining a set of conditional functions</td>
<td>563</td>
</tr>
<tr>
<td>46.3</td>
<td>The boolean data type</td>
<td>563</td>
</tr>
<tr>
<td>46.4</td>
<td>Internal auxiliaries</td>
<td>565</td>
</tr>
<tr>
<td>46.5</td>
<td>Boolean expressions</td>
<td>566</td>
</tr>
<tr>
<td>46.6</td>
<td>Logical loops</td>
<td>571</td>
</tr>
<tr>
<td>46.7</td>
<td>Producing multiple copies</td>
<td>572</td>
</tr>
<tr>
<td>46.8</td>
<td>Detecting \TeX’s mode</td>
<td>574</td>
</tr>
<tr>
<td>46.9</td>
<td>Internal programming functions</td>
<td>574</td>
</tr>
</tbody>
</table>
47 l3sys implementation
  47.1 Kernel code ........................................ 576
    47.1.1 Detecting the engine ............................ 576
    47.1.2 Randomness ...................................... 578
    47.1.3 Platform .......................................... 578
    47.1.4 Configurations .................................... 578
    47.1.5 Access to the shell ............................... 580
  47.2 Dynamic (every job) code .............................. 583
    47.2.1 The name of the job ................................ 583
    47.2.2 Time and date ..................................... 583
    47.2.3 Random numbers .................................... 584
    47.2.4 Access to the shell ............................... 585
    47.2.5 Held over from l3file ............................ 586
  47.3 Last-minute code ...................................... 586
    47.3.1 Detecting the output ............................. 587
    47.3.2 Configurations .................................... 587

48 l3msg implementation
  48.1 Internal auxiliaries ................................ 589
  48.2 Creating messages ................................... 589
  48.3 Messages: support functions and text ............... 591
  48.4 Showing messages: low level mechanism .............. 592
  48.5 Displaying messages .................................. 594
  48.6 Kernel-specific functions ............................ 603
  48.7 Internal messages .................................... 604
  48.8 Expandable errors .................................... 610
  48.9 Message formatting ................................... 611

49 l3file implementation
  49.1 Input operations ..................................... 612
    49.1.1 Variables and constants ......................... 612
    49.1.2 Stream management ............................... 613
    49.1.3 Reading input .................................... 616
  49.2 Output operations ................................... 619
    49.2.1 Variables and constants ......................... 619
    49.2.2 Internal auxiliaries ............................. 620
  49.3 Stream management ................................... 620
    49.3.1 Deferred writing .................................. 622
    49.3.2 Immediate writing ................................ 622
    49.3.3 Special characters for writing ................. 623
    49.3.4 Hard-wrapping lines to a character count ..... 624
  49.4 File operations ...................................... 633
    49.4.1 Internal auxiliaries ............................. 635
  49.5 GetIdInfo ............................................ 650
  49.6 Checking the version of kernel dependencies ....... 651
  49.7 Messages .............................................. 653
  49.8 Functions delayed from earlier modules ............ 654
59 l3clist implementation
  59.1 Removing spaces around items .................................. 835
  59.2 Allocation and initialisation .................................. 836
  59.3 Adding data to comma lists .................................... 838
  59.4 Comma lists as stacks ......................................... 839
  59.5 Modifying comma lists ......................................... 841
  59.6 Comma list conditionals ....................................... 844
  59.7 Mapping over comma lists ...................................... 845
  59.8 Using comma lists ............................................. 849
  59.9 Using a single item ........................................... 851
  59.10 Viewing comma lists .......................................... 853
  59.11 Scratch comma lists .......................................... 854

60 l3token implementation 855
  60.1 Internal auxiliaries ........................................... 855
  60.2 Manipulating and interrogating character tokens ............... 855
  60.3 Creating character tokens .................................... 858
  60.4 Generic tokens ................................................ 866
  60.5 Token conditionals ............................................ 867
  60.6 Peeking ahead at the next token ................................ 877

61 l3prop implementation 884
  61.1 Internal auxiliaries ........................................... 885
  61.2 Allocation and initialisation .................................. 886
  61.3 Accessing data in property lists ................................ 888
  61.4 Property list conditionals .................................... 893
  61.5 Recovering values from property lists with branching ........ 894
  61.6 Mapping over property lists ................................... 895
  61.7 Viewing property lists ....................................... 896

62 l3skip implementation 898
  62.1 Length primitives renamed ..................................... 898
  62.2 Internal auxiliaries ........................................... 898
  62.3 Creating and initialising dim variables ........................ 898
  62.4 Setting dim variables ......................................... 899
  62.5 Utilities for dimension calculations ........................... 900
  62.6 Dimension expression conditionals ............................. 901
  62.7 Dimension expression loops ................................... 903
  62.8 Dimension step functions ..................................... 904
  62.9 Using dim expressions and variables ........................... 906
  62.10 Viewing dim variables ........................................ 908
  62.11 Constant dimensions ......................................... 908
  62.12 Scratch dimensions .......................................... 908
  62.13 Creating and initialising skip variables ...................... 908
  62.14 Setting skip variables ....................................... 910
  62.15 Skip expression conditionals ................................. 910
  62.16 Using skip expressions and variables ......................... 911
  62.17 Inserting skips into the output ................................ 911
  62.18 Viewing skip variables ....................................... 912
  62.19 Constant skips .............................................. 912
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>62.20</td>
<td>Scratch skips</td>
<td>912</td>
</tr>
<tr>
<td>62.21</td>
<td>Creating and initialising \texttt{muskip} variables</td>
<td>912</td>
</tr>
<tr>
<td>62.22</td>
<td>Setting \texttt{muskip} variables</td>
<td>913</td>
</tr>
<tr>
<td>62.23</td>
<td>Using \texttt{muskip} expressions and variables</td>
<td>914</td>
</tr>
<tr>
<td>62.24</td>
<td>Viewing \texttt{muskip} variables</td>
<td>914</td>
</tr>
<tr>
<td>62.25</td>
<td>Constant \texttt{muskips}</td>
<td>915</td>
</tr>
<tr>
<td>62.26</td>
<td>Scratch \texttt{muskips}</td>
<td>915</td>
</tr>
<tr>
<td>63</td>
<td>\texttt{l3keys} Implementation</td>
<td>916</td>
</tr>
<tr>
<td>63.1</td>
<td>Low-level interface</td>
<td>916</td>
</tr>
<tr>
<td>63.2</td>
<td>Constants and variables</td>
<td>923</td>
</tr>
<tr>
<td>63.2.1</td>
<td>Internal auxiliaries</td>
<td>925</td>
</tr>
<tr>
<td>63.3</td>
<td>The key defining mechanism</td>
<td>926</td>
</tr>
<tr>
<td>63.4</td>
<td>Turning properties into actions</td>
<td>928</td>
</tr>
<tr>
<td>63.5</td>
<td>Creating key properties</td>
<td>935</td>
</tr>
<tr>
<td>63.6</td>
<td>Setting keys</td>
<td>941</td>
</tr>
<tr>
<td>63.7</td>
<td>Utilities</td>
<td>950</td>
</tr>
<tr>
<td>63.8</td>
<td>Messages</td>
<td>952</td>
</tr>
<tr>
<td>64</td>
<td>\texttt{l3intarray} implementation</td>
<td>954</td>
</tr>
<tr>
<td>64.1</td>
<td>Lua implementation</td>
<td>954</td>
</tr>
<tr>
<td>64.1.1</td>
<td>Allocating arrays</td>
<td>954</td>
</tr>
<tr>
<td>64.1.2</td>
<td>Array items</td>
<td>957</td>
</tr>
<tr>
<td>64.1.3</td>
<td>Working with contents of integer arrays</td>
<td>959</td>
</tr>
<tr>
<td>64.2</td>
<td>Font dimension based implementation</td>
<td>960</td>
</tr>
<tr>
<td>64.2.1</td>
<td>Allocating arrays</td>
<td>961</td>
</tr>
<tr>
<td>64.2.2</td>
<td>Array items</td>
<td>962</td>
</tr>
<tr>
<td>64.2.3</td>
<td>Working with contents of integer arrays</td>
<td>964</td>
</tr>
<tr>
<td>64.3</td>
<td>Common parts</td>
<td>966</td>
</tr>
<tr>
<td>64.3.1</td>
<td>Random arrays</td>
<td>966</td>
</tr>
<tr>
<td>65</td>
<td>\texttt{l3fp} implementation</td>
<td>968</td>
</tr>
<tr>
<td>66</td>
<td>\texttt{l3fp-aux} implementation</td>
<td>969</td>
</tr>
<tr>
<td>66.1</td>
<td>Access to primitives</td>
<td>969</td>
</tr>
<tr>
<td>66.2</td>
<td>Internal representation</td>
<td>969</td>
</tr>
<tr>
<td>66.3</td>
<td>Using arguments and semicolons</td>
<td>970</td>
</tr>
<tr>
<td>66.4</td>
<td>Constants, and structure of floating points</td>
<td>971</td>
</tr>
<tr>
<td>66.5</td>
<td>Overflow, underflow, and exact zero</td>
<td>974</td>
</tr>
<tr>
<td>66.6</td>
<td>Expanding after a floating point number</td>
<td>974</td>
</tr>
<tr>
<td>66.7</td>
<td>Other floating point types</td>
<td>975</td>
</tr>
<tr>
<td>66.8</td>
<td>Packing digits</td>
<td>978</td>
</tr>
<tr>
<td>66.9</td>
<td>Decimate (dividing by a power of 10)</td>
<td>981</td>
</tr>
<tr>
<td>66.10</td>
<td>Functions for use within primitive conditional branches</td>
<td>983</td>
</tr>
<tr>
<td>66.11</td>
<td>Integer floating points</td>
<td>984</td>
</tr>
<tr>
<td>66.12</td>
<td>Small integer floating points</td>
<td>985</td>
</tr>
<tr>
<td>66.13</td>
<td>Fast string comparison</td>
<td>986</td>
</tr>
<tr>
<td>66.14</td>
<td>Name of a function from its \texttt{l3fp-parse} name</td>
<td>986</td>
</tr>
<tr>
<td>66.15</td>
<td>Messages</td>
<td>986</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>67</td>
<td>l3fp-traps Implementation</td>
<td>987</td>
</tr>
<tr>
<td>67.1</td>
<td>Flags</td>
<td>987</td>
</tr>
<tr>
<td>67.2</td>
<td>Traps</td>
<td>987</td>
</tr>
<tr>
<td>67.3</td>
<td>Errors</td>
<td>991</td>
</tr>
<tr>
<td>67.4</td>
<td>Messages</td>
<td>991</td>
</tr>
<tr>
<td>68</td>
<td>l3fp-round implementation</td>
<td>993</td>
</tr>
<tr>
<td>68.1</td>
<td>Rounding tools</td>
<td>993</td>
</tr>
<tr>
<td>68.2</td>
<td>The <code>round</code> function</td>
<td>997</td>
</tr>
<tr>
<td>69</td>
<td>l3fp-parse implementation</td>
<td>1002</td>
</tr>
<tr>
<td>69.1</td>
<td>Work plan</td>
<td>1002</td>
</tr>
<tr>
<td>69.1.1</td>
<td>Storing results</td>
<td>1003</td>
</tr>
<tr>
<td>69.1.2</td>
<td>Precedence and infix operators</td>
<td>1004</td>
</tr>
<tr>
<td>69.1.3</td>
<td>Prefix operators, parentheses, and functions</td>
<td>1007</td>
</tr>
<tr>
<td>69.1.4</td>
<td>Numbers and reading tokens one by one</td>
<td>1008</td>
</tr>
<tr>
<td>69.2</td>
<td>Main auxiliary functions</td>
<td>1010</td>
</tr>
<tr>
<td>69.3</td>
<td>Helpers</td>
<td>1011</td>
</tr>
<tr>
<td>69.4</td>
<td>Parsing one number</td>
<td>1012</td>
</tr>
<tr>
<td>69.4.1</td>
<td>Numbers: trimming leading zeros</td>
<td>1018</td>
</tr>
<tr>
<td>69.4.2</td>
<td>Number: small significand</td>
<td>1019</td>
</tr>
<tr>
<td>69.4.3</td>
<td>Number: large significand</td>
<td>1021</td>
</tr>
<tr>
<td>69.4.4</td>
<td>Number: beyond 16 digits, rounding</td>
<td>1023</td>
</tr>
<tr>
<td>69.4.5</td>
<td>Number: finding the exponent</td>
<td>1026</td>
</tr>
<tr>
<td>69.5</td>
<td>Constants, functions and prefix operators</td>
<td>1029</td>
</tr>
<tr>
<td>69.5.1</td>
<td>Prefix operators</td>
<td>1029</td>
</tr>
<tr>
<td>69.5.2</td>
<td>Constants</td>
<td>1032</td>
</tr>
<tr>
<td>69.5.3</td>
<td>Functions</td>
<td>1033</td>
</tr>
<tr>
<td>69.6</td>
<td>Main functions</td>
<td>1034</td>
</tr>
<tr>
<td>69.7</td>
<td>Infix operators</td>
<td>1036</td>
</tr>
<tr>
<td>69.7.1</td>
<td>Closing parentheses and commas</td>
<td>1037</td>
</tr>
<tr>
<td>69.7.2</td>
<td>Usual infix operators</td>
<td>1039</td>
</tr>
<tr>
<td>69.7.3</td>
<td>Juxtaposition</td>
<td>1040</td>
</tr>
<tr>
<td>69.7.4</td>
<td>Multi-character cases</td>
<td>1041</td>
</tr>
<tr>
<td>69.7.5</td>
<td>Ternary operator</td>
<td>1041</td>
</tr>
<tr>
<td>69.7.6</td>
<td>Comparisons</td>
<td>1041</td>
</tr>
<tr>
<td>69.8</td>
<td>Tools for functions</td>
<td>1043</td>
</tr>
<tr>
<td>69.9</td>
<td>Messages</td>
<td>1046</td>
</tr>
<tr>
<td>70</td>
<td>l3fp-assign implementation</td>
<td>1047</td>
</tr>
<tr>
<td>70.1</td>
<td>Assigning values</td>
<td>1047</td>
</tr>
<tr>
<td>70.2</td>
<td>Updating values</td>
<td>1048</td>
</tr>
<tr>
<td>70.3</td>
<td>Showing values</td>
<td>1048</td>
</tr>
<tr>
<td>70.4</td>
<td>Some useful constants and scratch variables</td>
<td>1049</td>
</tr>
</tbody>
</table>
71 l3fp-logic Implementation 1050
  71.1 Syntax of internal functions ...................................... 1050
  71.2 Tests ............................................................ 1050
  71.3 Comparison ....................................................... 1051
  71.4 Floating point expression loops .................................. 1054
  71.5 Extrema .......................................................... 1057
  71.6 Boolean operations ............................................... 1059
  71.7 Ternary operator ................................................ 1060

72 l3fp-basics Implementation 1062
  72.1 Addition and subtraction .......................................... 1062
     72.1.1 Sign, exponent, and special numbers ....................... 1063
     72.1.2 Absolute addition ......................................... 1065
     72.1.3 Absolute subtraction ...................................... 1067
  72.2 Multiplication ................................................... 1071
     72.2.1 Signs, and special numbers ................................ 1071
     72.2.2 Absolute multiplication ................................... 1073
  72.3 Division .......................................................... 1075
     72.3.1 Signs, and special numbers ................................ 1075
     72.3.2 Work plan .................................................. 1076
     72.3.3 Implementing the significand division .................... 1079
  72.4 Square root ...................................................... 1084
  72.5 About the sign and exponent ..................................... 1091
  72.6 Operations on tuples ............................................. 1092

73 l3fp-extended implementation 1094
  73.1 Description of fixed point numbers ............................... 1094
  73.2 Helpers for numbers with extended precision .................... 1095
  73.3 Multiplying a fixed point number by a short one .................. 1096
  73.4 Dividing a fixed point number by a small integer ................ 1096
  73.5 Adding and subtracting fixed points ................................ 1097
  73.6 Multiplying fixed points ........................................ 1098
  73.7 Combining product and sum of fixed points ....................... 1099
  73.8 Extended-precision floating point numbers ...................... 1102
  73.9 Dividing extended-precision numbers ............................. 1104
  73.10 Inverse square root of extended precision numbers ............ 1108
  73.11 Converting from fixed point to floating point .................. 1110

74 l3fp-expo implementation 1112
  74.1 Logarithm ........................................................ 1112
     74.1.1 Work plan ................................................ 1112
     74.1.2 Some constants ............................................ 1113
     74.1.3 Sign, exponent, and special numbers ...................... 1113
     74.1.4 Absolute ln ............................................... 1113
  74.2 Exponential ...................................................... 1120
     74.2.1 Sign, exponent, and special numbers ...................... 1120
  74.3 Power ............................................................ 1125
  74.4 Factorial ........................................................ 1131
87 l3pdf implementation 1359
  87.1 Compression ............................................. 1359
  87.2 Objects .................................................. 1360
  87.3 Version .................................................. 1360
  87.4 Destinations ............................................ 1362

88 l3candidates Implementation 1363
  88.1 Additions to l3box ....................................... 1363
      88.1.1 Viewing part of a box ............................. 1363
  88.2 Additions to l3flag ..................................... 1365
  88.3 Additions to l3msg ...................................... 1366
  88.4 Additions to l3prg ...................................... 1367
  88.5 Additions to l3prop ..................................... 1368
  88.6 Additions to l3seq ...................................... 1368
  88.7 Additions to l3sys ...................................... 1373
  88.8 Additions to l3file ..................................... 1374
  88.9 Additions to l3tl ........................................ 1374
      88.9.1 Building a token list ............................. 1374
      88.9.2 Other additions to l3tl ......................... 1378
  88.10 Additions to l3token ................................... 1378

89 l3deprecation implementation 1381
  89.1 Patching definitions to deprecate .................... 1381
  89.2 Removed functions ...................................... 1383
  89.3 Deprecated l3str functions ............................ 1387
  89.4 Deprecated l3seq functions ............................ 1388
  89.5 Deprecated l3sys functions ............................ 1388
  89.6 Deprecated l3tl functions ............................. 1388
  89.7 Deprecated l3token functions ......................... 1389

Index 1391
Part I
Introduction
Chapter 1

Introduction to expl3 and this document

This document is intended to act as a comprehensive reference manual for the expl3 language. A general guide to the \LaTeX3 programming language is found in expl3.pdf.

1.1 Naming functions and variables

\LaTeX3 does not use @ as a “letter” for defining internal macros. Instead, the symbols _ and : are used in internal macro names to provide structure. The name of each function is divided into logical units using _, while : separates the name of the function from the argument specifier (“arg-spec”). This describes the arguments expected by the function.

In most cases, each argument is represented by a single letter. The complete list of arg-spec letters for a function is referred to as the signature of the function.

Each function name starts with the module to which it belongs. Thus apart from a small number of very basic functions, all expl3 function names contain at least one underscore to divide the module name from the descriptive name of the function. For example, all functions concerned with comma lists are in module clist and begin \clist_.

Every function must include an argument specifier. For functions which take no arguments, this will be blank and the function name will end :. Most functions take one or more arguments, and use the following argument specifiers:

N and n These mean no manipulation, of a single token for N and of a set of tokens given in braces for n. Both pass the argument through exactly as given. Usually, if you use a single token for an n argument, all will be well.

c This means csname, and indicates that the argument will be turned into a csname before being used. So \foo:c {ArgumentOne} will act in the same way as \foo:N \ArgumentOne. All macros that appear in the argument are expanded. An internal error will occur if the result of expansion inside a c-type argument is not a series of character tokens.

V and v These mean value of variable. The V and v specifiers are used to get the content of a variable without needing to worry about the underlying \TeX structure containing the data. A V argument will be a single token (similar to N), for example
\texttt{\textbackslash foo:V \textbackslash MyVariable}: on the other hand, using \texttt{v} a csname is constructed first, and then the value is recovered, for example \texttt{\textbackslash foo:v \{MyVariable\}}.

\textbullet{} This means \textit{expansion once}. In general, the \texttt{V} and \texttt{v} specifiers are favoured over \texttt{O} for recovering stored information. However, \texttt{O} is useful for correctly processing information with delimited arguments.

\textbullet{} The \texttt{x} specifier stands for \textit{exhaustive expansion}: every token in the argument is fully expanded until only unexpandable ones remain. The \TeX{} \texttt{\edef} primitive carries out this type of expansion. Functions which feature an \texttt{x}-type argument are \textit{not} expandable.

\textbullet{} The \texttt{e} specifier is in many respects identical to \texttt{x}, but with a very different implementation. Functions which feature an \texttt{e}-type argument may be expandable. The drawback is that \texttt{e} is extremely slow (often more than 200 times slower) in older engines, more precisely in non-Lua\TeX{} engines older than 2019.

\textbullet{} The \texttt{f} specifier stands for \textit{full expansion}, and in contrast to \texttt{x} stops at the first non-expandable token (reading the argument from left to right) without trying to expand it. If this token is a \texttt{⟨space token⟩}, it is gobbled, and thus won’t be part of the resulting argument. For example, when setting a token list variable (a macro used for storage), the sequence

\begin{verbatim}
\tl_set:Nn \l_mya_tl \{ A \}
\tl_set:Nn \l_myb_tl \{ B \}
\tl_set:Nf \l_mya_tl \l_myb_tl \}
\end{verbatim}

will leave \texttt{\l_mya_tl} with the content \texttt{A\l_myb_tl}, as \texttt{A} cannot be expanded and so terminates expansion before \texttt{\l_myb_tl} is considered.

\textbf{T} and \textbf{F} For logic tests, there are the branch specifiers \texttt{T} (true) and \texttt{F} (false). Both specifiers treat the input in the same way as \texttt{N} (no change), but make the logic much easier to see.

\textbullet{} The letter \texttt{p} indicates \TeX{} \textit{parameters}. Normally this will be used for delimited functions as expl3 provides better methods for creating simple sequential arguments.

\textbullet{} The \texttt{w} specifier is for \textit{weird} arguments. This covers everything else, but mainly applies to delimited values (where the argument must be terminated by some specified string).

\textbf{D} The \texttt{D} stands for \textbf{Do not use}. All of the \TeX{} primitives are initially \texttt{\let} to a \texttt{D} name, and some are then given a second name. These functions have no standardized syntax, they are engine dependent and their name can change without warning, thus their use is \textbf{strongly discouraged} in package code: programmers should instead use the interfaces documented in interface3.pdf\textsuperscript{1}.

Notice that the argument specifier describes how the argument is processed prior to being passed to the underlying function. For example, \texttt{\textbackslash foo:c} will take its argument, convert it to a control sequence and pass it to \texttt{\textbackslash foo:N}.

\textsuperscript{1}If a primitive offers a functionality not yet in the kernel, programmers and users are encouraged to write to the \LaTeX{}-L mailing list (mailto:LaTEX-L@listserv.uni-heidelberg.de) describing their use-case and intended behaviour, so that a possible interface can be discussed. Temporarily, while an interface is not provided, programmers may use the procedure described in the l3styleguide.pdf.
Variables are named in a similar manner to functions, but begin with a single letter to define the type of variable:

- **c** Constant: global parameters whose value should not be changed.
- **g** Parameters whose value should only be set globally.
- **l** Parameters whose value should only be set locally.

Each variable name is then built up in a similar way to that of a function, typically starting with the module\(^2\) name and then a descriptive part. Variables end with a short identifier to show the variable type:

- **clist** Comma separated list.
- **dim** “Rigid” lengths.
- **fp** Floating-point values;
- **int** Integer-valued count register.
- **muskip** “Rubber” lengths for use in mathematics.
- **seq** “Sequence”: a data-type used to implement lists (with access at both ends) and stacks.
- **skip** “Rubber” lengths.
- **str** String variables: contain character data.
- **tl** Token list variables: placeholder for a token list.

Applying V-type or v-type expansion to variables of one of the above types is supported, while it is not supported for the following variable types:

- **bool** Either true or false.
- **box** Box register.
- **coffin** A “box with handles” — a higher-level data type for carrying out box alignment operations.
- **flag** Integer that can be incremented expandably.
- **fparray** Fixed-size array of floating point values.
- **intarray** Fixed-size array of integers.
- **ior/iow** An input or output stream, for reading from or writing to, respectively.
- **prop** Property list: analogue of dictionary or associative arrays in other languages.
- **regex** Regular expression.

\(^2\)The module names are not used in case of generic scratch registers defined in the data type modules, e.g., the int module contains some scratch variables called \(l_{tmpa\_int}, l_{tmpb\_int}, \text{and so on.}\) In such a case adding the module name up front to denote the module and in the back to indicate the type, as in \(l\_int\_tmpa\_int\) would be very unreadable.
1.1.1 Scratch variables
 Modules focussed on variable usage typically provide four scratch variables, two local and two global, with names of the form \( (\text{scope})_\text{tmp}_a (\text{type}) / (\text{scope})_\text{tmp}_b (\text{type}) \). These are never used by the core code. The nature of \TeXX\ grouping means that as with any other scratch variable, these should only be set and used with no intervening third-party code.

1.1.2 Terminological inexactitude
 A word of warning. In this document, and others referring to the expl3 programming modules, we often refer to “variables” and “functions” as if they were actual constructs from a real programming language. In truth, \TeXX\ is a macro processor, and functions are simply macros that may or may not take arguments and expand to their replacement text. Many of the common variables are also macros, and if placed into the input stream will simply expand to their definition as well — a “function” with no arguments and a “token list variable” are almost the same.\(^3\) On the other hand, some “variables” are actually registers that must be initialised and their values set and retrieved with specific functions.

The conventions of the expl3 code are designed to clearly separate the ideas of “macros that contain data” and “macros that contain code”, and a consistent wrapper is applied to all forms of “data” whether they be macros or actually registers. This means that sometimes we will use phrases like “the function returns a value”, when actually we just mean “the macro expands to something”. Similarly, the term “execute” might be used in place of “expand” or it might refer to the more specific case of “processing in \TeXX\’s stomach” (if you are familiar with the \TeXX\book parlance).

If in doubt, please ask; chances are we’ve been hasty in writing certain definitions and need to be told to tighten up our terminology.

1.2 Documentation conventions
 This document is typeset with the experimental l3doc class; several conventions are used to help describe the features of the code. A number of conventions are used here to make the documentation clearer.

Each group of related functions is given in a box. For a function with a “user” name, this might read:

\[\texttt{\ExplSyntaxOn} \ldots \texttt{\ExplSyntaxOff}\]

The textual description of how the function works would appear here. The syntax of the function is shown in mono-spaced text to the right of the box. In this example, the function takes no arguments and so the name of the function is simply reprinted.

For programming functions, which use _ and : in their name there are a few additional conventions: If two related functions are given with identical names but different argument specifiers, these are termed variants of each other, and the latter functions are printed in grey to show this more clearly. They will carry out the same function but will take different types of argument:

\(^3\)\TeXX\nically, functions with no arguments are \texttt{\textbackslash long} while token list variables are not.
When a number of variants are described, the arguments are usually illustrated only for
the base function. Here, \textit{\textlangle sequence\textrangle} indicates that \texttt{\seq_new:N} expects the name of a
sequence. From the argument specifier, \texttt{\seq_new:c} also expects a sequence name, but
as a name rather than as a control sequence. Each argument given in the illustration
should be described in the following text.

\textbf{Fully expandable functions} Some functions are fully expandable, which allows them
to be used within an \texttt{x}-type or \texttt{e}-type argument (in plain \TeX terms, inside an \texttt{edef} or
\texttt{expanded}), as well as within an \texttt{f}-type argument. These fully expandable functions are
indicated in the documentation by a star:

\begin{itemize}
  \item \texttt{\cs_to_str:N \textstar}
\end{itemize}

As with other functions, some text should follow which explains how the function works.
Usually, only the star will indicate that the function is expandable. In this case, the
function expects a \textit{\textlangle cs\textrangle}, shorthand for a \textit{\textlangle control sequence\textrangle}.

\textbf{Restricted expandable functions} A few functions are fully expandable but cannot
be fully expanded within an \texttt{f}-type argument. In this case a hollow star is used to indicate
this:

\begin{itemize}
  \item \texttt{\seq_map_function:NN \textstar}
\end{itemize}

\textbf{Conditional functions} Conditional (if) functions are normally defined in three vari-
ants, with \texttt{T}, \texttt{F} and \texttt{TF} argument specifiers. This allows them to be used for different
“true”/“false” branches, depending on which outcome the conditional is being used to
test. To indicate this without repetition, this information is given in a shortened form:

\begin{itemize}
  \item \texttt{\sys_if_engine_xetex:TF \textstar}
\end{itemize}

The underlining and italic of \texttt{TF} indicates that three functions are available:

\begin{itemize}
  \item \texttt{\sys_if_engine_xetex:T}
  \item \texttt{\sys_if_engine_xetex:F}
  \item \texttt{\sys_if_engine_xetex:TF}
\end{itemize}

Usually, the illustration will use the \texttt{TF} variant, and so both \textit{(true code)} and \textit{(false code)}
will be shown. The two variant forms \texttt{T} and \texttt{F} take only \textit{(true code)} and \textit{(false code)},
respectively. Here, the star also shows that this function is expandable. With some minor
exceptions, all conditional functions in the \texttt{expl3} modules should be defined in this way.

Variables, constants and so on are described in a similar manner:

\begin{itemize}
  \item \texttt{l\_tmpa\_tl}
\end{itemize}

A short piece of text will describe the variable: there is no syntax illustration in this case.

In some cases, the function is similar to one in \texttt{etex2} or plain \TeX. In these cases,
the text will include an extra “\TeXhackers note” section:
The normal description text.

\textbf{\TeX hacklers note:} Detail for the experienced \TeX or \LaTeX\ programmer. In this case, it would point out that this function is the \TeX\ primitive \texttt{\string}.

\textbf{Changes to behaviour} When new functions are added to \texttt{expl3}, the date of first inclusion is given in the documentation. Where the documented behaviour of a function changes after it is first introduced, the date of the update will also be given. This means that the programmer can be sure that any release of \texttt{expl3} after the date given will contain the function of interest with expected behaviour as described. Note that changes to code internals, including bug fixes, are not recorded in this way \textit{unless} they impact on the expected behaviour.

\section{Formal language conventions which apply generally}

As this is a formal reference guide for \LaTeX\ programming, the descriptions of functions are intended to be reasonably “complete”. However, there is also a need to avoid repetition. Formal ideas which apply to general classes of function are therefore summarised here.

For tests which have a \texttt{TF} argument specification, the test if evaluated to give a logically \texttt{TRUE} or \texttt{FALSE} result. Depending on this result, either the \texttt{⟨true code⟩} or the \texttt{⟨false code⟩} will be left in the input stream. In the case where the test is expandable, and a predicate (_p) variant is available, the logical value determined by the test is left in the input stream: this will typically be part of a larger logical construct.

\section{\TeX concepts not supported by \LaTeX\}

The \TeX concept of an “\texttt{\outer}” macro is \textit{not supported} at all by \LaTeX. As such, the functions provided here may break when used on top of \LaTeX\ if \texttt{\outer} tokens are used in the arguments.
Part II

Bootstrapping
Chapter 2

The \texttt{l3bootstrap} package

Bootstrap code

2.1 Using the \texttt{\LaTeX3} modules

The modules documented in \texttt{source3} are designed to be used on top of \LaTeXe and are loaded all as one with the usual \texttt{\usepackage{expl3}} or \texttt{\RequirePackage{expl3}} instructions.

As the modules use a coding syntax different from standard \LaTeXe it provides a few functions for setting it up.

\begin{verbatim}
\ExplSyntaxOn \ExplSyntaxOff
\end{verbatim}

The \texttt{\ExplSyntaxOn} function switches to a category code regime in which spaces and new lines are ignored, and in which the colon (:) and underscore (_\texttt{)} are treated as “letters”, thus allowing access to the names of code functions and variables. Within this environment, - is used to input a space. The \texttt{\ExplSyntaxOff} reverts to the document category code regime.

\textbf{\texttt{\LaTeX}hackers note}: Spaces introduced by - behave much in the same way as normal space characters in the standard category code regime: they are ignored after a control word or at the start of a line, and multiple consecutive - are equivalent to a single one. However, - is \textit{not} ignored at the end of a line.

\begin{verbatim}
\ProvidesExplPackage \ProvidesExplClass \ProvidesExplFile
\end{verbatim}

These functions act broadly in the same way as the corresponding \texttt{\LaTeXe} kernel functions \texttt{\ProvidesPackage}, \texttt{\ProvidesClass} and \texttt{\ProvidesFile}. However, they also implicitly switch \texttt{\ExplSyntaxOn} for the remainder of the code with the file. At the end of the file, \texttt{\ExplSyntaxOff} will be called to reverse this. (This is the same concept as \texttt{\LaTeXe} provides in turning on \texttt{\makeatletter} within package and class code.) The \texttt{\langle date\rangle} should be given in the format \texttt{(year)/(month)/(day)} or in the ISO date format \texttt{(year)-(month)-(day)}. If the \texttt{\langle version\rangle} is given then it will be prefixed with \texttt{v} in the package identifier line.
\RequirePackage{l3bootstrap}
\GetIdInfo $Id: ⟨SVN info field⟩ $ {(description)}

Extracts all information from a SVN field. Spaces are not ignored in these fields. The information pieces are stored in separate control sequences with \ExplFileName for the part of the file name leading up to the period, \ExplFileDate for date, \ExplFileVersion for version and \ExplFileDescription for the description.

To summarize: Every single package using this syntax should identify itself using one of the above methods. Special care is taken so that every package or class file loaded with \RequirePackage or similar are loaded with usual \LaTeX\^2\LaTeX\^3 category codes and the \LaTeX\^3 category code scheme is reloaded when needed afterwards. See implementation for details. If you use the \GetIdInfo command you can use the information when loading a package with

\ProvidesExplPackage{\ExplFileName}
{\ExplFileDate}{\ExplFileVersion}{\ExplFileDescription}
Chapter 3

The l3names package
Namespace for primitives

3.1 Setting up the \LaTeX3 programming language

This module is at the core of the \LaTeX3 programming language. It performs the following tasks:

- defines new names for all \TeX primitives;
- emulate required primitives not provided by default in Lua\TeX;
- switches to the category code régime for programming;

This module is entirely dedicated to primitives (and emulations of these), which should not be used directly within \LaTeX3 code (outside of “kernel-level” code). As such, the primitives are not documented here: *The \TeXbook, \TeX by Topic* and the manuals for pdf\TeX, X\TeX, Lua\TeX, p\TeX and up\TeX should be consulted for details of the primitives. These are named \texttt{\textbackslash tex\_⟨name⟩} in pdf\TeX and omitting a leading \texttt{pdf} when the primitive is not related to pdf output.
Part III

Programming Flow
Chapter 4

The \texttt{l3basics} package

Basic definitions

As the name suggest this package holds some basic definitions which are needed by most or all other packages in this set.

Here we describe those functions that are used all over the place. With that we mean functions dealing with the construction and testing of control sequences. Furthermore the basic parts of conditional processing are covered; conditional processing dealing with specific data types is described in the modules specific for the respective data types.

4.1 No operation functions

\texttt{prg\_do\_nothing}: \star

An expandable function which does nothing at all: leaves nothing in the input stream after a single expansion.

\texttt{scan\_stop}:

A non-expandable function which does nothing. Does not vanish on expansion but produces no typeset output.

4.2 Grouping material

\texttt{group\_begin}: \texttt{group\_end}:

These functions begin and end a group for definition purposes. Assignments are local to groups unless carried out in a global manner. (A small number of exceptions to this rule will be noted as necessary elsewhere in this document.) Each \texttt{group\_begin} must be matched by a \texttt{group\_end}, although this does not have to occur within the same function. Indeed, it is often necessary to start a group within one function and finish it within another, for example when seeking to use non-standard category codes.
\group_insert_after:N \group_insert_after:N (token)

Adds (token) to the list of (tokens) to be inserted when the current group level ends. The list of (tokens) to be inserted is empty at the beginning of a group: multiple applications of \group_insert_after:N may be used to build the inserted list one (token) at a time. The current group level may be closed by a \group_end: function or by a token with category code 2 (close-group), namely a \} if standard category codes apply.

\group_show_list: \group_log_list:

Display (to the terminal or log file) a list of the groups that are currently opened. This is intended for tracking down problems.

\TeXhackers note: This is a wrapper around the \showgroups primitive.

4.3 Control sequences and functions

As \TeX is a macro language, creating new functions means creating macros. At point of use, a function is replaced by the replacement text (“code”) in which each parameter in the code (#1, #2, etc.) is replaced the appropriate arguments absorbed by the function. In the following, (code) is therefore used as a shorthand for “replacement text”.

Functions which are not “protected” are fully expanded inside an x expansion. In contrast, “protected” functions are not expanded within x expansions.

4.3.1 Defining functions

Functions can be created with no requirement that they are declared first (in contrast to variables, which must always be declared). Declaring a function before setting up the code means that the name chosen is checked and an error raised if it is already in use. The name of a function can be checked at the point of definition using the \cs_new... functions: this is recommended for all functions which are defined for the first time.

There are three ways to define new functions. All classes define a function to expand to the substitution text. Within the substitution text the actual parameters are substituted for the formal parameters (#1, #2, ...).

new Create a new function with the new scope, such as \cs_new:Npn. The definition is global and results in an error if it is already defined.

set Create a new function with the set scope, such as \cs_set:Npn. The definition is restricted to the current \TeX group and does not result in an error if the function is already defined.

gset Create a new function with the gset scope, such as \cs_gset:Npn. The definition is global and does not result in an error if the function is already defined.

Within each set of scope there are different ways to define a function. The differences depend on restrictions on the actual parameters and the expandability of the resulting function.

nopar Create a new function with the nopar restriction, such as \cs_set_nopar:Npn. The parameter may not contain \par tokens.
protected Create a new function with the protected restriction, such as \cs_set_-_protected:Npn. The parameter may contain \par tokens but the function will not expand within an x-type or e-type expansion.

Finally, the functions in Subsections 4.3.2 and 4.3.3 are primarily meant to define base functions only. Base functions can only have the following argument specifiers:

N and n No manipulation.

T and F Functionally equivalent to n (you are actually encouraged to use the family of \prg_new_conditional: functions described in Section 9.1).

p and w These are special cases.

The \cs_new: functions below (and friends) do not stop you from using other argument specifiers in your function names, but they do not handle expansion for you. You should define the base function and then use \cs_generate_variant:Nn to generate custom variants as described in Section 5.2.

4.3.2 Defining new functions using parameter text

\begin{verbatim}
\cs_new:Npn \cs_new:nopar:Npn \cs_new:cpn \cs_new:Npx \cs_new:cpx
\cs_new_protected:Npn \cs_new_protected:cpn \cs_new_protected:Npx \cs_new_protected:cpx
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \cs_new:cpn \cs_new:Npx \cs_new:cpx
\end{verbatim}

Creating (function) to expand to (code) as replacement text. Within the (code), the (parameters) (#1, #2, etc.) will be replaced by those absorbed by the function. The definition is global and an error results if the (function) is already defined.

\begin{verbatim}
\cs_new_nopar:Npn \cs_new_nopar:cpn \cs_new_nopar:Npx \cs_new_nopar:cpx
\end{verbatim}

Creating (function) to expand to (code) as replacement text. Within the (code), the (parameters) (#1, #2, etc.) will be replaced by those absorbed by the function. When the (function) is used the (parameters) absorbed cannot contain \par tokens. The definition is global and an error results if the (function) is already defined.

\begin{verbatim}
\cs_new_protected:Npn \cs_new_protected:cpn \cs_new_protected:Npx \cs_new_protected:cpx
\end{verbatim}

Creating (function) to expand to (code) as replacement text. Within the (code), the (parameters) (#1, #2, etc.) will be replaced by those absorbed by the function. The (function) will not expand within an x-type or e-type argument. The definition is global and an error results if the (function) is already defined.

\begin{verbatim}
\cs_new_protected_nopar:Npn \cs_new_protected_nopar:cpn \cs_new_protected_nopar:Npx \cs_new_protected_nopar:cpx
\end{verbatim}

Creating (function) to expand to (code) as replacement text. Within the (code), the (parameters) (#1, #2, etc.) will be replaced by those absorbed by the function. When the (function) is used the (parameters) absorbed cannot contain \par tokens. The (function) will not expand within an x-type or e-type argument. The definition is global and an error results if the (function) is already defined.
Globally sets (function) to expand to (code) as replacement text. Within the (code), the (parameters) (#1, #2, etc.) will be replaced by those absorbed by the function. The assignment of a meaning to the (function) is restricted to the current \TeX{} group level.

\texttt{\cs_set_protected:Npn (function) (parameters) \{}(code)\}\texttt{\} \}

\texttt{\cs_set_protected:cpn (function) (parameters) \{}(code)\}\texttt{\} \}

\texttt{\cs_set_protected:Npx (function) (parameters) \{}(code)\}\texttt{\} \}

\texttt{\cs_set_protected:cpn (function) (parameters) \{}(code)\}\texttt{\} \}

\texttt{\cs_set_protected:cpn (function) (parameters) \{}(code)\}\texttt{\} \}

Globally sets (function) to expand to (code) as replacement text. Within the (code), the (parameters) (#1, #2, etc.) will be replaced by those absorbed by the function. When the (function) is used the (parameters) absorbed cannot contain \texttt{\par} tokens. The assignment of a meaning to the (function) is restricted to the current \TeX{} group level. The (function) will not expand within an x-type or e-type argument.

\texttt{\cs_gset:Npn (function) (parameters) \{}(code)\}\texttt{\} \}

\texttt{\cs_gset:cpn (function) (parameters) \{}(code)\}\texttt{\} \}

\texttt{\cs_gset:Npx (function) (parameters) \{}(code)\}\texttt{\} \}

\texttt{\cs_gset:cpn (function) (parameters) \{}(code)\}\texttt{\} \}

\texttt{\cs_gset:Npx (function) (parameters) \{}(code)\}\texttt{\} \}

\texttt{\cs_gset:cpn (function) (parameters) \{}(code)\}\texttt{\} \}

Globally sets (function) to expand to (code) as replacement text. Within the (code), the (parameters) (#1, #2, etc.) will be replaced by those absorbed by the function. When the (function) is used the (parameters) absorbed cannot contain \texttt{\par} tokens. The assignment of a meaning to the (function) is not restricted to the current \TeX{} group level: the assignment is global.

\texttt{\cs_gset_protected:Npn (function) (parameters) \{}(code)\}\texttt{\} \}

\texttt{\cs_gset_protected:cpn (function) (parameters) \{}(code)\}\texttt{\} \}

\texttt{\cs_gset_protected:Npx (function) (parameters) \{}(code)\}\texttt{\} \}

\texttt{\cs_gset_protected:cpn (function) (parameters) \{}(code)\}\texttt{\} \}

\texttt{\cs_gset_protected:Npx (function) (parameters) \{}(code)\}\texttt{\} \}

Globally sets (function) to expand to (code) as replacement text. Within the (code), the (parameters) (#1, #2, etc.) will be replaced by those absorbed by the function. The assignment of a meaning to the (function) is not restricted to the current \TeX{} group level: the assignment is global. The (function) will not expand within an x-type or e-type argument.
\cs_set: Nn \cs_set: (cn|Nx|cx)

Sets \textit{(function)} to expand to \textit{(code)} as replacement text. Within the \textit{(code)}, the number of \textit{(parameters)} is detected automatically from the function signature. These \textit{(parameters)} \((#1, #2, \text{etc.})\) will be replaced by those absorbed by the function. The assignment of a meaning to the \textit{(function)} is restricted to the current \TeX{} group level: the assignment is global. The \textit{(function)} will not expand within an x-type or e-type argument.

\subsection*{4.3.3 Defining new functions using the signature}

\begin{itemize}
\item \textbackslash cs\new: Nn \textbackslash cs\new: (cn|Nx|cx)
\end{itemize}

\begin{itemize}
\item \textbackslash cs\new_nopar: Nn \textbackslash cs\new_nopar: (cn|Nx|cx)
\end{itemize}

\begin{itemize}
\item \textbackslash cs\new_protected: Nn \textbackslash cs\new_protected: (cn|Nx|cx)
\end{itemize}

\begin{itemize}
\item \textbackslash cs\new_protected_nopar: Nn \textbackslash cs\new_protected_nopar: (cn|Nx|cx)
\end{itemize}

\begin{itemize}
\item \textbackslash cs\gset_protected_nopar: cpx \textbackslash cs\gset_protected_nopar: Npx
\end{itemize}

Globally sets \textit{(function)} to expand to \textit{(code)} as replacement text. Within the \textit{(code)}, the \textit{(parameters)} \((#1, #2, \text{etc.})\) will be replaced by those absorbed by the function. When the \textit{(function)} is used the \textit{(parameters)} absorbed cannot contain \texttt{par} tokens. The assignment of a meaning to the \textit{(function)} is not restricted to the current \TeX{} group level: the assignment is global. The \textit{(function)} will not expand within an x-type or e-type argument.
\cs_set_protected:Nn
\cs_set_protected:(cn|Nx|cx)
\cs_set_protected:Nn \langle \text{function} \rangle \{\langle \text{code} \rangle \}

Sets \langle \text{function} \rangle to expand to \langle \text{code} \rangle as replacement text. Within the \langle \text{code} \rangle, the number of \langle \text{parameters} \rangle is detected automatically from the function signature. These \langle \text{parameters} \rangle (\#1, \#2, etc.) will be replaced by those absorbed by the function. When the \langle \text{function} \rangle is used the \langle \text{parameters} \rangle absorbed cannot contain \texttt{par} tokens. The assignment of a meaning to the \langle \text{function} \rangle is restricted to the current \TeX{} group level.

\cs_set_protected_nopar:Nn
\cs_set_protected_nopar:(cn|Nx|cx)
\cs_set_protected_nopar:Nn \langle \text{function} \rangle \{\langle \text{code} \rangle \}

Sets \langle \text{function} \rangle to expand to \langle \text{code} \rangle as replacement text. Within the \langle \text{code} \rangle, the number of \langle \text{parameters} \rangle is detected automatically from the function signature. These \langle \text{parameters} \rangle (\#1, \#2, etc.) will be replaced by those absorbed by the function. The \langle \text{function} \rangle will not expand within an \textit{x}-type or \textit{e}-type argument. The assignment of a meaning to the \langle \text{function} \rangle is restricted to the current \TeX{} group level.

\cs_set_protected_nopar:cn
\cs_set_protected_nopar:cn \langle \text{function} \rangle \{\langle \text{code} \rangle \}

Sets \langle \text{function} \rangle to expand to \langle \text{code} \rangle as replacement text. Within the \langle \text{code} \rangle, the number of \langle \text{parameters} \rangle is detected automatically from the function signature. These \langle \text{parameters} \rangle (\#1, \#2, etc.) will be replaced by those absorbed by the function. The \langle \text{function} \rangle will not expand within an \textit{x}-type or \textit{e}-type argument. The assignment of a meaning to the \langle \text{function} \rangle is restricted to the current \TeX{} group level.

\cs_gset:Nn
\cs_gset:(cn|Nx|cx)
\cs_gset:Nn \langle \text{function} \rangle \{\langle \text{code} \rangle \}

Sets \langle \text{function} \rangle to expand to \langle \text{code} \rangle as replacement text. Within the \langle \text{code} \rangle, the number of \langle \text{parameters} \rangle is detected automatically from the function signature. These \langle \text{parameters} \rangle (\#1, \#2, etc.) will be replaced by those absorbed by the function. The assignment of a meaning to the \langle \text{function} \rangle is global.

\cs_gset_nopar:Nn
\cs_gset_nopar:(cn|Nx|cx)
\cs_gset_nopar:Nn \langle \text{function} \rangle \{\langle \text{code} \rangle \}

Sets \langle \text{function} \rangle to expand to \langle \text{code} \rangle as replacement text. Within the \langle \text{code} \rangle, the number of \langle \text{parameters} \rangle is detected automatically from the function signature. These \langle \text{parameters} \rangle (\#1, \#2, etc.) will be replaced by those absorbed by the function. When the \langle \text{function} \rangle is used the \langle \text{parameters} \rangle absorbed cannot contain \texttt{par} tokens. The assignment of a meaning to the \langle \text{function} \rangle is global.

\cs_gset_protected:Nn
\cs_gset_protected:(cn|Nx|cx)
\cs_gset_protected:Nn \langle \text{function} \rangle \{\langle \text{code} \rangle \}

Sets \langle \text{function} \rangle to expand to \langle \text{code} \rangle as replacement text. Within the \langle \text{code} \rangle, the number of \langle \text{parameters} \rangle is detected automatically from the function signature. These \langle \text{parameters} \rangle (\#1, \#2, etc.) will be replaced by those absorbed by the function. The \langle \text{function} \rangle will not expand within an \textit{x}-type or \textit{e}-type argument. The assignment of a meaning to the \langle \text{function} \rangle is global.
Sets \textit{function} to expand to \textit{code} as replacement text. Within the \textit{code}, the number of \textit{parameters} is detected automatically from the function signature. These \textit{parameters} (#1, #2, \textit{etc.}) will be replaced by those absorbed by the function. When the \textit{function} is used the \textit{parameters} absorbed cannot contain \texttt{par} tokens. The \textit{function} will not expand within an \texttt{x}-type or \texttt{e}-type argument. The assignment of a meaning to the \textit{function} is global.

\begin{verbatim}
\cs_gset_protected_nopar:Nn \cs_gset_protected_nopar:cn
\cs_gset_protected_nopar:cn (cn|Nx|cx)
\end{verbatim}

\begin{verbatim}
\cs_generate_from_arg_count:NNnn \cs_generate_from_arg_count:NNnn (cNnn|Ncnn|cc)
\end{verbatim}

Updated: 2012-01-14

Uses the \textit{creator} function (which should have signature \texttt{Npn}, for example \texttt{\cs_new:Npn}) to define a \textit{function} which takes \textit{number} arguments and has \textit{code} as replacement text. The \textit{number} of arguments is an integer expression, evaluated as detailed for \texttt{\int_eval:n}.

\section*{4.3.4 Copying control sequences}

Control sequences (not just functions as defined above) can be set to have the same meaning using the functions described here. Making two control sequences equivalent means that the second control sequence is a \textit{copy} of the first (rather than a pointer to it). Thus the old and new control sequence are not tied together: changes to one are not reflected in the other.

In the following text “cs” is used as an abbreviation for “control sequence”.

\begin{verbatim}
\cs_new_eq:NN \cs_new_eq:(Nc|cN|cc)
\cs_set_eq:NN \cs_set_eq:(Nc|cN|cc)
\cs_gset_eq:NN \cs_gset_eq:(Nc|cN|cc)
\end{verbatim}

Globally creates \textit{control sequence} and sets it to have the same meaning as \textit{control sequence} or \textit{token}. The second control sequence may subsequently be altered without affecting the copy.

Sets \textit{control sequence} to have the same meaning as \textit{control sequence} (or \textit{token}). The second control sequence may subsequently be altered without affecting the copy. The assignment of a meaning to the \textit{control sequence} is restricted to the current \TeX group level.

Globally sets \textit{control sequence} to have the same meaning as \textit{control sequence} (or \textit{token}). The second control sequence may subsequently be altered without affecting the copy. The assignment of a meaning to the \textit{control sequence} is \textit{not} restricted to the current \TeX group level: the assignment is global.
4.3.5 Deleting control sequences

There are occasions where control sequences need to be deleted. This is handled in a very simple manner.

\cs_undefine:N \langle control sequence \rangle

Sets \langle control sequence \rangle to be globally undefined.

4.3.6 Showing control sequences

\cs_meaning:N \langle control sequence \rangle

This function expands to the meaning of the \langle control sequence \rangle control sequence. For a macro, this includes the \langle replacement text \rangle.

\TeXhackers note: This is \TeX's \meaning primitive. For tokens that are not control sequences, it is more logical to use \token_to_meaning:N. The c variant correctly reports undefined arguments.

\cs_show:N \langle control sequence \rangle

Displays the definition of the \langle control sequence \rangle on the terminal.

\TeXhackers note: This is similar to the \TeX primitive \show, wrapped to a fixed number of characters per line.

\cs_log:N \langle control sequence \rangle

Writes the definition of the \langle control sequence \rangle in the log file. See also \cs_show:N which displays the result in the terminal.

4.3.7 Converting to and from control sequences

\use:c \langle control sequence name \rangle

Expands the \langle control sequence name \rangle until only characters remain, and then converts this into a control sequence. This process requires two expansions. As in other c-type arguments the \langle control sequence name \rangle must, when fully expanded, consist of character tokens, typically a mixture of category code 10 (space), 11 (letter) and 12 (other).

As an example of the \use:c function, both

\use:c \{ a b c \}

and

\tl_new:N \l_my_tl
\tl_set:Nn \l_my_tl \{ a b c \}
\use:c \{ \tl_use:N \l_my_tl \}
would be equivalent to
\abc
after two expansions of \use:c.

\cs_if_exist_use:N * \cs_if_exist_use:c * \cs_if_exist_use:NTF * \cs_if_exist_use:cTF *
New: 2012-11-10

\cs_if_exist_use:N \cs_if_exist_use:c \cs_if_exist_use:NTF \cs_if_exist_use:cTF
Tests whether the \textit{control sequence} is currently defined according to the conditional
\cs_if_exist:NTF (whether as a function or another control sequence type), and if it is
inserts the \textit{control sequence} into the input stream followed by the \textit{true code}. Otherwise
the \textit{false code} is used.

\cs:w \cs_end: *
\cs:w \textit{control sequence name} \cs_end:
Converts the given \textit{control sequence name} into a single control sequence token. This
process requires one expansion. The content for \textit{control sequence name} may be literal
material or from other expandable functions. The \textit{control sequence name} must, when
fully expanded, consist of character tokens which are not active: typically of category
code 10 (space), 11 (letter) or 12 (other), or a mixture of these.

\TeXhacksnote: These are the \TeX primitives \csname and \endcsname.

As an example of the \texttt{cs:w} and \texttt{cs_end:} functions, both
\texttt{cs:w \{ \texttt{a b c} \}} \texttt{cs_end:}
and
\texttt{tl_new:N \_my_tl}
\texttt{tl_set:Nn \_my_tl { a b c }}
\texttt{cs:w \texttt{tl_use:N \_my_tl \cs_end:}}
would be equivalent to
\abc
after one expansion of \texttt{cs:w}.

\cs_to_str:N *
\cs_to_str:N \textit{control sequence}
Converts the given \textit{control sequence} into a series of characters with category code 12
(other), except spaces, of category code 10. The result does \textit{not} include the current
escape token, contrarily to \texttt{token_to_str:N}. Full expansion of this function requires
exactly 2 expansion steps, and so an x-type or e-type expansion, or two o-type expansions
are required to convert the \textit{control sequence} to a sequence of characters in the input
stream. In most cases, an f-expansion is correct as well, but this loses a space at the
start of the result.
4.4 Analysing control sequences

\cs_split_function:N \langle function \rangle

Splits the \langle function \rangle into the \langle name \rangle (i.e. the part before the colon) and the \langle signature \rangle (i.e. after the colon). This information is then placed in the input stream in three parts: the \langle name \rangle, the \langle signature \rangle and a logic token indicating if a colon was found (to differentiate variables from function names). The \langle name \rangle does not include the escape character, and both the \langle name \rangle and \langle signature \rangle are made up of tokens with category code 12 (other).

The next three functions decompose \LaTeX{} macros into their constituent parts: if the \langle token \rangle passed is not a macro then no decomposition can occur. In the latter case, all three functions leave \texttt{\textbackslash scan\_stop} in the input stream.

\cs_prefix_spec:N \langle token \rangle

If the \langle token \rangle is a macro, this function leaves the applicable \LaTeX{} prefixes in input stream as a string of tokens of category code 12 (with spaces having category code 10). Thus for example

\begin{verbatim}
\cs_set:Npn \next:nn \#1\#2 \{ x \#1-y \#2 \}
\cs_prefix_spec:N \next:nn
\end{verbatim}

leaves \texttt{\textbackslash long} in the input stream. If the \langle token \rangle is not a macro then \texttt{\textbackslash scan\_stop}: is left in the input stream.

\textbf{\LaTeX{}hackers note:} The prefix can be empty, \texttt{\textbackslash long}, \texttt{\textbackslash protected} or \texttt{\textbackslash protected\textbackslash long} with backslash replaced by the current escape character.

\cs_argument_spec:N \langle token \rangle

If the \langle token \rangle is a macro, this function leaves the primitive \LaTeX{} argument specification in input stream as a string of character tokens of category code 12 (with spaces having category code 10). Thus for example

\begin{verbatim}
\cs_set:Npn \next:nn \#1\#2 \{ x \#1 y \#2 \}
\cs_argument_spec:N \next:nn
\end{verbatim}

leaves \#1\#2 in the input stream. If the \langle token \rangle is not a macro then \texttt{\textbackslash scan\_stop}: is left in the input stream.

\textbf{\LaTeX{}hackers note:} If the argument specification contains the string -\rightarrow, then the function produces incorrect results.
\texttt{\cs_replacement_spec:N} \texttt{⟨token⟩}

If the \texttt{⟨token⟩} is a macro, this function leaves the replacement text in input stream as a string of character tokens of category code 12 (with spaces having category code 10).

Thus for example

\begin{verbatim}
\cs_set:Npn \next:nn #1 #2 { x #1 y #2 }
\cs_replacement_spec:N \next:nn
\end{verbatim}

leaves \texttt{x#1 \textbackslash{} y#2} in the input stream. If the \texttt{⟨token⟩} is not a macro then \texttt{\scan_stop:} is left in the input stream.

\textbf{\LaTeX}hackers note: If the argument specification contains the string ->, then the function produces incorrect results.

4.5 Using or removing tokens and arguments

Tokens in the input can be read and used or read and discarded. If one or more tokens are wrapped in braces then when absorbing them the outer set is removed. At the same time, the category code of each token is set when the token is read by a function (if it is read more than once, the category code is determined by the situation in force when first function absorbs the token).

\begin{verbatim}
\use:n \{⟨group1⟩\}
\use:nn \{⟨group1⟩\} \{⟨group2⟩\}
\use:nnn \{⟨group1⟩\} \{⟨group2⟩\} \{⟨group3⟩\}
\use:nnnn \{⟨group1⟩\} \{⟨group2⟩\} \{⟨group3⟩\} \{⟨group4⟩\}
\end{verbatim}

As illustrated, these functions absorb between one and four arguments, as indicated by the argument specifier. The braces surrounding each argument are removed and the remaining tokens are left in the input stream. The category code of these tokens is also fixed by this process (if it has not already been by some other absorption). All of these functions require only a single expansion to operate, so that one expansion of

\begin{verbatim}
\use:nn \{ abc \} \{ \{ def \} \}
\end{verbatim}

results in the input stream containing

\begin{verbatim}
abc \{ def \}
\end{verbatim}

\textit{i.e.} only the outer braces are removed.

\textbf{\LaTeX}hackers note: The \texttt{\use:n} function is equivalent to \LaTeX{}2ε's \texttt{\@firstofone}.  

23
\use_i:nn * \use_i:nn \{\texttt{arg}_1\} \{\texttt{arg}_2\}

These functions absorb two arguments from the input stream. The function \use_i:nn discards the second argument, and leaves the content of the first argument in the input stream. \use_i:nn discards the first argument and leaves the content of the second argument in the input stream. The category code of these tokens is also fixed (if it has not already been by some other absorption). A single expansion is needed for the functions to take effect.

\textbf{\LaTeX} hackers note: These are equivalent to \LaTeX's \texttt{@firstoftwo} and \texttt{@secondoftwo}.

\use_i:nnn * \use_i:nnn \{\texttt{arg}_1\} \{\texttt{arg}_2\} \{\texttt{arg}_3\}

These functions absorb three arguments from the input stream. The function \use_i:nnn discards the second and third arguments, and leaves the content of the first argument in the input stream. \use_i:nnn and \use_i:nnn work similarly, leaving the content of second or third arguments in the input stream, respectively. The category code of these tokens is also fixed (if it has not already been by some other absorption). A single expansion is needed for the functions to take effect.

\use_i:nnnn * \use_i:nnnn \{\texttt{arg}_1\} \{\texttt{arg}_2\} \{\texttt{arg}_3\} \{\texttt{arg}_4\}

These functions absorb four arguments from the input stream. The function \use_i:nnnn discards the second, third and fourth arguments, and leaves the content of the first argument in the input stream. \use_i:nnnn, \use_i:nnnn and \use_i:nnnn work similarly, leaving the content of second, third or fourth arguments in the input stream, respectively. The category code of these tokens is also fixed (if it has not already been by some other absorption). A single expansion is needed for the functions to take effect.

\use_i:ii:nnn * \use_i:ii:nnn \{\texttt{arg}_1\} \{\texttt{arg}_2\} \{\texttt{arg}_3\}

This function absorbs three arguments and leaves the content of the first and second in the input stream. The category code of these tokens is also fixed (if it has not already been by some other absorption). A single expansion is needed for the function to take effect. An example:

\begin{verbatim}
\use_i:ii:nnn \{\texttt{abc}\} \{\texttt{\{def\}\}} \{\texttt{ghi}\}
\end{verbatim}

results in the input stream containing

```
abc { def }
```

\textit{i.e.} the outer braces are removed and the third group is removed.

\use_ii:i:nnn * \use_ii:i:nn \{\texttt{arg}_1\} \{\texttt{arg}_2\}

This function absorbs two arguments and leaves the content of the second and first in the input stream. The category code of these tokens is also fixed (if it has not already been by some other absorption). A single expansion is needed for the function to take effect.

\textit{New: 2019-06-02}
These functions absorb between one and nine groups from the input stream, leaving nothing on the resulting input stream. These functions work after a single expansion. One or more of the \( n \) arguments may be an unbraced single token (i.e. an \( N \) argument).

\textbf{TeXhackers note:} These are equivalent to \LaTeX 2e’s \texttt{\textbackslash gobble}, \texttt{\textbackslash gobbbletwo}, etc.

\begin{itemize}
  \item \use:e \{\texttt{\textbackslash expandable tokens}\}
  \begin{itemize}
    \item Fully expands the \{\texttt{token list}\} in an \texttt{x}-type manner, \textit{but} the function remains fully expandable, and parameter character (usually \#) need not be doubled.
    \item \textbf{TeXhackers note:} \use:e is a wrapper around the primitive \texttt{\textbackslash expanded} where it is available: it requires two expansions to complete its action. When \texttt{\textbackslash expanded} is not available this function is very slow.
  \end{itemize}
\end{itemize}

\section{Selecting tokens from delimited arguments}

A different kind of function for selecting tokens from the token stream are those that use delimited arguments.

\begin{itemize}
  \item \use_i_delimit_by_q_nil:nw \{\texttt{\textbackslash inserted tokens}\} \{\texttt{balanced text}\} \q_nil
  \item \use_i_delimit_by_q_stop:nw \{\texttt{\textbackslash inserted tokens}\} \{\texttt{balanced text}\} \q_stop
  \item \use_i_delimit_by_q_recursion_stop:nw \{\texttt{\textbackslash inserted tokens}\} \{\texttt{balanced text}\} \q_recursion_stop
\end{itemize}

Absorb the \{\texttt{balanced text}\} from the input stream delimited by the marker given in the function name, leaving nothing in the input stream.

\begin{itemize}
  \item \use_i_delimit_by_q_nil:nw \{\texttt{\textbackslash inserted tokens}\} \{\texttt{balanced text}\} \q_nil
  \item \use_i_delimit_by_q_stop:nw \{\texttt{\textbackslash inserted tokens}\} \{\texttt{balanced text}\} \q_stop
  \item \use_i_delimit_by_q_recursion_stop:nw \{\texttt{\textbackslash inserted tokens}\} \{\texttt{balanced text}\} \q_recursion_stop
\end{itemize}

Absorb the \{\texttt{balanced text}\} from the input stream delimited by the marker given in the function name, leaving \{\texttt{inserted tokens}\} in the input stream for further processing.

\section{Predicates and conditionals}

\LaTeX 3 has three concepts for conditional flow processing:
Branching conditionals Functions that carry out a test and then execute, depending on its result, either the code supplied as the (true code) or the (false code). These arguments are denoted with \texttt{T} and \texttt{F}, respectively. An example would be

\begin{verbatim}
\cs_if_free:cTF \{abc\} \{\langle true\ code\ \rangle\} \{\langle false\ code\ \rangle\}
\end{verbatim}

a function that turns the first argument into a control sequence (since it’s marked as \texttt{c}) then checks whether this control sequence is still free and then depending on the result carries out the code in the second argument (true case) or in the third argument (false case).

These type of functions are known as “conditionals”; whenever a \texttt{TF} function is defined it is usually accompanied by \texttt{T} and \texttt{F} functions as well. These are provided for convenience when the branch only needs to go a single way. Package writers are free to choose which types to define but the kernel definitions always provide all three versions.

Important to note is that these branching conditionals with \texttt{\langle true\ code\ \rangle} and/or \texttt{\langle false\ code\ \rangle} are always defined in a way that the code of the chosen alternative can operate on following tokens in the input stream.

These conditional functions may or may not be fully expandable, but if they are expandable they are accompanied by a “predicate” for the same test as described below.

Predicates “Predicates” are functions that return a special type of boolean value which can be tested by the boolean expression parser. All functions of this type are expandable and have names that end with \texttt{_p} in the description part. For example,

\begin{verbatim}
\cs_if_free_p:N
\end{verbatim}

would be a predicate function for the same type of test as the conditional described above. It would return “true” if its argument (a single token denoted by \texttt{N}) is still free for definition. It would be used in constructions like

\begin{verbatim}
\bool_if:nTF {
   \cs_if_free_p:N \l_tmpz_tl || \cs_if_free_p:N \g_tmpz_tl
} {\langle true\ code\ \rangle} {\langle false\ code\ \rangle}
\end{verbatim}

For each predicate defined, a “branching conditional” also exists that behaves like a conditional described above.

Primitive conditionals There is a third variety of conditional, which is the original concept used in plain \TeX and \LaTeX2\epsilon. Their use is discouraged in expl3 (although still used in low-level definitions) because they are more fragile and in many cases require more expansion control (hence more code) than the two types of conditionals described above.

\begin{verbatim}
\c_true_bool \c_false_bool
\end{verbatim}

Constants that represent \texttt{true} and \texttt{false}, respectively. Used to implement predicates.
4.6.1 Tests on control sequences

\cs_if_eq_p:NN \cs_if_eq_p:NN \langle cs_1 \rangle \langle cs_2 \rangle
\cs_if_eq:NNTF \langle cs_1 \rangle \langle cs_2 \rangle \{\langle true code\rangle\} \{\langle false code\rangle\}

Compares the definition of two \langle control sequences \rangle and is logically \texttt{true} if they are the same, \textit{i.e.} if they have exactly the same definition when examined with \texttt{cs_show:N}.

\cs_if_exist_p:N \cs_if_exist:NTF \langle control sequence \rangle \{\langle true code\rangle\} \{\langle false code\rangle\}
\cs_if_exist_p:c \cs_if_exist:c \langle control sequence \rangle \langle true code\rangle \langle false code\rangle

Tests whether the \langle control sequence \rangle is currently defined (whether as a function or another control sequence type). Any definition of \langle control sequence \rangle other than \relax evaluates as \texttt{true}.

\cs_if_free_p:N \cs_if_free:NTF \langle control sequence \rangle \{\langle true code\rangle\} \{\langle false code\rangle\}
\cs_if_free_p:c \cs_if_free:c \langle control sequence \rangle \langle true code\rangle \langle false code\rangle

Tests whether the \langle control sequence \rangle is currently free to be defined. This test is \texttt{false} if the \langle control sequence \rangle currently exists (as defined by \texttt{cs_if_exist:NTF}).

4.6.2 Primitive conditionals

The \TeX engine itself provides many different conditionals. Some expand whatever comes after them and others don’t. Hence the names for these underlying functions often contains a \texttt{w} part but higher level functions are often available. See for instance \texttt{int_compare_p:nNn} which is a wrapper for \texttt{if_int_compare:w}.

Certain conditionals deal with specific data types like boxes and fonts and are described there. The ones described below are either the universal conditionals or deal with control sequences. We prefix primitive conditionals with \texttt{if_}.

\if_true: \if_true: \langle true code\rangle \else: \langle false code\rangle \fi:
\if_false: \if_false: \langle true code\rangle \else: \langle false code\rangle \fi:
\else: \else: \reverse_if:N \langle primitive conditional\rangle
\fi: \reverse_if:N
\reverse_if:N \if_true: always executes \langle true code\rangle, while \if_false: always executes \langle false code\rangle. \reverse_if:N reverses any two-way primitive conditional. \else: and \fi: delimit the branches of the conditional. The function \texttt{or} is documented in \texttt{int} and used in case switches.

\TeXhackers note: These are equivalent to their corresponding \TeX primitive conditionals; \reverse_if:N is \TeX’s \texttt{unless}.

\if_meaning:w \langle arg_1 \rangle \langle arg_2 \rangle \langle true code\rangle \else: \langle false code\rangle \fi:
\if_meaning:w \langle arg_1 \rangle \langle arg_2 \rangle \langle true code\rangle when \langle arg_1 \rangle and \langle arg_2 \rangle are the same, otherwise it executes \langle false code\rangle. \langle arg_1 \rangle and \langle arg_2 \rangle could be functions, variables, tokens; in all cases the \texttt{unexpanded} definitions are compared.

\TeXhackers note: This is \TeX’s \texttt{ifx}.
These conditionals expand any following tokens until two unexpandable tokens are left. If you wish to prevent this expansion, prefix the token in question with \exp_not:N. \if_catcode:w tests if the category codes of the two tokens are the same whereas \if:w tests if the character codes are identical. \if_charcode:w is an alternative name for \if:w.

\if_cs_exist:N \if_cs_exist:w

Check if \langle cs \rangle appears in the hash table or if the control sequence that can be formed from \langle tokens \rangle appears in the hash table. The latter function does not turn the control sequence in question into \scan_stop:! This can be useful when dealing with control sequences which cannot be entered as a single token.

\if_mode_horizontal: \if_mode_vertical: \if_mode_math: \if_mode_inner:

 Execute \langle true code \rangle if currently in horizontal mode, otherwise execute \langle false code \rangle. Similar for the other functions.

### 4.7 Starting a paragraph

\mode_leave_vertical:

Ensures that \TeX{} is not in vertical (inter-paragraph) mode. In horizontal or math mode this command has no effect, in vertical mode it switches to horizontal mode, and inserts a box of width \parindent, followed by the \everypar token list.

\TeX{}hackers note: This results in the contents of the \everypar token register being inserted, after \mode_leave_vertical: is complete. Notice that in contrast to the \TeX{}2e \leavevmode approach, no box is used by the method implemented here.
4.8 Debugging support

\debug_on:n { \langle comma-separated list \rangle }
\debug_off:n { \langle comma-separated list \rangle }

Turn on and off within a group various debugging code, some of which is also available as expl3 load-time options. The items that can be used in the \langle list \rangle are

- **check-declarations** that checks all expl3 variables used were previously declared and that local/global variables (based on their name or on their first assignment) are only locally/globally assigned;
- **check-expressions** that checks integer, dimension, skip, and muskip expressions are not terminated prematurely;
- **deprecation** that makes soon-to-be-deprecated commands produce errors;
- **log-functions** that logs function definitions;
- **all** that does all of the above.

Providing these as switches rather than options allows testing code even if it relies on other packages: load all other packages, call \debug_on:n, and load the code that one is interested in testing. These functions can only be used in \LaTeXe package mode loaded with \enabledebug or another option implying it.

\debug_suspend: ... \debug_resume:

Suppress (locally) errors and logging from debug commands, except for the deprecation errors or warnings. These pairs of commands can be nested. This can be used around pieces of code that are known to fail checks, if such failures should be ignored. See for instance \l3coffins.
Chapter 5

The l3expan package
Argument expansion

This module provides generic methods for expanding \TeX arguments in a systematic manner. The functions in this module all have prefix `\exp`.

Not all possible variations are implemented for every base function. Instead only those that are used within the \L3\TeX{} kernel or otherwise seem to be of general interest are implemented. Consult the module description to find out which functions are actually defined. The next section explains how to define missing variants.

5.1 Defining new variants

The definition of variant forms for base functions may be necessary when writing new functions or when applying a kernel function in a situation that we haven’t thought of before.

Internally preprocessing of arguments is done with functions of the form `\exp_- ...`. They all look alike; an example would be `\exp_args:NNo`. This function has three arguments, the first and the second are a single tokens, while the third argument should be given in braces. Applying `\exp_args:NNo` expands the content of third argument once before any expansion of the first and second arguments. If `\seq_gpush:No` was not defined it could be coded in the following way:

\begin{verbatim}
\exp_args:NNo \seq_gpush:Nn \\
\g_file_name_stack \\
{ \l_tmpa_tl }
\end{verbatim}

In other words, the first argument to `\exp_args:NNo` is the base function and the other arguments are preprocessed and then passed to this base function. In the example the first argument to the base function should be a single token which is left unchanged while the second argument is expanded once. From this example we can also see how the variants are defined. They just expand into the appropriate `\exp_` function followed by the desired base function, \textit{e.g.}:

\begin{verbatim}
\cs_generate_variant:Nn \seq_gpush:Nn { No }
\end{verbatim}

results in the definition of `\seq_gpush:No`
Providing variants in this way in style files is safe as the \texttt{\cs Generate Variant:Nn} function will only create new definitions if there is not already one available. Therefore adding such definition to later releases of the kernel will not make such style files obsolete.

The steps above may be automated by using the function \texttt{cs Generate Variant:Nn}, described next.

## 5.2 Methods for defining variants

We recall the set of available argument specifiers.

- \texttt{N} is used for single-token arguments while \texttt{c} constructs a control sequence from its name and passes it to a parent function as an \texttt{N}-type argument.

- Many argument types extract or expand some tokens and provide it as an \texttt{n}-type argument, namely a braced multiple-token argument: \texttt{V} extracts the value of a variable, \texttt{v} extracts the value from the name of a variable, \texttt{n} uses the argument as it is, \texttt{o} expands once, \texttt{f} expands fully the front of the token list, \texttt{e} and \texttt{x} expand fully all tokens (differences are explained later).

- A few odd argument types remain: \texttt{T} and \texttt{F} for conditional processing, otherwise identical to \texttt{n}-type arguments, \texttt{p} for the parameter text in definitions, \texttt{w} for arguments with a specific syntax, and \texttt{D} to denote primitives that should not be used directly.
This function is used to define argument-specifier variants of the \textit{(parent control sequence)} for \LaTeX3 code-level macros. The \textit{(parent control sequence)} is first separated into the \textit{(base name)} and \textit{(original argument specifier)}. The comma-separated list of \textit{(variant argument specifiers)} is then used to define variants of the \textit{(original argument specifier)} if these are not already defined. For each \textit{(variant)} given, a function is created that expands its arguments as detailed and passes them to the \textit{(parent control sequence)}. So for example

\begin{verbatim}
\cs_set:Npn \foo:Nn \texttt{#1#2} \{ \texttt{code here} \}
\cs_generate_variant:Nn \foo:Nn \{ \texttt{c} \}
\end{verbatim}

creates a new function \texttt{\foo:cn} which expands its first argument into a control sequence name and passes the result to \texttt{\foo:Nn}. Similarly

\begin{verbatim}
\cs_generate_variant:Nn \foo:Nn \{ \texttt{NV} , \texttt{cV} \}
\end{verbatim}

generates the functions \texttt{\foo:NV} and \texttt{\foo:cV} in the same way. The \texttt{\texttt{cs\_generate\_variant}:Nn} function can only be applied if the \textit{(parent control sequence)} is already defined. If the \textit{(parent control sequence)} is protected or if the \textit{(variant)} involves any \texttt{x} argument, then the \textit{(variant control sequence)} is also protected. The \textit{(variant)} is created globally, as is any \texttt{\exp_args:No} function needed to carry out the expansion.

Only \texttt{n} and \texttt{N} arguments can be changed to other types. The only allowed changes are

- \texttt{c} variant of an \texttt{N} parent;
- \texttt{o}, \texttt{V}, \texttt{v}, \texttt{f}, \texttt{e}, or \texttt{x} variant of an \texttt{n} parent;
- \texttt{N}, \texttt{n}, \texttt{T}, \texttt{F}, or \texttt{p} argument unchanged.

This means the \textit{(parent)} of a \textit{(variant)} form is always unambiguous, even in cases where both an \texttt{n}-type parent and an \texttt{N}-type parent exist, such as for \texttt{\tl\_count:n} and \texttt{\tl\_count:N}.

For backward compatibility it is currently possible to make \texttt{n}, \texttt{o}, \texttt{V}, \texttt{v}, \texttt{f}, \texttt{e}, or \texttt{x}-type variants of an \texttt{N}-type argument or \texttt{N} or \texttt{c}-type variants of an \texttt{n}-type argument. Both are deprecated. The first because passing more than one token to an \texttt{N}-type argument will typically break the parent function’s code. The second because programmers who use that most often want to access the value of a variable given its name, hence should use a \texttt{V}-type or \texttt{v}-type variant instead of \texttt{c}-type. In those cases, using the lower-level \texttt{\exp_args:No} or \texttt{\exp_args:Nc} functions explicitly is preferred to defining confusing variants.

### 5.3 Introduction the variants

The \texttt{V} type returns the value of a register, which can be one of \texttt{tl}, \texttt{clist}, \texttt{int}, \texttt{skip}, \texttt{dim}, \texttt{muskip}, or built-in \TeX registers. The \texttt{v} type is the same except it first creates a control sequence out of its argument before returning the value.

In general, the programmer should not need to be concerned with expansion control. When simply using the content of a variable, functions with a \texttt{V} specifier should be used. For those referred to by (cs)name, the \texttt{v} specifier is available for the same purpose. Only
when specific expansion steps are needed, such as when using delimited arguments, should the lower-level functions with o specifiers be employed.

The e type expands all tokens fully, starting from the first. More precisely the expansion is identical to that of \text{TEx}'s \texttt{\textbackslash message} (in particular \# needs not be doubled). It was added in May 2018. In recent enough engines (starting around 2019) it relies on the primitive \texttt{\textbackslash expanded} hence is fast. In older engines it is very much slower. As a result it should only be used in performance critical code if typical users will have a recent installation of the \text{TEx} ecosystem.

The x type expands all tokens fully, starting from the first. In contrast to e, all macro parameter characters \# must be doubled, and omitting this leads to low-level errors. In addition this type of expansion is not expandable, namely functions that have x in their signature do not themselves expand when appearing inside x or e expansion.

The f type is so special that it deserves an example. It is typically used in contexts where only expandable commands are allowed. Then x-expansion cannot be used, and f-expansion provides an alternative that expands the front of the token list as much as can be done in such contexts. For instance, say that we want to evaluate the integer expression $3 + 4$ and pass the result 7 as an argument to an expandable function \example:n. For this, one should define a variant using \texttt{\textbackslash cs\_generate\_variant:Nn} \example:n \{ f \}, then do

\example:f \{ \int_eval:n \{ 3 + 4 \} \}

Note that x-expansion would also expand \texttt{\textbackslash int\_eval:n} fully to its result 7, but the variant \example:x would not be expandable. Note also that o-expansion would not expand \texttt{\textbackslash int\_eval:n} fully to its result since that function requires several expansions. Besides the fact that x-expansion is protected rather than expandable, another difference between f-expansion and x-expansion is that f-expansion expands tokens from the beginning and stops as soon as a non-expandable token is encountered, while x-expansion continues expanding further tokens. Thus, for instance

\example:f \{ \int_eval:n \{ 1 + 2 \} , \int_eval:n \{ 3 + 4 \} \}

results in the call

\example:n \{ 3 , \int_eval:n \{ 3 + 4 \} \}

while using \example:x or \example:e instead results in

\example:n \{ 3 , 7 \}

at the cost of being protected (for x type) or very much slower in old engines (for e type). If you use f type expansion in conditional processing then you should stick to using TF type functions only as the expansion does not finish any \texttt{\textbackslash if... \textbackslash fi}: itself!

It is important to note that both f- and o-type expansion are concerned with the expansion of tokens from left to right in their arguments. In particular, o-type expansion applies to the first token in the argument it receives: it is conceptually similar to

\exp_after:wN <base function> \exp_after:wN \{ <argument> \}

At the same time, f-type expansion stops at the first non-expandable token. This means for example that both

\tl_set:No \l_tmpa_tl \{ \{ \g_tmpb_tl \} \}
leave \g_tmpb_tl unchanged: \{ is the first token in the argument and is non-expandable.

It is usually best to keep the following in mind when using variant forms.

- Variants with x-type arguments (that are fully expanded before being passed to the n-type base function) are never expandable even when the base function is. Such variants cannot work correctly in arguments that are themselves subject to expansion. Consider using f or e expansion.

- In contrast, e expansion (full expansion, almost like x except for the treatment of #) does not prevent variants from being expandable (if the base function is). The drawback is that e expansion is very much slower in old engines (before 2019). Consider using f expansion if that type of expansion is sufficient to perform the required expansion, or x expansion if the variant will not itself need to be expandable.

- Finally f expansion only expands the front of the token list, stopping at the first non-expandable token. This may fail to fully expand the argument.

When speed is essential (for functions that do very little work and whose variants are used numerous times in a document) the following considerations apply because internal functions for argument expansion come in two flavours, some faster than others.

- Arguments that might need expansion should come first in the list of arguments.

- Arguments that should consist of single tokens N, c, V, or v should come first among these.

- Arguments that appear after the first multi-token argument n, f, e, or o require slightly slower special processing to be expanded. Therefore it is best to use the optimized functions, namely those that contain only N, c, V, and v, and, in the last position, o, f, e, with possible trailing N or n or T or F, which are not expanded. Any x-type argument causes slightly slower processing.

5.4 Manipulating the first argument

These functions are described in detail: expansion of multiple tokens follows the same rules but is described in a shorter fashion.

\exp_args:Nc \exp_args:cc *\exp_args:Nc (function) {tokens}

This function absorbs two arguments (the (function) name and the (tokens)). The (tokens) are expanded until only characters remain, and are then turned into a control sequence. The result is inserted into the input stream after reinsertion of the (function). Thus the (function) may take more than one argument: all others are left unchanged.

The :cc variant constructs the (function) name in the same manner as described for the (tokens).
This function absorbs two arguments (the \textit{function} name and the \textit{tokens}). The \textit{tokens} are expanded once, and the result is inserted in braces into the input stream \textit{after} reinsertion of the \textit{function}. Thus the \textit{function} may take more than one argument: all others are left unchanged.

This function absorbs two arguments (the names of the \textit{function} and the \textit{variable}). The content of the \textit{variable} are recovered and placed inside braces into the input stream \textit{after} reinsertion of the \textit{function}. Thus the \textit{function} may take more than one argument: all others are left unchanged.

This function absorbs two arguments (the \textit{function} name and the \textit{tokens}). The \textit{tokens} are expanded until only characters remain, and are then turned into a control sequence. This control sequence should be the name of a \textit{variable}. The content of the \textit{variable} are recovered and placed inside braces into the input stream \textit{after} reinsertion of the \textit{function}. Thus the \textit{function} may take more than one argument: all others are left unchanged.

\textbf{\TeXhackers note:} This relies on the \texttt{\expanded} primitive when available (in \LaTeX{} and starting around 2019 in other engines). Otherwise it uses some fall-back code that is very much slower. As a result it should only be used in performance-critical code if typical users have a recent installation of the \TeX{} ecosystem.

This function absorbs two arguments (the \textit{function} name and the \textit{tokens}) and exhaustively expands the \textit{tokens}. The result is inserted in braces into the input stream \textit{after} reinsertion of the \textit{function}. Thus the \textit{function} may take more than one argument: all others are left unchanged.

This function absorbs two arguments (the \textit{function} name and the \textit{tokens}) and exhaustively expands the \textit{tokens}. The result is inserted in braces into the input stream \textit{after} reinsertion of the \textit{function}. Thus the \textit{function} may take more than one argument: all others are left unchanged.
5.5 Manipulating two arguments

These optimized functions absorb three arguments and expand the second and third as detailed by their argument specifier. The first argument of the function is then the next item on the input stream, followed by the expansion of the second and third arguments.

5.6 Manipulating three arguments

These optimized functions absorb four arguments and expand the second, third and fourth as detailed by their argument specifier. The first argument of the function is then the next item on the input stream, followed by the expansion of the second argument, etc.
These functions absorb four arguments and expand the second, third and fourth as detailed by their argument specifier. The first argument of the function is then the next item on the input stream, followed by the expansion of the second argument, etc. These functions need slower processing.

\exp_args:NNcf \star \exp_args:NNno \star \exp_args:NNV \star \exp_args:NNoo \star \exp_args:Ncno \star \exp_args:NCVV \star \exp_args:NcV \star \exp_args:Nncc \star \exp_args:Nnno \star \exp_args:Nnff \star \exp_args:Nnfo \star \exp_args:Nf \star \exp_args:NNo \star \exp_args:NV \star \exp_args:Nf \star \exp_args:NcV \star \exp_args:Nnno \star \exp_args:Nnff \star \exp_args:Nnfo \star \exp_args:Nf \star \exp_args:NNV \star \exp_args:NNf \star \exp_args:NcV \star \exp_args:Nnno \star \exp_args:NNV \star \exp_args:NNf \star \exp_args:Nnno \star \exp_args:NNNo \star \exp_args:NNNV \star \exp_args:NNNf \star \exp_args:Nnnf \star \exp_args:Nnff \star \exp_args:Nnfo \star \exp_args:Nf \star \exp_args:NNf \star \exp_args:Nnfo \star \exp_args:Nf \star \exp_args:NNV \star \exp_args:NNf \star \exp_args:Nnno \star \exp_args:NNV \star \exp_args:NNf \star \exp_args:Nnno \star \exp_args:NNNo \star \exp_args:NNNf \star

\exp_last_unbraced:No \star \exp_last_unbraced:NV \star \exp_last_unbraced:Ne \star \exp_last_unbraced:Nf \star \exp_last_unbraced:NNo \star \exp_last_unbraced:NNV \star \exp_last_unbraced:Ncv \star \exp_last_unbraced:Nco \star \exp_last_unbraced:Nf \star \exp_last_unbraced:NNV \star \exp_last_unbraced:Nff \star \exp_last_unbraced:Nfo \star \exp_last_unbraced:Nf \star \exp_last_unbraced:NNNo \star \exp_last_unbraced:NNV \star \exp_last_unbraced:Nff \star \exp_last_unbraced:Nfo \star \exp_last_unbraced:Nf \star

\exp_last_unbraced:Nno (token) \{ (tokens) \} \{ (tokens) \}

These functions absorb four arguments and expand the second, third and fourth as detailed by their argument specifier. The first argument of the function is then the next item on the input stream, followed by the expansion of the second argument, etc.

\exp_args:NNax (token) (token2) \{ (tokens1) \} \{ (tokens2) \}

These functions absorb four arguments and expand the second, third and fourth as detailed by their argument specifier. The first argument of the function is then the next item on the input stream, followed by the expansion of the second argument, etc.

5.7 Unbraced expansion

\exp_last_unbraced:Nno (token) \{ (tokens1) \} \{ (tokens2) \}

These functions absorb the number of arguments given by their specification, carry out the expansion indicated and leave the results in the input stream, with the last argument not surrounded by the usual braces. Of these, the :Nno, :Noo, :Nfo and :NnNo variants need slower processing.

**\TeX hackers note:** As an optimization, the last argument is unbraced by some of those functions before expansion. This can cause problems if the argument is empty: for instance, \exp_last_unbraced:Nf \\foo_bar:w \{ \} \q_stop leads to an infinite loop, as the quark is f-expanded.
\input{sections/58-expansion}

5.8 Preventing expansion

Despite the fact that the following functions are all about preventing expansion, they’re designed to be used in an expandable context and hence are all marked as being ‘expandable’ since they themselves disappear after the expansion has completed.

\exp_not:N \exp_not:N \{token\}

Prevents expansion of the \{token\} in a context where it would otherwise be expanded, for example an \texttt{x}-type argument or the first token in an \texttt{o} or \texttt{e} or \texttt{f} argument.

\TeXhacksnote{This is the \TeX \texttt{\noexpand} primitive. It only prevents expansion. At the beginning of an \texttt{f}-type argument, a space \{token\} is removed even if it appears as \texttt{\exp_not:N \c_space_token}. In an \texttt{x}-expanding definition (\texttt{\cs_new:NpX}), a macro parameter introduces an argument even if it appears as \texttt{\exp_not:N \# 1}. This differs from \texttt{\exp_not:n}.}

\exp_not:c \exp_not:c \{tokens\}

Expands the \{tokens\} until only characters remain, and then converts this into a control sequence. Further expansion of this control sequence is then inhibited using \exp_not:N.
\exp_not:n \{\text{tokens}\} 

Prevents expansion of the \text{tokens} in an \text{e} or \text{x}-type argument. In all other cases the \text{tokens} continue to be expanded, for example in the input stream or in other types of arguments such as \text{c}, \text{f}, \text{v}. The argument of \exp_not:n must be surrounded by braces.

\textbf{\textsc{TeX}hackers note:} This is the \texttt{\unexpanded} primitive. In an \text{x}-expanding definition (\texttt{\cs_new:Npx}), \exp_not:n \{\#1\} is equivalent to ##1 rather than to #1, namely it inserts the two characters # and 1. In an \text{e}-type argument \exp_not:n \{\#\} is equivalent to #, namely it inserts the character #.

\exp_not:o \{\text{tokens}\} 

Expands the \text{tokens} once, then prevents any further expansion in \text{x}-type or \text{e}-type arguments using \exp_not:n.

\exp_not:V \{\text{variable}\} 

Recovers the content of the \text{variable}, then prevents expansion of this material in \text{x}-type or \text{e}-type arguments using \exp_not:n.

\exp_not:V \{\text{tokens}\} 

Expands the \text{tokens} until only characters remains, and then converts this into a control sequence which should be a \text{variable} name. The content of the \text{variable} is recovered, and further expansion in \text{x}-type or \text{e}-type arguments is prevented using \exp_not:n.

\exp_not:e \{\text{tokens}\} 

Expands \text{tokens} exhaustively, then protects the result of the expansion (including any tokens which were not expanded) from further expansion in \text{e} or \text{x}-type arguments using \exp_not:n. This is very rarely useful but is provided for consistency.

\exp_not:f \{\text{tokens}\} 

Expands \text{tokens} fully until the first unexpandable token is found (if it is a space it is removed). Expansion then stops, and the result of the expansion (including any tokens which were not expanded) is protected from further expansion in \text{x}-type or \text{e}-type arguments using \exp_not:n.

\exp_stop_f: \{\text{tokens}\} \exp_stop_f: \{\text{more tokens}\} 

This function terminates an \text{f}-type expansion. Thus if a function \texttt{\foo_bar:f} starts an \text{f}-type expansion and all of \text{tokens} are expandable \exp_stop_f: terminates the expansion of \text{tokens} even if \text{more tokens} are also expandable. The function itself is an implicit space token. Inside an \text{x}-type or \text{e}-type expansion, it retains its form, but when typeset it produces the underlying space (\texttt{\textbackslash{\hspace{0.2mm}}}).

\section{Controlled expansion}

The \texttt{expl3} language makes all efforts to hide the complexity of \textsc{TeX} expansion from the programmer by providing concepts that evaluate/expand arguments of functions prior to
calling the “base” functions. Thus, instead of using many \expandafter calls and other
trickery it is usually a matter of choosing the right variant of a function to achieve a
desired result.

Of course, deep down \TeX{} is using expansion as always and there are cases where
a programmer needs to control that expansion directly; typical situations are basic data
manipulation tools. This section documents the functions for that level. These commands
are used throughout the kernel code, but we hope that outside the kernel there will be
little need to resort to them. Instead the argument manipulation methods document
above should usually be sufficient.

While \exp_after:wN expands one token (out of order) it is sometimes necessary to
expand several tokens in one go. The next set of commands provide this functionality.
Be aware that it is absolutely required that the programmer has full control over the
tokens to be expanded, i.e., it is not possible to use these functions to expand unknown
input as part of \langle expandable-tokens\rangle as that will break badly if unexpandable tokens are
encountered in that place!

While \exp_after:wN expands one token (out of order) it is sometimes necessary to
expand several tokens in one go. The next set of commands provide this functionality.
Be aware that it is absolutely required that the programmer has full control over the
tokens to be expanded, i.e., it is not possible to use these functions to expand unknown
input as part of \langle expandable-tokens\rangle as that will break badly if unexpandable tokens are
encountered in that place!

\begin{verbatim}
\exp:w {\expandafter tokens} \exp_end:
\end{verbatim}

Expands \langle expandable-tokens\rangle until reaching \exp_end: at which point expansion stops.
The full expansion of \langle expandable tokens\rangle has to be empty. If any token in \langle expandable
tokens\rangle or any token generated by expanding the tokens therein is not expandable the
expansion will end prematurely and as a result \exp_end: will be misinterpreted later on.

In typical use cases the \exp_end: is hidden somewhere in the replacement text of
\langle expandable-tokens\rangle rather than being on the same expansion level than \exp:w, e.g., you
may see code such as

\begin{verbatim}
\exp:w \@@_case:NnTF #1 {#2} { } { }
\end{verbatim}

where somewhere during the expansion of \@@_case:NnTF the \exp_end: gets generated.

\TeX{}hackers note: The current implementation uses \romannumeral hence ignores space
tokens and explicit signs + and − in the expansion of the \langle expandable tokens\rangle, but this should
not be relied upon.

\footnotetext[4]{Due to the implementation you might get the character in position 0 in the current font (typically
“*”) in the output without any error message!}
Expands \langle expandable-tokens \rangle until reaching \exp_end_continue_f:w at which point expansion continues as an f-type expansion expanding \langle further-tokens \rangle until an unexpandable token is encountered (or the f-type expansion is explicitly terminated by \exp_stop_f:). As with all f-type expansions a space ending the expansion gets removed.

The full expansion of \langle expandable-tokens \rangle has to be empty. If any token in \langle expandable-tokens \rangle or any token generated by expanding the tokens therein is not expandable the expansion will end prematurely and as a result \exp_end_continue_f:w will be misinterpreted later on.\footnote{In this particular case you may get a character into the output as well as an error message.}

In typical use cases \langle expandable-tokens \rangle contains no tokens at all, e.g., you will see code such as

\begin{verbatim}
\exp_after:wN { \exp:w \exp_end_continue_f:w #2 }
\end{verbatim}

where the \exp_after:wN triggers an f-expansion of the tokens in #2. For technical reasons this has to happen using two tokens (if they would be hidden inside another command \exp_after:wN would only expand the command but not trigger any additional f-expansion).

You might wonder why there are two different approaches available, after all the effect of

\begin{verbatim}
\exp:w \langle expandable-tokens \rangle \exp_end:
\end{verbatim}

can be alternatively achieved through an f-type expansion by using \exp_stop_f:, i.e.

\begin{verbatim}
\exp:w \exp_end_continue_f:w \langle expandable-tokens \rangle \exp_stop_f:
\end{verbatim}

The reason is simply that the first approach is slightly faster (one less token to parse and less expansion internally) so in places where such performance really matters and where we want to explicitly stop the expansion at a defined point the first form is preferable.

\begin{verbatim}
\exp:w \langle expandable-tokens \rangle \exp_end_continue_f:nw \langle further-tokens \rangle
\end{verbatim}

The difference to \exp_end_continue_f:w is that we first we pick up an argument which is then returned to the input stream. If \langle further-tokens \rangle starts with space tokens then these space tokens are removed while searching for the argument. If it starts with a brace group then the braces are removed. Thus such spaces or braces will not terminate the f-type expansion.
### 5.10 Internal functions

\[ \text{\texttt{\textbackslash cs\_new\_Npn \textbackslash exp\_args\_Ncof \{ \textless c \textless o \textless f \textless \}} \]

Internal forms for the base expansion types. These names do not conform to the general \LaTeX{}3 approach as this makes them more readily visible in the log and so forth. They should not be used outside this module.

\[ \text{\texttt{\textbackslash :\textbackslash :n \textbackslash :\textbackslash :N \textbackslash :\textbackslash :p \textbackslash :\textbackslash :c \textbackslash :\textbackslash :o \textbackslash :\textbackslash :e \textbackslash :\textbackslash :f \textbackslash :\textbackslash :x \textbackslash :\textbackslash :v \textbackslash :\textbackslash :v \textbackslash :\textbackslash :V \textbackslash :\textbackslash ::}} \]

\[ \text{\texttt{\textbackslash cs\_new\_Npn \textbackslash exp\_last\_unbraced\_Nno \{ \textbackslash :\textbackslash :n \textbackslash :\textbackslash :o\_unbraced \textbackslash ::}} \]

Internal forms for the expansion types which leave the terminal argument unbraced. These names do not conform to the general \LaTeX{}3 approach as this makes them more readily visible in the log and so forth. They should not be used outside this module.

\[ \text{\texttt{\textbackslash :\textbackslash :o\_unbraced \textbackslash :\textbackslash :e\_unbraced \textbackslash :\textbackslash :f\_unbraced \textbackslash :\textbackslash :x\_unbraced \textbackslash :\textbackslash :v\_unbraced \textbackslash :\textbackslash :V\_unbraced}} \]
Chapter 6

The \texttt{l3sort} package

Sorting functions

6.1 Controlling sorting

\LaTeX{} comes with a facility to sort list variables (sequences, token lists, or comma-lists) according to some user-defined comparison. For instance,

\begin{verbatim}
\clist_set:Nn \l_foo_clist { 3 , 01 , -2 , 5 , +1 }
\clist_sort:Nn \l_foo_clist
{ \int_compare:nNnTF { #1 } > { #2 }
{ \sort_return_swapped: }
{ \sort_return_same: }
}
\end{verbatim}

results in \texttt{\l_foo_clist} holding the values \{ -2 , 01 , +1 , 3 , 5 \} sorted in non-decreasing order.

The code defining the comparison should call \texttt{\sort_return_swapped:} if the two items given as \texttt{#1} and \texttt{#2} are not in the correct order, and otherwise it should call \texttt{\sort_return_same:} to indicate that the order of this pair of items should not be changed.

For instance, a \texttt{⟨comparison code⟩} consisting only of \texttt{\sort_return_same:} with no test yields a trivial sort: the final order is identical to the original order. Conversely, using a \texttt{⟨comparison code⟩} consisting only of \texttt{\sort_return_swapped:} reverses the list (in a fairly inefficient way).

\textbf{\texttt{\LaTeX} hackers note:} The current implementation is limited to sorting approximately 20000 items (40000 in \LaTeX{}), depending on what other packages are loaded.

Internally, the code from \texttt{l3sort} stores items in \texttt{\toks} registers allocated locally. Thus, the \texttt{⟨comparison code⟩} should not call \texttt{\newtoks} or other commands that allocate new \texttt{\toks} registers. On the other hand, altering the value of a previously allocated \texttt{\toks} register is not a problem.
\texttt{\textbackslash sort\_return\_same:} \texttt{\seq\_sort:Nn (seq var)}
\texttt{\textbackslash sort\_return\_swapped:} \{ ... \texttt{\sort\_return\_same:} or \texttt{\sort\_return\_swapped:} ... \}

Indicates whether to keep the order or swap the order of two items that are compared in the sorting code. Only one of the \texttt{\sort\_return\_...} functions should be used by the code, according to the results of some tests on the items \texttt{#1} and \texttt{#2} to be compared.
Chapter 7

The \texttt{l3tl-analysis} package: Analysing token lists

This module provides functions that are particularly useful in the \texttt{l3regex} module for mapping through a token list one \texttt{(token)} at a time (including begin-group/end-group tokens). For \texttt{l_analysis_map_inline:Nn} or \texttt{l_analysis_map_inline:nn}, the token list is given as an argument; the analogous function \texttt{peek_analysis_map_inline:n} documented in \texttt{l3token} finds tokens in the input stream instead. In both cases the user provides \texttt{(inline code)} that receives three arguments for each \texttt{(token)}:

- \texttt{(tokens)}, which both \texttt{o}-expand and \texttt{x}-expand to the \texttt{(token)}. The detailed form of \texttt{(tokens)} may change in later releases.
- \texttt{(char code)}, a decimal representation of the character code of the \texttt{(token)}, $-1$ if it is a control sequence.
- \texttt{(catcode)}, a capital hexadecimal digit which denotes the category code of the \texttt{(token)} ($0$: control sequence, $1$: begin-group, $2$: end-group, $3$: math shift, $4$: alignment tab, $6$: parameter, $7$: superscript, $8$: subscript, A: space, B: letter, C: other, D: active). This can be converted to an integer by writing \texttt{"(catcode)}.

In addition, there is a debugging function \texttt{l_analysis_show:n}, very similar to the \texttt{ShowTokens} macro from the \texttt{ted} package.

\begin{verbatim}
\tl_analysis_show:N \tl_analysis_show:n \tl_analysis_log:N \tl_analysis_log:n
\end{verbatim}

\begin{verbatim}
\tl_analysis_map_inline:nn \tl_analysis_map_inline:Nn
\end{verbatim}

\texttt{l_analysis_show:n} \{\texttt{(token list)}\}
\texttt{l_analysis_log:n} \{\texttt{(token list)}\}

Displays to the terminal (or log) the detailed decomposition of the \texttt{(token list)} into tokens, showing the category code of each character token, the meaning of control sequences and active characters, and the value of registers.

\texttt{l_analysis_map_inline:nn} \{\texttt{(token list)}\} \{\texttt{(inline function)}\}

Applies the \texttt{(inline function)} to each individual \texttt{(token)} in the \texttt{(token list)}. The \texttt{(inline function)} receives three arguments as explained above. As all other mappings the mapping is done at the current group level, \texttt{i.e.} any local assignments made by the \texttt{(inline function)} remain in effect after the loop.
Chapter 8

The \texttt{l3regex} package: Regular expressions in \TeX

The \texttt{l3regex} package provides regular expression testing, extraction of submatches, splitting, and replacement, all acting on token lists. The syntax of regular expressions is mostly a subset of the \texttt{pcre} syntax (and very close to \texttt{POSIX}), with some additions due to the fact that \TeX\ manipulates tokens rather than characters. For performance reasons, only a limited set of features are implemented. Notably, back-references are not supported.

Let us give a few examples. After
\begin{verbatim}
\tl_set:Nn \l_my_tl { That-cat. }
\regex_replace_once:nnN { at } { is } \l_my_tl
\end{verbatim}
the token list variable \texttt{\l_my_tl} holds the text “This cat.”, where the first occurrence of “at” was replaced by “is”. A more complicated example is a pattern to emphasize each word and add a comma after it:
\begin{verbatim}
\regex_replace_all:nnN { \w+ } { \c{emph}\cB\{ \0 \cE\} , } \l_my_tl
\end{verbatim}
The \texttt{\w} sequence represents any “word” character, and + indicates that the \texttt{\w} sequence should be repeated as many times as possible (at least once), hence matching a word in the input token list. In the replacement text, \texttt{\0} denotes the full match (here, a word). The command \texttt{\emph} is inserted using \texttt{\c{emph}}, and its argument \texttt{\0} is put between braces \texttt{\cB\{ and \cE\}}.

If a regular expression is to be used several times, it can be compiled once, and stored in a regex variable using \texttt{\regex_set:Nn}. For example,
\begin{verbatim}
\regex_new:N \l_foo_regex
\regex_set:Nn \l_foo_regex { \c{begin} \cB\{ \c[BE]\* \cE\} }\end{verbatim}
stores in \texttt{\l_foo_regex} a regular expression which matches the starting marker for an environment: \texttt{\begin}, followed by a begin-group token (\texttt{\cB\}), then any number of tokens which are neither begin-group nor end-group character tokens (\texttt{\c[BE]\*}), ending with an end-group token (\texttt{\cE\}). As explained in the next section, the parentheses “capture” the result of \texttt{\c[BE]\*}, giving us access to the name of the environment when doing replacements.
8.1 Syntax of regular expressions

8.1.1 Regular expression examples

We start with a few examples, and encourage the reader to apply \regex_show:n to these regular expressions.

- **Cat** matches the word “Cat” capitalized in this way, but also matches the beginning of the word “Cattle”: use \bCat\b to match a complete word only.

- \[abc\] matches one letter among “a”, “b”, “c”; the pattern (a|b|c) matches the same three possible letters (but see the discussion of submatches below).

- \[A-Za-z\]* matches any number (due to the quantifier *) of Latin letters (not accented).

- \c{[A-Za-z\]*} matches a control sequence made of Latin letters.

- \_\[^\_\]*\_ matches an underscore, any number of characters other than underscore, and another underscore; it is equivalent to \_.\_*\_ where . matches arbitrary characters and the lazy quantifier *? means to match as few characters as possible, thus avoiding matching underscores.

- \[\+\-]?\d+ matches an explicit integer with at most one sign.

- \[\+\-\.]\*\d+\.* matches an explicit integer with any number of + and – signs, with spaces allowed except within the mantissa, and surrounded by spaces.

- \[\+\-\.]\*(\d+|\d+\.)\.* matches an explicit integer or decimal number; using [.,] instead of \. would allow the comma as a decimal marker.

- \[\+\-\.]\*(\d+|\d+\.)\.*((?i)pt|in|mex|[cem]m|\[bs\]p|\[dn\]d|\[pcn\]c)\.* matches an explicit dimension with any unit that \TeX{} knows, where (?i) means to treat lowercase and uppercase letters identically.

- \[\+\-\.]\*((?i)nan|inf|((\d+|\d+\.)((?i)e\[\+\-\.]\*\d+)?)\.* matches an explicit floating point number or the special values nan and inf (with signs and spaces allowed).

- \[\+\-\.]\*(\d+|\cC.)\.* matches an explicit integer or control sequence (without checking whether it is an integer variable).

- \G.*?\K at the beginning of a regular expression matches and discards (due to \K) everything between the end of the previous match (\G) and what is matched by the rest of the regular expression; this is useful in \regex_replace_all:nnN when the goal is to extract matches or submatches in a finer way than with \regex_extract_all:nnN.

While it is impossible for a regular expression to match only integer expressions, \[\+\-\.\*\d+\.*((?i)e\[\+\-\.]\*\d+)\.* matches among other things all valid integer expressions (made only with explicit integers). One should follow it with further testing.
8.1.2 Characters in regular expressions

Most characters match exactly themselves, with an arbitrary category code. Some characters are special and must be escaped with a backslash (e.g., \* matches a star character). Some escape sequences of the form backslash–letter also have a special meaning (for instance \d matches any digit). As a rule,

- every alphanumeric character (A–Z, a–z, 0–9) matches exactly itself, and should not be escaped, because \A, \B, . . . have special meanings;
- non-alphanumeric printable ascii characters can (and should) always be escaped: many of them have special meanings (e.g., use \(, \), \?, \., \^);
- spaces should always be escaped (even in character classes);
- any other character may be escaped or not, without any effect: both versions match exactly that character.

Note that these rules play nicely with the fact that many non-alphanumeric characters are difficult to input into \TeX\ under normal category codes. For instance, \abc\% matches the characters \abc\% (with arbitrary category codes), but does not match the control sequence \abc followed by a percent character. Matching control sequences can be done using the \c{⟨regex⟩} syntax (see below).

Any special character which appears at a place where its special behaviour cannot apply matches itself instead (for instance, a quantifier appearing at the beginning of a string), after raising a warning.

Characters.

\x{hh...} Character with hex code hh...

\xhh Character with hex code hh.

\a Alarm (hex 07).
\e Escape (hex 1B).
\f Form-feed (hex 0C).
\n New line (hex 0A).
\r Carriage return (hex 0D).
\t Horizontal tab (hex 09).

8.1.3 Characters classes

Character types.

. A single period matches any token.
\d Any decimal digit.
\h Any horizontal space character, equivalent to \[\ \^\I\]: space and tab.
\s Any space character, equivalent to \[\ \^\I\^\J\^\L\^\M\].
Any vertical space character, equivalent to `\^^J\^^K\^^L\^^M`. Note that `\^^K` is a vertical space, but not a space, for compatibility with Perl.

Any word character, *i.e.*, alphanumerics and underscore, equivalent to the explicit class `*[A-Za-z0-9\_]`.

Any token not matched by `\d`.

Any token not matched by `\h`.

Any token other than the `\n` character (hex 0A).

Any token not matched by `\s`.

Any token not matched by `\v`.

Any token not matched by `\w`.

Of those, `\D`, `\H`, `\N`, `\S`, `\V`, and `\W` match arbitrary control sequences.

Character classes match exactly one token in the subject.

`[...]` Positive character class. Matches any of the specified tokens.

`[^...]` Negative character class. Matches any token other than the specified characters.

`[^a-z]` Matches any token other than the lowercase Latin letters.

`[^a-z]` Negative character class. Matches any token other than the specified characters.

`[^a-z]` Negative character class. Matches any token other than the specified characters.

Within a character class, this denotes a range (can be used with escaped characters).

Within a character class (one more set of brackets), this denotes the POSIX character class `(name)`, which can be `alnum`, `alpha`, `ascii`, `blank`, `cntrl`, `digit`, `graph`, `lower`, `print`, `punct`, `space`, `upper`, `word`, or `xdigit`.

For instance, `*[a-oq-z\cC.]` matches any lowercase Latin letter except `p`, as well as control sequences (see below for a description of `\c`).

In character classes, only `[^a-z]`, `\` and spaces are special, and should be escaped. Other non-alphanumeric characters can still be escaped without harm. Any escape sequence which matches a single character (`\d`, `\D`, etc.) is supported in character classes.

If the first character is `^`, then the meaning of the character class is inverted; `^` appearing anywhere else in the range is not special. If the first character (possibly following a leading `^`) is `)]` then it does not need to be escaped since ending the range there would make it empty. Ranges of characters can be expressed using `^`, for instance, `*[\d 0-5]` and `[\^6-9]` are equivalent.

### 8.1.4 Structure: alternatives, groups, repetitions

Quantifiers (repetition).

`?` 0 or 1, greedy.

`??` 0 or 1, lazy.

`*` 0 or more, greedy.

`*?` 0 or more, lazy.

`+` 1 or more, greedy.
1 or more, lazy.

{n} Exactly n.

{n,} n or more, greedy.

{n,}? n or more, lazy.

{n, m} At least n, no more than m, greedy.

{n, m}? At least n, no more than m, lazy.

For greedy quantifiers the regex code will first investigate matches that involve as many repetitions as possible, while for lazy quantifiers it investigates matches with as few repetitions as possible first.

Alternation and capturing groups.

A|B|C Either one of A, B, or C, investigating A first.

(....) Capturing group.

(?::...) Non-capturing group.

(?|....) Non-capturing group which resets the group number for capturing groups in each alternative. The following group is numbered with the first unused group number.

Capturing groups are a means of extracting information about the match. Parenthesized groups are labelled in the order of their opening parenthesis, starting at 1. The contents of those groups corresponding to the “best” match (leftmost longest) can be extracted and stored in a sequence of token lists using for instance `\regex_extract_once:nnNTF`.

The `\K` escape sequence resets the beginning of the match to the current position in the token list. This only affects what is reported as the full match. For instance,

`\regex_extract_all:nnN \{ a \K . \} \{ a123aaxyz \} \l_foo_seq`

results in `\l_foo_seq` containing the items `{1}` and `{a}`: the true matches are `{a1}` and `{aa}`, but they are trimmed by the use of `\K`. The `\K` command does not affect capturing groups: for instance,

`\regex_extract_once:nnN \{ (. \K c)+ \d \} \{ acbc3 \} \l_foo_seq`

results in `\l_foo_seq` containing the items `{c3}` and `{bc}`: the true match is `{acbc3}`, with first submatch `{bc}`, but `\K` resets the beginning of the match to the last position where it appears.

8.1.5 Matching exact tokens

The `\c` escape sequence allows to test the category code of tokens, and match control sequences. Each character category is represented by a single uppercase letter:

- C for control sequences;
- B for begin-group tokens;
- E for end-group tokens;
- M for math shift;
- T for alignment tab tokens;
- P for macro parameter tokens;
- U for superscript tokens (up);
- D for subscript tokens (down);
- S for spaces;
- L for letters;
- O for others; and
- A for active characters.

The \c escape sequence is used as follows.

\c{⟨regex⟩} A control sequence whose csname matches the ⟨regex⟩, anchored at the beginning and end, so that \c{begin} matches exactly \begin, and nothing else.

\cX Applies to the next object, which can be a character, escape character sequence such as \x{0A}, character class, or group, and forces this object to only match tokens with category X (any of CBEMTPUDSLOA). For instance, \cL[A-Z\d] matches uppercase letters and digits of category code letter, \cC matches any control sequence, and \cO(abc) matches abc where each character has category other.6

\c[XYZ] Applies to the next object, and forces it to only match tokens with category X, Y, or Z (each being any of CBEMTPUDSLOA). For instance, \c[LSO](abc) matches two tokens of category letter, space, or other.

\c[¬XYZ] Applies to the next object and prevents it from matching any token with category X, Y, or Z (each being any of CBEMTPUDSLOA). For instance, \c[¬O]\d matches digits which have any category different from other.

The category code tests can be used inside classes; for instance, [\cO\d \c[LO]\{A-F\}] matches what TpX considers as hexadecimal digits, namely digits with category other, or uppercase letters from A to F with category either letter or other. Within a group affected by a category code test, the outer test can be overridden by a nested test; for instance, \cL(ab\cO\*cd) matches ab\*cd where all characters are of category letter, except * which has category other.

The \u escape sequence allows to insert the contents of a token list directly into a regular expression or a replacement, avoiding the need to escape special characters. Namely, \u\{⟨var name⟩\} matches the exact contents (both character codes and category codes) of the variable \(\langle\text{var name}\rangle\), which are obtained by applying \exp_not:v \\(\langle\text{var name}\rangle\) at the time the regular expression is compiled. Within a \c{...} control sequence matching, the \u escape sequence only expands its argument once, in effect performing \tl_to_str:v. Quantifiers are supported.

The \ur escape sequence allows to insert the contents of a regex variable into a larger regular expression. For instance, A\ur{l_tma_regex}D matches the tokens A and

---

6This last example also captures “abc” as a regex group; to avoid this use a non-capturing group \cO(?:abc).

51
\d separated by something that matches the regular expression \l_tmpa_regex. This behaves as if a non-capturing group were surrounding \l_tmpa_regex, and any group contained in \l_tmpa_regex is converted to a non-capturing group. Quantifiers are supported.

For instance, if \l_tmpa_regex has value B|C, then A\ur{l_tmpa_regex}D is equivalent to A(?:B|C)D (matching ABD or ACD) and not to AB|CD (matching AB or CD). To get the latter effect, it is simplest to use \TeX’s expansion machinery directly: if \l_\-
_\mymodule_BC_tl contains B|C then the following two lines show the same result:

\regex_show:n { A \u{l_mymodule_BC_tl} D }
\regex_show:n { A B | C D }

8.1.6 Miscellaneous

Anchors and simple assertions.

\b Word boundary: either the previous token is matched by \w and the next by \W, or the opposite. For this purpose, the ends of the token list are considered as \W.

\B Not a word boundary: between two \w tokens or two \W tokens (including the boundary).

\^ or \A Start of the subject token list.

\$ \Z or \z End of the subject token list.

\G Start of the current match. This is only different from ^ in the case of multiple matches: for instance \regex_count:nnN { \G a } { aaba } \l_tmpa_int yields 2, but replacing \G by ^ would result in \l_tmpa_int holding the value 1.

The option (?i) makes the match case insensitive (identifying A–Z with a–z; no Unicode support yet). This applies until the end of the group in which it appears, and can be reverted using (?-i). For instance, in (?i)(a(?-i)b|c)d, the letters a and d are affected by the i option. Characters within ranges and classes are affected individually: (?i)[Y-\] is equivalent to [YZ\[\yz], and (?i)[^aeiou] matches any character which is not a vowel. Neither character properties, nor \c{…} nor \u{…} are affected by the i option.

8.2 Syntax of the replacement text

Most of the features described in regular expressions do not make sense within the replacement text. Backslash introduces various special constructions, described further below:

- \0 is the whole match;
- \1 is the submatch that was matched by the first (capturing) group (…); similarly for \2, …, \9 and \g{\langle number\rangle};
- \␣ inserts a space (spaces are ignored when not escaped);
- \a, \e, \f, \n, \r, \t, \xhh, \x{hhh} correspond to single characters as in regular expressions;
\c\{\langle cs\ name\ \rangle\} \inserts\ a\ control\ sequence;
\c\langle\\text{category}\rangle\langle\text{character}\rangle\ (\text{see}\ \text{below});
\u\{\langle\tl\ var\ name\ \rangle\} \inserts\ \text{the}\ contents\ of\ \text{the}\ \langle\tl\ var\ \rangle\ (\text{see}\ \text{below}).

Characters other than backslash and space are simply inserted in the result (but since the replacement text is first converted to a string, one should also escape characters that are special for \TeX, for instance use \#). Non-alphanumeric characters can always be safely escaped with a backslash.

For instance,
\begin{verbatim}
\tl_set:Nn \l_my_tl { Hello,-world! }
\regex_replace_all:nnN { \[^{,]+ } { \u{\l_my_{\0}_tl} } \l_my_tl
\end{verbatim}
results in \l_my_tl holding H(ell--el)(o,--o) w(or--o)(ld--l)!

The submatches are numbered according to the order in which the opening parenthesis of capturing groups appear in the regular expression to match. The \text{n}-th submatch is empty if there are fewer than \text{n} capturing groups or for capturing groups that appear in alternatives that were not used for the match. In case a capturing group matches several times during a match (due to quantifiers) only the last match is used in the replacement text. Submatches always keep the same category codes as in the original token list.

By default, the category code of characters inserted by the replacement are determined by the prevailing category code regime at the time where the replacement is made, with two exceptions:

- space characters (with character code 32) inserted with \c\ or \x20 or \x{20} have category code 10 regardless of the prevailing category code regime;
- if the category code would be 0 (escape), 5 (newline), 9 (ignore), 14 (comment) or 15 (invalid), it is replaced by 12 (other) instead.

The escape sequence \c allows to insert characters with arbitrary category codes, as well as control sequences.
\cX\{\ldots\}\ Produces\ the\ characters \ldots \ with\ category \text{X},\ which\ must\ be\ one\ of\ \CBEMTPUDSLOA\ as\ in\ regular\ expressions.\ Parentheses\ are\ optional\ for\ a\ single\ character\ (which\ can\ be\ an\ escape\ sequence).\ When\ nested,\ the\ innermost\ category\ code\ applies,\ for\ instance\ \cL(Hello\cS\ world)!\ gives\ this\ text\ with\ standard\ category\ codes.

\c\{\text{text}\}\ Produces\ the\ control\ sequence\ with\ csname\ \text{text}.\ The\ \langle\text{text}\rangle\ may\ contain\ references\ to\ the\ submatches\ \\0,\ \\1,\ and\ so\ on,\ as\ in\ the\ example\ for\ \u\ below.

The escape sequence \u\{\langle\var\ name\ \rangle\} \allows\ to\ insert\ the\ contents\ of\ the\ variable\ with\ name\ \langle\var\ name\ \rangle\ directly\ into\ the\ replacement,\ giving\ an\ easier\ control\ of\ category\ codes.\ When\ nested\ in\ \c\{\ldots\}\ and\ \u\{\ldots\}\ constructions,\ the\ \u\ and\ \c\ escape\ sequences\ perform\ \tl_to_str:v,\ namely\ extract\ the\ value\ of\ the\ control\ sequence\ and\ turn\ it\ into\ a\ string.\ Matches\ can\ also\ be\ used\ within\ the\ arguments\ of\ \c\ and\ \u.\ For\ instance,
\begin{verbatim}
\tl_set:Nn \l_my_one_tl { first }
\tl_set:Nn \l_my_two_tl { \emph{second} }
\tl_set:Nn \l_my_tl { one , two , one , one }
\regex_replace_all:nnN { [\-,]+ } { \u{\l_{my_\0_tl}} } \l_my_tl
\end{verbatim}
results in \texttt{\_my\_tl} holding \texttt{first,\texttt{second},first,first}.

Regexp replacement is also a convenient way to produce token lists with arbitrary category codes. For instance

\begin{verbatim}
\tl_clear:N \l_tmpa_tl
\regex_replace_all:nnN { } { \cU\% \cA\~ } \l_tmpa_tl
\end{verbatim}

results in \texttt{\_tmpa\_tl} containing the percent character with category code 7 (superscript) and an active tilde character.

### 8.3 Pre-compiling regular expressions

If a regular expression is to be used several times, it is better to compile it once rather than doing it each time the regular expression is used. The compiled regular expression is stored in a variable. All of the \texttt{l3regex} module's functions can be given their regular expression argument either as an explicit string or as a compiled regular expression.

\begin{verbatim}
\regex_new:N \l_my_regex
\regex_set:Nn \l_my_regex { my\ (simple\ )? reg\(ex\|ar\ulem{e}\ expression) }
\end{verbatim}

\begin{verbatim}
\regex_show:n \l_my_regex
\regex_log:n \l_my_regex
\end{verbatim}

indicating that the anchor \texttt{\_A} only applies to the first branch: the second branch is not anchored to the beginning of the match.
8.4 Matching

All regular expression functions are available in both \texttt{:n} and \texttt{:N} variants. The former require a “standard” regular expression, while the later require a compiled expression as generated by \texttt{\regex_set:Nn}.

\begin{verbatim}
\regex_match:nnTF \{\regex\} \{\token\list\} \{\true\ code\} \{\false\ code\}
\end{verbatim}

Tests whether the \texttt{regular expression} matches any part of the \texttt{token list}. For instance,

\begin{verbatim}
\regex_match:nnTF \{b \{cde\}^* \} \{abecdx\} \{TRUE\} \{FALSE\}
\regex_match:nnTF \{[b-dq-w]\} \{example\} \{TRUE\} \{FALSE\}
\end{verbatim}

leaves TRUE then FALSE in the input stream.

\begin{verbatim}
\regex_count:nnN \{\regex\} \{\token\list\} \{\int\ var\}
\end{verbatim}

Sets \texttt{\int\ var} within the current \TeX{} group level equal to the number of times \texttt{regular expression} appears in \texttt{token list}. The search starts by finding the left-most longest match, respecting greedy and lazy (non-greedy) operators. Then the search starts again from the character following the last character of the previous match, until reaching the end of the token list. Infinite loops are prevented in the case where the regular expression can match an empty token list: then we count one match between each pair of characters. For instance,

\begin{verbatim}
\int_new:N \l_foo_int
\regex_count:nnN \{(b+|c)\} \{abbababcbb\} \l_foo_int
\end{verbatim}

results in \texttt{\l_foo_int} taking the value 5.

\begin{verbatim}
\regex_match_case:nnTF \{\regex\} \{\token\list\} \{\true\ code\} \{\false\ code\}
\end{verbatim}

Determines which of the \texttt{regular expressions} matches at the earliest point in the \texttt{token list}, and leaves the corresponding \texttt{\true\ code} followed by the \texttt{\false\ code} in the input stream. If several \texttt{regex} match starting at the same point, then the first one in the list is selected and the others are discarded. If none of the \texttt{regex} match, the \texttt{false code} is left in the input stream. Each \texttt{regex} can either be given as a regex variable or as an explicit regular expression.

In detail, for each starting position in the \texttt{token list}, each of the \texttt{regex} is searched in turn. If one of them matches then the corresponding \texttt{code} is used and everything else is discarded, while if none of the \texttt{regex} match at a given position then the next starting position is attempted. If none of the \texttt{regex} match anywhere in the \texttt{token list} then nothing is left in the input stream. Note that this differs from nested \texttt{\regex_match:nnTF} statements since all \texttt{regex} are attempted at each position rather than attempting to match \texttt{regex\_1} at every position before moving on to \texttt{regex\_2}.

55
8.5 Submatch extraction

\regex_extract_once:nnN {⟨regex⟩} {⟨token list⟩} ⟨seq var⟩
\regex_extract_once:nnNTF {⟨regex⟩} {⟨token list⟩} ⟨seq var⟩ {⟨true code⟩} {⟨false code⟩}

Finds the first match of the ⟨regular expression⟩ in the ⟨token list⟩. If it exists, the match is stored as the first item of the ⟨seq var⟩, and further items are the contents of capturing groups, in the order of their opening parenthesis. The ⟨seq var⟩ is assigned locally. If there is no match, the ⟨seq var⟩ is cleared. The testing versions insert the ⟨true code⟩ into the input stream if a match was found, and the ⟨false code⟩ otherwise.

For instance, assume that you type

\regex_extract_once:nnNTF { \A(La)?TeX(!*)\Z } { LaTeX!!! } \l_foo_seq
{ true } { false }

Then the regular expression (anchored at the start with \A and at the end with \Z) must match the whole token list. The first capturing group, (La)?, matches La, and the second capturing group, (!*), matches !!! . Thus, \l_foo_seq contains as a result the items {LaTeX!!!}, {La}, and {!!!}, and the true branch is left in the input stream. Note that the n-th item of \l_foo_seq, as obtained using \seq_item:Nn, correspond to the submatch numbered (n − 1) in functions such as \regex_replace_once:nnN.

\regex_extract_all:nnN
\regex_extract_all:nnNTF
\regex_extract_all:NnN
\regex_extract_all:NnNTF

Finds all matches of the ⟨regular expression⟩ in the ⟨token list⟩, and stores all the sub-match information in a single sequence (concatenating the results of multiple \regex_extract_once:nnN calls). The ⟨seq var⟩ is assigned locally. If there is no match, the ⟨seq var⟩ is cleared. The testing versions insert the ⟨true code⟩ into the input stream if a match was found, and the ⟨false code⟩ otherwise. For instance, assume that you type

\regex_extract_all:nnNTF { \w+ } { Hello,-world! } \l_foo_seq
{ true } { false }

Then the regular expression matches twice, the resulting sequence contains the two items {Hello} and {world}, and the true branch is left in the input stream.
Splits the \textit{token list} into a sequence of parts, delimited by matches of the \textit{regular expression}. If the \textit{regular expression} has capturing groups, then the token lists that they match are stored as items of the sequence as well. The assignment to \textit{seq var} is local. If no match is found the resulting \textit{seq var} has the \textit{token list} as its sole item. If the \textit{regular expression} matches the empty token list, then the \textit{token list} is split into single tokens. The testing versions insert the \textit{true code} into the input stream if a match was found, and the \textit{false code} otherwise. For example, after
\begin{verbatim}
\seq_new:N \l_path_seq
\regex_split:nnNTF { / } { the/path/for/this/file.tex } \l_path_seq
  { true } { false }
\end{verbatim}
the sequence \l_path_seq contains the items \{the\}, \{path\}, \{for\}, \{this\}, and \{file.tex\}, and the \textit{true} branch is left in the input stream.

\subsection{Replacement}

Searches for the \textit{regular expression} in the contents of the \textit{tl var} and replaces the first match with the \textit{replacement}. In the \textit{replacement}, \textbackslash 0 represents the full match, \textbackslash 1 represents the contents of the first capturing group, \textbackslash 2 of the second, \textit{etc}. The result is assigned locally to \textit{tl var}. Every match is treated independently, and matches cannot overlap. The result is assigned locally to \textit{tl var}.
\regex_replace_case_once:nN \regex_replace_case_once:nNTF

Replaces the earliest match of the regular expression \(?1\langle regex1 \rangle \ldots 1\langle regex_n \rangle \)? in the \(token list variable\) by the \(replacement\) corresponding to which \(regex_i \) matched, then leaves the \(true code\) in the input stream. If none of the \(regex\) match, then the \(tl var\) is not modified, and the \(false code\) is left in the input stream. Each \(regex\) can either be given as a regex variable or as an explicit regular expression.

In detail, for each starting position in the \(token list\), each of the \(regex\) is searched in turn. If one of them matches then it is replaced by the corresponding \(replacement\) as described for \regex_replace_once:nnN. This is equivalent to checking with \regex_match_case:nn which \(regex\) matches, then performing the replacement with \regex_replace_once:nnN.

\regex_replace_case_all:nN \regex_replace_case_all:nNTF

Replaces all occurrences of all \(regex\) in the \(token list\) by the corresponding \(replacement\). Every match is treated independently, and matches cannot overlap. The result is assigned locally to \(tl var\), and the \(true code\) or \(false code\) is left in the input stream depending on whether any replacement was made or not.

In detail, for each starting position in the \(token list\), each of the \(regex\) is searched in turn. If one of them matches then it is replaced by the corresponding \(replacement\), and the search resumes at the position that follows this match (and replacement). For instance

\tl_set:Nn \l_tmpa_tl { Hello,-world! }
\regex_replace_case_all:nN

results in \l_tmpa_tl having the contents ‘‘Hello’’---[,][,]‘‘world’’---[!]. Note in particular that the word-boundary assertion \b did not match at the start of words because the case \[A-Za-z]+ matched at these positions. To change this, one could simply swap the order of the two cases in the argument of \regex_replace_case_all:nN.
8.7 Scratch regular expressions

\l_tmpa_regex \l_tmpb_regex

Scratch regex for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_regex \g_tmpb_regex

Scratch regex for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

8.8 Bugs, misfeatures, future work, and other possibilities

The following need to be done now.

- Rewrite the documentation in a more ordered way, perhaps add a BNF?

  Additional error-checking to come.

- Clean up the use of messages.

- Cleaner error reporting in the replacement phase.

- Add tracing information.

- Detect attempts to use back-references and other non-implemented syntax.

- Test for the maximum register \c_max_register_int.

- Find out whether the fact that \W and friends match the end-marker leads to bugs. Possibly update \_\_regex_item_reverse:n.

- The empty cs should be matched by \c{}, not by \c{csname.?endcsname\s?}.

  Code improvements to come.

- Shift arrays so that the useful information starts at position 1.

- Only build \c{...} once.

- Use arrays for the left and right state stacks when compiling a regex.

- Should \_\_regex_action_free_group:n only be used for greedy \{n,\} quantifier? (I think not.)

- Quantifiers for \u and assertions.

- When matching, keep track of an explicit stack of curr_state and curr_-

  submatches.

- If possible, when a state is reused by the same thread, kill other subthreads.
• Use an array rather than `\g__regex_balance_tl` to build the function `\__regex_replacement_balance_one_match:n`.
• Reduce the number of epsilon-transitions in alternatives.
• Optimize simple strings: use less states (`abcade` should give two states, for `abc` and `ade`). [Does that really make sense?]
• Optimize groups with no alternative.
• Optimize states with a single `\__regex_action_free:n`.
• Optimize the use of `\__regex_action_success:` by inserting it in state 2 directly instead of having an extra transition.
• Optimize the use of `\int_step...` functions.
• Groups don’t capture within regexes for csnames; optimize and document.
• Better “show” for anchors, properties, and catcode tests.
• Does `\K` really need a new state for itself?
• When compiling, use a boolean `in_cs` and less magic numbers.
• Instead of checking whether the character is special or alphanumeric using its character code, check if it is special in regexes with `\cs_if_exist` tests.

The following features are likely to be implemented at some point in the future.
• General look-ahead/behind assertions.
• Regex matching on external files.
• Conditional subpatterns with look ahead/behind: “if what follows is […], then […].”
• `(*...)` and `(?...)` sequences to set some options.
• UTF-8 mode for pdfTeX.
• Newline conventions are not done. In particular, we should have an option for . not to match newlines. Also, `\A` should differ from `^`, and `\Z`, `\z`, and `$` should differ.
• Unicode properties: `\p{..}` and `\P{..};` `\X` which should match any “extended” Unicode sequence. This requires to manipulate a lot of data, probably using treeboxes.

The following features of PCRE or Perl may or may not be implemented.
• Callout with `(?C...)` or other syntax: some internal code changes make that possible, and it can be useful for instance in the replacement code to stop a regex replacement when some marker has been found; this raises the question of a potential `\regex_break:`, and then of playing well with `\tl_map_break:`, called from within the code in a regex. It also raises the question of nested calls to the regex machinery, which is a problem since `\fontdimen` are global.
• Conditional subpatterns (other than with a look-ahead or look-behind condition): this is non-regular, isn’t it?

• Named subpatterns: \TeX\ programmers have lived so far without any need for named macro parameters.

The following features of \texttt{PCRE} or \texttt{Perl} will definitely not be implemented.

• Back-references: non-regular feature, this requires backtracking, which is prohibitively slow.

• Recursion: this is a non-regular feature.

• Atomic grouping, possessive quantifiers: those tools, mostly meant to fix catastrophic backtracking, are unnecessary in a non-backtracking algorithm, and difficult to implement.

• Subroutine calls: this syntactic sugar is difficult to include in a non-backtracking algorithm, in particular because the corresponding group should be treated as atomic.

• Backtracking control verbs: intrinsically tied to backtracking.

• \texttt{\textbackslash ddd}, matching the character with octal code \texttt{ddd}: we already have \texttt{\textbackslash x{}{...}} and the syntax is confusingly close to what we could have used for backreferences (\texttt{\textbackslash 1, \textbackslash 2, ...}), making it harder to produce useful error message.

• \texttt{\textbackslash cx}, similar to \TeX’s own \texttt{\textbackslash ^{~}x}.

• Comments: \TeX\ already has its own system for comments.

• \texttt{\textbackslash Q...\textbackslash E} escaping: this would require to read the argument verbatim, which is not in the scope of this module.

• \texttt{\textbackslash C} single byte in UTF-8 mode: \TeX and \texttt{Lua\TeX} serve us characters directly, and splitting those into bytes is tricky, encoding dependent, and most likely not useful anyways.
Chapter 9

The l3prg package

Control structures

Conditional processing in \LaTeX{}3 is defined as something that performs a series of tests, possibly involving assignments and calling other functions that do not read further ahead in the input stream. After processing the input, a state is returned. The states returned are \texttt{⟨true⟩} and \texttt{⟨false⟩}.

\LaTeX{}3 has two forms of conditional flow processing based on these states. The first form is predicate functions that turn the returned state into a boolean \texttt{⟨true⟩} or \texttt{⟨false⟩}. For example, the function \texttt{\cs_if_free_p:NN} checks whether the control sequence given as its argument is free and then returns the boolean \texttt{⟨true⟩} or \texttt{⟨false⟩} values to be used in testing with \texttt{\if_predicate:w} or in functions to be described below. The second form is the kind of functions choosing a particular argument from the input stream based on the result of the testing as in \texttt{\cs_if_free:NTF} which also takes one argument (the \texttt{N}) and then executes either \texttt{true} or \texttt{false} depending on the result.

\texttt{T\TeX{}hackers note:} The arguments are executed after exiting the underlying \texttt{\if...\fi:} structure.

9.1 Defining a set of conditional functions

\verb|\prg_new_conditional:Npnn \prg_set_conditional:Npnn \prg_new_conditional:Nnn \prg_set_conditional:Nnn|

\verb|\prg_new_conditional:Npnn \prg_set_conditional:Npnn \prg_new_conditional:Nnn \prg_set_conditional:Nnn|

These functions create a family of conditionals using the same \texttt{(code)} to perform the test created. Those conditionals are expandable if \texttt{(code)} is. The \texttt{new} versions check for existing definitions and perform assignments globally (\textit{cf.} \texttt{\cs_new:Npn}) whereas the \texttt{set} versions do no check and perform assignments locally (\textit{cf.} \texttt{\cs_set:Npn}). The conditionals created are dependent on the comma-separated list of \texttt{(conditions)}, which should be one or more of \texttt{p}, \texttt{T}, \texttt{F} and \texttt{TF}. 
These functions create a family of protected conditionals using the same \{(code)\} to perform the test created. The \{(code)\} does not need to be expandable. The \texttt{new} version check for existing definitions and perform assignments globally (cf. \texttt{cs_new:Npn}) whereas the \texttt{set} version do not (cf. \texttt{cs_set:Npn}). The conditionals created are depended on the comma-separated list of \{(conditions)\}, which should be one or more of \texttt{T}, \texttt{F} and \texttt{TF} (not \texttt{p}).

The conditionals are defined by \texttt{prg_new_conditional:Npnn} and friends as:

- \texttt{\langle name \rangle_p:\langle arg spec \rangle} — a predicate function which will supply either a logical \texttt{true} or logical \texttt{false}. This function is intended for use in cases where one or more logical tests are combined to lead to a final outcome. This function cannot be defined for protected conditionals.

- \texttt{\langle name \rangle:\langle arg spec \rangle T} — a function with one more argument than the original \langle arg spec \rangle demands. The \langle true branch \rangle code in this additional argument will be left on the input stream only if the test is \texttt{true}.

- \texttt{\langle name \rangle:\langle arg spec \rangle F} — a function with one more argument than the original \langle arg spec \rangle demands. The \langle false branch \rangle code in this additional argument will be left on the input stream only if the test is \texttt{false}.

- \texttt{\langle name \rangle:\langle arg spec \rangle TF} — a function with two more argument than the original \langle arg spec \rangle demands. The \langle true branch \rangle code in the first additional argument will be left on the input stream if the test is \texttt{true}, while the \langle false branch \rangle code in the second argument will be left on the input stream if the test is \texttt{false}.

The \{(code)\} of the test may use \langle parameters \rangle as specified by the second argument to \texttt{prg_set_conditional:Npnn}: this should match the \langle argument specification \rangle but this is not enforced. The \texttt{Nnn} versions infer the number of arguments from the argument specification given (cf. \texttt{cs_new:Nn}, etc.). Within the \langle code \rangle, the functions \texttt{prg_return_true:} and \texttt{prg_return_false:} are used to indicate the logical outcomes of the test.

An example can easily clarify matters here:

\begin{verbatim}
\prg_set_conditional:Npnn \foo_if_bar:NN #1#2 { p , T , TF }
{\fi:}
\end{verbatim}
This defines the function \(\texttt{foo_if_bar_p:NN}, \texttt{foo_if_bar:NNTF}\) and \(\texttt{foo_if_bar:NNT}\) but not \(\texttt{foo_if_bar:NNF}\) (because F is missing from the \(\langle\text{conditions}\rangle\) list). The return statements take care of resolving the remaining \texttt{\textbackslash{else}}: and \texttt{\textbackslash{fi}}: before returning the state. There must be a return statement for each branch; failing to do so will result in erroneous output if that branch is executed.

\[\texttt{\textbackslash{prg_new_eq_conditional}:NNn} \langle \text{name}1 \rangle: \langle \text{arg spec1} \rangle \langle \text{name}2 \rangle: \langle \text{arg spec2} \rangle\{\langle\text{conditions}\rangle}\]

\[\texttt{\textbackslash{prg_set_eq_conditional}:NNn}\]

These functions copy a family of conditionals. The \texttt{new} version checks for existing definitions (cf. \texttt{\textbackslash{cs_new_eq}:NN}) whereas the \texttt{set} version does not (cf. \texttt{\textbackslash{cs_set_eq}:NN}). The conditionals copied are depended on the comma-separated list of \(\langle\text{conditions}\rangle\), which should be one or more of \texttt{p}, \texttt{T}, \texttt{F} and \texttt{TF}.

\[\texttt{\textbackslash{prg_return_true}: *}\]
\[\texttt{\textbackslash{prg_return_false}: *}\]

These “return” functions define the logical state of a conditional statement. They appear within the code for a conditional function generated by \(\texttt{\textbackslash{prg_set_conditional}:Npn}\), etc. to indicate when a true or false branch should be taken. While they may appear multiple times each within the code of such conditionals, the execution of the conditional must result in the expansion of one of these two functions \textit{exactly once}.

The return functions trigger what is internally an \(f\)-expansion process to complete the evaluation of the conditional. Therefore, after \texttt{\textbackslash{prg_return_true}:} or \texttt{\textbackslash{prg_return_false}:} there must be no non-expandable material in the input stream for the remainder of the expansion of the conditional code. This includes other instances of either of these functions.

\[\texttt{\textbackslash{prg_generate_conditional_variant}:Nnn} \langle \text{name} \rangle: \langle \text{arg spec} \rangle\{\langle\text{variant argument specifiers}\rangle\} \{\langle\text{condition specifiers}\rangle\}\]

\[\texttt{\textbackslash{prg_generate_conditional_variant}:Nnn}\]

New: 2017-12-12

Defines argument-specifier variants of conditionals. This is equivalent to running \texttt{\textbackslash{cs_generate_variant}:Na} \(\langle\text{conditional}\rangle\) \{\(\langle\text{variant argument specifiers}\rangle\}\} on each \(\langle\text{conditional}\rangle\) described by the \(\langle\text{condition specifiers}\rangle\). These base-form \(\langle\text{conditionals}\rangle\) are obtained from the \(\langle\text{name}1\rangle\) and \(\langle\text{arg spec}\rangle\) as described for \texttt{\textbackslash{prg_new_conditional}:Npnn}, and they should be defined.

### 9.2 The boolean data type

This section describes a boolean data type which is closely connected to conditional processing as sometimes you want to execute some code depending on the value of a switch (e.g., draft/final) and other times you perhaps want to use it as a predicate function in an \texttt{\textbackslash{if_predicate}:w} test. The problem of the primitive \texttt{\textbackslash{if_false}:} and \texttt{\textbackslash{if_true}:} tokens is that it is not always safe to pass them around as they may interfere with scanning for termination of primitive conditional processing. Therefore, we employ two canonical booleans: \texttt{\textbackslash{c_true_bool}} or \texttt{\textbackslash{c_false_bool}}. Besides preventing problems as described above, it also allows us to implement a simple boolean parser supporting the logical operations And, Or, Not, \textit{etc.} which can then be used on both the boolean type and predicate functions.
All conditional \bool functions except assignments are expandable and expect the input to also be fully expandable (which generally means being constructed from predicate functions and booleans, possibly nested).

\TeX hackers note: The \bool data type is not implemented using the \iffalse/\iftrue primitives, in contrast to \neviz, etc., in plain \TeX. \IfThenElse\TeX and so on. Programmers should not base use of \bool switches on any particular expectation of the implementation.

\begin{verbatim}
\bool_new:N \bool_new:c
\bool_const:Nn \bool_const:cn \\
\bool_set_false:N \bool_set_false:c \\
\bool_set_true:N \bool_set_true:c \\
\bool_set_eq:NN \bool_set_eq:Nn \bool_set_eq:cn \\
\bool_set:Nn \bool_set:cn \bool_gset:Nn \bool_gset:cn \\
\bool_if_p:N \bool_if_p:c \bool_if:N \bool_if:NTF \bool_if:TF \\
\bool_to_str:N \bool_to_str:c \bool_to_str:n
\end{verbatim}

\begin{itemize}
\item \bool_new:N \bool_new:c
\begin{itemize}
\item Creates a new \langle boolean \rangle or raises an error if the name is already taken. The declaration is global. The \langle boolean \rangle is initially false.
\end{itemize}
\item \bool_const:Nn \bool_const:cn \\
\begin{itemize}
\item Creates a new constant \langle boolean \rangle or raises an error if the name is already taken. The value of the \langle boolean \rangle is set globally to the result of evaluating the \langle boolexpr \rangle.
\end{itemize}
\item \bool_set_false:N \bool_set_false:c \\
\begin{itemize}
\item Sets \langle boolean \rangle logically false.
\end{itemize}
\item \bool_set_true:N \bool_set_true:c \\
\begin{itemize}
\item Sets \langle boolean \rangle logically true.
\end{itemize}
\item \bool_set_eq:NN \bool_set_eq:Nn \bool_set_eq:cn \\
\begin{itemize}
\item Sets \langle boolean \rangle to the current value of \langle boolean\rangle2.
\end{itemize}
\item \bool_set:Nn \bool_set:cn \\
\begin{itemize}
\item Evaluates the \langle boolean expression \rangle as described for \bool_if:nTF, and sets the \langle boolean \rangle variable to the logical truth of this evaluation.
\end{itemize}
\item \bool_if_p:N \bool_if_p:c \bool_if:N \bool_if:NTF \bool_if:TF \\
\begin{itemize}
\item Tests the current truth of \langle boolean \rangle, and continues expansion based on this result.
\end{itemize}
\item \bool_to_str:N \bool_to_str:c \bool_to_str:n \\
\begin{itemize}
\item Expands to the letters true or false depending on the logical truth of the \langle boolean \rangle or \langle boolean expression \rangle.
\end{itemize}
\end{itemize}
\bool_show:N \bool_show:N \bool_show:c
\bool_show:n \bool_show:n \bool_show:n \bool_show:n
\bool_log:N \bool_log:c \bool_log:n \bool_log:n
\bool_if_exist_p:N \bool_if_exist_p:c \bool_if_exist:N \bool_if_exist:c
\bool_if_exist:N \bool_if_exist:N \bool_if_exist:N \bool_if_exist:N

\bool_show:N \bool_show:N \bool_show:c
New: 2012-02-09
Updated: 2021-04-29

\bool_show:n \bool_show:n \bool_show:n \bool_show:n
New: 2012-02-09
Updated: 2017-07-15

\bool_log:N \bool_log:c \bool_log:n \bool_log:n
New: 2014-08-22
Updated: 2021-04-29

\bool_if_exist_p:N \bool_if_exist_p:c \bool_if_exist:N \bool_if_exist:c
New: 2012-03-03

Displays the logical truth of the \textit{boolean} on the terminal.

Displays the logical truth of the \textit{boolean expression} on the terminal.

Writes the logical truth of the \textit{boolean} in the log file.

Writes the logical truth of the \textit{boolean expression} in the log file.

Tests whether the \textit{boolean} is currently defined. This does not check that the \textit{boolean} really is a boolean variable.

9.2.1 Scratch booleans

\l_tmpa_bool \l_tmpb_bool
\g_tmpa_bool \g_tmpb_bool

A scratch boolean for local assignment. It is never used by the kernel code, and so is safe for use with any \LaTeX\-defined function. However, it may be overwritten by other non-kernel code and so should only be used for short-term storage.

A scratch boolean for global assignment. It is never used by the kernel code, and so is safe for use with any \LaTeX\-defined function. However, it may be overwritten by other non-kernel code and so should only be used for short-term storage.

9.3 Boolean expressions

As we have a boolean datatype and predicate functions returning boolean \textit{true} or \textit{false} values, it seems only fitting that we also provide a parser for \textit{boolean expressions}.

A boolean expression is an expression which given input in the form of predicate functions and boolean variables, return boolean \textit{true} or \textit{false}. It supports the logical operations And, Or and Not as the well-known infix operators \&\& and ||| and prefix ! with their usual precedences (namely, \&\& binds more tightly than |||). In addition to this, parentheses can be used to isolate sub-expressions. For example,
\int_compare_p:n \{ \ 1 = 1 \} \&\&
( \int_compare_p:n \{ \ 2 = 3 \} ||
\int_compare_p:n \{ \ 4 <= 4 \} ||
\str_if_eq_p:nn \{ \ abc \} \{ \ def \}
) \&\&
! \int_compare_p:n \{ \ 2 = 4 \}

is a valid boolean expression.

Contrarily to some other programming languages, the operators \&\& and || evaluate both operands in all cases, even when the first operand is enough to determine the result. This “eager” evaluation should be contrasted with the “lazy” evaluation of \bool_lazy_- ...

**\TeXhackers note:** The eager evaluation of boolean expressions is unfortunately necessary in \TeX. Indeed, a lazy parser can get confused if \&\& or || or parentheses appear as (unbraced) arguments of some predicates. For instance, the innocuous-looking expression below would break (in a lazy parser) if \#1 were a closing parenthesis and \l_tmpa_bool were true.

( \l_tmpa_bool || \token_if_eq_meaning_p:NN X \#1 )

Minimal (lazy) evaluation can be obtained using the conditionals \bool_lazy_all:nTF, \bool_lazy_and:nnTF, \bool_lazy_any:nTF, or \bool_lazy_or:nnTF, which only evaluate their boolean expression arguments when they are needed to determine the resulting truth value. For example, when evaluating the boolean expression

\bool_lazy_and_p:nn
{
  \bool_lazy_any_p:n
  {
    \int_compare_p:n \{ \ 2 = 3 \}
    \int_compare_p:n \{ \ 4 <= 4 \}
  }
  \int_compare_p:n \{ \ 1 = \error \} % skipped
}

{ ! \int_compare_p:n \{ \ 2 = 4 \}

the line marked with skipped is not expanded because the result of \bool_lazy_any_p:n is known once the second boolean expression is found to be logically true. On the other hand, the last line is expanded because its logical value is needed to determine the result of \bool_lazy_and_p:nn.

\bool_if_p:n \{ ⟨boolean expression⟩\}
\bool_if:nTF \{ ⟨boolean expression⟩\} \{⟨true code⟩\} \{⟨false code⟩\}

Tests the current truth of ⟨boolean expression⟩, and continues expansion based on this result. The ⟨boolean expression⟩ should consist of a series of predicates or boolean variables with the logical relationship between these defined using \&\& (“And”), || (“Or”), ! (“Not”) and parentheses. The logical Not applies to the next predicate or group.

---

\bool_if_p:n *
\bool_if:nTF *

Updated: 2017-07-15
\bool_lazy_and_p:n * \bool_lazy_and_p:nn \bool_lazy_and_p:nn \bool_lazy_and:n \bool_lazy_all_p:n \bool_lazy_all_p:nn \bool_lazy_all:n \bool_lazy_or_p:nn \bool_lazy_or_p:nn \bool_lazy_or:nn \bool_lazy_or:nn \bool_lazy_or:n \bool_lazy_or:n \bool_xor:nn \bool_xor_p:nn \bool_xor_p:nn \bool_xor:nn \bool_xor:nn

Updated: 2017-07-15
Updated: 2017-07-15
Updated: 2017-07-15
Updated: 2017-07-15
Updated: 2017-07-15
Updated: 2017-07-15
Updated: 2017-07-15
Updated: 2017-07-15

\bool_lazy_and:nn \bool_lazy_and:nn \bool_lazy_and:nn \bool_lazy_and:nn \bool_lazy_and:nn \bool_lazy_and:nn
\bool_lazy_all:nn \bool_lazy_all:nn \bool_lazy_all:nn \bool_lazy_all:nn \bool_lazy_all:nn \bool_lazy_all:nn
\bool_lazy_or:nn \bool_lazy_or:nn \bool_lazy_or:nn \bool_lazy_or:nn \bool_lazy_or:nn \bool_lazy_or:nn
\bool_lazy_any:n \bool_lazy_any:n \bool_lazy_any:n \bool_lazy_any:n \bool_lazy_any:n \bool_lazy_any:n
\bool_not_p:n \bool_not_p:n \bool_not_p:n
\bool_xor:p:nn \bool_xor:p:nn \bool_xor:p:nn
\bool_xor:n:nn \bool_xor:n:nn \bool_xor:n:nn

Updated: 2015-11-15
Updated: 2015-11-15
Updated: 2015-11-15
Updated: 2015-11-15
Updated: 2015-11-15
Updated: 2018-05-09

\bool_do_until:Nn \bool_do_until:cn

Updated: 2017-07-15
Updated: 2017-07-15

\bool_do_until:Nn \bool_do_until:cn

Implements the “And” operation on the \langle boolean expressions \rangle, hence is true if all of them are true and false if any of them is false. Contrarily to the infix operator &&, only the \langle boolean expressions \rangle which are needed to determine the result of \bool_lazy_all:nTF are evaluated. See also \bool_lazy_all:nnTF when there are only two \langle boolean expressions \rangle.

\bool_lazy_and:nn \bool_lazy_and:nn \bool_lazy_and:nn \bool_lazy_and:nn \bool_lazy_and:nn
\bool_lazy_all:nn \bool_lazy_all:nn \bool_lazy_all:nn \bool_lazy_all:nn \bool_lazy_all:nn
\bool_lazy_or:nn \bool_lazy_or:nn \bool_lazy_or:nn \bool_lazy_or:nn \bool_lazy_or:nn
\bool_lazy_any:n \bool_lazy_any:n \bool_lazy_any:n \bool_lazy_any:n \bool_lazy_any:n

Updated: 2015-11-15
Updated: 2015-11-15
Updated: 2015-11-15
Updated: 2015-11-15
Updated: 2015-11-15

\bool_not:p:n

Updated: 2017-07-15

\bool_xor:p:nn \bool_xor:p:nn \bool_xor:p:nn
\bool_xor:n:nn \bool_xor:n:nn \bool_xor:n:nn

Updated: 2018-05-09

\bool_do_until:Nn

\bool_do_until:Nn (boolean) \{ (code) \}

Function version of !\langle boolean expression \rangle within a boolean expression.

9.4 Logical loops

Loops using either boolean expressions or stored boolean values.

\bool_do_until:Nn (boolean) \{ (code) \}

Places the \langle code \rangle in the input stream for \TeX{} to process, and then checks the logical value of the \langle boolean \rangle. If it is false then the \langle code \rangle is inserted into the input stream again and the process loops until the \langle boolean \rangle is true.
\bool_do_while:Nn \bool_do_while:cn \ Updated: 2017-07-15

Places the \langle code\rangle in the input stream for \TeX{} to process, and then checks the logical value of the \langle boolean\rangle. If it is \texttt{true} then the \langle code\rangle is inserted into the input stream again and the process loops until the \langle boolean\rangle is \texttt{false}.

\bool_until_do:Nn \bool_until_do:cn \ Updated: 2017-07-15

This function firsts checks the logical value of the \langle boolean\rangle. If it is \texttt{false} the \langle code\rangle is placed in the input stream and expanded. After the completion of the \langle code\rangle the truth of the \langle boolean\rangle is re-evaluated. The process then loops until the \langle boolean\rangle is \texttt{true}.

\bool_while_do:Nn \bool_while_do:cn \ Updated: 2017-07-15

This function firsts checks the logical value of the \langle boolean\rangle. If it is \texttt{true} the \langle code\rangle is placed in the input stream and expanded. After the completion of the \langle code\rangle the truth of the \langle boolean\rangle is re-evaluated. The process then loops until the \langle boolean\rangle is \texttt{false}.

\bool_do_until:nn \ Updated: 2017-07-15

Places the \langle code\rangle in the input stream for \TeX{} to process, and then checks the logical value of the \langle boolean expression\rangle as described for \verb|\bool_if:nTF|. If it is \texttt{false} then the \langle code\rangle is inserted into the input stream again and the process loops until the \langle boolean expression\rangle evaluates to \texttt{true}.

\bool_until_do:nn \ Updated: 2017-07-15

This function firsts checks the logical value of the \langle boolean expression\rangle (as described for \verb|\bool_if:nTF|). If it is \texttt{false} the \langle code\rangle is placed in the input stream and expanded. After the completion of the \langle code\rangle the truth of the \langle boolean expression\rangle is re-evaluated. The process then loops until the \langle boolean expression\rangle is \texttt{true}.

\bool_while_do:nn \ Updated: 2017-07-15

This function firsts checks the logical value of the \langle boolean expression\rangle (as described for \verb|\bool_if:nTF|). If it is \texttt{true} the \langle code\rangle is placed in the input stream and expanded. After the completion of the \langle code\rangle the truth of the \langle boolean expression\rangle is re-evaluated. The process then loops until the \langle boolean expression\rangle is \texttt{false}.

\prg_replicate:nn \ Updated: 2011-07-04

Evaluates the \langle integer expression\rangle (which should be zero or positive) and creates the resulting number of copies of the \langle tokens\rangle. The function is both expandable and safe for nesting. It yields its result after two expansion steps.

\section{Producing multiple copies}

\prg_replicate:nn \ Updated: 2011-07-04

Evaluates the \langle integer expression\rangle (which should be zero or positive) and creates the resulting number of copies of the \langle tokens\rangle. The function is both expandable and safe for nesting. It yields its result after two expansion steps.
9.6 Detecting \TeX’s mode

\mode_if_horizontal_p: \* \mode_if_horizontal:TF {⟨true code⟩} {⟨false code⟩}
Detects if \TeX{} is currently in horizontal mode.

\mode_if_inner_p: \* \mode_if_inner:TF {⟨true code⟩} {⟨false code⟩}
Detects if \TeX{} is currently in inner mode.

\mode_if_math_p: \* \mode_if_math:TF {⟨true code⟩} {⟨false code⟩}
Detects if \TeX{} is currently in maths mode.

\mode_if_vertical_p: \* \mode_if_vertical:TF {⟨true code⟩} {⟨false code⟩}
Detects if \TeX{} is currently in vertical mode.

9.7 Primitive conditionals

\if_predicate:w ⟨predicate⟩ ⟨true code⟩ \else: ⟨false code⟩ \fi:
This function takes a predicate function and branches according to the result. (In practice
this function would also accept a single boolean variable in place of the ⟨predicate⟩ but
to make the coding clearer this should be done through \if_bool:N.)

\if_bool:N ⟨boolean⟩ ⟨true code⟩ \else: ⟨false code⟩ \fi:
This function takes a boolean variable and branches according to the result.

9.8 Nestable recursions and mappings

There are a number of places where recursion or mapping constructs are used in expl3.
At a low-level, these typically require insertion of tokens at the end of the content to
allow “clean up”. To support such mappings in a nestable form, the following functions
are provided.

\prg_break_point:Nn ⟨type⟩ _map_break: {⟨code⟩}
Used to mark the end of a recursion or mapping: the functions ⟨type⟩ _map_break:
and ⟨type⟩ _map_break:n use this to break out of the loop (see \prg_map_break:Nn
for how to set these up). After the loop ends, the ⟨code⟩ is inserted into the input
stream. This occurs even if the break functions are not applied: \prg_break_point:Nn
is functionally-equivalent in these cases to \use_ii:nn.
\texttt{\textbackslash prg\_map\_break:Nn} \{\langle\textit{user code}\rangle\}

\texttt{\ldots} \texttt{\textbackslash prg\_break\_point:Nn} \{\langle\textit{ending code}\rangle\}

Breaks a recursion in mapping contexts, inserting in the input stream the \langle\textit{user code}\rangle after the \langle\textit{ending code}\rangle for the loop. The function breaks loops, inserting their \langle\textit{ending code}\rangle, until reaching a loop with the same \langle\textit{type}\rangle as its first argument. This \langle\textit{type}\rangle\_\texttt{map\_break:} argument must be defined; it is simply used as a recognizable marker for the \langle\textit{type}\rangle.

For types with mappings defined in the kernel, \langle\textit{type}\rangle\_\texttt{map\_break:} and \langle\textit{type}\rangle\_\texttt{map\_break:n} are defined as \texttt{\prg\_map\_break:Nn} \{\} and the same with \{\} omitted.

### 9.8.1 Simple mappings

In addition to the more complex mappings above, non-nestable mappings are used in a number of locations and support is provided for these.

\texttt{\textbackslash prg\_break:n} \{\langle\textit{code}\rangle\} \ldots \texttt{\textbackslash prg\_break\_point:}

This copy of \texttt{\prg\_do\_nothing:} is used to mark the end of a fast short-term recursion: the function \texttt{\prg\_break:n} uses this to break out of the loop.

\texttt{\prg\_break:n} \{\langle\textit{code}\rangle\} \ldots \texttt{\prg\_break\_point:}

Breaks a recursion which has no \langle\textit{ending code}\rangle and which is not a user-breakable mapping (see for instance \texttt{\prop\_get:Nn}), and inserts the \langle\textit{code}\rangle in the input stream.

### 9.9 Internal programming functions

\texttt{\group\_align\_safe\_begin:} \ldots \texttt{\group\_align\_safe\_end:}

These functions are used to enclose material in a \TeX{} alignment environment within a specially-constructed group. This group is designed in such a way that it does not add brace groups to the output but does act as a group for the \& token inside \texttt{\halign}. This is necessary to allow grabbing of tokens for testing purposes, as \TeX{} uses group level to determine the effect of alignment tokens. Without the special grouping, the use of a function such as \texttt{\peek\_after:Nw} would result in a forbidden comparison of the internal \texttt{\endtemplate} token, yielding a fatal error. Each \texttt{\group\_align\_safe\_begin:} must be matched by a \texttt{\group\_align\_safe\_end:}, although this does not have to occur within the same function.
Chapter 10

The \texttt{l3sys} package: System/runtime functions

10.1 The name of the job

\texttt{\_c\_sys\_jobname\_str}

Constant that gets the “job name” assigned when \TeX\ starts.

\texttt{\TeXhackers\ note}: This copies the contents of the primitive \texttt{\jobname}. For technical reasons, the string here is not of the same internal form as other, but may be manipulated using normal string functions.

10.2 Date and time

\texttt{\_c\_sys\_minute\_int}
\texttt{\_c\_sys\_hour\_int}
\texttt{\_c\_sys\_day\_int}
\texttt{\_c\_sys\_month\_int}
\texttt{\_c\_sys\_year\_int}

The date and time at which the current job was started: these are all reported as integers.

\texttt{\TeXhackers\ note}: Whilst the underlying primitives can be altered by the user, this interface to the time and date is intended to be the “real” values.


### 10.3 Engine

Conditionals which allow engine-specific code to be used. The names follow naturally from those of the engine binaries: note that the \( \mathrm{up} \)\TeX\ tests are for \( \varepsilon \)-\TeX and \( \varepsilon \)-up\TeX\ as expl3 requires the \( \varepsilon \)-\TeX\ extensions. Each conditional is true for exactly one supported engine. In particular, \texttt{\sys_if_engine_ptex_p}: is true for \( \varepsilon \)-\TeX but false for \( \varepsilon \)-up\TeX\.

\begin{verbatim}
\sys_if_engine_luatex_p: \star
\sys_if_engine_luatex: \star\star
\sys_if_engine_pdfTeX_p: \star
\sys_if_engine_pdfTeX: \star\star
\sys_if_engine_ptex_p: \star
\sys_if_engine_ptex: \star\star
\sys_if_engine_upTeX_p: \star
\sys_if_engine_upTeX: \star\star
\sys_if_engine_xetex_p: \star
\sys_if_engine_xetex: \star\star
\end{verbatim}

- \texttt{\sys_if_engine_luatex_p}\: The current engine given as a lower case string: one of \texttt{luatex}, \texttt{pdftex}, \texttt{ptex}, \texttt{uptex} or \texttt{xetex}.
- \texttt{\sys_if_engine_exec_str}\: The name of the standard executable for the current \TeX\ engine given as a lower case string: one of \texttt{luatex}, \texttt{luahbtex}, \texttt{pdftex}, \texttt{eptex}, \texttt{euptex} or \texttt{xetex}.
- \texttt{\sys_if_engine_format_str}\: The name of the preloaded format for the current \TeX\ run given as a lower case string: one of \texttt{lualatex} (or \texttt{dvilualatex}), \texttt{pdflatex} (or \texttt{latex}), \texttt{platex}, \texttt{uplatex} or \texttt{xelatex} for \( \varepsilon \)\TeX, similar names for plain \TeX\ (except \texttt{pdf}\TeX\ in DVI mode yields \texttt{etex}), and \texttt{cont-en} for Con\TeX\ (i.e. the \texttt{\fmtname}).
- \texttt{\sys_timer}: \star
- \texttt{\sys_timer}: \star\star

Expands to the current value of the engine’s timer clock, a non-negative integer. This function is only defined for engines with timer support. This command measures not just CPU time but real time (including time waiting for user input). The unit are scaled seconds (\( 2^{-16} \) seconds).

### 10.4 Output format

Conditionals which give the current output mode the \TeX\ run is operating in. This is always one of two outcomes, DVI mode or PDF mode. The two sets of conditionals are thus complementary and are both provided to allow the programmer to emphasise the most appropriate case.

\begin{verbatim}
\sys_if_output_dvi_p: \star
\sys_if_output_dvi: \star\star
\sys_if_output_pdf_p: \star
\sys_if_output_pdf: \star\star
\end{verbatim}

- \texttt{\sys_if_output_dvi_p}\: The current output mode given as a lower case string: one of \texttt{dvi} or \texttt{pdf}.

---

73
10.5 Platform

\sys_if_platform_unix_p: * \sys_if_platform_unix:TF \{(true code) \{(false code)\}
\sys_if_platform_windows_p: *
\sys_if_platform_windows:TF *

New: 2018-07-27

Conditionals which allow platform-specific code to be used. The names follow the Lua \texttt{os.type()} function, \textit{i.e.} all Unix-like systems are \texttt{unix} (including Linux and MacOS).

\c_sys_platform_str

New: 2018-07-27

The current platform given as a lower case string: one of \texttt{unix}, \texttt{windows} or \texttt{unknown}.

10.6 Random numbers

\sys_rand_seed:

\sys_rand_seed:

Expands to the current value of the engine’s random seed, a non-negative integer. In engines without random number support this expands to 0.

\sys_gset_rand_seed:n \{(intexpr)\}

New: 2017-05-27

Globally sets the seed for the engine’s pseudo-random number generator to the \langle integer expression\rangle. This random seed affects all \ldots_rand functions (such as \int_rand:nn or \clist_rand_item:n) as well as other packages relying on the engine’s random number generator. In engines without random number support this produces an error.

\textbf{TeXhacker note:} While a 32-bit (signed) integer can be given as a seed, only the absolute value is used and any number beyond \(2^{28}\) is divided by an appropriate power of 2. We recommend using an integer in \([0, 2^{28} - 1]\).

10.7 Access to the shell

\sys_get_shell:nnN \{(shell command) \{(setup)\} \{tl var\}\}
\sys_get_shell:nnNTF \{(shell command) \{(setup)\} \{tl var\} \{\{true code\}\} \{\{false code\}\}

New: 2019-09-20

Defines \langle tl var\rangle to the text returned by the \langle shell command\rangle. The \langle shell command\rangle is converted to a string using \texttt{\tl_to_str:n}. Category codes may need to be set appropriately via the \langle setup\rangle argument, which is run just before running the \langle shell command\rangle (in a group). If shell escape is disabled, the \langle tl var\rangle will be set to \texttt{\q_no_value} in the non-branching version. Note that quote characters (*) cannot be used inside the \langle shell command\rangle. The \texttt{\sys_get_shell:nnNTF} conditional inserts the true code if the shell is available and no quote is detected, and the false code otherwise.
This variable exposes the internal triple of the shell escape status. The possible values are:

0    Shell escape is disabled
1    Unrestricted shell escape is enabled
2    Restricted shell escape is enabled

\sys_if_shell_p:
\sys_if_shell:TF {\langle true code\rangle} {\langle false code\rangle}

Performs a check for whether shell escape is enabled. This returns true if either of restricted or unrestricted shell escape is enabled.

\sys_if_shell_unrestricted_p:
\sys_if_shell_unrestricted:TF {\langle true code\rangle} {\langle false code\rangle}

Performs a check for whether unrestricted shell escape is enabled.

\sys_if_shell_restricted_p:
\sys_if_shell_restricted:TF {\langle true code\rangle} {\langle false code\rangle}

Performs a check for whether restricted shell escape is enabled. This returns false if unrestricted shell escape is enabled. Unrestricted shell escape is not considered a superset of restricted shell escape in this case. To find whether any shell escape is enabled use \sys_if_shell:.

\sys_shell_now:n
\sys_shell_now:x

\sys_shell_now:n {\langle tokens\rangle}

Execute {\langle tokens\rangle} through shell escape immediately.

\sys_shell_shipout:n
\sys_shell_shipout:x

\sys_shell_shipout:n {\langle tokens\rangle}

Execute {\langle tokens\rangle} through shell escape at shipout.

10.8  Loading configuration data

\sys_load_backend:n
\sys_load_backend:n {\langle backend\rangle}

Loads the additional configuration file needed for backend support. If the {\langle backend\rangle} is empty, the standard backend for the engine in use will be loaded. This command may only be used once.

\c_sys_backend_str

Set to the name of the backend in use by \sys_load_backend:n when issued.
Load the additional configuration file for debugging support.

10.8.1 Final settings

Finalises all system-dependent functionality: required before loading a backend.
Chapter 11

The l3msg package

Messages

Messages need to be passed to the user by modules, either when errors occur or to indicate how the code is proceeding. The l3msg module provides a consistent method for doing this (as opposed to writing directly to the terminal or log).

The system used by l3msg to create messages divides the process into two distinct parts. Named messages are created in the first part of the process; at this stage, no decision is made about the type of output that the message will produce. The second part of the process is actually producing a message. At this stage a choice of message class has to be made, for example error, warning or info.

By separating out the creation and use of messages, several benefits are available. First, the messages can be altered later without needing details of where they are used in the code. This makes it possible to alter the language used, the detail level and so on. Secondly, the output which results from a given message can be altered. This can be done on a message class, module or message name basis. In this way, message behaviour can be altered and messages can be entirely suppressed.

11.1 Creating new messages

All messages have to be created before they can be used. The text of messages is automatically wrapped to the length available in the console. As a result, formatting is only needed where it helps to show meaning. In particular, \ may be used to force a new line and \ forces an explicit space. Additionally, \, \#, \%, \ and \ can be used to produce the corresponding character.

Messages may be subdivided by one level using the / character. This is used within the message filtering system to allow for example the \TeX kernel messages to belong to the module \TeX while still being filterable at a more granular level. Thus for example

\msg_new:nnnn { mymodule } { submodule / message } ...

will allow to filter out specifically messages from the submodule.
\msg_new:nnnn
\msg_new:nnnn

Updated: 2011-08-16

Creates a \langle message \rangle for a given \langle module \rangle. The message is defined to first give \langle text \rangle and then \langle more text \rangle if the user requests it. If no \langle more text \rangle is available then a standard text is given instead. Within \langle text \rangle and \langle more text \rangle four parameters (#1 to #4) can be used: these will be supplied at the time the message is used. An error is raised if the \langle message \rangle already exists.

\msg_set:nnnn
\msg_set:nnn
\msg_set:nnnn
\msg_gset:nnnn

Sets up the text for a \langle message \rangle for a given \langle module \rangle. The message is defined to first give \langle text \rangle and then \langle more text \rangle if the user requests it. If no \langle more text \rangle is available then a standard text is given instead. Within \langle text \rangle and \langle more text \rangle four parameters (#1 to #4) can be used: these will be supplied at the time the message is used.

\msg_if_exist_p:nn
\msg_if_exist:nnTF

Tests whether the \langle message \rangle for the \langle module \rangle is currently defined.

\msg_module_name:n
\msg_module_name:n

Expands to the public name of the \langle module \rangle as defined by \g_msg_module_name_prop (or otherwise leaves the \langle module \rangle unchanged).

\msg_module_type:n
\msg_module_type:n

Expands to the description which applies to the \langle module \rangle, for example a Package or Class. The information here is defined in \g_msg_module_type_prop, and will default to Package if an entry is not present.

\g_msg_module_name_prop
\g_msg_module_type_prop

Provides a mapping between the module name used for messages, and that for documentation. For example, \LaTeX{}3 core messages are stored in the reserved \LaTeX{} tree, but are printed as \LaTeX{}3.

\g_msg_module_name_prop
\g_msg_module_type_prop

Provides a mapping between the module name used for messages, and that type of module. For example, for \LaTeX{}3 core messages, an empty entry is set here meaning that they are not described using the standard Package text.

\msg_line_context:
\msg_line_context:

Prints the current line number when a message is given, and thus suitable for giving context to messages. The number itself is proceeded by the text on line.

---

11.2 Customizable information for message modules

\msg_module_name:n

Expands to the public name of the \langle module \rangle as defined by \g_msg_module_name_prop (or otherwise leaves the \langle module \rangle unchanged).

\msg_module_type:n

Expands to the description which applies to the \langle module \rangle, for example a Package or Class. The information here is defined in \g_msg_module_type_prop, and will default to Package if an entry is not present.

\g_msg_module_name_prop
\g_msg_module_type_prop

Provides a mapping between the module name used for messages, and that for documentation. For example, \LaTeX{}3 core messages are stored in the reserved \LaTeX{} tree, but are printed as \LaTeX{}3.

11.3 Contextual information for messages

\msg_line_context:

Prints the current line number when a message is given, and thus suitable for giving context to messages. The number itself is proceeded by the text on line.
\texttt{\textbackslash msg\_line\_number: \textbackslash *}

Prints the current line number when a message is given.

\texttt{\textbackslash msg\_fatal\_text:n \{\textit{\textbackslash (module)}\} \textbackslash *

Produces the standard text

\textbf{Fatal Package \textit{\textbackslash (module)} Error}

This function can be redefined to alter the language in which the message is given, using #1 as the name of the \textit{\textbackslash (module)} to be included.

\texttt{\textbackslash msg\_critical\_text:n \{\textit{\textbackslash (module)}\} \textbackslash *

Produces the standard text

\textbf{Critical Package \textit{\textbackslash (module)} Error}

This function can be redefined to alter the language in which the message is given, using #1 as the name of the \textit{\textbackslash (module)} to be included.

\texttt{\textbackslash msg\_error\_text:n \{\textit{\textbackslash (module)}\} \textbackslash *

Produces the standard text

\textbf{Package \textit{\textbackslash (module)} Error}

This function can be redefined to alter the language in which the message is given, using #1 as the name of the \textit{\textbackslash (module)} to be included.

\texttt{\textbackslash msg\_warning\_text:n \{\textit{\textbackslash (module)}\} \textbackslash *

Produces the standard text

\textbf{Package \textit{\textbackslash (module)} Warning}

This function can be redefined to alter the language in which the message is given, using #1 as the name of the \textit{\textbackslash (module)} to be included. The \textit{\textless type\textgreater} of \textit{\textbackslash (module)} may be adjusted: \textbf{Package} is the standard outcome: see \texttt{\textbackslash msg\_module\_type:n}.

\texttt{\textbackslash msg\_info\_text:n \{\textit{\textbackslash (module)}\} \textbackslash *

Produces the standard text:

\textbf{Package \textit{\textbackslash (module)} Info}

This function can be redefined to alter the language in which the message is given, using #1 as the name of the \textit{\textbackslash (module)} to be included. The \textit{\textless type\textgreater} of \textit{\textbackslash (module)} may be adjusted: \textbf{Package} is the standard outcome: see \texttt{\textbackslash msg\_module\_type:n}. 

79
\msg_see_documentation_text:n \module

Produces the standard text
See the \module documentation for further information.

This function can be redefined to alter the language in which the message is given, using #1 as the name of the \module to be included. The name of the \module is produced using \msg_module_name:n.

11.4 Issuing messages

Messages behave differently depending on the message class. In all cases, the message may be issued supplying 0 to 4 arguments. If the number of arguments supplied here does not match the number in the definition of the message, extra arguments are ignored, or empty arguments added (of course the sense of the message may be impaired). The four arguments are converted to strings before being added to the message text: the x-type variants should be used to expand material. Note that this expansion takes place with the standard definitions in effect, which means that shorthands such as \- or \ are not available; instead one should use \iow_char:N \- and \iow_newline:, respectively. The following message classes exist:

- fatal, ending the TEX run;
- critical, ending the file being input;
- error, interrupting the TEX run without ending it;
- warning, written to terminal and log file, for important messages that may require corrections by the user;
- note (less common than info) for important information messages written to the terminal and log file;
- info for normal information messages written to the log file only;
- term and log for un-decorated messages written to the terminal and log file, or to the log file only;
- none for suppressed messages.

\msg_fatal:nnnnnn \module \message \arg one \arg two \arg three \arg four

Issues \module error \message, passing \arg one to \arg four to the text-creating functions. After issuing a fatal error the TEX run halts. No PDF file will be produced in this case (DVI mode runs may produce a truncated DVI file).

Updated: 2012-08-11
\msg_critical:nnnnn {\langle module \rangle \langle message \rangle \langle arg one \rangle \langle arg two \rangle \langle arg three \rangle \langle arg four \rangle}

Issues \langle module \rangle error \langle message \rangle, passing \langle arg one \rangle to \langle arg four \rangle to the text-creating functions. After issuing a critical error, \TeX\ stops reading the current input file. This may halt the \TeX\ run (if the current file is the main file) or may abort reading a sub-file.

\TeX\ hackers note: The \TeX\ \texttt{\endinput} primitive is used to exit the file. In particular, the rest of the current line remains in the input stream.

\msg_error:nnnnn {\langle module \rangle \langle message \rangle \langle arg one \rangle \langle arg two \rangle \langle arg three \rangle \langle arg four \rangle}

Issues \langle module \rangle error \langle message \rangle, passing \langle arg one \rangle to \langle arg four \rangle to the text-creating functions. The error interrupts processing and issues the text at the terminal. After user input, the run continues.

\msg_warning:nnnnn {\langle module \rangle \langle message \rangle \langle arg one \rangle \langle arg two \rangle \langle arg three \rangle \langle arg four \rangle}

Issues \langle module \rangle warning \langle message \rangle, passing \langle arg one \rangle to \langle arg four \rangle to the text-creating functions. The warning text is added to the log file and the terminal, but the \TeX\ run is not interrupted.
Issues \langle \text{module} \rangle \text{information} \langle \text{message} \rangle, passing \langle \text{arg one} \rangle to \langle \text{arg four} \rangle to the text-creating functions. For the more common \msg_info:nnnnnn, the information text is added to the log file only, while \msg_note:nnnnnn adds the info text to both the log file and the terminal. The \TeX{} run is not interrupted.

\msg_note:nnnnnn \{\langle \text{module} \rangle\} \{\langle \text{message} \rangle\} \{\langle \text{arg one} \rangle\} \{\langle \text{arg two} \rangle\} \{\langle \text{arg three} \rangle\} \{\langle \text{arg four} \rangle\}
\msg_info:nnnnnn \{\langle \text{module} \rangle\} \{\langle \text{message} \rangle\} \{\langle \text{arg one} \rangle\} \{\langle \text{arg two} \rangle\} \{\langle \text{arg three} \rangle\} \{\langle \text{arg four} \rangle\}

Updated: 2012-08-11

Issues \langle \text{module} \rangle \text{information} \langle \text{message} \rangle, passing \langle \text{arg one} \rangle to \langle \text{arg four} \rangle to the text-creating functions. The output is briefer than \msg_info:nnnnnn, omitting for instance the module name. It is added to the log file by \msg_log:nnnnnn while \msg_term:nnnnnn also prints it on the terminal.

\msg_term:nnnnnn \{\langle \text{module} \rangle\} \{\langle \text{message} \rangle\} \{\langle \text{arg one} \rangle\} \{\langle \text{arg two} \rangle\} \{\langle \text{arg three} \rangle\} \{\langle \text{arg four} \rangle\}
\msg_log:nnnnnn \{\langle \text{module} \rangle\} \{\langle \text{message} \rangle\} \{\langle \text{arg one} \rangle\} \{\langle \text{arg two} \rangle\} \{\langle \text{arg three} \rangle\} \{\langle \text{arg four} \rangle\}

Updated: 2012-08-11

\msg_none:nnnnnn \{\langle \text{module} \rangle\} \{\langle \text{message} \rangle\} \{\langle \text{arg one} \rangle\} \{\langle \text{arg two} \rangle\} \{\langle \text{arg three} \rangle\} \{\langle \text{arg four} \rangle\}

Does nothing: used as a message class to prevent any output at all (see the discussion of message redirection).

\msg_none:nnnnnn \{\langle \text{module} \rangle\} \{\langle \text{message} \rangle\} \{\langle \text{arg one} \rangle\} \{\langle \text{arg two} \rangle\} \{\langle \text{arg three} \rangle\} \{\langle \text{arg four} \rangle\}

Updated: 2012-08-11
11.4.1 Messages for showing material

\msg{nnnn}{module} {message} {arg one} {arg two} {arg three} {arg four}

Issues \langle module \rangle information \langle message \rangle, passing \langle arg one \rangle to \langle arg four \rangle to the text-creating functions. The information text is shown on the terminal and the \TeX{} run is interrupted in a manner similar to \texttt{tl\_show:n}. This is used in conjunction with \msg{nnnn}{item:n} and similar functions to print complex variable contents completely. If the formatted text does not contain >\, at the start of a line, an additional line >-, will be put at the end. In addition, a final period is added if not present.

11.4.2 Expandable error messages

In very rare cases it may be necessary to produce errors in an expansion-only context. The functions in this section should only be used if there is no alternative approach using \texttt{msg\_error:nnnnn} or other non-expandable commands from the previous section. Despite having a similar interface as non-expandable messages, expandable errors must be handled internally very differently from normal error messages, as none of the tools to print to the terminal or the log file are expandable. As a result, short-hands such as \texttt{\{ or \\} do not work, and messages must be very short (with default settings, they are truncated after approximately 50 characters). It is advisable to ensure that the message is understandable even when truncated, by putting the most important information up front. Another particularity of expandable messages is that they cannot be redirected or turned off by the user.

\msg{nnnnn}{expandable_error}{module} {message} {arg one} {arg two} {arg three} {arg four}

Issues an “Undefined error” message from \TeX{} itself using the undefined control sequence \texttt{\::error} then prints “! \langle module \rangle: "\langle error message \rangle", which should be short. With default settings, anything beyond approximately 60 characters long (or bytes in some engines) is cropped. A leading space might be removed as well.

11.5 Redirecting messages

Each message has a “name”, which can be used to alter the behaviour of the message when it is given. Thus we might have

\msg{nnnn}{new}{module} {my-message} {Some-text} {Some-more-text}
to define a message, with

\msg_error:nn { module } { my-message }

when it is used. With no filtering, this raises an error. However, we could alter the
behaviour with

\msg_redirect_class:nn { error } { warning }

to turn all errors into warnings, or with

\msg_redirect_module:nnn { module } { error } { warning }

to alter only messages from that module, or even

\msg_redirect_name:nnn { module } { my-message } { warning }

to target just one message. Redirection applies first to individual messages, then to
messages from one module and finally to messages of one class. Thus it is possible to
select out an individual message for special treatment even if the entire class is already
redirected.

Multiple redirections are possible. Redirections can be cancelled by providing an
empty argument for the target class. Redirection to a missing class raises an error
immediately. Infinite loops are prevented by eliminating the redirection starting from
the target of the redirection that caused the loop to appear. Namely, if redirections are
requested as \(A \rightarrow B, B \rightarrow C\) and \(C \rightarrow A\) in this order, then the \(A \rightarrow B\) redirection is
cancelled.

---

\msg_redirect_class:nn

Updated: 2012-04-27

\msg_redirect_class:nn {⟨class one⟩} {⟨class two⟩}
Changes the behaviour of messages of ⟨class one⟩ so that they are processed using the
code for those of ⟨class two⟩. Each ⟨class⟩ can be one of fatal, critical, error, warning, note, info, term, log, none.

\msg_redirect_module:nnn

Updated: 2012-04-27

\msg_redirect_module:nnn {⟨module⟩} {⟨class one⟩} {⟨class two⟩}
Redirects message of ⟨class one⟩ for ⟨module⟩ to act as though they were from ⟨class
two⟩. Messages of ⟨class one⟩ from sources other than ⟨module⟩ are not affected by this
redirection. This function can be used to make some messages “silent” by default. For
example, all of the warning messages of ⟨module⟩ could be turned off with:

\msg_redirect_module:nnn { module } { warning } { none }

---

\msg_redirect_name:nnn

Updated: 2012-04-27

\msg_redirect_name:nnn {⟨module⟩} {⟨message⟩} {⟨class⟩}
Redirects a specific ⟨message⟩ from a specific ⟨module⟩ to act as a member of ⟨class⟩ of
messages. No further redirection is performed. This function can be used to make a
selected message “silent” without changing global parameters:

\msg_redirect_name:nnn { module } { annoying-message } { none }

84
Chapter 12

The l3file package
File and I/O operations

This module provides functions for working with external files. Some of these functions apply to an entire file, and have prefix `\file_...`, while others are used to work with files on a line by line basis and have prefix `\ior_...` (reading) or `\iow_...` (writing).

It is important to remember that when reading external files \TeX\ attempts to locate them using both the operating system path and entries in the \TeX\ file database (most \TeX\ systems use such a database). Thus the “current path” for \TeX\ is somewhat broader than that for other programs.

For functions which expect a ⟨file name⟩ argument, this argument may contain both literal items and expandable content, which should on full expansion be the desired file name. Active characters (as declared in `\l_char_active_seq`) are not expanded, allowing the direct use of these in file names. Quote tokens (") are not permitted in file names as they are reserved for internal use by some \TeX\ primitives.

Spaces are trimmed at the beginning and end of the file name: this reflects the fact that some file systems do not allow or interact unpredictably with spaces in these positions. When no extension is given, this will trim spaces from the start of the name only.

12.1 Input–output stream management

As \TeX\ engines have a limited number of input and output streams, direct use of the streams by the programmer is not supported in \LaTeX3. Instead, an internal pool of streams is maintained, and these are allocated and deallocated as needed by other modules. As a result, the programmer should close streams when they are no longer needed, to release them for other processes.

Note that I/O operations are global: streams should all be declared with global names and treated accordingly.
Globally reserves the name of the \langle stream\rangle, either for reading or for writing as appropriate. The \langle stream\rangle is not opened until the appropriate \texttt{\\ldots open:Nn} function is used. Attempting to use a \langle stream\rangle which has not been opened is an error, and the \langle stream\rangle will behave as the corresponding \texttt{\_term\ldots}.

\ior_open:Nn\{\langle file name\rangle\}
Opens \langle file name\rangle for reading using \langle stream\rangle as the control sequence for file access. If the \langle stream\rangle was already open it is closed before the new operation begins. The \langle stream\rangle is available for access immediately and will remain allocated to \langle file name\rangle until a \ior_close:N instruction is given or the T\TeX\ run ends. If the file is not found, an error is raised.

\ior_open:NnTF\{\langle file name\rangle\}\{\langle true code\rangle\}\{\langle false code\rangle\}
Opens \langle file name\rangle for reading using \langle stream\rangle as the control sequence for file access. If the \langle stream\rangle was already open it is closed before the new operation begins. The \langle stream\rangle is available for access immediately and will remain allocated to \langle file name\rangle until a \ior_close:N instruction is given or the T\TeX\ run ends. The \langle true code\rangle is then inserted into the input stream. If the file is not found, no error is raised and the \langle false code\rangle is inserted into the input stream.

\ior_open:cn\{\langle file name\rangle\}
Opens \langle file name\rangle for writing using \langle stream\rangle as the control sequence for file access. If the \langle stream\rangle was already open it is closed before the new operation begins. The \langle stream\rangle is available for access immediately and will remain allocated to \langle file name\rangle until a \iow_close:N instruction is given or the T\TeX\ run ends. Opening a file for writing clears any existing content in the file (\textit{i.e.} writing is not additive).

\ior_close:N\{\langle stream\rangle\}
Closes the \langle stream\rangle. Streams should always be closed when they are finished with as this ensures that they remain available to other programmers.

\ior_show:N\{\langle stream\rangle\}
Display (to the terminal or log file) the file name associated to the \langle read or write\rangle \langle stream\rangle.
Display (to the terminal or log file) a list of the file names associated with each open (read or write) stream. This is intended for tracking down problems.

12.1.1 Reading from files

Reading from files and reading from the terminal are separate processes in expl3. The functions \ior_get:NN and \ior_str_get:NN, and their branching equivalents, are designed to work with files.

\ior_get:NN \ior_get:NNTF

\ior_get:NN \langle stream \rangle \langle token list variable \rangle
\ior_get:NNTF \langle stream \rangle \langle token list variable \rangle \langle true code \rangle \langle false code \rangle

Function that reads one or more lines (until an equal number of left and right braces are found) from the file input \langle stream \rangle and stores the result locally in the \langle token list \rangle variable. The material read from the \langle stream \rangle is tokenized by \TeX\ according to the category codes and \endlinechar in force when the function is used. Assuming normal settings, any lines which do not end in a comment character \% have the line ending converted to a space, so for example input

\begin{verbatim}
a b c
\end{verbatim}

results in a token list \texttt{a,b,c}. Any blank line is converted to the token \texttt{\par}. Therefore, blank lines can be skipped by using a test such as

\begin{verbatim}
\ior_get:NN \l_my_stream \l_tmpa_tl
\tl_set:Nn \l_tmpb_tl { \par }
\tl_if_eq:NNF \l_tmpa_tl \l_tmpb_tl
\end{verbatim}

Also notice that if multiple lines are read to match braces then the resulting token list can contain \texttt{\par} tokens. In the non-branching version, where the \langle stream \rangle is not open the \langle tl var \rangle is set to \texttt{\q_no_value}.

\TeX\hackers note: This protected macro is a wrapper around the \TeX\ primitive \texttt{\read}. Regardless of settings, \TeX\ replaces trailing space and tab characters (character codes 32 and 9) in each line by an end-of-line character (character code \texttt{\endlinechar}, omitted if \texttt{\endlinechar} is negative or too large) before turning characters into tokens according to current category codes. With default settings, spaces appearing at the beginning of lines are also ignored.
\ior_str_get:NN \ior_str_get:NN \tokenlist \ior_str_get:NNTF \ior_str_get:NN \tokenlist \true \false

Function that reads one line from the file input \stream and stores the result locally in the \tokenlist variable. The material is read from the \stream as a series of tokens with category code 12 (other), with the exception of space characters which are given category code 10 (space). Multiple whitespace characters are retained by this process. It always only reads one line and any blank lines in the input result in the \tokenlist variable being empty. Unlike \ior_get:NN, line ends do not receive any special treatment. Thus input

\begin{verbatim}
a b c
\end{verbatim}

results in a token list \verb+a b c+ with the letters a, b, and c having category code 12. In the non-branching version, where the \stream is not open the \tlvar is set to \texttt{q_no_value}.

\textbf{\TeXhacksnote:} This protected macro is a wrapper around the \TeX primitive \texttt{\readline}. Regardless of settings, \TeX removes trailing space and tab characters (character codes 32 and 9). However, the end-line character normally added by this primitive is not included in the result of \ior_str_get:NN.

All mappings are done at the current group level, \ie any local assignments made by the \langle function \rangle or \langle code \rangle discussed below remain in effect after the loop.

\ior_map_inline:Nn \ior_map_inline:Nn \langle inline function \rangle

Applies the \langle inline function \rangle to each set of \langle lines \rangle obtained by calling \ior_get:NN until reaching the end of the file. \TeX ignores any trailing new-line marker from the file it reads. The \langle inline function \rangle should consist of code which receives the \langle line \rangle as \#1.

\ior_str_map_inline:Nn \ior_str_map_inline:Nn \langle inline function \rangle

Applies the \langle inline function \rangle to every \langle line \rangle in the \langle stream \rangle. The material is read from the \langle stream \rangle as a series of tokens with category code 12 (other), with the exception of space characters which are given category code 10 (space). The \langle inline function \rangle should consist of code which receives the \langle line \rangle as \#1. Note that \TeX removes trailing space and tab characters (character codes 32 and 9) from every line upon input. \TeX also ignores any trailing new-line marker from the file it reads.

\ior_map_variable:NNn \ior_map_variable:NNn \langle stream \rangle \tlvar \langle code \rangle

For each set of \langle lines \rangle obtained by calling \ior_get:NN until reaching the end of the file, stores the \langle lines \rangle in the \tlvar then applies the \langle code \rangle. The \langle code \rangle will usually make use of the \langle variable \rangle, but this is not enforced. The assignments to the \langle variable \rangle are local. Its value after the loop is the last set of \langle lines \rangle, or its original value if the \langle stream \rangle is empty. \TeX ignores any trailing new-line marker from the file it reads. This function is typically faster than \ior_map_inline:Nn.
\ior_str_map_variable:NNn \ior_str_map_variable:NNn \langle stream \rangle \langle variable \rangle \{ \langle code \rangle \}

For each \langle line \rangle in the \langle stream \rangle, stores the \langle line \rangle in the \langle variable \rangle then applies the \langle code \rangle. The material is read from the \langle stream \rangle as a series of tokens with category code 12 (other), with the exception of space characters which are given category code 10 (space). The \langle code \rangle will usually make use of the \langle variable \rangle, but this is not enforced. The assignments to the \langle variable \rangle are local. Its value after the loop is the last \langle line \rangle, or its original value if the \langle stream \rangle is empty. Note that \TeX\ removes trailing space and tab characters (character codes 32 and 9) from every line upon input. \TeX\ also ignores any trailing new-line marker from the file it reads. This function is typically faster than \ior_str_map_inline:Nn.

\ior_map_break: \ior_map_break:

Used to terminate a \ior_map_... function before all lines from the \langle stream \rangle have been processed. This normally takes place within a conditional statement, for example

\ior_map_inline:Nn \l_my_ior
{ \str_if_eq:nnTF { #1 } { bingo } { \ior_map_break: } { \% Do something useful } }

Use outside of a \ior_map_... scenario leads to low level \TeX\ errors.

\TeXhackers note: When the mapping is broken, additional tokens may be inserted before further items are taken from the input stream. This depends on the design of the mapping function.

\ior_map_break:n \ior_map_break:n \langle code \rangle

Used to terminate a \ior_map_... function before all lines in the \langle stream \rangle have been processed, inserting the \langle code \rangle after the mapping has ended. This normally takes place within a conditional statement, for example

\ior_map_inline:Nn \l_my_ior
{ \str_if_eq:nnTF { #1 } { bingo } { \ior_map_break:n \langle code \rangle } { \% Do something useful } }

Use outside of a \ior_map_... scenario leads to low level \TeX\ errors.

\TeXhackers note: When the mapping is broken, additional tokens may be inserted before the \langle code \rangle is inserted into the input stream. This depends on the design of the mapping function.
Tests if the end of a file \langle stream \rangle has been reached during a reading operation. The test also returns a true value if the \langle stream \rangle is not open.

### 12.1.2 Writing to files

\texttt{\textbackslash iow\_now:Nn} \langle \texttt{\textbackslash iow\_now:}(\texttt{Nx}\mid \texttt{cn}\mid \texttt{cx}) \rangle

This functions writes \langle tokens \rangle to the specified \langle stream \rangle immediately (\textit{i.e.} the write operation is called on expansion of \texttt{\textbackslash iow\_now:Nn}).

\texttt{\textbackslash iow\_log:n} \langle \texttt{\textbackslash iow\_log:x} \rangle

This function writes the given \langle tokens \rangle to the log (transcript) file immediately: it is a dedicated version of \texttt{\textbackslash iow\_now:Nn}.

\texttt{\textbackslash iow\_term:n} \langle \texttt{\textbackslash iow\_term:x} \rangle

This function writes the given \langle tokens \rangle to the terminal file immediately: it is a dedicated version of \texttt{\textbackslash iow\_now:Nn}.

\texttt{\textbackslash iow\_shipout:Nn} \langle \texttt{\textbackslash iow\_shipout:}(\texttt{Nx}\mid \texttt{cn}\mid \texttt{cx}) \rangle

This functions writes \langle tokens \rangle to the specified \langle stream \rangle when the current page is finalised (\textit{i.e.} at shipout). The x-type variants expand the \langle tokens \rangle at the point where the function is used but \textit{not} when the resulting tokens are written to the \langle stream \rangle (cf. \texttt{\textbackslash iow\_shipout\_x:Nn}).

\textbf{\TeX}Xhackers note: When using expl3 with a format other than \LaTeX, new line characters inserted using \texttt{\textbackslash iow\_newline}: or using the line-wrapping code \texttt{\textbackslash iow\_wrap:mmN} are not recognized in the argument of \texttt{\textbackslash iow\_shipout:Nn}. This may lead to the insertion of additional unwanted line-breaks.

\texttt{\textbackslash iow\_shipout\_x:Nn} \langle \texttt{\textbackslash iow\_shipout\_x:}(\texttt{Nx}\mid \texttt{cn}\mid \texttt{cx}) \rangle

This functions writes \langle tokens \rangle to the specified \langle stream \rangle when the current page is finalised (\textit{i.e.} at shipout). The \langle tokens \rangle are expanded at the time of writing in addition to any expansion when the function is used. This makes these functions suitable for including material finalised during the page building process (such as the page number integer).

\textbf{\TeX}Xhackers note: This is a wrapper around the \TeX\ primitive \texttt{\textbackslash write}. When using expl3 with a format other than \LaTeX, new line characters inserted using \texttt{\textbackslash iow\_newline}: or using the line-wrapping code \texttt{\textbackslash iow\_wrap:mmN} are not recognized in the argument of \texttt{\textbackslash iow\_shipout:Nn}. This may lead to the insertion of additional unwanted line-breaks.
\texttt{\iochar{\textbackslash char}} \texttt{\iowchar{\textbackslash char}}

Inserts \texttt{\textbackslash char} into the output stream. Useful when trying to write difficult characters such as \texttt{\%}, \texttt{\{}\texttt{\}}, \textit{etc.} in messages, for example:

\texttt{\iownowx{} \g_my_iow \{ \iochar{\textbackslash} \{ text \iochar{\textbackslash} \} \} }

The function has no effect if writing is taking place without expansion (\textit{e.g.} in the second argument of \texttt{\iownow{n}}).

\texttt{\newline}

\texttt{\newline}

Function to add a new line within the \texttt{(tokens)} written to a file. The function has no effect if writing is taking place without expansion (\textit{e.g.} in the second argument of \texttt{\iownow{n}}).

\textbf{\LaTeX{}hackers note}: When using expl3 with a format other than \LaTeX{}, the character inserted by \texttt{\newline} is not recognized by \LaTeX{}, which may lead to the insertion of additional unwanted line-breaks. This issue only affects \texttt{\shipout{N}{n}}, \texttt{\shipoutx{X}{N}} and direct uses of primitive operations.
12.1.3 Wrapping lines in output

This function wraps the ⟨text⟩ to a fixed number of characters per line. At the start of each line which is wrapped, the ⟨run-on text⟩ is inserted. The line character count targeted is the value of \l_{\text{iow\_line\_count\_int}} minus the number of characters in the ⟨run-on text⟩ for all lines except the first, for which the target number of characters is simply \l_{\text{iow\_line\_count\_int}} since there is no run-on text. The ⟨text⟩ and ⟨run-on text⟩ are exhaustively expanded by the function, with the following substitutions:

- \ or \texttt{\textbackslash new\_line}: may be used to force a new line,
- \texttt{\textbackslash}: may be used to represent a forced space (for example after a control sequence),
- \#, \%, \{, \}, \texttt{\textbackslash -} may be used to represent the corresponding character,
- \texttt{\textbackslash allow\_break}: may be used to allow a line-break without inserting a space (this is experimental),
- \texttt{\textbackslash indent:n} may be used to indent a part of the ⟨text⟩ (not the ⟨run-on text⟩).

Additional functions may be added to the wrapping by using the ⟨set up⟩, which is executed before the wrapping takes place: this may include overriding the substitutions listed.

Any expandable material in the ⟨text⟩ which is not to be expanded on wrapping should be converted to a string using \token_to_str:N, \tl_to_str:n, \tl_to_str:N, etc.

The result of the wrapping operation is passed as a braced argument to the ⟨function⟩, which is typically a wrapper around a write operation. The output of \iow_wrap:nnnN (i.e. the argument passed to the ⟨function⟩) consists of characters of category “other” (category code 12), with the exception of spaces which have category “space” (category code 10). This means that the output does \texttt{not} expand further when written to a file.

\texttt{\TeXhacks{note}}: Internally, \iow_wrap:nnnN carries out an \texttt{x}-type expansion on the ⟨text⟩ to expand it. This is done in such a way that \exp_not:N or \exp_not:n \texttt{could} be used to prevent expansion of material. However, this is less conceptually clear than conversion to a string, which is therefore the supported method for handling expandable material in the ⟨text⟩.

\begin{align*}
\texttt{\textbackslash indent:n} & \texttt{(⟨text⟩)} \\
\texttt{\textbackslash indent:n} & \texttt{(⟨text⟩)}
\end{align*}

In the first argument of \iow_wrap:nnnN (for instance in messages), indents ⟨text⟩ by four spaces. This function does not cause a line break, and only affects lines which start within the scope of the ⟨text⟩. In case the indented ⟨text⟩ should appear on separate lines from the surrounding text, use \texttt{\textbackslash} to force line breaks.

\begin{align*}
\texttt{\textbackslash l_{\text{iow\_line\_count\_int}}} & \\
\texttt{\textbackslash l_{\text{iow\_line\_count\_int}}}
\end{align*}

The maximum number of characters in a line to be written by the \iow_wrap:nnnN function. This value depends on the \TeX system in use: the standard value is 78, which is typically correct for unmodified \TeX\ Live and MiK\TeX systems.
12.1.4 Constant input–output streams, and variables

Scratch input stream for global use. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_ior
\g_tmpb_ior

New: 2017-12-11

Constant output streams for writing to the log and to the terminal (plus the log), respectively.

\c_log_iow
\c_term_iow

Scratch output stream for global use. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_iow
\g_tmpb_iow

New: 2017-12-11

12.1.5 Primitive conditionals

\if_eof:w  \if_eof:w ⟨stream⟩
 ⟨true code⟩
\else:
 ⟨false code⟩
\fi:
Tests if the ⟨stream⟩ returns “end of file”, which is true for non-existent files. The \else: branch is optional.

\TeXhackers note: This is the \TeX primitive \ifeof.

12.2 File operation functions

Contain the directory, name and extension of the current file. The directory is empty if the file was loaded without an explicit path (i.e. if it is in the \TeX search path), and does not end in / other than the case that it is exactly equal to the root directory. The ⟨name⟩ and ⟨ext⟩ parts together make up the file name, thus the ⟨name⟩ part may be thought of as the “job name” for the current file. Note that \TeX does not provide information on the ⟨ext⟩ part for the main (top level) file and that this file always has an empty ⟨dir⟩ component. Also, the ⟨name⟩ here will be equal to \c_sys_jobname_str, which may be different from the real file name (if set using --jobname, for example).
Each entry is the path to a directory which should be searched when seeking a file. Each path can be relative or absolute, and should not include the trailing slash. The entries are not expanded when used so may contain active characters but should not feature any variable content. Spaces need not be quoted.

\TeXhackers note: When working as a package in \LaTeX\ expl3 will automatically append the current \input@path to the set of values from \l_file_search_path_seq.

\file_if_exist:nTF \{⟨file name⟩\} \{⟨true code⟩\} \{⟨false code⟩\}

Searches for ⟨file name⟩ using the current \TeX search path and the additional paths controlled by \l_file_search_path_seq.

\file_get:nnN \{⟨filename⟩\} \{⟨setup⟩\} \{⟨tl⟩\}
\file_get:nnNTF \{⟨filename⟩\} \{⟨setup⟩\} \{⟨true code⟩\} \{⟨false code⟩\}

Defines ⟨tl⟩ to the contents of ⟨filename⟩. Category codes may need to be set appropriately via the ⟨setup⟩ argument. The non-branching version sets the ⟨tl⟩ to \texttt{\textendash no_value} if the file is not found. The branching version runs the ⟨true code⟩ after the assignment to ⟨tl⟩ if the file is found, and ⟨false code⟩ otherwise.

\file_get_full_name:nN \{⟨file name⟩\} \{⟨tl⟩\}
\file_get_full_name:nNTF \{⟨file name⟩\} \{⟨true code⟩\} \{⟨false code⟩\}

Searches for ⟨file name⟩ in the path as detailed for \file_if_exist:nTF, and if found sets the ⟨tl var⟩ the fully-qualified name of the file, \textit{i.e.} the path and file name. This includes an extension \texttt{.tex} when the given ⟨file name⟩ has no extension but the file found has that extension. In the non-branching version, the ⟨tl var⟩ will be set to \texttt{\textendash no_value} in the case that the file does not exist.

\file_full_name:n \{⟨file name⟩\}

Searches for ⟨file name⟩ in the path as detailed for \file_if_exist:nTF, and if found leaves the fully-qualified name of the file, \textit{i.e.} the path and file name, in the input stream. This includes an extension \texttt{.tex} when the given ⟨file name⟩ has no extension but the file found has that extension. If the file is not found on the path, the expansion is empty.
\file_parse_full_name:nNNN \file_parse_full_name:nNNN \file_parse_full_name:VNNN

Parses the \texttt{\langle full name\rangle} and splits it into three parts, each of which is returned by setting the appropriate local string variable:

- The \texttt{\langle dir\rangle}: everything up to the last / (path separator) in the \texttt{\langle file path\rangle}. As with system PATH variables and related functions, the \texttt{\langle dir\rangle} does not include the trailing / unless it points to the root directory. If there is no path (only a file name), \texttt{\langle dir\rangle} is empty.

- The \texttt{\langle name\rangle}: everything after the last / up to the last ., where both of those characters are optional. The \texttt{\langle name\rangle} may contain multiple . characters. It is empty if \texttt{\langle full name\rangle} consists only of a directory name.

- The \texttt{\langle ext\rangle}: everything after the last . (including the dot). The \texttt{\langle ext\rangle} is empty if there is no . after the last /.

Before parsing, the \texttt{\langle full name\rangle} is expanded until only non-expandable tokens remain, except that active characters are also not expanded. Quotes (") are invalid in file names and are discarded from the input.

\file_parse_full_name:n \file_parse_full_name:n

\file_parse_full_name_apply:nN \file_parse_full_name_apply:nN

Parses the \texttt{\langle full name\rangle} as described for \file_parse_full_name:nNNN, and leaves \texttt{\langle dir\rangle}, \texttt{\langle name\rangle}, and \texttt{\langle ext\rangle} in the input stream, each inside a pair of braces.

\file_hex_dump:n \file_hex_dump:nnn

\file_hex_dump:n \file_hex_dump:nnn

Searches for \texttt{\langle file name\rangle} using the current \TeX{} search path and the additional paths controlled by \file_search_path_seq. It then expands to leave the hexadecimal dump of the file content in the input stream. The file is read as bytes, which means that in contrast to most \TeX{} behaviour there will be a difference in result depending on the line endings used in text files. The same file will produce the same result between different engines: the algorithm used is the same in all cases. When the file is not found, the result of expansion is empty. The \texttt{\langle start index\rangle} and \texttt{\langle end index\rangle} values work as described for \texttt{\langle start index\rangle} and \texttt{\langle end index\rangle}.

\file_get_hex_dump:nN \file_get_hex_dump:nnnN

Sets the \texttt{\langle tl var\rangle} to the result of applying \file_hex_dump:n/\file_hex_dump:nnn to the \texttt{\langle file\rangle}. If the file is not found, the \texttt{\langle tl var\rangle} will be set to \texttt{\langle q_no_value\rangle}.
\file_mdfive_hash:n \(\langle\text{name}\rangle\)  

Searches for \(\langle\text{name}\rangle\) using the current \TeX{} search path and the additional paths controlled by \l_file_search_path_seq. It then expands to leave the MD5 sum generated from the contents of the file in the input stream. The file is read as bytes, which means that in contrast to most \TeX{} behaviour there will be a difference in result depending on the line endings used in text files. The same file will produce the same result between different engines: the algorithm used is the same in all cases. When the file is not found, the result of expansion is empty.

\file_get_mdfive_hash:nN \(\langle\text{name}\rangle\) \(\langle\text{tl var}\rangle\)  

Sets the \(\langle\text{tl var}\rangle\) to the result of applying \file_mdfive_hash:n to the \(\langle\text{file}\rangle\). If the file is not found, the \(\langle\text{tl var}\rangle\) will be set to \q_no_value.

\file_size:n \(\langle\text{name}\rangle\)  

Searches for \(\langle\text{name}\rangle\) using the current \TeX{} search path and the additional paths controlled by \l_file_search_path_seq. It then expands to leave the size of the file in bytes in the input stream. When the file is not found, the result of expansion is empty.

\file_get_size:nN \(\langle\text{name}\rangle\) \(\langle\text{tl var}\rangle\)  

Sets the \(\langle\text{tl var}\rangle\) to the result of applying \file_size:n to the \(\langle\text{file}\rangle\). If the file is not found, the \(\langle\text{tl var}\rangle\) will be set to \q_no_value. This is not available in older versions of Xe\TeX{}.

\file_timestamp:n \(\langle\text{name}\rangle\)  

Searches for \(\langle\text{name}\rangle\) using the current \TeX{} search path and the additional paths controlled by \l_file_search_path_seq. It then expands to leave the modification timestamp of the file in the input stream. The timestamp is of the form \(D:\langle\text{year}\rangle:\langle\text{month}\rangle:\langle\text{day}\rangle:\langle\text{hour}\rangle:\langle\text{minute}\rangle:\langle\text{second}\rangle:\langle\text{offset}\rangle\), where the latter may be \(Z\) (UTC) or \(\langle\text{plus-minus}\rangle:\langle\text{hours}\rangle:\langle\text{minutes}\rangle\). When the file is not found, the result of expansion is empty. This is not available in older versions of Xe\TeX{}.

\file_get_timestamp:nN \(\langle\text{name}\rangle\) \(\langle\text{tl var}\rangle\)  

Sets the \(\langle\text{tl var}\rangle\) to the result of applying \file_timestamp:n to the \(\langle\text{file}\rangle\). If the file is not found, the \(\langle\text{tl var}\rangle\) will be set to \q_no_value. This is not available in older versions of Xe\TeX{}.

96
\file_compare_timestamp:nNn \file_compare_timestamp:nNnTF

\file_compare_timestamp:nNn \{file-1\} \{comparator\} \{file-2\} \{true code\} \{false code\}

Compares the file stamps on the two \textit{files} as indicated by the \textit{comparator}, and inserts either the \textit{true code} or \textit{false code} as required. A file which is not found is treated as older than any file which is found. This allows for example the construct

\begin{verbatim}
\file_compare_timestamp:nNnT \{ source-file \} > \{ derived-file \}
{ % Code to regenerate derived file }
\end{verbatim}

to work when the derived file is entirely absent. The timestamp of two absent files is regarded as different. This is not available in older versions of Xe\TeX.

\file_compare_timestamp:nNn\* \file_compare_timestamp:nNn\* TF

\file_input:n
\texttt{(file name)}

Searches for \texttt{(file name)} in the path as detailed for \file_if_exist:nTF, and if found reads in the file as additional \LaTeX source. All files read are recorded for information and the file name stack is updated by this function. An error is raised if the file is not found.

\file_input:nF
\{false code\}

Searches for \texttt{(file name)} using the current \TeX search path and the additional paths included in \l_file_search_path_seq. If found then reads in the file as additional \LaTeX source as described for \file_input:n, otherwise inserts the \texttt{false code}. Note that these functions do not raise an error if the file is not found, in contrast to \file_input:n.

\file_input_stop:

Ends the reading of a file started by \file_input:n or similar before the end of the file is reached. Where the file reading is being terminated due to an error, \msg_critical should be preferred.

\textbf{\TeXhackers note:} This function must be used on a line on its own: \TeX reads files line-by-line and so any additional tokens in the “current” line will still be read. This is also true if the function is hidden inside another function (which will be the normal case), i.e., all tokens on the same line in the source file are still processed. Putting it on a line by itself in the definition doesn’t help as it is the line where it is used that counts!

\file_show_list:
\file_log_list:

These functions list all files loaded by \LaTeX commands that populate \@filelist or by \file_input:n. While \file_show_list: displays the list in the terminal, \file_log_list: outputs it to the log file only.
Chapter 13

The l3luatex package: Lua\TeX-specific functions

The Lua\TeX engine provides access to the Lua programming language, and with it access to the “internals” of \TeX. In order to use this within the framework provided here, a family of functions is available. When used with pdf\TeX, \pdf\TeX, up\TeX or Xe\TeX these raise an error: use \texttt{\sys_if_engine_luatex:T} to avoid this. Details on using Lua with the Lua\TeX engine are given in the Lua\TeX manual.

13.1 Breaking out to Lua

\texttt{\lua_now:n \langle token list \rangle}

The \langle token list \rangle is first tokenized by \TeX, which includes converting line ends to spaces in the usual \TeX manner and which respects currently-applicable \TeX category codes. The resulting \langle Lua input \rangle is passed to the Lua interpreter for processing. Each \texttt{\lua_now:n} block is treated by Lua as a separate chunk. The Lua interpreter executes the \langle Lua input \rangle immediately, and in an expandable manner.

\textbf{\LaTeXhackers note:} \texttt{\lua_now:e} is a macro wrapper around \texttt{\directlua}: when Lua\TeX is in use two expansions are required to yield the result of the Lua code.

\texttt{\lua_shipout:n \langle token list \rangle}

\texttt{\lua_shipout:e \langle token list \rangle}

The \langle token list \rangle is first tokenized by \TeX, which includes converting line ends to spaces in the usual \TeX manner and which respects currently-applicable \TeX category codes. The resulting \langle Lua input \rangle is passed to the Lua interpreter when the current page is finalised (\textit{i.e.} at shipout). Each \texttt{\lua_shipout:n} block is treated by Lua as a separate chunk. The Lua interpreter will execute the \langle Lua input \rangle during the page-building routine: no \TeX expansion of the \langle Lua input \rangle will occur at this stage.

In the case of the \texttt{\lua_shipout_e:n} version the input is fully expanded by \TeX in an \texttt{e}-type manner during the shipout operation.

\textbf{\LaTeXhackers note:} At a \TeX level, the \langle Lua input \rangle is stored as a “whatsit.”

98
\lua_escape:n \lua_escape:e

\lua_escape:n \{\textit{token list}\}

Converts the \textit{token list} such that it can safely be passed to Lua: embedded backslashes, double and single quotes, and newlines and carriage returns are escaped. This is done by prepending an extra token consisting of a backslash with category code 12, and for the line endings, converting them to \texttt{\textbackslash n} and \texttt{\textbackslash r}, respectively.

\textbf{\textsc{\LaTeX}hackers note}: \texttt{\lua_escape:e} is a macro wrapper around \texttt{\luaescapestring}: when \textsc{Lua}\LaTeX{} is in use two expansions are required to yield the result of the Lua code.

### 13.2 Lua interfaces

As well as interfaces for \TeX{}, there are a small number of Lua functions provided here.

**ltx.utils**

Most public interfaces provided by the module are stored within the \texttt{ltx.utils} table.

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\texttt{ltx.utils.filedump}</td>
<td>{dump} = \texttt{ltx.utils.filedump}((file),{offset},{length})</td>
</tr>
<tr>
<td></td>
<td>Returns the uppercase hexadecimal representation of the content of the {file} read as bytes. If the {length} is given, only this part of the file is returned; similarly, one may specify the {offset} from the start of the file. If the {length} is not given, the entire file is read starting at the {offset}.</td>
</tr>
<tr>
<td>\texttt{ltx.utils.filemd5sum}</td>
<td>{hash} = \texttt{ltx.utils.filemd5sum}((file))</td>
</tr>
<tr>
<td></td>
<td>Returns the MD5 sum of the file contents read as bytes; note that the result will depend on the nature of the line endings used in the file, in contrast to normal \TeX{} behaviour. If the {file} is not found, nothing is returned with no error raised.</td>
</tr>
<tr>
<td>\texttt{ltx.utils.filemoddate}</td>
<td>{date} = \texttt{ltx.utils.filemoddate}((file))</td>
</tr>
<tr>
<td></td>
<td>Returns the date/time of last modification of the {file} in the format \texttt{D:{year}{month}{day}{hour}{minute}{second}{offset}} where the latter may be \texttt{Z} (UTC) or \texttt{{plus-minus}{hours}{minutes}}. If the {file} is not found, nothing is returned with no error raised.</td>
</tr>
<tr>
<td>\texttt{ltx.utils.filesize}</td>
<td>\texttt{size} = \texttt{ltx.utils.filesize}((file))</td>
</tr>
<tr>
<td></td>
<td>Returns the size of the {file} in bytes. If the {file} is not found, nothing is returned with no error raised.</td>
</tr>
</tbody>
</table>
Chapter 14

The \texttt{l3legacy} package

Interfaces to legacy concepts

There are a small number of \TeX or \LaTeX2ε concepts which are not used in \texttt{expl3} code but which need to be manipulated when working as a \LaTeX2ε package. To allow these to be integrated cleanly into \texttt{expl3} code, a set of legacy interfaces are provided here.

```latex
\legacy_if_p:n \legacy_if:nTF \legacy_if:nTF
```

Tests if the \LaTeX2ε/plain \TeX conditional (generated by `\newif`) if \texttt{true} or \texttt{false} and branches accordingly. The \texttt{name} of the conditional should \textit{omit} the leading \texttt{if}.

```latex
\legacy_if_set_true:n \legacy_if_set_false:n \legacy_if_gset_true:n \legacy_if_gset_false:n
```

Sets the \LaTeX2ε/plain \TeX conditional `\if\texttt{name}` (generated by `\newif`) to be \texttt{true} or \texttt{false}.

```latex
\legacy_if_set:nn \legacy_if_gset:nn
```

Sets the \LaTeX2ε/plain \TeX conditional `\if\texttt{name}` (generated by `\newif`) to the result of evaluating the `\texttt{boolean expression}`.
Part IV

Data types
Chapter 15

The l3tl package
Token lists

\TeX\ works with tokens, and \LaTeX3\ therefore provides a number of functions to deal with lists of tokens. Token lists may be present directly in the argument to a function:

```
\foo:n \{ a collection of \texttt{\textbf{tokens}} \}
```

or may be stored in a so-called “token list variable”, which have the suffix \texttt{tl}: a token list variable can also be used as the argument to a function, for example

```
\foo:N \l\_some\_tl
```

In both cases, functions are available to test and manipulate the lists of tokens, and these have the module prefix \texttt{tl}. In many cases, functions which can be applied to token list variables are paired with similar functions for application to explicit lists of tokens: the two “views” of a token list are therefore collected together here.

A token list (explicit, or stored in a variable) can be seen either as a list of “items”, or a list of “tokens”. An item is whatever \texttt{\textbf{\use:n}} would grab as its argument: a single non-space token or a brace group, with optional leading explicit space characters (each item is thus itself a token list). A token is either a normal \normaltextit{N} argument, or \texttt{\l}, {, or } (assuming normal \TeX\ category codes). Thus for example

```
\{ \texttt{Hello} \} - \texttt{world}
```

contains six items (\texttt{Hello}, \texttt{w}, \texttt{o}, \texttt{r}, \texttt{l} and \texttt{d}), but thirteen tokens \{\texttt{, H, e, l, l, o, \l, w, o, r, l and d}\}. Functions which act on items are often faster than their analogue acting directly on tokens.

15.1 Creating and initialising token list variables

```
\tl_new:N \tl_new:N \tl_new:c
```

Creates a new \texttt{(tl var)} or raises an error if the name is already taken. The declaration is global. The \texttt{(tl var)} is initially empty.
\tlconst:Nn
\tlconst:(Nx|cn|cx)

\tlconst:Nn \tlvar \{\langle \text{token list} \rangle\}

Creates a new constant \langle \tlvar \rangle or raises an error if the name is already taken. The value of the \langle \tlvar \rangle is set globally to the \langle \text{token list} \rangle.

\tlclear:N
\tlclear:c
\tlgclear:N
\tlgclear:c

\tlclear:N \tlvar

Clears all entries from the \langle \tlvar \rangle.

\tlclear_new:N
\tlclear_new:c
\tlgclear_new:N
\tlgclear_new:c

\tlclear_new:N \tlvar

Ensures that the \langle \tlvar \rangle exists globally by applying \tlnew:N if necessary, then applies \tl_(g)clear:N to leave the \langle \tlvar \rangle empty.

\tlset_eq:NN
\tlset_eq:(cN|Nc|cc)
\tlgset_eq:NN
\tlgset_eq:(cN|Nc|cc)

\tlset_eq:NN \tlvar_{1} \tlvar_{2}

Sets the content of \langle \tlvar_{1} \rangle equal to that of \langle \tlvar_{2} \rangle.

\tlconcat:NNN
\tlconcat:ccc
\tlgconcat:NNN
\tlgconcat:ccc

\tlconcat:NNN \tlvar_{1} \tlvar_{2} \tlvar_{3}

Concatenates the content of \langle \tlvar_{2} \rangle and \langle \tlvar_{3} \rangle together and saves the result in \langle \tlvar_{1} \rangle. The \langle \tlvar_{3} \rangle is placed at the left side of the new token list.

\tlif_exist_p:N
\tlif_exist_p:c
\tlif_exist:NTF
\tlif_exist:cTF

New: 2012-03-03

\tlif_exist_p:N \tlvar

Tests whether the \langle \tlvar \rangle is currently defined. This does not check that the \langle \tlvar \rangle really is a token list variable.

15.2 Adding data to token list variables

\tlset:Nn
\tlset:(NV|Nv|No|Nf|Nx|cn|cv|co|cf|cx)
\tlgset:Nn
\tlgset:(NV|Nv|No|Nf|Nx|cn|cv|co|cf|cx)

\tlset:Nn \tlvar \{\langle \text{tokens} \rangle\}

Sets \langle \tlvar \rangle to contain \langle \text{tokens} \rangle, removing any previous content from the variable.

\tlput_left:Nn
\tlput_left:(NV|Nv|No|Nx|cn|cv|co|cf|cx)
\tlgput_left:Nn
\tlgput_left:(NV|Nv|No|Nx|cn|cv|co|cf|cx)

\tlput_left:Nn \tlvar \{\langle \text{tokens} \rangle\}

Appends \langle \text{tokens} \rangle to the left side of the current content of \langle \tlvar \rangle.
15.3 Token list conditionals

\texttt{\textbackslash tl\_put\_right:Nn}\{\texttt{token\_list}\}\{\texttt{true\_code}\}\{\texttt{false\_code}\}

Tests if the \texttt{token\_list} consists only of blank spaces (i.e. contains no item). The test is true if \texttt{token\_list} is zero or more explicit space characters (explicit tokens with character code 32 and category code 10), and is false otherwise.

\texttt{\textbackslash tl\_if\_empty:p:N}\{\texttt{true\_code}\}\{\texttt{false\_code}\}

Tests if the \texttt{token\_list} is entirely empty (i.e. contains no tokens at all).

\texttt{\textbackslash tl\_if\_eq:p:NN}\{\texttt{true\_code}\}\{\texttt{false\_code}\}

Compares the content of two \texttt{token\_list} variables and is logically true if the two contain the same list of tokens (i.e. identical in both the list of characters they contain and the category codes of those characters). Thus for example

\texttt{\textbackslash tl\_set:Nn \_l\_tmpa\_tl \{ abc \}}
\texttt{\textbackslash tl\_set:Nx \_l\_tmpb\_tl \{ \tl\_to\_str:n \{ abc \} \}}
\texttt{\textbackslash tl\_if\_eq:NN \_l\_tmpa\_tl \_l\_tmpb\_tl \{ true \} \{ false \}}

yields false. See also \texttt{\str\_if\_eq:nnTF} for a comparison that ignores category codes.

\texttt{\textbackslash tl\_if\_eq:NnTF}\{\texttt{token\_list1}\}\{\texttt{true\_code}\}\{\texttt{false\_code}\}

Tests if the \texttt{token\_list1} and the \texttt{token\_list2} contain the same list of tokens, both in respect of character codes and category codes. This conditional is not expandable: see \texttt{\textbackslash tl\_if\_eq:NNTF} for an expandable version when both token lists are stored in variables, or \texttt{\str\_if\_eq:nnTF} if category codes are not important.
\tl_if_eq:nnTF \{\langle \text{token list}_1 \rangle \} \{\langle \text{token list}_2 \rangle \} \{\langle \text{true code} \rangle \} \{\langle \text{false code} \rangle \}

Tests if \{\langle \text{token list}_1 \rangle \} and \{\langle \text{token list}_2 \rangle \} contain the same list of tokens, both in respect of character codes and category codes. This conditional is not expandable: see \tl_if_eq:NNTF for an expandable version when token lists are stored in variables, or \str_if_eq:nnTF if category codes are not important.

\tl_if_in:NnTF \{\langle \text{token list} \rangle \} \{\langle \text{tl var} \rangle \} \{\langle \text{true code} \rangle \} \{\langle \text{false code} \rangle \}

Tests if the \{\langle \text{token list} \rangle \} is found in the content of the \{\langle \text{tl var} \rangle \}. The \{\langle \text{token list} \rangle \} cannot contain the tokens \{\{, \} or \# (more precisely, explicit character tokens with category code 1 (begin-group) or 2 (end-group), and tokens with category code 6).

\tl_if_in:nnTF \{\langle \text{token list}_1 \rangle \} \{\langle \text{token list}_2 \rangle \} \{\langle \text{true code} \rangle \} \{\langle \text{false code} \rangle \}

Tests if \{\langle \text{token list}_2 \rangle \} is found inside \{\langle \text{token list}_1 \rangle \}. The \{\langle \text{token list}_2 \rangle \} cannot contain the tokens \{\{, \} or \# (more precisely, explicit character tokens with category code 1 (begin-group) or 2 (end-group), and tokens with category code 6). The search does not enter brace (category code 1/2) groups.

\tl_if_novalue_p:n \tl_if_novalue:nTF \{\langle \text{token list} \rangle \} \{\langle \text{true code} \rangle \} \{\langle \text{false code} \rangle \}

Tests if the \{\langle \text{token list} \rangle \} is exactly equal to the special \c_novalue_tl marker. This function is intended to allow construction of flexible document interface structures in which missing optional arguments are detected.

\tl_if_single_p:N \tl_if_single:NTF \langle \text{tl var} \rangle \{\langle \text{true code} \rangle \} \{\langle \text{false code} \rangle \}

Tests if the content of the \{\langle \text{tl var} \rangle \} consists of a single \langle \text{item} \rangle, i.e. is a single normal token (neither an explicit space character nor a begin-group character) or a single brace group, surrounded by optional spaces on both sides. In other words, such a token list has token count 1 according to \tl_count:N.

\tl_if_single_token_p:n \tl_if_single_token:nTF \{\langle \text{token list} \rangle \} \{\langle \text{true code} \rangle \} \{\langle \text{false code} \rangle \}

Tests if the \{\langle \text{token list} \rangle \} consists of exactly one \langle \text{item} \rangle, i.e. is either a single space character or a single normal token. Token groups \{\ldots\} are not single tokens.
This function compares the \textit{test token list variable} in turn with each of the \textit{token list variable cases}. If the two are equal (as described for \texttt{tl_if_eq:NNTF}) then the associated \textit{code} is left in the input stream and other cases are discarded. If any of the cases are matched, the \textit{true code} is also inserted into the input stream (after the code for the appropriate case), while if none match then the \textit{false code} is inserted. The function \texttt{tl_case:Nn}, which does nothing if there is no match, is also available.

15.3.1 Testing the first token

Tests if the first \textit{token} in the \textit{token list} has the same category code as the \textit{test token}. In the case where the \textit{token list} is empty, the test is always \textit{false}.

Tests if the first \textit{token} in the \textit{token list} has the same character code as the \textit{test token}. In the case where the \textit{token list} is empty, the test is always \textit{false}.

Tests if the first \textit{token} in the \textit{token list} has the same meaning as the \textit{test token}. In the case where \textit{token list} is empty, the test is always \textit{false}.

Tests if the first \textit{token} in the \textit{token list} is an explicit begin-group character (with category code 1 and any character code), in other words, if the \textit{token list} starts with a brace group. In particular, the test is \textit{false} if the \textit{token list} starts with an implicit token such as \texttt{vc\_group\_begin\_token}, or if it is empty. This function is useful to implement actions on token lists on a token by token basis.
Tests if the first \langle token \rangle in the \langle token list \rangle is a normal $N$-type argument. In other words, it is neither an explicit space character (explicit token with character code 32 and category code 10) nor an explicit begin-group character (with category code 1 and any character code). An empty argument yields false, as it does not have a normal first token. This function is useful to implement actions on token lists on a token by token basis.

Tests if the first \langle token \rangle in the \langle token list \rangle is an explicit space character (explicit token with character code 12 and category code 10). In particular, the test is false if the \langle token list \rangle starts with an implicit token such as \texttt{\cs{space token}}, or if it is empty. This function is useful to implement actions on token lists on a token by token basis.

15.4 Working with token lists as a whole

15.4.1 Using token lists

Converts the \langle token list \rangle to a \langle string \rangle, leaving the resulting character tokens in the input stream. A \langle string \rangle is a series of tokens with category code 12 (other) with the exception of spaces, which retain category code 10 (space). This function requires only a single expansion. Its argument must be braced.

\textbf{\TeX}hacker note: This is the \TeX primitive \texttt{\detokenize}. Converting a \langle token list \rangle to a \langle string \rangle yields a concatenation of the string representations of every token in the \langle token list \rangle. The string representation of a control sequence is

- an escape character, whose character code is given by the internal parameter \texttt{\escapechar}, absent if the \texttt{\escapechar} is negative or greater than the largest character code;
- the control sequence name, as defined by \texttt{\cs_to_str:N};
- a space, unless the control sequence name is a single character whose category at the time of expansion of \texttt{\tl_to_str:n} is not “letter”.

The string representation of an explicit character token is that character, doubled in the case of (explicit) macro parameter characters (normally #). In particular, the string representation of a token list may depend on the category codes in effect when it is evaluated, and the value of the \texttt{\escapechar}: for instance \texttt{\tl_to_str:n \{\texttt{a}\}} normally produces the three character “backslash”, “lower-case a”, “space”, but it may also produce a single “lower-case a” if the escape character is negative and a is currently not a letter.

Converts the content of the \langle tl var \rangle into a series of characters with category code 12 (other) with the exception of spaces, which retain category code 10 (space). This \langle string \rangle is then left in the input stream. For low-level details, see the notes given for \texttt{\tl_to_\-str:n}. 

107
Recovers the content of a \langle\text{tl var}\rangle and places it directly in the input stream. An error is raised if the variable does not exist or if it is invalid. Note that it is possible to use a \langle\text{tl var}\rangle directly without an accessor function.

15.4.2 Counting and reversing token lists

```
\tl_count:n \{\langle\text{tokens}\rangle\}
```

Counts the number of \langle\text{items}\rangle in \langle\text{tokens}\rangle and leaves this information in the input stream. Unbraced tokens count as one element as do each token group (\{\ldots\}). This process ignores any unprotected spaces within \langle\text{tokens}\rangle. See also \tl_count:n. This function requires three expansions, giving an \langle\text{integer denotation}\rangle.

```
\tl_count_tokens:n \{\langle\text{tokens}\rangle\}
```

Counts the number of \TeX tokens in the \langle\text{tokens}\rangle and leaves this information in the input stream. Every token, including spaces and braces, contributes one to the total; thus for instance, the token count of \texttt{a\{-bc\}} is 6.

```
\tl_reverse:n \{\langle\text{token list}\rangle\}
```

Reverses the order of the \langle\text{items}\rangle in the \langle\text{token list}\rangle, so that \langle\text{item}_1\rangle\langle\text{item}_2\rangle\langle\text{item}_3\rangle \ldots \langle\text{item}_n\rangle becomes \langle\text{item}_n\rangle\ldots\langle\text{item}_2\rangle\langle\text{item}_1\rangle. This process preserves unprotected space within the \langle\text{token list}\rangle. Tokens are not reversed within braced token groups, which keep their outer set of braces. In situations where performance is important, consider \tl_reverse_items:n. See also \tl_reverse:n.

\TeXhackers note: The result is returned within \unexpanded, which means that the token list does not expand further when appearing in an \x-type or \e-type argument expansion.

```
\tl_reverse:N \langle\text{tl var}\rangle
```

Sets the \langle\text{tl var}\rangle to contain the result of reversing the order of its \langle\text{items}\rangle, so that \langle\text{item}_1\rangle\langle\text{item}_2\rangle\langle\text{item}_3\rangle \ldots \langle\text{item}_n\rangle becomes \langle\text{item}_n\rangle\ldots\langle\text{item}_2\rangle\langle\text{item}_1\rangle. This process preserves unprotected spaces within the \langle\text{token list variable}\rangle. Braced token groups are copied without reversing the order of tokens, but keep the outer set of braces. This is equivalent to a combination of an assignment and \tl_reverse:V. See also \tl_reverse_items:n for improved performance.
\tl_reverse_items:n \{(token list)\}  \nReverses the order of the \langle items \rangle stored in \langle tl var \rangle, so that \{\langle item_1 \rangle\}{\langle item_2 \rangle}{\langle item_3 \rangle} \ldots{\langle item_n \rangle} becomes \{\langle item_n \rangle\} \ldots \{\langle item_3 \rangle\}{\langle item_2 \rangle}{\langle item_1 \rangle}. This process removes any unprotected space within the \langle token list \rangle. Braced token groups are copied without reversing the order of tokens, and keep the outer set of braces. Items which are initially not braced are copied with braces in the result. In cases where preserving spaces is important, consider the slower function \tl_reverse:n.

\textbf{\TeX hackers note:} The result is returned within \texttt{\unexpanded}, which means that the token list does not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.

\tl_trim_spaces:n \tl_trim_spaces:o \tl_trim_spaces_apply:nN \tl_trim_spaces_apply:oN \tl_trim_spaces:N \tl_trim_spaces:c \tl_gtrim_spaces:N \tl_gtrim_spaces:c
\tl_trim_spaces:n \{(token list)\} \nRemoves any leading and trailing explicit space characters (explicit tokens with character code 32 and category code 10) from the \langle token list \rangle and leaves the result in the input stream.

\textbf{\TeX hackers note:} The result is returned within \texttt{\unexpanded}, which means that the token list does not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.

\tl_trim_spaces_apply:nN \tl_trim_spaces_apply:oN \tl_trim_spaces:N \tl_trim_spaces:c \tl_gtrim_spaces:N \tl_gtrim_spaces:c
\tl_trim_spaces:n \{(token list)\} \{function\}  \nRemoves any leading and trailing explicit space characters (explicit tokens with character code 32 and category code 10) from the \langle token list \rangle and passes the result to the \langle function \rangle as an \texttt{n}-type argument.

\tl_trim_spaces:N \tl_trim_spaces:c \tl_gtrim_spaces:N \tl_gtrim_spaces:c  \nSets the \langle tl var \rangle to contain the result of removing any leading and trailing explicit space characters (explicit tokens with character code 32 and category code 10) from its contents.

\subsection{15.4.3 Viewing token lists}
\tl_show:N \tl_show:c  \nDisplays the content of the \langle tl var \rangle on the terminal.

\textbf{\TeX hackers note:} This is similar to the \TeX primitive \texttt{\show}, wrapped to a fixed number of characters per line.

\tl_show:n \tl_show:c  \nDisplays the \langle token list \rangle on the terminal.

\textbf{\TeX hackers note:} This is similar to the \texttt{\showtokens} primitive in \eTeX, wrapped to a fixed number of characters per line.
\tl_log:N \tl_log:c

New: 2014-08-22
Updated: 2021-04-29

\tl_log:n

New: 2014-08-22
Updated: 2015-08-07

15.5 Manipulating items in token lists

15.5.1 Mapping over token lists

All mappings are done at the current group level, \textit{i.e.} any local assignments made by the \langle function \rangle or \langle code \rangle discussed below remain in effect after the loop.

\tl_map_function:NN \tl_map_function:cN

Updated: 2012-06-29

\tl_map_function:nN

Updated: 2012-06-29

\tl_map_inline:Nn \tl_map_inline:cn

Updated: 2012-06-29

\tl_map_inline:nn

Updated: 2012-06-29

\tl_log:N \langle tl var \rangle

Writes the content of the \langle tl var \rangle in the log file. See also \tl_show:N which displays the result in the terminal.

\tl_log:n \{(token list)\}

Writes the \langle token list \rangle in the log file. See also \tl_show:n which displays the result in the terminal.

\tl_map_function:NN \langle tl var \rangle \langle function \rangle

Applies \langle function \rangle to every \langle item \rangle in the \langle tl var \rangle. The \langle function \rangle receives one argument for each iteration. This may be a number of tokens if the \langle item \rangle was stored within braces. Hence the \langle function \rangle should anticipate receiving n-type arguments. See also \tl_map_function:nN.

\tl_map_function:nN \{(token list)\} \langle function \rangle

Applies \langle function \rangle to every \langle item \rangle in the \langle token list \rangle. The \langle function \rangle receives one argument for each iteration. This may be a number of tokens if the \langle item \rangle was stored within braces. Hence the \langle function \rangle should anticipate receiving n-type arguments. See also \tl_map_function:NN.

\tl_map_inline:Nn \langle tl var \rangle \{(inline function)\}

Applies the \langle inline function \rangle to every \langle item \rangle stored within the \langle tl var \rangle. The \langle inline function \rangle should consist of code which receives the \langle item \rangle as \#1. See also \tl_map_function:nn.

\tl_map_inline:nn \{(token list)\} \{(inline function)\}

Applies the \langle inline function \rangle to every \langle item \rangle stored within the \langle token list \rangle. The \langle inline function \rangle should consist of code which receives the \langle item \rangle as \#1. See also \tl_map_function:nn.
Analogue of \texttt{\tl_map_function:NN} which maps several tokens instead of a single function. The \texttt{⟨code⟩} receives each \texttt{⟨item⟩} in the \texttt{⟨tl var⟩} or in \texttt{⟨tokens⟩} as a trailing brace group. For instance,

\begin{verbatim}
\tl_map_tokens:Nn \l_my_tl \{ \prg_replicate:nn \{ 2 \} \}
\end{verbatim}

expands to twice each \texttt{⟨item⟩} in the \texttt{⟨tl var⟩}: for each \texttt{⟨item⟩} in \texttt{\l_my_tl} the function \texttt{\prg_replicate:nn} receives 2 and \texttt{⟨item⟩} as its two arguments. The function \texttt{\tl_map_inline:Nn} is typically faster but is not expandable.

\begin{verbatim}
\tl_map_variable:NNn \tl_map_variable:cNn
\end{verbatim}

Stores each \texttt{⟨item⟩} of the \texttt{⟨tl var⟩} in turn in the (token list) \texttt{⟨variable⟩} and applies the \texttt{⟨code⟩}. The \texttt{⟨code⟩} will usually make use of the \texttt{⟨variable⟩}, but this is not enforced. The assignments to the \texttt{⟨variable⟩} are local. Its value after the loop is the last \texttt{⟨item⟩} in the \texttt{⟨tl var⟩}, or its original value if the \texttt{⟨tl var⟩} is blank. See also \texttt{\tl_map_inline:Nn}.

\begin{verbatim}
\tl_map_break:
\end{verbatim}

Used to terminate a \texttt{\tl_map...} function before all entries in the \texttt{⟨token list variable⟩} have been processed. This normally takes place within a conditional statement, for example

\begin{verbatim}
\tl_map_inline:Nn \l_my_tl
\{\str_if_eq:nnT \{ #1 \} \{ bingo \} \{ \tl_map_break: \}
% Do something useful
\}
\end{verbatim}

See also \texttt{\tl_map_break:n}. Use outside of a \texttt{\tl_map...} scenario leads to low level \TeX{} errors.

\textbf{\TeX{}hackers note:} When the mapping is broken, additional tokens may be inserted before the \texttt{⟨tokens⟩} are inserted into the input stream. This depends on the design of the mapping function.
\tl_map_break:n \{\textit{code}\}\}

Used to terminate a \tl_map_... function before all entries in the \textit{token list variable} have been processed, inserting the \textit{code} after the mapping has ended. This normally takes place within a conditional statement, for example

\begin{verbatim}
\tl_map_inline:Nn \l_my_tl
{\str_if_eq:nnT { #1 } { bingo } { \tl_map_break:n \{ <code> \} { % Do something useful }\}
\end{verbatim}

Use outside of a \tl_map_... scenario leads to low level \TeX{} errors.

\texttt{\TeX{}hackers note:} When the mapping is broken, additional tokens may be inserted before the \textit{code} is inserted into the input stream. This depends on the design of the mapping function.

---

15.5.2 Head and tail of token lists

Functions which deal with either only the very first item (balanced text or single normal token) in a token list, or the remaining tokens.

\begin{verbatim}
\tl_head:n \{\textit{token list}\}\}
\end{verbatim}

Leaves in the input stream the first \textit{item} in the \textit{token list}, discarding the rest of the \textit{token list}. All leading explicit space characters (explicit tokens with character code 32 and category code 10) are discarded; for example

\begin{verbatim}
\tl_head:n \{ abc \}
\end{verbatim}

and

\begin{verbatim}
\tl_head:n \{ - abc \}
\end{verbatim}

both leave \texttt{a} in the input stream. If the “head” is a brace group, rather than a single token, the braces are removed, and so

\begin{verbatim}
\tl_head:n \{ - \{ - ab \} c \}
\end{verbatim}

yields \texttt{ab}. A blank \textit{token list} (see \texttt{\tl_if_blank:nTF}) results in \texttt{\tl_head:n} leaving nothing in the input stream.

\texttt{\TeX{}hackers note:} The result is returned within \texttt{\exp_not:n}, which means that the token list does not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.
Leaves in the input stream the first ⟨item⟩ in the ⟨token list⟩, discarding the rest of the ⟨token list⟩. All leading explicit space characters (explicit tokens with character code 32 and category code 10) are discarded. A blank ⟨token list⟩ (which consists only of space characters) results in a low-level \TeX{} error, which may be avoided by the inclusion of an empty group in the input (as shown), without the need for an explicit test. Alternatively, \tl_if_blank:nF may be used to avoid using the function with a “blank” argument. This function requires only a single expansion, and thus is suitable for use within an \o-type expansion. In general, \tl_head:n should be preferred if the number of expansions is not critical.

Discards all leading explicit space characters (explicit tokens with character code 32 and category code 10) and the first ⟨item⟩ in the ⟨token list⟩, and leaves the remaining tokens in the input stream. Thus for example

\begin{verbatim}
\tl_tail:n { a ~ {bc} d }
\end{verbatim}

and

\begin{verbatim}
\tl_tail:n { - a ~ {bc} d }
\end{verbatim}

both leave ⟨bc⟩d in the input stream. A blank ⟨token list⟩ (see \tl_if_blank:nTF) results in \tl_tail:n leaving nothing in the input stream.

\textbf{\TeX{}hackers note:} The result is returned within \exp_not:n, which means that the token list does not expand further when appearing in an \x-type or \e-type argument expansion.

### 15.5.3 Items and ranges in token lists

Indexing items in the ⟨token list⟩ from 1 on the left, this function evaluates the ⟨integer expression⟩ and leaves the appropriate item from the ⟨token list⟩ in the input stream. If the ⟨integer expression⟩ is negative, indexing occurs from the right of the token list, starting at −1 for the right-most item. If the index is out of bounds, then the function expands to nothing.

\textbf{\TeX{}hackers note:} The result is returned within the \unexpanded primitive (\exp_not:n), which means that the ⟨item⟩ does not expand further when appearing in an \x-type or \e-type argument expansion.
\tl_rand_item:N \{ tl \ var \}
\tl_rand_item:c \{ \langle \text{token list} \rangle \}
\tl_rand_item:n \{ \langle \text{token list} \rangle \}

Selects a pseudo-random item of the \langle \text{token list} \rangle. If the \langle \text{token list} \rangle is blank, the result is empty. This is not available in older versions of Xe\TeX.

\TeXhacksnote{The result is returned within the \texttt{\unexpanded} primitive (\texttt{\exp_not:n}), which means that the \langle \text{item} \rangle does not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.}
Leaves in the input stream the items from the \langle start index \rangle to the \langle end index \rangle inclusive. Spaces and braces are preserved between the items returned (but never at either end of the list). Here \langle start index \rangle and \langle end index \rangle should be \langle integer expressions \rangle. For describing in detail the functions’ behavior, let \( m \) and \( n \) be the start and end index respectively. If either is 0, the result is empty. A positive index means ‘start counting from the left end’, and a negative index means ‘from the right end’. Let \( l \) be the count of the token list.

The actual start point is determined as \( M = m \) if \( m > 0 \) and as \( M = l + m + 1 \) if \( m < 0 \). Similarly the actual end point is \( N = n \) if \( n > 0 \) and \( N = l + n + 1 \) if \( n < 0 \). If \( M > N \), the result is empty. Otherwise it consists of all items from position \( M \) to position \( N \) inclusive; for the purpose of this rule, we can imagine that the token list extends at infinity on either side, with void items at positions \( s \) for \( s \leq 0 \) or \( s > l \).

Spaces in between items in the actual range are preserved. Spaces at either end of the token list will be removed anyway (think to the token list being passed to \tl_trim_spaces:n \) to begin with.

Thus, with \( l = 7 \) as in the examples below, all of the following are equivalent and result in the whole token list

\[
\texttt{\tl_range:nnn \{ abcd{-e{}}fg \} \{ 1 \} \{ 7 \}}
\]
\[
\texttt{\tl_range:nnn \{ abcd{-e{}}fg \} \{ 1 \} \{ 12 \}}
\]
\[
\texttt{\tl_range:nnn \{ abcd{-e{}}fg \} \{ -7 \} \{ 7 \}}
\]
\[
\texttt{\tl_range:nnn \{ abcd{-e{}}fg \} \{ -12 \} \{ 7 \}}
\]

Here are some more interesting examples. The calls

\[
\texttt{\iow_term:x \{ \tl_range:nnn \{ abcd{e{}}fg \} \{ 2 \} \{ 5 \} \}}
\]
\[
\texttt{\iow_term:x \{ \tl_range:nnn \{ abcd{e{}}fg \} \{ 2 \} \{ -3 \} \}}
\]
\[
\texttt{\iow_term:x \{ \tl_range:nnn \{ abcd{e{}}fg \} \{ -6 \} \{ 5 \} \}}
\]
\[
\texttt{\iow_term:x \{ \tl_range:nnn \{ abcd{e{}}fg \} \{ -6 \} \{ -3 \} \}}
\]

are all equivalent and will print \( bcd{e{}} \) on the terminal; similarly

\[
\texttt{\iow_term:x \{ \tl_range:nnn \{ abcd{-e{}}fg \} \{ 2 \} \{ 5 \} \}}
\]
\[
\texttt{\iow_term:x \{ \tl_range:nnn \{ abcd{-e{}}fg \} \{ 2 \} \{ -3 \} \}}
\]
\[
\texttt{\iow_term:x \{ \tl_range:nnn \{ abcd{-e{}}fg \} \{ -6 \} \{ 5 \} \}}
\]
\[
\texttt{\iow_term:x \{ \tl_range:nnn \{ abcd{-e{}}fg \} \{ -6 \} \{ -3 \} \}}
\]

are all equivalent and will print \( bcd \ \{e{}} \) on the terminal (note the space in the middle).

To the contrary,

\[
\texttt{\tl_range:nnn \{ abcd{-e{}}f \} \{ 2 \} \{ 4 \}}
\]

will discard the space after ‘d’.

If we want to get the items from, say, the third to the last in a token list <tl>, the call is \( \texttt{\tl_range:nnn \{ <tl> \} \{ 3 \} \{ -1 \}} \). Similarly, for discarding the last item, we can do \( \texttt{\tl_range:nnn \{ <tl> \} \{ 1 \} \{ -2 \}} \).

For better performance, see \( \texttt{\tl_range_braced:nnn} \) and \( \texttt{\tl_range_unbraced:nnn} \).

**TeXhackers note:** The result is returned within the \unexpanded primitive \exp_not:n, which means that the \langle item \rangle does not expand further when appearing in an \x-type or \e-type argument expansion.
15.5.4 Sorting token lists

\tl_sort:Nn \tl_var \langle \text{comparison code} \rangle
Sorts the items in the \( \langle \text{tl var} \rangle \) according to the \( \langle \text{comparison code} \rangle \), and assigns the result to \( \langle \text{tl var} \rangle \). The details of sorting comparison are described in Section 6.1.

\tl_sort:cn
\tl_gsort:Nn
\tl_gsort:cn

New: 2017-02-06
\tl_sort:nN \{ \langle \text{token list} \rangle \} \langle \text{conditional} \rangle
Sorts the items in the \( \langle \text{token list} \rangle \), using the \( \langle \text{conditional} \rangle \) to compare items, and leaves the result in the input stream. The \( \langle \text{conditional} \rangle \) should have signature :nnTF, and return true if the two items being compared should be left in the same order, and false if the items should be swapped. The details of sorting comparison are described in Section 6.1.

\textbf{TeXhackers note:} The result is returned within \texttt{\exp_not:n}, which means that the token list does not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.

15.6 Manipulating tokens in token lists

15.6.1 Replacing tokens

Within token lists, replacement takes place at the top level: there is no recursion into brace groups (more precisely, within a group defined by a categroiy code 1/2 pair).

\tl_replace_once:NNn \tl_var \{ \langle \text{old tokens} \rangle \} \{ \langle \text{new tokens} \rangle \}
Replaces the first (leftmost) occurrence of \( \langle \text{old tokens} \rangle \) in the \( \langle \text{tl var} \rangle \) with \( \langle \text{new tokens} \rangle \). \( \langle \text{Old tokens} \rangle \) cannot contain {}, {} or # (more precisely, explicit character tokens with category code 1 (begin-group) or 2 (end-group), and tokens with category code 6).

\tl_replace_all:NNn \tl_var \{ \langle \text{old tokens} \rangle \} \{ \langle \text{new tokens} \rangle \}
Replaces all occurrences of \( \langle \text{old tokens} \rangle \) in the \( \langle \text{tl var} \rangle \) with \( \langle \text{new tokens} \rangle \). \( \langle \text{Old tokens} \rangle \) cannot contain {}, {} or # (more precisely, explicit character tokens with category code 1 (begin-group) or 2 (end-group), and tokens with category code 6). As this function operates from left to right, the pattern \( \langle \text{old tokens} \rangle \) may remain after the replacement (see \texttt{\tl_remove_all:Nn} for an example).

\tl_remove_once:Nn \tl_var \{ \langle \text{tokens} \rangle \}
Removes the first (leftmost) occurrence of \( \langle \text{tokens} \rangle \) from the \( \langle \text{tl var} \rangle \). \( \langle \text{Tokens} \rangle \) cannot contain {}, {} or # (more precisely, explicit character tokens with category code 1 (begin-group) or 2 (end-group), and tokens with category code 6).
\begin{verbatim}
\tl_remove_all:Nn \tl_remove_all:cn \tl_gremove_all:Nn \tl_gremove_all:cn
\end{verbatim}

Updated: 2011-08-11

15.6.2 Reassigning category codes

These functions allow the rescanning of tokens: re-apply \TeX{}’s tokenization process to apply category codes different from those in force when the tokens were absorbed. Whilst this functionality is supported, it is often preferable to find alternative approaches to achieving outcomes rather than rescanning tokens (for example construction of token lists token-by-token with intervening category code changes or using \texttt{\char_generate:nn}).

\begin{verbatim}
\tl_set_rescan:Nnn \tl_set_rescan:Nnn \tl_set_rescan:cn \tl_set_rescan:cn
\tl_gset_rescan:Nnn \tl_gset_rescan:cn \tl_gset_rescan:cn \tl_gset_rescan:cn
\end{verbatim}

Updated: 2015-08-11

Sets \texttt{\tl var} to contain \texttt{\tokens}, applying the category code régime specified in the \texttt{\setup} before carrying out the assignment. (Category codes applied to tokens not explicitly covered by the \texttt{\setup} are those in force at the point of use of \texttt{\tl_set_rescan:Nnn}.) This allows the \texttt{\tl var} to contain material with category codes other than those that apply when \texttt{\tokens} are absorbed. The \texttt{\setup} is run within a group and may contain any valid input, although only changes in category codes, such as uses of \texttt{\cctab_select:N}, are relevant. See also \texttt{\tl_rescan:nn}.

\textbf{\TeX{}hackers note:} The \texttt{\tokens} are first turned into a string (using \texttt{\tl_to_str:n}). If the string contains one or more characters with character code \texttt{\newlinechar} (set equal to \texttt{\endlinechar} unless that is equal to 32, before the user \texttt{\setup}), then it is split into lines at these characters, then read as if reading multiple lines from a file, ignoring spaces (catcode 10) at the beginning and spaces and tabs (character code 32 or 9) at the end of every line. Otherwise, spaces (and tabs) are retained at both ends of the single-line string, as if it appeared in the middle of a line read from a file.
Rescans (tokens) applying the category code régime specified in the \langle setup\rangle, and leaves the resulting tokens in the input stream. (Category codes applied to tokens not explicitly covered by the \langle setup\rangle are those in force at the point of use of \tl_rescan:nn.) The \langle setup\rangle is run within a group and may contain any valid input, although only changes in category codes, such as uses of \cctab_select:N, are relevant. See also \tl_set_:rescan:Nnn, which is more robust than using \tl_set:Nn in the \langle tokens\rangle argument of \tl_rescan:nn.

\textbf{\TeX}hackers note: The \langle tokens\rangle are first turned into a string (using \tl_to_str:n). If the string contains one or more characters with character code \newlinechar (set equal to \endlinechar unless that is equal to 32, before the user \langle setup\rangle), then it is split into lines at these characters, then read as if reading multiple lines from a file, ignoring spaces (catcode 10) at the beginning and spaces and tabs (character code 32 or 9) at the end of every line. Otherwise, spaces (and tabs) are retained at both ends of the single-line string, as if it appeared in the middle of a line read from a file.

Contrarily to the \scantokens primitive, \tl_rescan:nn tokenizes the whole string in the same category code régime rather than one token at a time, so that directives such as \verb that rely on changing category codes will not function properly.

\section{Constant token lists}

\c_empty_tl Constant that is always empty.

\c_novaluel_tl A marker for the absence of an argument. This constant tl can safely be typeset (cf. \q_:nil), with the result being \texttt{-NoValue-}. It is important to note that \c_novaluel_tl is constructed such that it will \textit{not} match the simple text input \texttt{-NoValue-}, \textit{i.e.} that

\tl_if_eq:NnTF \c_novaluel_tl { -NoValue- } is logically false. The \c_novaluel_tl marker is intended for use in creating document-level interfaces, where it serves as an indicator that an (optional) argument was omitted. In particular, it is distinct from a simple empty tl.

\c_space_tl An explicit space character contained in a token list (compare this with \c_space_token). For use where an explicit space is required.

\section{Scratch token lists}

\l_tmpa_tl \l_tmpb_tl Scratch token lists for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
Scratch token lists for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX\-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_tl
\g_tmpb_tl
Chapter 16

The \texttt{l3str} package: Strings

\TeX\ associates each character with a category code: as such, there is no concept of a “string” as commonly understood in many other programming languages. However, there are places where we wish to manipulate token lists while in some sense “ignoring” category codes: this is done by treating token lists as strings in a \TeX\ sense.

A \TeX\ string (and thus an \texttt{expl3} string) is a series of characters which have category code 12 (“other”) with the exception of space characters which have category code 10 (“space”). Thus at a technical level, a \TeX\ string is a token list with the appropriate category codes. In this documentation, these are simply referred to as strings.

String variables are simply specialised token lists, but by convention should be named with the suffix \texttt{...str}. Such variables should contain characters with category code 12 (other), except spaces, which have category code 10 (blank space). All the functions in this module which accept a token list argument first convert it to a string using \texttt{\tl_to_str:n} for internal processing, and do not treat a token list or the corresponding string representation differently.

As a string is a subset of the more general token list, it is sometimes unclear when one should be used over the other. Use a string variable for data that isn’t primarily intended for typesetting and for which a level of protection from unwanted expansion is suitable. This data type simplifies comparison of variables since there are no concerns about expansion of their contents.

The functions \texttt{\cs_to_str:N}, \texttt{\tl_to_str:n}, \texttt{\tl_to_str:N} and \texttt{\token_to_str:N} (and variants) generate strings from the appropriate input: these are documented in \texttt{l3\basics}, \texttt{l3tl} and \texttt{l3token}, respectively.

Most expandable functions in this module come in three flavours:

• \texttt{\str\_\_\_\_\_:N}, which expect a token list or string variable as their argument;

• \texttt{\str\_\_\_:n}, taking any token list (or string) as an argument;

• \texttt{\str\_\_\_\_\_ignore_spaces:n}, which ignores any space encountered during the operation: these functions are typically faster than those which take care of escaping spaces appropriately.
16.1 Creating and initialising string variables

\texttt{\str_new:N} \langle str var \rangle

Creates a new \langle str var \rangle or raises an error if the name is already taken. The declaration is global. The \langle str var \rangle is initially empty.

\texttt{\str_new:c}

\texttt{\str_new:Nn} \langle str var \rangle \{\langle token list \rangle\}

Creates a new constant \langle str var \rangle or raises an error if the name is already taken. The value of the \langle str var \rangle is set globally to the \langle token list \rangle, converted to a string.

\texttt{\str_clear:N} \langle str var \rangle

Clears the content of the \langle str var \rangle.

\texttt{\str_clear:c}

\texttt{\str_gclear:N} \langle str var \rangle

\texttt{\str_gclear:c}

Ensures that the \langle str var \rangle exists globally by applying \texttt{\str_new:N} if necessary, then applies \texttt{\str_(g)clear:N} to leave the \langle str var \rangle empty.

\texttt{\str_set_eq:NN} \langle str var_1 \rangle \langle str var_2 \rangle

Sets the content of \langle str var_1 \rangle equal to that of \langle str var_2 \rangle.

\texttt{\str_set_eq:cc}

\texttt{\str_gset_eq:NN} \langle str var_1 \rangle \langle str var_2 \rangle \langle str var_3 \rangle

\texttt{\str_gset_eq:cc}

Concatenates the content of \langle str var_2 \rangle and \langle str var_3 \rangle together and saves the result in \langle str var_1 \rangle. The \langle str var_2 \rangle is placed at the left side of the new string variable. The \langle str var_2 \rangle and \langle str var_3 \rangle must indeed be strings, as this function does not convert their contents to a string.

\texttt{\str_if_exist_p:N} \langle str var \rangle

\texttt{\str_if_exist_p:c}

\texttt{\str_if_exist:NTF} \langle str var \rangle \{\langle true code \rangle\} \{\langle false code \rangle\}

Tests whether the \langle str var \rangle is currently defined. This does not check that the \langle str var \rangle really is a string.
### 16.2 Adding data to string variables

```latex
\texttt{\textbf{\texttt{str}}\texttt{\_set:Nn}}\langle \texttt{str} \texttt{var} \rangle \{\langle \texttt{token} \texttt{list} \rangle\} \\
\texttt{\textbf{\texttt{str}}\texttt{\_set:}}(\texttt{NV}|\texttt{Nx}|\texttt{cn}|\texttt{cV}|\texttt{cx}) \\
\texttt{\textbf{\texttt{str}}\texttt{\_gset:}}(\texttt{NV}|\texttt{Nx}|\texttt{cn}|\texttt{cV}|\texttt{cx})
```

Converts the \langle token list \rangle to a \langle string \rangle, and stores the result in \langle str var \rangle.

New: 2015-09-18  
Updated: 2018-07-28

```latex
\texttt{\textbf{\texttt{str}}\texttt{\_put\_left:Nn}}\langle \texttt{str} \texttt{var} \rangle \{\langle \texttt{token} \texttt{list} \rangle\} \\
\texttt{\textbf{\texttt{str}}\texttt{\_put\_left:}}(\texttt{NV}|\texttt{Nx}|\texttt{cn}|\texttt{cV}|\texttt{cx}) \\
\texttt{\textbf{\texttt{str}}\texttt{\_gput\_left:Nn}}\langle \texttt{str} \texttt{var} \rangle \{\langle \texttt{token} \texttt{list} \rangle\} \\
\texttt{\textbf{\texttt{str}}\texttt{\_gput\_left:}}(\texttt{NV}|\texttt{Nx}|\texttt{cn}|\texttt{cV}|\texttt{cx})
```

Converts the \langle token list \rangle to a \langle string \rangle, and prepends the result to \langle str var \rangle. The current contents of the \langle str var \rangle are not automatically converted to a string.

New: 2015-09-18  
Updated: 2018-07-28

```latex
\texttt{\textbf{\texttt{str}}\texttt{\_put\_right:Nn}}\langle \texttt{str} \texttt{var} \rangle \{\langle \texttt{token} \texttt{list} \rangle\} \\
\texttt{\textbf{\texttt{str}}\texttt{\_put\_right:}}(\texttt{NV}|\texttt{Nx}|\texttt{cn}|\texttt{cV}|\texttt{cx}) \\
\texttt{\textbf{\texttt{str}}\texttt{\_gput\_right:Nn}}\langle \texttt{str} \texttt{var} \rangle \{\langle \texttt{token} \texttt{list} \rangle\} \\
\texttt{\textbf{\texttt{str}}\texttt{\_gput\_right:}}(\texttt{NV}|\texttt{Nx}|\texttt{cn}|\texttt{cV}|\texttt{cx})
```

Converts the \langle token list \rangle to a \langle string \rangle, and appends the result to \langle str var \rangle. The current contents of the \langle str var \rangle are not automatically converted to a string.

New: 2015-09-18  
Updated: 2018-07-28

### 16.3 String conditionals

```latex
\texttt{\textbf{\texttt{str}}\texttt{\_if\_empty:p:}}(\texttt{N} | \texttt{c} | \texttt{n} | \texttt{cN} | \texttt{cc} | \texttt{cV} | \texttt{cx})
```

Tests if the \langle string variable \rangle is entirely empty (i.e. contains no characters at all).

New: 2015-09-18  
Updated: 2022-03-21

```latex
\texttt{\textbf{\texttt{str}}\texttt{\_if\_eq:p:}}(\texttt{NN} | \texttt{Nc} | \texttt{cN} | \texttt{cc} | \texttt{cV} | \texttt{cx})
```

Compares the content of two \langle string variables \rangle and is logically true if the two contain the same characters in the same order. See \texttt{\textbf{\texttt{tl}}\texttt{\_if\_eq:}}(\texttt{NN} | \texttt{Nc} | \texttt{cN} | \texttt{cc} | \texttt{cV} | \texttt{cx}) to compare tokens (including their category codes) rather than characters.

New: 2015-09-18
\str_if_eq_p:nn \\
\str_if_eq_p:{(Vn|on|nV|VV|vn|nv|ee)} \\
\str_if_eq:{nnTF} \\
\str_if_eq:(Vn|on|nV|VV|vn|nv|ee)TF *

Updated: 2018-06-18

Compares the two ⟨token lists⟩ on a character by character basis (namely after converting them to strings), and is true if the two ⟨strings⟩ contain the same characters in the same order. Thus for example

\str_if_eq_p:no \{ abc \} \{ \tl_to_str:n \{ abc \} \}

is logically true. See \tl_if_eq:nnTF to compare tokens (including their category codes) rather than characters.

\str_if_in:NnTF \str_if_in:cnTF

Converts the ⟨token list⟩ to a ⟨string⟩ and tests if that ⟨string⟩ is found in the content of the ⟨str var⟩.

\str_if_in:nnTF

Converts both ⟨token lists⟩ to ⟨strings⟩ and tests whether ⟨string⟩ is found inside ⟨string1⟩.

\str_case:nn TF \str_case:Nn TF

Updated: 2022-03-21

Compares the ⟨test string⟩ in turn with each of the ⟨string cases⟩ (all token lists are converted to strings). If the two are equal (as described for \str_if_eq:nnTF) then the associated ⟨code⟩ is left in the input stream and other cases are discarded. If any of the cases are matched, the ⟨true code⟩ is also inserted into the input stream (after the code for the appropriate case), while if none match then the ⟨false code⟩ is inserted. The function \str_case:nn, which does nothing if there is no match, is also available.
\str_case_e:nn \str_case_e:nnTF {
⟨test string⟩
}
{
⟨string case1⟩ ⟨code case1⟩
⟨string case2⟩ ⟨code case2⟩
...
⟨string case_n⟩ ⟨code case_n⟩
}
⟨true code⟩
⟨false code⟩

Compares the full expansion of the ⟨test string⟩ in turn with the full expansion of the ⟨string cases⟩ (all token lists are converted to strings). If the two full expansions are equal (as described for \str_if_eq:nnTF) then the associated ⟨code⟩ is left in the input stream and other cases are discarded. If any of the cases are matched, the ⟨true code⟩ is also inserted into the input stream (after the code for the appropriate case), while if none match then the ⟨false code⟩ is inserted. The function \str_case_e:nn, which does nothing if there is no match, is also available. The ⟨test string⟩ is expanded in each comparison, and must always yield the same result: for example, random numbers must not be used within this string.

\str_compare_p:nNn \str_compare_p:eNe \str_compare:nNnTF \str_compare:eNeTF

Compares the two ⟨token lists⟩ on a character by character basis (namely after converting them to strings) in a lexicographic order according to the character codes of the characters. The ⟨relation⟩ can be <, =, or > and the test is true under the following conditions:

- for <, if the first string is earlier than the second in lexicographic order;
- for =, if the two strings have exactly the same characters;
- for >, if the first string is later than the second in lexicographic order.

Thus for example the following is logically true:
\str_compare_p:nNn { ab } < { abc }

\TeX{hackers note}: This is a wrapper around the \TeX primitive ⟨pdf⟩\strcmp. It is meant for programming and not for sorting textual contents, as it simply considers character codes and not more elaborate considerations of grapheme clusters, locale, etc.

16.4 Mapping over strings

All mappings are done at the current group level, i.e. any local assignments made by the ⟨function⟩ or ⟨code⟩ discussed below remain in effect after the loop.

\str_map_function:nNstr_map_function:NNstr_map_function:cN

Converts the ⟨token list⟩ to a ⟨string⟩ then applies ⟨function⟩ to every ⟨character⟩ in the ⟨string⟩ including spaces.
Converts the ⟨token list⟩ to a ⟨string⟩ then applies the ⟨inline function⟩ to every ⟨character⟩ in the ⟨str var⟩ including spaces. The ⟨inline function⟩ should consist of code which receives the ⟨character⟩ as #1.

Conversely to ⟨token list⟩ then applies ⟨code⟩ to every ⟨character⟩ in the ⟨string⟩ including spaces. The ⟨code⟩ receives each character as a trailing brace group. This is equivalent to \str_map_function:nN if the ⟨code⟩ consists of a single function.

Converses the ⟨token list⟩ to a ⟨string⟩ then stores each ⟨character⟩ in the ⟨string⟩ (including spaces) in turn in the ⟨string⟩ ⟨variable⟩ and applies the ⟨code⟩. The ⟨code⟩ will usually make use of the ⟨variable⟩, but this is not enforced. The assignments to the ⟨variable⟩ are local. Its value after the loop is the last ⟨character⟩ in the ⟨string⟩, or its original value if the ⟨string⟩ is empty. See also \str_map_inline:Nn.

Used to terminate a \str_map_... function before all characters in the ⟨string⟩ have been processed. This normally takes place within a conditional statement, for example

\str_map_inline:Nn \l_my_str
{\str_if_eq:nnT { #1 } { bingo } { \str_map_break: }% Do something useful}

See also \str_map_break:n. Use outside of a \str_map_... scenario leads to low level \TeX errors.

\textbf{\TeXhackers note:} When the mapping is broken, additional tokens may be inserted before continuing with the code that follows the loop. This depends on the design of the mapping function.
\str_map_break:n \{\texttt{<code>}\} \\
\text{Used to terminate a } \str_map\ldots \text{ function before all characters in the } \langle\text{string}\rangle \text{ have been processed, inserting the } \langle\text{code}\rangle \text{ after the mapping has ended. This normally takes place within a conditional statement, for example}

\str_map_inline:Nn \l_my_str \\
{ \str_if_eq:nnT { #1 } { bingo } { \str_map_break:n \{ <\texttt{code}> \} } \% \text{ Do something useful} }

Use outside of a } \str_map\ldots \text{ scenario leads to low level \TeX\ errors.}

\TeX\text{}hackers note: When the mapping is broken, additional tokens may be inserted before the } \langle\text{code}\rangle \text{ is inserted into the input stream. This depends on the design of the mapping function.}

\section{Working with the content of strings}

\str_use:N \str_use:c \\
\text{Recovers the content of a } \langle \text{str var} \rangle \text{ and places it directly in the input stream. An error is raised if the variable does not exist or if it is invalid. Note that it is possible to use a } \langle \text{str} \rangle \text{ directly without an accessor function.}

\str_count:n \str_count:N \str_count:c \str_count:n \str_count_ignore_spaces:n \\
\text{Leaves in the input stream the number of characters in the string representation of } \langle \text{token list} \rangle \text{, as an integer denotation. The functions differ in their treatment of spaces. In the case of } \str_count:N \text{ and } \str_count:n \text{, all characters including spaces are counted. The } \str_count_ignore_spaces:n \text{ function leaves the number of non-space characters in the input stream.}

\str_count_spaces:n \str_count_spaces:N \str_count_spaces:c \str_count_spaces:n \\
\text{Leaves in the input stream the number of space characters in the string representation of } \langle \text{token list} \rangle \text{, as an integer denotation. Of course, this function has no } \_\text{ignore_spaces} \text{ variant.}
Converting a token list into a string:

\texttt{\textbackslash str\_head:N} * \texttt{\textbackslash str\_head:n \{token list\}}
\texttt{\textbackslash str\_head:c} *
\texttt{\textbackslash str\_head:n}
\texttt{\textbackslash str\_head\_ignore\_spaces:n} *

Rev: 2015-09-18

Converts the \langle token list \rangle into a \langle string \rangle. The first character in the \langle string \rangle is then left in the input stream, with category code “other”. The functions differ if the first character is a space: \texttt{\textbackslash str\_head:N} and \texttt{\textbackslash str\_head:n} return a space token with category code 10 (blank space), while the \texttt{\textbackslash str\_head\_ignore\_spaces:n} function ignores this space character and leaves the first non-space character in the input stream. If the \langle string \rangle is empty (or only contains spaces in the case of the \_ignore\_spaces function), then nothing is left on the input stream.

Converting a token list to a string, removing the first character, and leaving the remaining characters (if any) in the input stream, with category codes 12 and 10 (for spaces). The functions differ in the case where the first character is a space: \texttt{\textbackslash str\_tail:N} and \texttt{\textbackslash str\_tail:n} only trim that space, while \texttt{\textbackslash str\_tail\_ignore\_spaces:n} removes the first non-space character and any space before it. If the \langle token list \rangle is empty (or blank in the case of the \_ignore\_spaces variant), then nothing is left on the input stream.

Converting the \langle token list \rangle to a \langle string \rangle, and leaves in the input stream the character in position \langle integer expression \rangle of the \langle string \rangle, starting at 1 for the first (left-most) character. In the case of \texttt{\textbackslash str\_item:Nn} and \texttt{\textbackslash str\_item:nn}, all characters including spaces are taken into account. The \texttt{\textbackslash str\_item\_ignore\_spaces:nn} function skips spaces when counting characters. If the \langle integer expression \rangle is negative, characters are counted from the end of the \langle string \rangle. Hence, \texttt{-1} is the right-most character, \textit{etc.}
\str_range:Nnn \str_range:nnn \{\text{token list}\} \{\text{start index}\} \{\text{end index}\}
\str_range:nnn \str_range:nnn \str_range_ignore_spaces:nnn

Converts the \textit{token list} to a \textit{(string)}, and leaves in the input stream the characters from the \textit{start index} to the \textit{(end index)} inclusive. Spaces are preserved and counted as items (contrast this with \texttt{tl\_range:nnn} where spaces are not counted as items and are possibly discarded from the output).

Here \textit{(start index)} and \textit{(end index)} should be integer denotations. For describing in detail the functions’ behavior, let \(m\) and \(n\) be the start and end index respectively. If either is 0, the result is empty. A positive index means ‘start counting from the left end’, a negative index means ‘start counting from the right end’. Let \(l\) be the count of the token list.

The \textit{actual start point} is determined as \(M = m\) if \(m > 0\) and as \(M = l + m + 1\) if \(m < 0\). Similarly the \textit{actual end point} is \(N = n\) if \(n > 0\) and \(N = l + n + 1\) if \(n < 0\). If \(M > N\), the result is empty. Otherwise it consists of all items from position \(M\) to position \(N\) inclusive; for the purpose of this rule, we can imagine that the token list extends at infinity on either side, with void items at positions \(s\) for \(s \leq 0\) or \(s > l\). For instance,

\begin{verbatim}
\io_term:{ \str_range:nnn { abcdef } { 2 } { 5 } }\n\io_term:{ \str_range:nnn { abcdef } { -4 } { -1 } }\n\io_term:{ \str_range:nnn { abcdef } { -2 } { -1 } }\n\io_term:{ \str_range:nnn { abcdef } { 0 } { -1 } }
\end{verbatim}

prints \texttt{bcde}, \texttt{cdef}, \texttt{ef}, and an empty line to the terminal. The \textit{(start index)} must always be smaller than or equal to the \textit{(end index)}: if this is not the case then no output is generated. Thus

\begin{verbatim}
\io_term:{ \str_range:nnn { abcdef } { 5 } { 2 } }\n\io_term:{ \str_range:nnn { abcdef } { -1 } { -4 } }
\end{verbatim}

both yield empty strings.

The behavior of \texttt{\str\_range\_ignore\_spaces:nnn} is similar, but spaces are removed before starting the job. The input

\begin{verbatim}
\io_term:{ \str_range:nnn { abcdefg } { 2 } { 5 } }\n\io_term:{ \str_range:nnn { abcdef } { 2 } { -3 } }\n\io_term:{ \str_range:nnn { abcdef } { -6 } { 5 } }\n\io_term:{ \str_range:nnn { abcdefg } { -6 } { -3 } }
\io_term:{ \str_range:nnn { abc-efg } { 2 } { 5 } }\n\io_term:{ \str_range:nnn { abc-efg } { 2 } { -3 } }\n\io_term:{ \str_range:nnn { abc-efg } { -6 } { 5 } }\n\io_term:{ \str_range:nnn { abc-efg } { -6 } { -3 } }
\io_term:{ \str_range:nnn { abcdefg } { 2 } { 5 } }\n\io_term:{ \str_range:nnn { abcdef } { 2 } { -3 } }\n\io_term:{ \str_range:nnn { abcdef } { -6 } { 5 } }\n\io_term:{ \str_range:nnn { abcdef } { -6 } { -3 } }
\end{verbatim}

128
\iow_term:x { \str_range_ignore_spaces:nnn { abcd-efg } { 2 } { 5 } }
\iow_term:x { \str_range_ignore_spaces:nnn { abcd-efg } { 2 } { -3 } }
\iow_term:x { \str_range_ignore_spaces:nnn { abcd-efg } { -6 } { 5 } }
\iow_term:x { \str_range_ignore_spaces:nnn { abcd-efg } { -6 } { -3 } }

will print four instances of bgde, four instances of bc e and eight instances of bcde.

16.6 Modifying string variables

\str_replace_once:Nnn \str_replace_once:cnn \str_greplace_once:Nnn \str_greplace_once:cnn

\str_replace_all:Nnn \str_replace_all:cnn \str_greplace_all:Nnn \str_greplace_all:cnn

\str_remove_once:Nn \str_remove_once:cn \str_gremove_once:Nn \str_gremove_once:cn

\str_set:Nn \l_tmpa_str {abbccd} \str_remove_all:Nn \l_tmpa_str {bc}

results in \l_tmpa_str containing abcd.
16.7 String manipulation

\texttt{\str\_lowercase:n \{\textit{tokens}\}} \hspace{1cm} \texttt{\str\_uppercase:n \{\textit{tokens}\}}

Converts the input (\textit{tokens}) to their string representation, as described for \texttt{\tl\_to\_-str:n}, and then to the lower or upper case representation using a one-to-one mapping as described by the Unicode Consortium file \texttt{UnicodeData.txt}.

These functions are intended for case changing programmatic data in places where upper/lower case distinctions are meaningful. One example would be automatically generating a function name from user input where some case changing is needed. In this situation the input is programmatic, not textual, case does have meaning and a language-independent one-to-one mapping is appropriate. For example

\begin{verbatim}
\cs_new_protected:Npn \myfunc:nn #1#2
{\cs_set_protected:cpn
 {user
 \str\_uppercase:f \{ \tl\_head:n \{#1\} \}
 \str\_lowercase:f \{ \tl\_tail:n \{#1\} \}
}
 {#2}
}
\end{verbatim}

would be used to generate a function with an auto-generated name consisting of the upper case equivalent of the supplied name followed by the lower case equivalent of the rest of the input.

These functions should not be used for

- Caseless comparisons: use \texttt{\str\_foldcase:n} for this situation (case folding is distinct from lower casing).

- Case changing text for typesetting: see the \texttt{\text\_lowercase:n(n)}, \texttt{\text\_uppercase:n(n)} and \texttt{\text\_titlecase:n(n)} functions which correctly deal with context-dependence and other factors appropriate to text case changing.
\str_foldcase{n} \\ \str_foldcase{V} \\ New: 2019-11-26

\str_foldcase{n} \{\langle tokens\rangle\}

Converts the input \langle tokens\rangle to their string representation, as described for \tl_to_str:n, and then folds the case of the resulting \langle string\rangle to remove case information. The result of this process is left in the input stream.

String folding is a process used for material such as identifiers rather than for “text”. The folding provided by \str_foldcase:n follows the mappings provided by the Unicode Consortium, who state:

Case folding is primarily used for caseless comparison of text, such as identifiers in a computer program, rather than actual text transformation. Case folding in Unicode is based on the lowercase mapping, but includes additional changes to the source text to help make it language-insensitive and consistent. As a result, case-folded text should be used solely for internal processing and generally should not be stored or displayed to the end user.

The folding approach implemented by \str_foldcase:n follows the “full” scheme defined by the Unicode Consortium (e.g. SSfolds to SS). As case-folding is a language-insensitive process, there is no special treatment of Turkic input (i.e. İ always folds to i and not to ı).

### 16.8 Viewing strings

\str_show:n \\ \str_show:c \\ \str_show:n \\ New: 2015-09-18 \\ Updated: 2021-04-29

\str_show:n \{str var\}

Displays the content of the \langle str var\rangle on the terminal.

\str_log:n \\ \str_log:c \\ \str_log:n \\ New: 2019-02-15 \\ Updated: 2021-04-29

\str_log:n \{str var\}

Writes the content of the \langle str var\rangle in the log file.
16.9 Constant strings

```
c_andersand_str
\c_atsign_str
\c_backslash_str
\c_left_brace_str
\c_right_brace_str
\c_circumflex_str
\c_colon_str
\c_dollar_str
\c_hash_str
\c_percent_str
\c_tilde_str
\c_underscore_str
\c_zero_str
```

Constant strings, containing a single character token, with category code 12.

New: 2015-09-19
Updated: 2020-12-22

16.10 Scratch strings

```
l_tmpa_str
l_tmpb_str
```

Scratch strings for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

```
g_tmpa_str
g_tmpb_str
```

Scratch strings for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
Chapter 17

The \texttt{l3str-convert} package: string encoding conversions

17.1 Encoding and escaping schemes

Traditionally, string encodings only specify how strings of characters should be stored as bytes. However, the resulting lists of bytes are often to be used in contexts where only a restricted subset of bytes are permitted (e.g., PDF string objects, URLs). Hence, storing a string of characters is done in two steps.

- The code points (“character codes”) are expressed as bytes following a given “encoding”. This can be \texttt{utf-16}, \texttt{iso 8859-1}, etc. See Table 1 for a list of supported encodings.\footnote{Encodings and escapings will be added as they are requested.}

- Bytes are translated to \TeX tokens through a given “escaping”. Those are defined for the most part by the PDF file format. See Table 2 for a list of escaping methods supported.\footnote{Encodings and escapings will be added as they are requested.}
Table 1: Supported encodings. Non-alphanumeric characters are ignored, and capital letters are lower-cased before searching for the encoding in this list.

<table>
<thead>
<tr>
<th>Encoding</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>utf8</td>
<td>UTF-8</td>
</tr>
<tr>
<td>utf16</td>
<td>UTF-16, with byte-order mark</td>
</tr>
<tr>
<td>utf16be</td>
<td>UTF-16, big-endian</td>
</tr>
<tr>
<td>utf16le</td>
<td>UTF-16, little-endian</td>
</tr>
<tr>
<td>utf32</td>
<td>UTF-32, with byte-order mark</td>
</tr>
<tr>
<td>utf32be</td>
<td>UTF-32, big-endian</td>
</tr>
<tr>
<td>utf32le</td>
<td>UTF-32, little-endian</td>
</tr>
<tr>
<td>iso88591, latin1</td>
<td>ISO 8859-1</td>
</tr>
<tr>
<td>iso88592, latin2</td>
<td>ISO 8859-2</td>
</tr>
<tr>
<td>iso88593, latin3</td>
<td>ISO 8859-3</td>
</tr>
<tr>
<td>iso88594, latin4</td>
<td>ISO 8859-4</td>
</tr>
<tr>
<td>iso88595</td>
<td>ISO 8859-5</td>
</tr>
<tr>
<td>iso88596</td>
<td>ISO 8859-6</td>
</tr>
<tr>
<td>iso88597</td>
<td>ISO 8859-7</td>
</tr>
<tr>
<td>iso88598</td>
<td>ISO 8859-8</td>
</tr>
<tr>
<td>iso88599, latin5</td>
<td>ISO 8859-9</td>
</tr>
<tr>
<td>iso885910, latin6</td>
<td>ISO 8859-10</td>
</tr>
<tr>
<td>iso885911</td>
<td>ISO 8859-11</td>
</tr>
<tr>
<td>iso885913, latin7</td>
<td>ISO 8859-13</td>
</tr>
<tr>
<td>iso885914, latin8</td>
<td>ISO 8859-14</td>
</tr>
<tr>
<td>iso885915, latin9</td>
<td>ISO 8859-15</td>
</tr>
<tr>
<td>iso885916, latin10</td>
<td>ISO 8859-16</td>
</tr>
<tr>
<td>cllist</td>
<td>Comma-list of integers</td>
</tr>
<tr>
<td>empty</td>
<td>Native (Unicode) string</td>
</tr>
<tr>
<td>default</td>
<td>Like utf8 with 8-bit engines, and like native with unicode-engines</td>
</tr>
</tbody>
</table>

Table 2: Supported escapings. Non-alphanumeric characters are ignored, and capital letters are lower-cased before searching for the escaping in this list.

<table>
<thead>
<tr>
<th>Escaping</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>bytes, or empty</td>
<td>Arbitrary bytes</td>
</tr>
<tr>
<td>hex, hexadecimal</td>
<td>Byte = two hexadecimal digits</td>
</tr>
<tr>
<td>name</td>
<td>See <code>\pdfescapename</code></td>
</tr>
<tr>
<td>string</td>
<td>See <code>\pdfescapestring</code></td>
</tr>
<tr>
<td>url</td>
<td>Encoding used in URLs</td>
</tr>
</tbody>
</table>
17.2 Conversion functions

```
\str_set_convert:Nnnn
\str_gset_convert:Nnnn
```

This function converts the \langle string \rangle from the encoding given by \langle name 1 \rangle to the encoding given by \langle name 2 \rangle, and stores the result in the \langle str var \rangle. Each \langle name \rangle can have the form \langle encoding \rangle or \langle encoding ⟩/\langle escaping \rangle, where the possible values of \langle encoding \rangle and \langle escaping \rangle are given in Tables 1 and 2, respectively. The default escaping is to input and output bytes directly. The special case of an empty \langle name \rangle indicates the use of “native” strings, 8-bit for pdf\TeX, and Unicode strings for the other two engines.

For example,

```
\str_set_convert:Nnnn \l_foo_str { Hello! } { } { utf16/hex }
```

results in the variable \l_foo_str holding the string \texttt{FEFF00480065006C006F0021}. This is obtained by converting each character in the (native) string \texttt{Hello!} to the UTF-16 encoding, and expressing each byte as a pair of hexadecimal digits. Note the presence of a (big-endian) byte order mark "FEFF, which can be avoided by specifying the encoding utf16be/hex.

An error is raised if the \langle string \rangle is not valid according to the \langle escaping 1 \rangle and \langle encoding 1 \rangle, or if it cannot be reencoded in the \langle encoding 2 \rangle and \langle escaping 2 \rangle (for instance, if a character does not exist in the \langle encoding 2 \rangle). Erroneous input is replaced by the Unicode replacement character \texttt{FFFD}, and characters which cannot be reencoded are replaced by either the replacement character \texttt{FFFD} if it exists in the \langle encoding 2 \rangle, or an encoding-specific replacement character, or the question mark character.

```
\str_set_convert:NnnnTF
\str_gset_convert:NnnnTF
```

As \str_set_convert:Nnnn, converts the \langle string \rangle from the encoding given by \langle name 1 \rangle to the encoding given by \langle name 2 \rangle, and assigns the result to \langle str var \rangle. Contrarily to \str_set_convert:Nnnn, the conditional variant does not raise errors in case the \langle string \rangle is not valid according to the \langle name 1 \rangle encoding, or cannot be expressed in the \langle name 2 \rangle encoding. Instead, the \langle false code \rangle is performed.

17.3 Conversion by expansion (for PDF contexts)

A small number of expandable functions are provided for use in PDF string/name contexts. These assume \texttt{UTF-8} and no escaping in the input.

```
\str_convert_pdfname:n *
```

As \str_convert_pdfname:n, converts the \langle string \rangle on a byte-by-byte basis with non-ASCII codepoints escaped using hashes.

17.4 Possibilities, and things to do

Encoding/escaping-related tasks.
• In Xe\TeX/Lua\TeX, would it be better to use the \ldots approach to build a string from a given list of character codes? Namely, within a group, assign 0–9a–f and all characters we want to category “other”, then assign ~ the category superscript, and use \scantokens.

• Change \str_set_convert:Nnnn to expand its last two arguments.

• Describe the internal format in the code comments. Refuse code points in [\texttt{\textasciitilde}D800, \texttt{\textasciitilde}DFFF] in the internal representation?

• Add documentation about each encoding and escaping method, and add examples.

• The \texttt{hex} unescaping should raise an error for odd-token count strings.

• Decide what bytes should be escaped in the url escaping. Perhaps the characters !’\(\texttt{-./}0123456789_\texttt{}\) are safe, and all other characters should be escaped?

• Automate generation of 8-bit mapping files.

• Change the framework for 8-bit encodings: for decoding from 8-bit to Unicode, use 256 integer registers; for encoding, use a tree-box.

• More encodings (see Heiko’s \texttt{stringenc}). CESU?

• More escapings: \texttt{ascii85}, shell escapes, lua escapes, \textit{etc.}? 
Chapter 18

The \texttt{l3quark} package

Quarks

Two special types of constants in \LaTeX are “quarks” and “scan marks”. By convention all constants of type quark start out with $\q_-$ and scan marks start with $\s_-$.

18.1 Quarks

Quarks are control sequences (and in fact, token lists) that expand to themselves and should therefore \textit{never} be executed directly in the code. This would result in an endless loop!

They are meant to be used as delimiter in weird functions, the most common use case being the ‘stop token’ (\textit{i.e.} $\q\texttt{\_stop}$). For example, when writing a macro to parse a user-defined date

\begin{verbatim}
\date_parse:n {19/June/1981}
\end{verbatim}

one might write a command such as

\begin{verbatim}
\cs_new:Npn \date_parse:n #1 { \date_parse_aux:w #1 \q_stop }
\cs_new:Npn \date_parse_aux:w #1 / #2 / #3 \q_stop
{ <do something with the date> }
\end{verbatim}

Quarks are sometimes also used as error return values for functions that receive erroneous input. For example, in the function $\texttt{\prop\_get:NnN}$ to retrieve a value stored in some key of a property list, if the key does not exist then the return value is the quark $\q\texttt{\_no\_value}$. As mentioned above, such quarks are extremely fragile and it is imperative when using such functions that code is carefully written to check for pathological cases to avoid leakage of a quark into an uncontrolled environment.

Quarks also permit the following ingenious trick when parsing tokens: when you pick up a token in a temporary variable and you want to know whether you have picked up a particular quark, all you have to do is compare the temporary variable to the quark using $\texttt{\tl\_if\_eq:NNTF}$. A set of special quark testing functions is set up below. All the quark testing functions are expandable although the ones testing only single tokens are much faster.
18.2 Defining quarks

\texttt{\textbackslash quark\_new:N}\quad \texttt{\textbackslash quark\_new:N (quark)}

Creates a new \texttt{\textlangle quark \textrangle} which expands only to \texttt{\langle quark \rangle}. The \texttt{\langle quark \rangle} is defined globally, and an error message is raised if the name was already taken.

\texttt{\textbackslash q\_stop}\quad \texttt{\textbackslash q\_stop}

Used as a marker for delimited arguments, such as

\texttt{\cs_set:Npn \tmp:w \#1\#2 \textbackslash q\_stop \{\#1\}}

\texttt{\textbackslash q\_mark}\quad \texttt{\textbackslash q\_mark}

Used as a marker for delimited arguments when \texttt{\textbackslash q\_stop} is already in use.

\texttt{\textbackslash q\_nil}\quad \texttt{\textbackslash q\_nil}

Quark to mark a null value in structured variables or functions. Used as an end delimiter when this may itself need to be tested (in contrast to \texttt{\textbackslash q\_stop}, which is only ever used as a delimiter).

\texttt{\textbackslash q\_no\_value}\quad \texttt{\textbackslash q\_no\_value}

A canonical value for a missing value, when one is requested from a data structure. This is therefore used as a “return” value by functions such as \texttt{\prop_get:NnN} if there is no data to return.

18.3 Quark tests

The method used to define quarks means that the single token (N) tests are faster than the multi-token (n) tests. The latter should therefore only be used when the argument can definitely take more than a single token.

\texttt{\textbackslash quark\_if\_nil\_p:N} \star \texttt{\textbackslash quark\_if\_nil\_p:N (token)}

Tests if the \texttt{(token)} is equal to \texttt{\textbackslash q\_nil}.

\texttt{\textbackslash quark\_if\_nil\_p:n} \star \texttt{\textbackslash quark\_if\_nil\_p:n (token list)}

Tests if the \texttt{(token list)} contains only \texttt{\textbackslash q\_nil} (distinct from \texttt{(token list) being empty or containing \textbackslash q\_nil plus one or more other tokens}).

\texttt{\textbackslash quark\_if\_no\_value\_p:N} \star \texttt{\textbackslash quark\_if\_no\_value\_p:N (token)}

Tests if the \texttt{(token)} is equal to \texttt{\textbackslash q\_no\_value}.

\texttt{\textbackslash quark\_if\_no\_value\_p:n} \star \texttt{\textbackslash quark\_if\_no\_value\_p:n (token list)}

Tests if the \texttt{(token list)} contains only \texttt{\textbackslash q\_no\_value} (distinct from \texttt{(token list) being empty or containing \textbackslash q\_no\_value plus one or more other tokens}).
18.4 Recursion

This module provides a uniform interface to intercepting and terminating loops as when one is doing tail recursion. The building blocks follow below and an example is shown in Section 18.4.1.

\texttt{\textbackslash q\_recursion\_tail} This quark is appended to the data structure in question and appears as a real element there. This means it gets any list separators around it.

\texttt{\textbackslash q\_recursion\_stop} This quark is added after the data structure. Its purpose is to make it possible to terminate the recursion at any point easily.

\texttt{\textbackslash quark\_if\_recursion\_tail\_stop:N} \langle token \rangle \texttt{\textbackslash quark\_if\_recursion\_tail\_stop:N \star} Tests if \langle token \rangle contains only the marker \texttt{\textbackslash q\_recursion\_tail}, and if so uses \texttt{\textbackslash use\_none\_delimit\_by\_q\_recursion\_stop:w} to terminate the recursion that this belongs to. The recursion input must include the marker tokens \texttt{\textbackslash q\_recursion\_tail} and \texttt{\textbackslash q\_recursion\_stop} as the last two items.

\texttt{\textbackslash quark\_if\_recursion\_tail\_stop:n} \star \texttt{\textbackslash quark\_if\_recursion\_tail\_stop:o} \star Updated: 2011-09-06

Tests if the \langle token list \rangle contains only \texttt{\textbackslash q\_recursion\_tail}, and if so uses \texttt{\textbackslash use\_i\_delimit\_by\_q\_recursion\_stop:w} to terminate the recursion that this belongs to. The recursion input must include the marker tokens \texttt{\textbackslash q\_recursion\_tail} and \texttt{\textbackslash q\_recursion\_stop} as the last two items. The \langle insertion \rangle code is then added to the input stream after the recursion has ended.

\texttt{\textbackslash quark\_if\_recursion\_tail\_stop:nn} \star \texttt{\textbackslash quark\_if\_recursion\_tail\_stop:on} \star Updated: 2011-09-06

Tests if the \langle token list \rangle contains only \texttt{\textbackslash q\_recursion\_tail}, and if so uses \texttt{\textbackslash use\_i\_delimit\_by\_q\_recursion\_stop:w} to terminate the recursion that this belongs to. The recursion input must include the marker tokens \texttt{\textbackslash q\_recursion\_tail} and \texttt{\textbackslash q\_recursion\_stop} as the last two items. The \langle insertion \rangle code is then added to the input stream after the recursion has ended.
Tests if \textit{(token list)} contains only \texttt{\textbackslash q\_recursion\_tail}, and if so terminates the recursion using \texttt{(\textit{type})\_map\_break}. The recursion end should be marked by \texttt{\textbackslash prg\_break\_point}:\texttt{\textbackslash (type)\_map\_break}.

18.4.1 An example of recursion with quarks

Quarks are mainly used internally in the expl3 code to define recursion functions such as \texttt{\textit{tl\_map\_inline}:nn} and so on. Here is a small example to demonstrate how to use quarks in this fashion. We shall define a command called \texttt{\textit{my\_map\_dbl}:nn} which takes a token list and applies an operation to every pair of tokens. For example, \texttt{\textit{my\_map\_dbl}:nn \{abcd\} \{\texttt{\textbackslash [-\#1--\#2--]}\}} would produce "\texttt{[-a-b-] [-c-d-]} "\texttt{. Using quarks to define such functions simplifies their logic and ensures robustness in many cases.}

Here’s the definition of \texttt{\textit{my\_map\_dbl}:nn}. First of all, define the function that does the processing based on the inline function argument \texttt{#2}. Then initiate the recursion using an internal function. The token list \texttt{#1} is terminated using \texttt{\textbf{\textbackslash q\_recursion\_tail}}, with delimiters according to the type of recursion (here a pair of \texttt{\textbf{\textbackslash q\_recursion\_tail}}), concluding with \texttt{\textbf{\textbackslash q\_recursion\_stop}}. These quarks are used to mark the end of the token list being operated upon.

\begin{verbatim}
\cs_new:Npn \my_map_dbl:nn \#1 \#2 
{
  \cs_set:Npn \__my_map_dbl_fn:nn ##1 ##2 {#2}
  \__my_map_dbl:nn \#1 \q_recursion_tail \q_recursion_tail 
  \q_recursion_stop
}
\end{verbatim}

The definition of the internal recursion function follows. First check if either of the input tokens are the termination quarks. Then, if not, apply the inline function to the two arguments.

\begin{verbatim}
\cs_new:Nn \__my_map_dbl:nn
{
  \quark_if_recursion_tail_stop:n \{#1\}
  \quark_if_recursion_tail_stop:n \{#2\}
  \__my_map_dbl_fn:nn \{#1\} \{#2\}
}\end{verbatim}

Finally, recurse:

\begin{verbatim}
\__my_map_dbl:nn
\end{verbatim}

Note that contrarily to \textsc{Ikex3} built-in mapping functions, this mapping function cannot be nested, since the second map would overwrite the definition of \texttt{\__my_map_dbl\_fn:nn}.
18.5 Scan marks

Scan marks are control sequences set equal to \(\texttt{\textbackslash scan\_stop}:\), hence never expand in an expansion context and are (largely) invisible if they are encountered in a typesetting context.

Like quarks, they can be used as delimiters in weird functions and are often safer to use for this purpose. Since they are harmless when executed by \TeX\ in non-expandable contexts, they can be used to mark the end of a set of instructions. This allows to skip to that point if the end of the instructions should not be performed (see \texttt{l3regex}).

\(\texttt{\textbackslash scan\_new:N} \langle \text{scan mark} \rangle\)

Creates a new \(\langle \text{scan mark} \rangle\) which is set equal to \(\texttt{\textbackslash scan\_stop}:\). The \(\langle \text{scan mark} \rangle\) is defined globally, and an error message is raised if the name was already taken by another scan mark.

\(\texttt{\textbackslash s\_stop}\)

Used at the end of a set of instructions, as a marker that can be jumped to using \(\texttt{\use\_none\_delimit\_by\_s\_stop:w}\).

\(\texttt{\use\_none\_delimit\_by\_s\_stop:w \ast} \langle \text{tokens} \rangle \texttt{\textbackslash s\_stop}\)

Removes the \(\langle \text{tokens} \rangle\) and \texttt{\textbackslash s\_stop} from the input stream. This leads to a low-level \TeX\ error if \texttt{\textbackslash s\_stop} is absent.
Chapter 19

The \texttt{l3seq} package

Sequences and stacks

\LaTeX{}X3 implements a “sequence” data type, which contain an ordered list of entries which may contain any \emph{(balanced text)}. It is possible to map functions to sequences such that the function is applied to every item in the sequence.

Sequences are also used to implement stack functions in \LaTeX{}X3. This is achieved using a number of dedicated stack functions.

19.1 Creating and initialising sequences

\texttt{\seq_new:N} \texttt{\seq_new:c}

\texttt{\seq_clear:N} \texttt{\seq_clear:c}

\texttt{\seq_gclear:N} \texttt{\seq_gclear:c}

\texttt{\seq_clear_new:N} \texttt{\seq_clear_new:c}

\texttt{\seq_gclear_new:N} \texttt{\seq_gclear_new:c}

\texttt{\seq_set_eq:NN} \texttt{\seq_set_eq:cnnc} \texttt{\seq_gset_eq:NN} \texttt{\seq_gset_eq:cnnc}

\texttt{\seq_set_eq:NW} \texttt{\seq_set_eq:(cNc)c} \texttt{\seq_gset_eq:NW} \texttt{\seq_gset_eq:(cNc)c}

\texttt{\seq_new:N} \texttt{(sequence)}

Creates a new \texttt{(sequence)} or raises an error if the name is already taken. The declaration is global. The \texttt{(sequence)} initially contains no items.

\texttt{\seq_clear:N} \texttt{(sequence)}

Clears all items from the \texttt{(sequence)}.

\texttt{\seq_clear_new:N} \texttt{(sequence)}

Ensures that the \texttt{(sequence)} exists globally by applying \texttt{\seq_new:N} if necessary, then applies \texttt{\seq_(g)clear:N} to leave the \texttt{(sequence)} empty.

\texttt{\seq_set_eq:NW} \texttt{\seq_set_eq:(sequence}_1) \texttt{\seq_set_eq:(sequence}_2)

Sets the content of \texttt{(sequence}_1) equal to that of \texttt{(sequence}_2).
Converts the data in the ⟨comma list⟩ into a ⟨sequence⟩: the original ⟨comma list⟩ is unchanged.

Creates a new constant ⟨seq var⟩ or raises an error if the name is already taken. The ⟨seq var⟩ is set globally to contain the items in the ⟨comma list⟩.

Splits the ⟨token list⟩ into ⟨items⟩ separated by ⟨delimiter⟩, and assigns the result to the ⟨sequence⟩. Spaces on both sides of each ⟨item⟩ are ignored, then one set of outer braces is removed (if any): this space trimming behaviour is identical to that of \l3clist functions. Empty ⟨items⟩ are preserved by \seq_set_split:Nnn, and can be removed afterwards using \seq_remove_all:Nn ⟨sequence⟩ \0. The ⟨delimiter⟩ may not contain {, } or # (assuming \TeX’s normal category code régime). If the ⟨delimiter⟩ is empty, the ⟨token list⟩ is split into ⟨items⟩ as a ⟨token list⟩. See also \seq_set_split_keep_spaces:Nnn, which omits space stripping.

Splits the ⟨token list⟩ into ⟨items⟩ separated by ⟨delimiter⟩, and assigns the result to the ⟨sequence⟩. One set of outer braces is removed (if any) but any surrounding spaces are retained: any braces inside one or more spaces are therefore kept. Empty ⟨items⟩ are preserved by \seq_set_split_keep_spaces:Nnn, and can be removed afterwards using \seq_remove_all:Nn ⟨sequence⟩ \0. The ⟨delimiter⟩ may not contain {, } or # (assuming \TeX’s normal category code régime). If the ⟨delimiter⟩ is empty, the ⟨token list⟩ is split into ⟨items⟩ as a ⟨token list⟩. See also \seq_set_split:Nnn, which removes spaces around the delimiters.

Concatenates the content of ⟨sequence⟩ and ⟨sequence⟩ together and saves the result in ⟨sequence⟩. The items in ⟨sequence⟩ are placed at the left side of the new sequence.
19.2 Appending data to sequences

\[\text{\texttt{\textbackslash seq\_put\_left:Nn}}\] \texttt{\textbackslash seq\_put\_left:Nn \langle sequence\rangle \{\langle item\rangle\}}

\[\text{\texttt{\textbackslash seq\_put\_left:Nn}}\] \texttt{\texttt{\textbackslash seq\_put\_left:Nn \langle sequence\rangle \langle item\rangle}}

\[\text{\texttt{\textbackslash seq\_gput\_left:Nn}}\] \texttt{\texttt{\textbackslash seq\_gput\_left:Nn \langle sequence\rangle \langle item\rangle}}

Appends the \langle item\rangle to the left of the \langle sequence\rangle.

\[\text{\texttt{\textbackslash seq\_put\_right:Nn}}\] \texttt{\texttt{\textbackslash seq\_put\_right:Nn \langle sequence\rangle \{\langle item\rangle\}}

\[\text{\texttt{\textbackslash seq\_put\_right:Nn}}\] \texttt{\texttt{\textbackslash seq\_put\_right:Nn \langle sequence\rangle \langle item\rangle}}

\[\text{\texttt{\textbackslash seq\_gput\_right:Nn}}\] \texttt{\texttt{\textbackslash seq\_gput\_right:Nn \langle sequence\rangle \langle item\rangle}}

Appends the \langle item\rangle to the right of the \langle sequence\rangle.

19.3 Recovering items from sequences

Items can be recovered from either the left or the right of \textit{sequences}. For implementation reasons, the actions at the left of the sequence are faster than those acting on the right. These functions all assign the recovered material locally, \textit{i.e.} setting the \langle token list variable\rangle used with \texttt{\textbackslash tl\_set:Nn} and \textit{never} \texttt{\textbackslash tl\_gset:Nn}.

\[\text{\texttt{\textbackslash seq\_get\_left:NN}}\] \texttt{\texttt{\textbackslash seq\_get\_left:NN \langle sequence\rangle \langle token list variable\rangle}}

\[\text{\texttt{\textbackslash seq\_get\_left:NN}}\] \texttt{\texttt{\textbackslash seq\_get\_left:NN \langle sequence\rangle \langle token list variable\rangle}}

Stores the left-most item from a \langle sequence\rangle in the \langle token list variable\rangle without removing it from the \langle sequence\rangle. The \langle token list variable\rangle is assigned locally. If \langle sequence\rangle is empty the \langle token list variable\rangle is set to the special marker \texttt{\textbackslash q\_no\_value}.

\[\text{\texttt{\textbackslash seq\_pop\_left:NN}}\] \texttt{\texttt{\textbackslash seq\_pop\_left:NN \langle sequence\rangle \langle token list variable\rangle}}

\[\text{\texttt{\textbackslash seq\_pop\_left:NN}}\] \texttt{\texttt{\textbackslash seq\_pop\_left:NN \langle sequence\rangle \langle token list variable\rangle}}

Pops the left-most item from a \langle sequence\rangle into the \langle token list variable\rangle, \textit{i.e.} removes the item from the sequence and stores it in the \langle token list variable\rangle. Both of the variables are assigned locally. If \langle sequence\rangle is empty the \langle token list variable\rangle is set to the special marker \texttt{\textbackslash q\_no\_value}.
Pops the left-most item from a \textit{sequence} into the \textit{token list variable}, i.e. removes the item from the sequence and stores it in the \textit{token list variable}. The \textit{sequence} is modified globally, while the assignment of the \textit{token list variable} is local. If \textit{sequence} is empty the \textit{token list variable} is set to the special marker \texttt{\textbackslash q\_no\_value}.

Pops the right-most item from a \textit{sequence} into the \textit{token list variable}, i.e. removes the item from the sequence and stores it in the \textit{token list variable}. Both of the variables are assigned locally. If \textit{sequence} is empty the \textit{token list variable} is set to the special marker \texttt{\textbackslash q\_no\_value}.

Indexing items in the \textit{sequence} from 1 at the top (left), this function evaluates the \textit{integer expression} and leaves the appropriate item from the sequence in the input stream. If the \textit{integer expression} is negative, indexing occurs from the bottom (right) of the sequence. If the \textit{integer expression} is larger than the number of items in the \textit{sequence} (as calculated by \texttt{\textbackslash seq\_count:N}) then the function expands to nothing.

\textbf{Texhackers note:} The result is returned within the \texttt{\unexpanded} primitive (\texttt{\exp_not:n}), which means that the \textit{item} does not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.

Selects a pseudo-random item of the \textit{sequence}. If the \textit{sequence} is empty the result is empty. This is not available in older versions of Xe\TeX.

\textbf{Texhackers note:} The result is returned within the \texttt{\unexpanded} primitive (\texttt{\exp_not:n}), which means that the \textit{item} does not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.

### 19.4 Recovering values from sequences with branching

The functions in this section combine tests for non-empty sequences with recovery of an item from the sequence. They offer increased readability and performance over separate testing and recovery phases.
If the \( \text{sequence} \) is empty, leaves the \( \text{false code} \) in the input stream. The value of the \( \text{token list variable} \) is not defined in this case and should not be relied upon. If the \( \text{sequence} \) is non-empty, stores the left-most item from the \( \text{sequence} \) in the \( \text{token list variable} \) without removing it from the \( \text{sequence} \), then leaves the \( \text{true code} \) in the input stream. The \( \text{token list variable} \) is assigned locally.

If the \( \text{sequence} \) is empty, leaves the \( \text{false code} \) in the input stream. The value of the \( \text{token list variable} \) is not defined in this case and should not be relied upon. If the \( \text{sequence} \) is non-empty, stores the right-most item from the \( \text{sequence} \) in the \( \text{token list variable} \) without removing it from the \( \text{sequence} \), then leaves the \( \text{true code} \) in the input stream. The \( \text{token list variable} \) is assigned locally.

If the \( \text{sequence} \) is empty, leaves the \( \text{false code} \) in the input stream. The value of the \( \text{token list variable} \) is not defined in this case and should not be relied upon. If the \( \text{sequence} \) is non-empty, stores the left-most item from the \( \text{sequence} \) in the \( \text{token list variable} \), i.e. removes the item from the \( \text{sequence} \), then leaves the \( \text{true code} \) in the input stream. Both the \( \text{sequence} \) and the \( \text{token list variable} \) are assigned locally.

If the \( \text{sequence} \) is empty, leaves the \( \text{false code} \) in the input stream. The value of the \( \text{token list variable} \) is not defined in this case and should not be relied upon. If the \( \text{sequence} \) is non-empty, pops the left-most item from the \( \text{sequence} \) in the \( \text{token list variable} \), i.e. removes the item from the \( \text{sequence} \), then leaves the \( \text{true code} \) in the input stream. Both the \( \text{sequence} \) and the \( \text{token list variable} \) are assigned locally.

If the \( \text{sequence} \) is empty, leaves the \( \text{false code} \) in the input stream. The value of the \( \text{token list variable} \) is not defined in this case and should not be relied upon. If the \( \text{sequence} \) is non-empty, pops the right-most item from the \( \text{sequence} \) in the \( \text{token list variable} \), i.e. removes the item from the \( \text{sequence} \), then leaves the \( \text{true code} \) in the input stream. Both the \( \text{sequence} \) and the \( \text{token list variable} \) are assigned locally.
19.5 Modifying sequences

While sequences are normally used as ordered lists, it may be necessary to modify the content. The functions here may be used to update sequences, while retaining the order of the unaffected entries.

\seq_remove_duplicates:N \langle sequence \rangle

Removes duplicate items from the \langle sequence \rangle, leaving the left most copy of each item in the \langle sequence \rangle. The \langle item \rangle comparison takes place on a token basis, as for \tl_if_eq:nnTF.

\TeXhackers note: This function iterates through every item in the \langle sequence \rangle and does a comparison with the \langle items \rangle already checked. It is therefore relatively slow with large sequences.

\seq_remove_all:Nn \langle sequence \rangle \{\langle item \rangle\}

Removes every occurrence of \langle item \rangle from the \langle sequence \rangle. The \langle item \rangle comparison takes place on a token basis, as for \tl_if_eq:nnTF.

\seq_reverse:N \langle sequence \rangle

Reverses the order of the items stored in the \langle sequence \rangle.

\seq_sort:Nn \langle sequence \rangle \{\langle comparison code \rangle\}

Sorts the items in the \langle sequence \rangle according to the \langle comparison code \rangle, and assigns the result to \langle sequence \rangle. The details of sorting comparison are described in Section 6.1.

\seq_shuffle:N \langle seq var \rangle

Sets the \langle seq var \rangle to the result of placing the items of the \langle seq var \rangle in a random order. Each item is (roughly) as likely to end up in any given position.

\TeXhackers note: For sequences with more than 13 items or so, only a small proportion of all possible permutations can be reached, because the random seed \sys_rand_seed: only has 28-bits. The use of \toks internally means that sequences with more than 32767 or 65535 items (depending on the engine) cannot be shuffled.

19.6 Sequence conditionals

\seq_if_empty_p:N \star
\seq_if_empty_p:c \star
\seq_if_empty:N \star
\seq_if_empty:NTF \star
\seq_if_empty:NTF \langle sequence \rangle \{\langle true code \rangle\} \{\langle false code \rangle\}

Tests if the \langle sequence \rangle is empty (containing no items).
Tests if the \langle item \rangle is present in the \langle sequence \rangle.

19.7 Mapping over sequences

All mappings are done at the current group level, \textit{i.e.} any local assignments made by the \langle function \rangle or \langle code \rangle discussed below remain in effect after the loop.

\seq_map_function:NN
\seq_map_function:cn

Applies \langle function \rangle to every \langle item \rangle stored in the \langle sequence \rangle. The \langle function \rangle will receive one argument for each iteration. The \langle items \rangle are returned from left to right. To pass further arguments to the \langle function \rangle, see \seq_map_tokens:Nn. The function \seq_map_inline:Nn is faster than \seq_map_function:NN for sequences with more than about 10 items.

\seq_map_inline:Nn
\seq_map_inline:cn

Displays \langle inline function \rangle to every \langle item \rangle stored within the \langle sequence \rangle. The \langle inline function \rangle should consist of code which will receive the \langle item \rangle as \#1. The \langle items \rangle are returned from left to right.

\seq_map_tokens:Nn
\seq_map_tokens:cn

Analogue of \seq_map_function:NN which maps several tokens instead of a single function. The \langle code \rangle receives each item in the \langle sequence \rangle as a trailing brace group. For instance,

\seq_map_tokens:Nn \l_my_seq { \prg_replicate:nn { 2 } }

expands to twice each item in the \langle sequence \rangle: for each item in \l_my_seq the function \prg_replicate:nn receives 2 and \langle item \rangle as its two arguments. The function \seq_map_inline:Nn is typically faster but it is not expandable.

\seq_map_variable:NNn
\seq_map_variable:({Ncn|cNn|ccn})

Stores each \langle item \rangle of the \langle sequence \rangle in turn in the (token list) \langle variable \rangle and applies the \langle code \rangle. The \langle code \rangle will usually make use of the \langle variable \rangle, but this is not enforced. The assignments to the \langle variable \rangle are local. Its value after the loop is the last \langle item \rangle in the \langle sequence \rangle, or its original value if the \langle sequence \rangle is empty. The \langle items \rangle are returned from left to right.

\seq_map_indexed_function:NN

Applies \langle function \rangle to every entry in the \langle sequence variable \rangle. The \langle function \rangle should have signature :nn. It receives two arguments for each iteration: the \langle index \rangle (namely 1 for the first entry, then 2 and so on) and the \langle item \rangle.
\seq_map_indexed_inline:Nn \seq_map_indexed_inline:Nn (seq var) \{ (inline function) \}

Applies \textit{(inline function)} to every entry in the \textit{(sequence variable)}. The \textit{(inline function)} should consist of code which receives the \textit{(index)} (namely 1 for the first entry, then 2 and so on) as \#1 and the \textit{(item)} as \#2.

\seq_map_break: ☆

Used to terminate a \texttt{\seq_map...} function before all entries in the \textit{(sequence)} have been processed. This normally takes place within a conditional statement, for example

\begin{verbatim}
\seq_map_inline:Nn \l_my_seq
 { \str_if_eq:nnTF { #1 } { bingo } { \seq_map_break: } { % Do something useful } }
\end{verbatim}

Use outside of a \texttt{\seq_map...} scenario leads to low level \TeX errors.

\textbf{\TeX hackers note:} When the mapping is broken, additional tokens may be inserted before further items are taken from the input stream. This depends on the design of the mapping function.

\seq_map_break:n ✪

\seq_map_break:n \{ (code) \}

Used to terminate a \texttt{\seq_map...} function before all entries in the \textit{(sequence)} have been processed, inserting the \textit{(code)} after the mapping has ended. This normally takes place within a conditional statement, for example

\begin{verbatim}
\seq_map_inline:Nn \l_my_seq
 { \str_if_eq:nnTF { #1 } { bingo } { \seq_map_break:n { <code> } } { % Do something useful } }
\end{verbatim}

Use outside of a \texttt{\seq_map...} scenario leads to low level \TeX errors.

\textbf{\TeX hackers note:} When the mapping is broken, additional tokens may be inserted before the \textit{(code)} is inserted into the input stream. This depends on the design of the mapping function.
\seq_set_map:NNn \seq_gset_map:NNn
\seq_set_map_x:NNn \seq_gset_map_x:NNn

Applies \textit{inline function} to every \textit{item} stored within the \textit{sequence2}. The \textit{inline function} should consist of code which will receive the \textit{item} as \#1. The sequence resulting from \textit{x}-expanding \textit{inline function} applied to each \textit{item} is assigned to \textit{sequence1}.

\textbf{\TeXhackers note:} Contrarily to other mapping functions, \texttt{\seq_map_break:} cannot be used in this function, and would lead to low-level \TeX errors.

\seq_count:N \seq_count:c

Leaves the number of items in the \textit{sequence} in the input stream as an \textit{integer denotation}. The total number of items in a \textit{sequence} includes those which are empty and duplicates, \textit{i.e.} every item in a \textit{sequence} is unique.

\seq_use:Nnnn \seq_use:cnnn

Places the contents of the \textit{seq var} in the input stream, with the appropriate \textit{separator} between the items. Namely, if the sequence has more than two items, the \textit{separator between more than two} is placed between each pair of items except the last, for which the \textit{separator between final two} is used. If the sequence has exactly two items, then they are placed in the input stream separated by the \textit{separator between two}. If the sequence has a single item, it is placed in the input stream, and an empty sequence produces no output. An error is raised if the variable does not exist or if it is invalid.

For example,

\seq_set_split:Nnn \l_tmpa_seq { | } { a | b | c | \{de\} | f }\seq_use:Nnnn \l_tmpa_seq { -and- } { , } { , -and- }

inserts “a, b, c, de, and f” in the input stream. The first separator argument is not used in this case because the sequence has more than 2 items.

\textbf{\TeXhackers note:} The result is returned within the \texttt{\unexpanded primitive (\exp_not:n)}, which means that the \textit{items} do not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.

19.8 Using the content of sequences directly
\seq_use:Nn * \seq_use:cn * NeX: 2013-05-26

\seq_use:Nn \seq var \{ \textit{separator} \} Places the contents of the \textit{seq var} in the input stream, with the \textit{separator} between the items. If the sequence has a single item, it is placed in the input stream with no \textit{separator}, and an empty sequence produces no output. An error is raised if the variable does not exist or if it is invalid.

For example,

\seq_set_split:Nnn \l_tmpa_seq { | } { a | b | c | \{de\} | f }
\seq_use:Nn \l_tmpa_seq { \textit{and} }

inserts “a and b and c and de and f” in the input stream.

\TeXhackers note: The result is returned within the \texttt{\textbackslash unexpanded} primitive (\texttt{\exp_not:n}), which means that the \textit{items} do not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.

19.9 Sequences as stacks

Sequences can be used as stacks, where data is pushed to and popped from the top of the sequence. (The left of a sequence is the top, for performance reasons.) The stack functions for sequences are not intended to be mixed with the general ordered data functions detailed in the previous section: a sequence should either be used as an ordered data type or as a stack, but not in both ways.

\seq_get:NN \seq_get:cn

\seq_get:NN \texttt{\textbackslash sequence} \texttt{\texttt{\textbackslash token list variable}}

Reads the top item from a \textit{sequence} into the \textit{token list variable} without removing it from the \textit{sequence}. The \textit{token list variable} is assigned locally. If \textit{sequence} is empty the \textit{token list variable} is set to the special marker \texttt{\textbackslash q\_no\_value}.

\seq_pop:NN \seq_pop:cn

\seq_pop:NN \texttt{\textbackslash sequence} \texttt{\texttt{\textbackslash token list variable}}

Pops the top item from a \textit{sequence} into the \textit{token list variable}. Both of the variables are assigned locally. If \textit{sequence} is empty the \textit{token list variable} is set to the special marker \texttt{\textbackslash q\_no\_value}.

\seq_gpop:NN \seq_gpop:cn

\seq_gpop:NN \texttt{\textbackslash sequence} \texttt{\texttt{\textbackslash token list variable}}

Pops the top item from a \textit{sequence} into the \textit{token list variable}. The \textit{sequence} is modified globally, while the \textit{token list variable} is assigned locally. If \textit{sequence} is empty the \textit{token list variable} is set to the special marker \texttt{\textbackslash q\_no\_value}.

\seq_get:NNTF \seq_get:cnT

\seq_get:NNTF \texttt{\textbackslash sequence} \texttt{\texttt{\textbackslash token list variable}} \{ \textit{true code} \} \{ \textit{false code} \}

If the \textit{sequence} is empty, leaves the \textit{false code} in the input stream. The value of the \textit{token list variable} is not defined in this case and should not be relied upon. If the \textit{sequence} is non-empty, stores the top item from a \textit{sequence} in the \textit{token list variable} without removing it from the \textit{sequence}. The \textit{token list variable} is assigned locally.

151
\seq_pop:NNTF  ⟨sequence⟩ ⟨token list variable⟩ {⟨true code⟩} {⟨false code⟩}

If the ⟨sequence⟩ is empty, leaves the ⟨false code⟩ in the input stream. The value of the ⟨token list variable⟩ is not defined in this case and should not be relied upon. If the ⟨sequence⟩ is non-empty, pops the top item from the ⟨sequence⟩ in the ⟨token list variable⟩, i.e. removes the item from the ⟨sequence⟩. Both the ⟨sequence⟩ and the ⟨token list variable⟩ are assigned locally.

\seq_gpop:NNTF ⟨sequence⟩ ⟨token list variable⟩ {⟨true code⟩} {⟨false code⟩}

If the ⟨sequence⟩ is empty, leaves the ⟨false code⟩ in the input stream. The value of the ⟨token list variable⟩ is not defined in this case and should not be relied upon. If the ⟨sequence⟩ is non-empty, pops the top item from the ⟨sequence⟩ in the ⟨token list variable⟩, i.e. removes the item from the ⟨sequence⟩. The ⟨sequence⟩ is modified globally, while the ⟨token list variable⟩ is assigned locally.

\seq_push:Nn ⟨sequence⟩ {⟨item⟩}
\seq_push:(NV|Nv|No|Nx|cn|cV|cv|co|cx)
\seq_gpush:Nn ⟨sequence⟩ {⟨item⟩}
\seq_gpush:(NV|Nv|No|Nx|cn|cV|cv|co|cx)

Adds the {⟨item⟩} to the top of the ⟨sequence⟩.

19.10 Sequences as sets

Sequences can also be used as sets, such that all of their items are distinct. Usage of sequences as sets is not currently widespread, hence no specific set function is provided. Instead, it is explained here how common set operations can be performed by combining several functions described in earlier sections. When using sequences to implement sets, one should be careful not to rely on the order of items in the sequence representing the set.

Sets should not contain several occurrences of a given item. To make sure that a ⟨sequence variable⟩ only has distinct items, use \seq_remove_duplicates:N ⟨sequence variable⟩. This function is relatively slow, and to avoid performance issues one should only use it when necessary.

Some operations on a set ⟨seq var⟩ are straightforward. For instance, \seq_count:N ⟨seq var⟩ expands to the number of items, while \seq_if_in:NnTF ⟨seq var⟩ {⟨item⟩} tests if the ⟨item⟩ is in the set.

Adding an ⟨item⟩ to a set ⟨seq var⟩ can be done by appending it to the ⟨seq var⟩ if it is not already in the ⟨seq var⟩:

\seq_if_in:NnF ⟨seq var⟩ {⟨item⟩}
{ \seq_put_right:Nn ⟨seq var⟩ {⟨item⟩} }

Removing an ⟨item⟩ from a set ⟨seq var⟩ can be done using \seq_remove_all:Nn,

\seq_remove_all:Nn ⟨seq var⟩ {⟨item⟩}

The intersection of two sets ⟨seq var1⟩ and ⟨seq var2⟩ can be stored into ⟨seq var3⟩ by collecting items of ⟨seq var1⟩ which are in ⟨seq var2⟩.
\seq_clear:N \seq_var_3
\seq_map_inline:Nn \seq_var_1{
  \seq_if_in:NnT \seq_var_2 \#1
  \seq_put_right:Nn \seq_var_3 \#1
}

The code as written here only works if \seq_var_3 is different from the other two sequence variables. To cover all cases, items should first be collected in a sequence \l__⟨pkg⟩_internal_seq, then \seq_var_3 should be set equal to this internal sequence.

The same remark applies to other set functions.

The union of two sets \seq_var_1 and \seq_var_2 can be stored into \seq_var_3 through
\seq_concat:NNN \seq_var_3 \seq_var_1 \seq_var_2
or by adding items to (a copy of) \seq_var_1 one by one
\seq_set_eq:NN \seq_var_3 \seq_var_1
\seq_map_inline:Nn \seq_var_2{
  \seq_if_in:NnF \seq_var_3 \#1
  \seq_put_right:Nn \seq_var_3 \#1
}

The second approach is faster than the first when the \seq_var_2 is short compared to \seq_var_1.

The difference of two sets \seq_var_1 and \seq_var_2 can be stored into \seq_var_3 by removing items of the \seq_var_2 from (a copy of) the \seq_var_1 one by one.
\seq_set_eq:NN \seq_var_3 \seq_var_1
\seq_map_inline:Nn \seq_var_2{
  \seq_remove_all:Nn \seq_var_3 {#1}
}

The symmetric difference of two sets \seq_var_1 and \seq_var_2 can be stored into \seq_var_3 by computing the difference between \seq_var_1 and \seq_var_2 and storing the result as \l__⟨pkg⟩_internal_seq, then the difference between \seq_var_2 and \seq_var_1, and finally concatenating the two differences to get the symmetric differences.
\seq_set_eq:NN \l__⟨pkg⟩_internal_seq \seq_var_1
\seq_map_inline:Nn \seq_var_2{
  \seq_remove_all:Nn \l__⟨pkg⟩_internal_seq {#1}
}
\seq_set_eq:NN \seq_var_3 \seq_var_1
\seq_map_inline:Nn \seq_var_2{
  \seq_remove_all:Nn \seq_var_3 {#1}
}
\seq_concat:NNN \seq_var_3 \seq_var_3 \l__⟨pkg⟩_internal_seq

19.11 Constant and scratch sequences

\c_empty_seq Constant that is always empty.

\seq: 2012-07-02
Scratch sequences for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

Scratch sequences for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

## 19.12 Viewing sequences

\seq_show:N \seq_show:c
Displays the entries in the \texttt{<sequence>} in the terminal.

\seq_log:N \seq_log:c
 Writes the entries in the \texttt{<sequence>} in the log file.
Chapter 20

The \texttt{l3int} package

Integers

Calculation and comparison of integer values can be carried out using literal numbers, \texttt{int} registers, constants and integers stored in token list variables. The standard operators +, −, \textbackslash{}/ and \texttt{*} and parentheses can be used within such expressions to carry arithmetic operations. This module carries out these functions on integer expressions ("\texttt{intexpr}").

20.1 Integer expressions

Throughout this module, (almost) all \texttt{n}-type argument allow for an \texttt{(intexpr)} argument with the following syntax. The \texttt{(integer expression)} should consist, after expansion, of +, −, *, \textbackslash{}/, (,) and of course integer operands. The result is calculated by applying standard mathematical rules with the following peculiarities:

- \texttt{/} denotes division rounded to the closest integer with ties rounded away from zero;
- there is an error and the overall expression evaluates to zero whenever the absolute value of any intermediate result exceeds $2^{31} - 1$, except in the case of scaling operations $a \texttt{*} b / c$, for which $a \texttt{*} b$ may be arbitrarily large (but the operands $a$, $b$, $c$ are still constrained to an absolute value at most $2^{31} - 1$);
- parentheses may not appear after unary + or −, namely placing +( or −( at the start of an expression or after +, −, *, \textbackslash{}/ or ( leads to an error.

Each integer operand can be either an integer variable (with no need for \texttt{\int\_use:N}) or an integer denotation. For example both

\texttt{\int\_show:n \{ 5 + 4 * 3 - ( 3 + 4 * 5 ) \}}

and

\texttt{\tl\_new:N \l\_my\_tl}
\texttt{\tl\_set:Nn \l\_my\_tl \{ 5 \}}
\texttt{\int\_new:N \l\_my\_int}
\texttt{\int\_set:Nn \l\_my\_int \{ 4 \}}
\texttt{\int\_show:n \{ \l\_my\_tl + \l\_my\_int * 3 - ( 3 + 4 * 5 ) \}}
show the same result \(-6\) because \(\_\_my\_tl\) expands to the integer denotation \(5\) while the integer variable \(\_\_my\_int\) takes the value \(4\). As the \(\langle\text{integer expression}\rangle\) is fully expanded from left to right during evaluation, fully expandable and restricted-expandable functions can both be used, and \(\expnot:n\) and its variants have no effect while \(\expnot:N\) may incorrectly interrupt the expression.

\textbf{\LaTeX}hacks note: Exactly two expansions are needed to evaluate \texttt{\inteval:n}. The result is \textit{not} an \(\langle\text{internal integer}\rangle\), and therefore should be terminated by a space if used in \texttt{\inteval:w} or in a \TeX{}-style integer assignment.

As all \TeX{} integers, integer operands can also be: \texttt{\value{⟨LATEX2ε counter⟩}}; dimension or skip variables, converted to integers in \sp{}; the character code of some character given as \`⟨char⟩\' or \`\langle\texttt{char}\rangle\'; octal numbers given as \`\texttt{0} followed by digits from 0 to 7; or hexadecimal numbers given as \`\texttt{A} to \texttt{F}.}

\begin{itemize}
\item \texttt{\inteval:n} \{\langle\text{integer expression}\rangle\}
\hspace{1em} Evaluates the \(\langle\text{integer expression}\rangle\) and leaves the result in the input stream as an integer denotation: for positive results an explicit sequence of decimal digits not starting with \(0\), for negative results \(-\) followed by such a sequence, and \(0\) for zero.
\item \texttt{\inteval:w} \{\langle\text{integer expression}\rangle\}
\hspace{1em} Evaluates the \(\langle\text{integer expression}\rangle\) as described for \texttt{\inteval:n}. The end of the expression is the first token encountered that cannot form part of such an expression. If that token is \texttt{\scanskip}: it is removed, otherwise not. Spaces do \textit{not} terminate the expression. However, spaces terminate explicit integers, and this may terminate the expression: for instance, \texttt{\inteval:w \texttt{1 ∆ 1 ∆ 9}} (with explicit space tokens inserted using \texttt{-} in a code setting) expands to \texttt{29} since the digit \texttt{9} is not part of the expression.
\item \texttt{\intsign:n} \{\langle\text{intexpr}\rangle\}
\hspace{1em} Evaluates the \(\langle\text{integer expression}\rangle\) then leaves \(1\) or \(0\) or \(-1\) in the input stream according to the sign of the result.
\item \texttt{\intabs:n} \{\langle\text{integer expression}\rangle\}
\hspace{1em} Evaluates the \(\langle\text{integer expression}\rangle\) as described for \texttt{\inteval:n} and leaves the absolute value of the result in the input stream as an \(\langle\text{integer denotation}\rangle\) after two expansions.
\item \texttt{\intdivround:nn} \{\langle\text{intexpr}\rangle\} \{\langle\text{intexpr}\rangle\}
\hspace{1em} Evaluates the two \(\langle\text{integer expressions}\rangle\) as described earlier, then divides the first value by the second, and rounds the result to the closest integer. Ties are rounded away from zero. Note that this is identical to using \texttt{/} directly in an \(\langle\text{integer expression}\rangle\). The result is left in the input stream as an \(\langle\text{integer denotation}\rangle\) after two expansions.
\item \texttt{\intdivtruncate:nn} \{\langle\text{intexpr}\rangle\} \{\langle\text{intexpr}\rangle\}
\hspace{1em} Evaluates the two \(\langle\text{integer expressions}\rangle\) as described earlier, then divides the first value by the second, and rounds the result towards zero. Note that division using \texttt{/} rounds to the closest integer instead. The result is left in the input stream as an \(\langle\text{integer denotation}\rangle\) after two expansions.
\end{itemize}
\int_max:nn \int_min:nn
\int_mod:nn
\int_new:N \int_new:c
\int_const:Nn \int_const:cn
\int_zero:N \int_zero:c \int_gzero:N \int_gzero:c
\int_zero_new:N \int_zero_new:c \int_gzero_new:N \int_gzero_new:c
\int_set_eq:NN \int_set_eq:(cN|Nc|cc) \int_gset_eq:NN
\int_if_exist_p:N \int_if_exist_p:c \int_if_exist:NTF \int_if_exist:cc

\int_max:nn \{\langle \text{intexpr}_1 \rangle \} \{\langle \text{intexpr}_2 \rangle \}
\int_min:nn \{\langle \text{intexpr}_1 \rangle \} \{\langle \text{intexpr}_2 \rangle \}
\int_mod:nn \{\langle \text{intexpr}_1 \rangle \} \{\langle \text{intexpr}_2 \rangle \}
\int_new:N \langle \text{integer} \rangle
\int_new:c \langle \text{integer} \rangle
\int_const:Nn \langle \text{integer} \rangle \{\langle \text{integer expression} \rangle \}
\int_zero:N \langle \text{integer} \rangle \int_zero:c \int_gzero:N \int_gzero:c
\int_set_eq:NN \langle \text{integer} \rangle_1 \langle \text{integer} \rangle_2
\int_if_exist_p:N \langle \text{int} \rangle \int_if_exist:NTF \langle \text{int} \rangle

Evaluates the \langle \text{integer expressions} \rangle as described for \int_eval:n and leaves either the larger or smaller value in the input stream as an \langle \text{integer denotation} \rangle after two expansions.

Evaluates the two \langle \text{integer expressions} \rangle as described earlier, then calculates the integer remainder of dividing the first expression by the second. This is obtained by subtracting \int_div_truncate:nn \{\langle \text{intexpr}_1 \rangle \} \{\langle \text{intexpr}_2 \rangle \} times \langle \text{intexpr}_2 \rangle from \langle \text{intexpr}_1 \rangle. Thus, the result has the same sign as \langle \text{intexpr}_1 \rangle and its absolute value is strictly less than that of \langle \text{intexpr}_2 \rangle. The result is left in the input stream as an \langle \text{integer denotation} \rangle after two expansions.

\section{20.2 Creating and initialising integers}

\int_new:N \langle \text{integer} \rangle
\int_new:c \langle \text{integer} \rangle
\int_const:Nn \langle \text{integer} \rangle \{\langle \text{integer expression} \rangle \}
\int_zero:N \langle \text{integer} \rangle \int_zero:c \int_gzero:N \int_gzero:c
\int_set_eq:NN \langle \text{integer} \rangle_1 \langle \text{integer} \rangle_2
\int_if_exist_p:N \langle \text{int} \rangle \int_if_exist:NTF \langle \text{int} \rangle

Creates a new \langle \text{integer} \rangle or raises an error if the name is already taken. The declaration is global. The \langle \text{integer} \rangle is initially equal to 0.

Creates a new constant \langle \text{integer} \rangle or raises an error if the name is already taken. The value of the \langle \text{integer} \rangle is set globally to the \langle \text{integer expression} \rangle.

Sets \langle \text{integer} \rangle to 0.

Ensures that the \langle \text{integer} \rangle exists globally by applying \int_new:N if necessary, then applies \int_(g)zero:N to leave the \langle \text{integer} \rangle set to zero.

Sets the content of \langle \text{integer}_1 \rangle equal to that of \langle \text{integer}_2 \rangle.

Tests whether the \langle \text{int} \rangle is currently defined. This does not check that the \langle \text{int} \rangle really is an integer variable.

Updated: 2012-09-26
Updated: 2012-09-26
Updated: 2012-09-26
Updated: 2011-10-22
New: 2011-12-13
New: 2012-03-03
20.3 Setting and incrementing integers

\texttt{\textbackslash int\_add:Nn} \texttt{\textbackslash int\_add:cn} \texttt{\textbackslash int\_gadd:Nn} \texttt{\textbackslash int\_gadd:cn}

\texttt{\textbackslash int\_add:Nn \{integer\} \{\texttt{\textbackslash int\_expression}\}}

Adds the result of the \texttt{\textbackslash int\_expression} to the current content of the \texttt{\textbackslash integer}.

\texttt{\textbackslash int\_decr:N} \texttt{\textbackslash int\_decr:c} \texttt{\textbackslash int\_gdecr:N} \texttt{\textbackslash int\_gdecr:c}

\texttt{\textbackslash int\_decr:N \{integer\}}

Decreases the value stored in \texttt{\textbackslash integer} by 1.

\texttt{\textbackslash int\_incr:N} \texttt{\textbackslash int\_incr:c} \texttt{\textbackslash int\_gincr:N} \texttt{\textbackslash int\_gincr:c}

\texttt{\textbackslash int\_incr:N \{integer\}}

Increases the value stored in \texttt{\textbackslash integer} by 1.

\texttt{\textbackslash int\_set:Nn} \texttt{\textbackslash int\_set:cn} \texttt{\textbackslash int\_gset:Nn} \texttt{\textbackslash int\_gset:cn}

\texttt{\textbackslash int\_set:Nn \{integer\} \{\texttt{\textbackslash int\_expression}\}}

Sets \texttt{\textbackslash integer} to the value of \texttt{\textbackslash int\_expression}, which must evaluate to an integer (as described for \texttt{\textbackslash int\_eval:n}).

\texttt{\textbackslash int\_sub:Nn} \texttt{\textbackslash int\_sub:cn} \texttt{\textbackslash int\_gsub:Nn} \texttt{\textbackslash int\_gsub:cn}

\texttt{\textbackslash int\_sub:Nn \{integer\} \{\texttt{\textbackslash int\_expression}\}}

Subtracts the result of the \texttt{\textbackslash int\_expression} from the current content of the \texttt{\textbackslash integer}.

20.4 Using integers

\texttt{\textbackslash int\_use:N} \texttt{\textbackslash int\_use:c}

\texttt{\textbackslash int\_use:N \{integer\}}

Recover the content of an \texttt{\textbackslash integer} and places it directly in the input stream. An error is raised if the variable does not exist or if it is invalid. Can be omitted in places where an \texttt{\textbackslash integer} is required (such as in the first and third arguments of \texttt{\textbackslash int\_compare:nNnTF}).

\texttt{\small \TeX\textbackslash hack\textbackslash e\textbackslash rs\textbackslash note: \textbackslash int\_use:N} is the \TeX primitive \texttt{\the: this is one of several \LaTeX3} names for this primitive.

158
20.5 Integer expression conditionals

This function first evaluates each of the \( \text{integer expressions} \) as described for \texttt{int_eval:n}. The two results are then compared using the \( \text{relation} \):

\[
\begin{align*}
\text{Equal} &= \,= \\
\text{Greater than} &> \\
\text{Less than} &<
\end{align*}
\]

This function is less flexible than \texttt{int_compare:nTF} but around 5 times faster.

This function evaluates the \( \text{integer expressions} \) as described for \texttt{int_eval:n} and compares consecutive result using the corresponding \( \text{relation} \), namely it compares \( \text{expr}_1 \) and \( \text{expr}_2 \) using the \( \text{relation}_1 \), then \( \text{expr}_2 \) and \( \text{expr}_3 \) using the \( \text{relation}_2 \), until finally comparing \( \text{expr}_N \) and \( \text{expr}_{N+1} \) using the \( \text{relation}_N \). The test yields true if all comparisons are true. Each \( \text{integer expression} \) is evaluated only once, and the evaluation is lazy, in the sense that if one comparison is false, then no other \( \text{integer expression} \) is evaluated and no other comparison is performed. The \( \text{relations} \) can be any of the following:

\[
\begin{align*}
\text{Equal} &= \,= \\
\text{Greater than or equal to} &> \\
\text{Greater than} &> \\
\text{Less than or equal to} &< \\
\text{Less than} &< \\
\text{Not equal} &!=
\end{align*}
\]

This function is more flexible than \texttt{int_compare:nNnTF} but around 5 times slower.
\int_case:nn \star \ \int_case:nnTF \{\text{test integer expression}\}
\{ 
\{\text{intexpr case}_1\} \{\text{code case}_1\}
\{\text{intexpr case}_2\} \{\text{code case}_2\}
\ldots
\{\text{intexpr case}_n\} \{\text{code case}_n\}
\}
\{\text{true code}\}
\{\text{false code}\}

This function evaluates the \textit{(test integer expression)} and compares this in turn to each of the \textit{(integer expression cases)}. If the two are equal then the associated \textit{(code)} is left in the input stream and other cases are discarded. If any of the cases are matched, the \textit{(true code)} is also inserted into the input stream (after the code for the appropriate case), while if none match then the \textit{(false code)} is inserted. The function \texttt{\int_case:nn}, which does nothing if there is no match, is also available. For example

\int_case:nnF
\{ 2 \times 5 \}
\{ 
\{ 5 \} \{ \text{Small} \}
\{ 4 \times 6 \} \{ \text{Medium} \}
\{ -2 \times 10 \} \{ \text{Negative} \}
\}
\{ \text{No idea!} \}

leaves “Medium” in the input stream.

\texttt{\int_if_odd_p:n \star \int_if_odd:nTF \star \int_if_even_p:n \star \int_if_even:nTF \star}

This function first evaluates the \textit{(integer expression)} as described for \texttt{\int_eval:n}. It then evaluates if this is odd or even, as appropriate.

\textbf{20.6 Integer expression loops}

\int_do_until:nNnn \star \int_do_until:nNnn \{\text{intexpr}_1\} \{\text{relation}\} \{\text{intexpr}_2\} \{\text{code}\}

Places the \textit{(code)} in the input stream for \TeX{} to process, and then evaluates the relationship between the two \textit{(integer expressions)} as described for \texttt{\int_compare:nNnTF}. If the test is \texttt{false} then the \textit{(code)} is inserted into the input stream again and a loop occurs until the \textit{(relation)} is \texttt{true}.

\int_do_while:nNnn \star \int_do_while:nNnn \{\text{intexpr}_1\} \{\text{relation}\} \{\text{intexpr}_2\} \{\text{code}\}

Places the \textit{(code)} in the input stream for \TeX{} to process, and then evaluates the relationship between the two \textit{(integer expressions)} as described for \texttt{\int_compare:nNnTF}. If the test is \texttt{true} then the \textit{(code)} is inserted into the input stream again and a loop occurs until the \textit{(relation)} is \texttt{false}.

\texttt{\textbackslash int\_until\_do:nnn} \textasciitilde \texttt{\textbackslash int\_until\_do:nnn \{\texttt{integer\ expression}_1\} \{\texttt{relation}\} \{\texttt{integer\ expression}_2\} \{\texttt{code}\}}

Evaluates the relationship between the two \textit{integer expressions} as described for \texttt{\textbackslash int\_compare:nNnTF}, and then places the \texttt{code} in the input stream if the \texttt{relation} is \texttt{false}. After the \texttt{code} has been processed by \TeX the test is repeated, and a loop occurs until the test is \texttt{true}.

\texttt{\textbackslash int\_while\_do:nnn} \textasciitilde \texttt{\textbackslash int\_while\_do:nnn \{\texttt{integer\ expression}_1\} \{\texttt{relation}\} \{\texttt{integer\ expression}_2\} \{\texttt{code}\}}

Evaluates the relationship between the two \textit{integer expressions} as described for \texttt{\textbackslash int\_compare:nNnTF}, and then places the \texttt{code} in the input stream if the \texttt{relation} is \texttt{true}. After the \texttt{code} has been processed by \TeX the test is repeated, and a loop occurs until the test is \texttt{false}.

\texttt{\textbackslash int\_do\_until:nn} \textasciitilde \texttt{\textbackslash int\_do\_until:nn \{\texttt{integer\ relation}\} \{\texttt{code}\}}

Places the \texttt{code} in the input stream for \TeX to process, and then evaluates the \texttt{integer\ relation} as described for \texttt{\textbackslash int\_compare:nNnTF}. If the test is \texttt{false} then the \texttt{code} is inserted into the input stream again and a loop occurs until the \texttt{relation} is \texttt{true}.

\texttt{\textbackslash int\_do\_while:nn} \textasciitilde \texttt{\textbackslash int\_do\_while:nn \{\texttt{integer\ relation}\} \{\texttt{code}\}}

Places the \texttt{code} in the input stream for \TeX to process, and then evaluates the \texttt{integer\ relation} as described for \texttt{\textbackslash int\_compare:nNnTF}. If the test is \texttt{true} then the \texttt{code} is inserted into the input stream again and a loop occurs until the \texttt{relation} is \texttt{false}.

\texttt{\textbackslash int\_until\_do:nn} \textasciitilde \texttt{\textbackslash int\_until\_do:nn \{\texttt{integer\ relation}\} \{\texttt{code}\}}

Evaluates the \texttt{integer\ relation} as described for \texttt{\textbackslash int\_compare:nNnTF}, and then places the \texttt{code} in the input stream if the \texttt{relation} is \texttt{false}. After the \texttt{code} has been processed by \TeX the test is repeated, and a loop occurs until the test is \texttt{true}.

\texttt{\textbackslash int\_while\_do:nn} \textasciitilde \texttt{\textbackslash int\_while\_do:nn \{\texttt{integer\ relation}\} \{\texttt{code}\}}

Evaluates the \texttt{integer\ relation} as described for \texttt{\textbackslash int\_compare:nNnTF}, and then places the \texttt{code} in the input stream if the \texttt{relation} is \texttt{true}. After the \texttt{code} has been processed by \TeX the test is repeated, and a loop occurs until the test is \texttt{false}. 

Updated: 2013-01-13
### 20.7 Integer step functions

\[ \text{int_step_function:nN} \quad \text{(function)} \]
\[ \text{int_step_function:nnN} \quad \text{(function)} \]
\[ \text{int_step_function:nnNn} \quad \text{(function)} \]

This function first evaluates the \( \langle \text{initial value} \rangle \), \( \langle \text{step} \rangle \) and \( \langle \text{final value} \rangle \), all of which should be integer expressions. The \( \langle \text{function} \rangle \) is then placed in front of each \( \langle \text{value} \rangle \) from the \( \langle \text{initial value} \rangle \) to the \( \langle \text{final value} \rangle \) in turn (using \( \langle \text{step} \rangle \) between each \( \langle \text{value} \rangle \)). The \( \langle \text{step} \rangle \) must be non-zero. If the \( \langle \text{step} \rangle \) is positive, the loop stops when the \( \langle \text{value} \rangle \) becomes larger than the \( \langle \text{final value} \rangle \). If the \( \langle \text{step} \rangle \) is negative, the loop stops when the \( \langle \text{value} \rangle \) becomes smaller than the \( \langle \text{final value} \rangle \). The \( \langle \text{function} \rangle \) should absorb one numerical argument. For example:

\begin{verbatim}
\cs_set:Npn \my_func:n #1 { \[I~saw~#1\] \quad }
\int_step_function:nNn { 1 } { 1 } { 5 } \my_func:n
\end{verbatim}

would print:

\begin{verbatim}
[I saw 1] [I saw 2] [I saw 3] [I saw 4] [I saw 5]
\end{verbatim}

The functions \( \text{int_step_function:nN} \) and \( \text{int_step_function:nnN} \) both use a fixed \( \langle \text{step} \rangle \) of 1, and in the case of \( \text{int_step_function:nN} \) the \( \langle \text{initial value} \rangle \) is also fixed as 1. These functions are provided as simple short-cuts for code clarity.

\[ \text{int_step_inline:nn} \quad \text{(code)} \]
\[ \text{int_step_inline:nn} \quad \text{(code)} \]
\[ \text{int_step_inline:nn} \quad \text{(code)} \]

This function first evaluates the \( \langle \text{initial value} \rangle \), \( \langle \text{step} \rangle \) and \( \langle \text{final value} \rangle \), all of which should be integer expressions. Then for each \( \langle \text{value} \rangle \) from the \( \langle \text{initial value} \rangle \) to the \( \langle \text{final value} \rangle \) in turn (using \( \langle \text{step} \rangle \) between each \( \langle \text{value} \rangle \)), the \( \langle \text{code} \rangle \) is inserted into the input stream with \#1 replaced by the current \( \langle \text{value} \rangle \). Thus the \( \langle \text{code} \rangle \) should define a function of one argument (\#1).

The functions \( \text{int_step_inline:nn} \) and \( \text{int_step_inline:nn} \) both use a fixed \( \langle \text{step} \rangle \) of 1, and in the case of \( \text{int_step_inline:nn} \) the \( \langle \text{initial value} \rangle \) is also fixed as 1. These functions are provided as simple short-cuts for code clarity.

\[ \text{int_step_variable:nN} \quad \text{(tl var) \{code\}} \]
\[ \text{int_step_variable:nN} \quad \text{(tl var) \{code\}} \]
\[ \text{int_step_variable:nNn} \quad \text{(tl var) \{code\}} \]

This function first evaluates the \( \langle \text{initial value} \rangle \), \( \langle \text{step} \rangle \) and \( \langle \text{final value} \rangle \), all of which should be integer expressions. Then for each \( \langle \text{value} \rangle \) from the \( \langle \text{initial value} \rangle \) to the \( \langle \text{final value} \rangle \) in turn (using \( \langle \text{step} \rangle \) between each \( \langle \text{value} \rangle \)), the \( \langle \text{code} \rangle \) is inserted into the input stream, with the \( \langle \text{tl var} \rangle \) defined as the current \( \langle \text{value} \rangle \). Thus the \( \langle \text{code} \rangle \) should make use of the \( \langle \text{tl var} \rangle \).

The functions \( \text{int_step_variable:nN} \) and \( \text{int_step_variable:nNn} \) both use a fixed \( \langle \text{step} \rangle \) of 1, and in the case of \( \text{int_step_variable:nN} \) the \( \langle \text{initial value} \rangle \) is also fixed as 1. These functions are provided as simple short-cuts for code clarity.
20.8 Formatting integers

Integers can be placed into the output stream with formatting. These conversions apply to any integer expressions.

\int_to_arabic:n \{\langle integer expression\rangle\}
Places the value of the \langle integer expression\rangle in the input stream as digits, with category code 12 (other).

\int_to_alph:n \{\langle integer expression\rangle\}
Evaluates the \langle integer expression\rangle and converts the result into a series of letters, which are then left in the input stream. The conversion rule uses the 26 letters of the English alphabet, in order, adding letters when necessary to increase the total possible range of representable numbers. Thus

\int_to_alph:n \{ 1 \}
places a in the input stream,

\int_to_alph:n \{ 26 \}
is represented as z and

\int_to_alph:n \{ 27 \}
is converted to aa. For conversions using other alphabets, use \int_to_symbols:nnn to define an alphabet-specific function. The basic \int_to_alph:n and \int_to_Alph:n functions should not be modified. The resulting tokens are digits with category code 12 (other) and letters with category code 11 (letter).

\int_to_symbols:nnn \{\langle integer expression\rangle\} \{(total symbols)\}
\{\langle value to symbol mapping\rangle\}
This is the low-level function for conversion of an \langle integer expression\rangle into a symbolic form (often letters). The \langle total symbols\rangle available should be given as an integer expression. Values are actually converted to symbols according to the \langle value to symbol mapping\rangle. This should be given as \langle total symbols\rangle pairs of entries, a number and the appropriate symbol. Thus the \int_to_alph:n function is defined as

\cs_new:Npn \int_to_alph:n #1
\int_to_symbols:nnn \{#1\} \{ 26 \}
\{ 1 \} { a }\{ 2 \} { b }\ldots\{ 26 \} { z }
\texttt{\int_to_bin:n} \star \texttt{\int_to_bin:n} \{\textit{integer expression}\}

Calculates the value of the \textit{integer expression} and places the binary representation of the result in the input stream.

\texttt{\int_to_hex:n} \star \texttt{\int_to_Hex:n} \{\textit{integer expression}\}

Calculates the value of the \textit{integer expression} and places the hexadecimal (base 16) representation of the result in the input stream. Letters are used for digits beyond 9: lower case letters for \texttt{\int_to_hex:n} and upper case ones for \texttt{\int_to_Hex:n}. The resulting tokens are digits with category code 12 (other) and letters with category code 11 (letter).

\texttt{\int_to_oct:n} \star \texttt{\int_to_Base:n} \{\textit{integer expression}\}

Calculates the value of the \textit{integer expression} and places the octal (base 8) representation of the result in the input stream. The resulting tokens are digits with category code 12 (other) and letters with category code 11 (letter).

\texttt{\int_to_base:nn} \star \texttt{\int_to_Base:nn} \{\textit{integer expression}\} \{\textit{base}\}

Calculates the value of the \textit{integer expression} and converts it into the appropriate representation in the \textit{base}; the later may be given as an integer expression. For bases greater than 10 the higher “digits” are represented by letters from the English alphabet: lower case letters for \texttt{\int_to_base:n} and upper case ones for \texttt{\int_to_Base:n}. The maximum \textit{base} value is 36. The resulting tokens are digits with category code 12 (other) and letters with category code 11 (letter).

\texttt{\int_to_roman:n} \dagger \texttt{\int_to_Roman:n} \dagger \{\textit{integer expression}\}

Places the value of the \textit{integer expression} in the input stream as Roman numerals, either lower case \texttt{\int_to_roman:n} or upper case \texttt{\int_to_Roman:n}. If the value is negative or zero, the output is empty. The Roman numerals are letters with category code 11 (letter). The letters used are \texttt{mdclxvi}, repeated as needed: the notation with bars (such as \texttt{\textbar{v}} for 5000) is not used. For instance \texttt{\int_to_roman:n} \{8249\} expands to \texttt{mmmmmmmccxlix}.

\texttt{\int_from_alph:n} \star \{\textit{letters}\}

Converts the \textit{letters} into the integer (base 10) representation and leaves this in the input stream. The \textit{letters} are first converted to a string, with no expansion. Lower and upper case letters from the English alphabet may be used, with “a” equal to 1 through to “z” equal to 26. The function also accepts a leading sign, made of + and −. This is the inverse function of \texttt{\int_to_alph:n} and \texttt{\int_to_Alph:n}.

\texttt{\int_to_base:nn} \star \texttt{\int_to_Base:nn} \{\textit{integer expression}\} \{\textit{base}\}

Calculates the value of the \textit{integer expression} and converts it into the appropriate representation in the \textit{base}; the later may be given as an integer expression. For bases greater than 10 the higher “digits” are represented by letters from the English alphabet: lower case letters for \texttt{\int_to_base:n} and upper case ones for \texttt{\int_to_Base:n}. The maximum \textit{base} value is 36. The resulting tokens are digits with category code 12 (other) and letters with category code 11 (letter).

\texttt{\int_to_roman:n} \dagger \texttt{\int_to_Roman:n} \dagger \{\textit{integer expression}\}

Places the value of the \textit{integer expression} in the input stream as Roman numerals, either lower case \texttt{\int_to_roman:n} or upper case \texttt{\int_to_Roman:n}. If the value is negative or zero, the output is empty. The Roman numerals are letters with category code 11 (letter). The letters used are \texttt{mdclxvi}, repeated as needed: the notation with bars (such as \texttt{\textbar{v}} for 5000) is not used. For instance \texttt{\int_to_roman:n} \{8249\} expands to \texttt{mmmmmmmccxlix}.

\textbf{20.9 Converting from other formats to integers}

\texttt{\int_from_alph:n} \{\textit{letters}\}

Converts the \textit{letters} into the integer (base 10) representation and leaves this in the input stream. The \textit{letters} are first converted to a string, with no expansion. Lower and upper case letters from the English alphabet may be used, with “a” equal to 1 through to “z” equal to 26. The function also accepts a leading sign, made of + and −. This is the inverse function of \texttt{\int_to_alph:n} and \texttt{\int_to_Alph:n}.
\int_{from\ bin:n} \{\langle\text{binary number}\rangle\}

Converts the \langle\text{binary number}\rangle into the integer (base 10) representation and leaves this in the input stream. The \langle\text{binary number}\rangle is first converted to a string, with no expansion. The function accepts a leading sign, made of + and -, followed by binary digits. This is the inverse function of \int_{to\ bin:n}.

\int_{from\ hex:n} \{\langle\text{hexadecimal number}\rangle\}

Converts the \langle\text{hexadecimal number}\rangle into the integer (base 10) representation and leaves this in the input stream. Digits greater than 9 may be represented in the \langle\text{hexadecimal number}\rangle by upper or lower case letters. The \langle\text{hexadecimal number}\rangle is first converted to a string, with no expansion. The function also accepts a leading sign, made of + and -. This is the inverse function of \int_{to\ hex:n} and \int_{to\ Hex:n}.

\int_{from\ oct:n} \{\langle\text{octal number}\rangle\}

Converts the \langle\text{octal number}\rangle into the integer (base 10) representation and leaves this in the input stream. The \langle\text{octal number}\rangle is first converted to a string, with no expansion. The function accepts a leading sign, made of + and -, followed by octal digits. This is the inverse function of \int_{to\ oct:n}.

\int_{from\ roman:n} \{\langle\text{roman numeral}\rangle\}

Converts the \langle\text{roman numeral}\rangle into the integer (base 10) representation and leaves this in the input stream. The \langle\text{roman numeral}\rangle is first converted to a string, with no expansion. The \langle\text{roman numeral}\rangle may be in upper or lower case; if the numeral contains characters besides mdclxvi or MDCLXVI then the resulting value is $-1$. This is the inverse function of \int_{to\ roman:n} and \int_{to\ Roman:n}.

\int_{from\ base:nn} \{\langle\text{number}\rangle\} \{\langle\text{base}\rangle\}

Converts the \langle\text{number}\rangle expressed in \langle\text{base}\rangle into the appropriate value in base 10. The \langle\text{number}\rangle is first converted to a string, with no expansion. The \langle\text{number}\rangle should consist of digits and letters (either lower or upper case), plus optionally a leading sign. The maximum \langle\text{base}\rangle value is 36. This is the inverse function of \int_{to\ base:nn} and \int_{-\ to\ Base:nn}.

### 20.10 Random integers

\int_{rand:nn} \{\langle\text{intexpr}_1\rangle\} \{\langle\text{intexpr}_2\rangle\}

Evaluates the two \langle\text{integer expressions}\rangle and produces a pseudo-random number between the two (with bounds included). This is not available in older versions of Xe\TeX.

\int_{rand:n} \{\langle\text{intexpr}\rangle\}

Evaluates the \langle\text{integer expression}\rangle then produces a pseudo-random number between 1 and the \langle\text{intexpr}\rangle (included). This is not available in older versions of Xe\TeX.
20.11 Viewing integers

\int_show:N \langle integer \rangle
Displays the value of the \langle integer \rangle on the terminal.

\int_show:n \langle integer expression \rangle
Displays the result of evaluating the \langle integer expression \rangle on the terminal.

\int_log:N \langle integer \rangle
\int_log:c
\int_log:n \langle integer expression \rangle
Writes the value of the \langle integer \rangle in the log file.
Writes the result of evaluating the \langle integer expression \rangle in the log file.

20.12 Constant integers

\c_zero_int \c_one_int
New: 2018-05-07

\c_max_int
The maximum value that can be stored as an integer.

\c_max_register_int
Maximum number of registers.

\c_max_char_int
Maximum character code completely supported by the engine.

20.13 Scratch integers

\l_tmpa_int \l_tmpb_int
Scratch integer for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_int \g_tmpb_int
Scratch integer for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
20.14 Direct number expansion

\texttt{\int_value:w \star \int_value:w \langle \text{integer} \rangle} \\
\texttt{\int_value:w \langle \text{integer denotation} \rangle \ (optional \ space)}

Expands the following tokens until an \langle \text{integer} \rangle is formed, and leaves a normalized form (no leading sign except for negative numbers, no leading digit 0 except for zero) in the input stream as category code 12 (other) characters. The \langle \text{integer} \rangle can consist of any number of signs (with intervening spaces) followed by

- an integer variable (in fact, any \TeX{} register except \texttt{\toks}) or
- explicit digits (or by \texttt{\textbackslash (octal \ digits)} or \texttt{\textbackslash (hexadecimal \ digits)} or \texttt{\textbackslash \langle \text{character} \rangle}).

In this last case expansion stops once a non-digit is found; if that is a space it is removed as in \texttt{f-expansion}, and so \texttt{\exp_stop_f:} may be employed as an end marker. Note that protected functions are expanded by this process.

This function requires exactly one expansion to produce a value, and so is suitable for use in cases where a number is required “directly”. In general, \texttt{\int_eval:n} is the preferred approach to generating numbers.

\textbf{\TeX{}hackers note:} This is the \TeX{} primitive \texttt{\number}.

20.15 Primitive conditionals

\texttt{\if_int_compare:w \star \if_int_compare:w \langle \text{integer}_1 \rangle \ (relation) \langle \text{integer}_2 \rangle} \\
\texttt{\langle \text{true \ code} \rangle} \texttt{\else:} \texttt{\langle \text{false \ code} \rangle} \texttt{\fi:}

Compare two integers using \langle \text{relation} \rangle, which must be one of =, < or > with category code 12. The \texttt{\else:} branch is optional.

\textbf{\TeX{}hackers note:} These are both names for the \TeX{} primitive \texttt{\ifnum}.

\texttt{\if_case:w \star \if_case:w \langle \text{integer} \rangle \ (case}_0 \rangle} \\
\texttt{\or:} \texttt{\langle case}_1 \rangle \texttt{\or:} \ldots \texttt{\or:} \texttt{\langle default \rangle} \texttt{\fi:}

Selects a case to execute based on the value of the \langle \text{integer} \rangle. The first case ((\langle case}_0 \rangle) is executed if \langle \text{integer} \rangle is 0, the second ((\langle case}_1 \rangle) if the \langle \text{integer} \rangle is 1, etc. The \langle \text{integer} \rangle may be a literal, a constant or an integer expression (\textit{e.g.} using \texttt{\int_eval:n}).

\textbf{\TeX{}hackers note:} These are the \TeX{} primitives \texttt{\ifcase} and \texttt{\or.}
\if_int_odd:w \if_int_odd:w \langle tokens \rangle \langle \text{optional space} \rangle
\langle \text{true code} \rangle
\else:\n\langle \text{true code} \rangle
\fi:

Expands \langle tokens \rangle until a non-numeric token or a space is found, and tests whether the resulting \langle integer \rangle is odd. If so, \langle true code \rangle is executed. The \else: branch is optional.

\textbf{\TeXXhackers note:} This is the \TeX primitive \texttt{\ifodd}. 
Chapter 21

The \l3flag package: Expandable flags

Flags are the only data-type that can be modified in expansion-only contexts. This module is meant mostly for kernel use: in almost all cases, booleans or integers should be preferred to flags because they are very significantly faster.

A flag can hold any non-negative value, which we call its \langle height \rangle. In expansion-only contexts, a flag can only be “raised”: this increases the \langle height \rangle by 1. The \langle height \rangle can also be queried expandably. However, decreasing it, or setting it to zero requires non-expandable assignments.

Flag variables are always local. They are referenced by a \langle flag name \rangle such as \texttt{str\_missing}. The \langle flag name \rangle is used as part of \texttt{\textbackslash use: c} constructions hence is expanded at point of use. It must expand to character tokens only, with no spaces.

A typical use case of flags would be to keep track of whether an exceptional condition has occurred during expandable processing, and produce a meaningful (non-expandable) message after the end of the expandable processing. This is exemplified by \l3str-convert, which for performance reasons performs conversions of individual characters expandably and for readability reasons produces a single error message describing incorrect inputs that were encountered.

Flags should not be used without carefully considering the fact that raising a flag takes a time and memory proportional to its height. Flags should not be used unless unavoidable.

21.1 Setting up flags

\begin{verbatim}
\flag_new:n \flag_new:n \{\langle flag name \rangle\}  
\end{verbatim}

Creates a new flag with a name given by \langle flag name \rangle, or raises an error if the name is already taken. The \langle flag name \rangle may not contain spaces. The declaration is global, but flags are always local variables. The \langle flag \rangle initially has zero height.

\begin{verbatim}
\flag_clear:n \flag_clear:n \{\langle flag name \rangle\}  
\end{verbatim}

The \langle flag \rangle’s height is set to zero. The assignment is local.
\flag_clear_new:n \flag_clear_new:n \{flag\ name\} 

Ensures that the \(<\text{flag}\>\) exists globally by applying \flag_new:n if necessary, then applies \flag_clear:n, setting the height to zero locally.

\flag_show:n \flag_show:n \{\text{\textit{(flag name)}}\} 

Displays the \(<\text{flag}\>\)>’s height in the terminal.

\flag_log:n \flag_log:n \{\text{\textit{(flag name)}}\} 

Writes the \(<\text{flag}\>\)>’s height to the log file.

### 21.2 Expandable flag commands

\flag_if_exist:n \star \flag_if_exist:n \{\text{\textit{(flag name)}}\} 

This function returns true if the \(<\text{flag name}\>\) references a flag that has been defined previously, and false otherwise.

\flag_if_raised:n \star \flag_if_raised:n \{\text{\textit{(flag name)}}\} 

This function returns true if the \(<\text{flag}\>\) has non-zero height, and false if the \(<\text{flag}\>\) has zero height.

\flag_height:n \star \flag_height:n \{\text{\textit{(flag name)}}\} 

Expands to the height of the \(<\text{flag}\>\) as an integer denotation.

\flag_raise:n \star \flag_raise:n \{\text{\textit{(flag name)}}\} 

The \(<\text{flag}\>\)>’s height is increased by 1 locally.
Chapter 22

The \texttt{l3clist} package
Comma separated lists

Comma lists (in short, \texttt{clist}) contain ordered data where items can be added to the left or right end of the list. This data type allows basic list manipulations such as adding/removing items, applying a function to every item, removing duplicate items, extracting a given item, using the comma list with specified separators, and so on. Sequences (defined in \texttt{l3seq}) are safer, faster, and provide more features, so they should often be preferred to comma lists. Comma lists are mostly useful when interfacing with \LaTeX{} or other code that expects or provides items separated by commas.

Several items can be added at once. To ease input of comma lists from data provided by a user outside an \texttt{\ExplSyntaxOn ... \ExplSyntaxOff} block, spaces are removed from both sides of each comma-delimited argument upon input. Blank arguments are ignored, to allow for trailing commas or repeated commas (which may otherwise arise when concatenating comma lists “by hand”). In addition, a set of braces is removed if the result of space-trimming is braced: this allows the storage of any item in a comma list. For instance,
\begin{verbatim}
\clist_new:N \l_my_clist
\clist_put_left:Nn \l_my_clist { ~a~ , ~{b}~ , c~\d }
\clist_put_right:Nn \l_my_clist { ~{e~} , , {{f}} , }
\end{verbatim}
results in \texttt{\l_my_clist} containing \texttt{a,b,c~\d,\{e\},\{f\}} namely the five items \texttt{a, b, c~\d, e~} and \texttt{f}. Comma lists normally do not contain empty or blank items so the following gives an empty comma list:
\begin{verbatim}
\clist_clear_new:N \l_my_clist
\clist_set:Nn \l_my_clist { , - , , }
\clist_if_empty:NTF \l_my_clist { true } { false }
\end{verbatim}
and it leaves \texttt{true} in the input stream. To include an “unsafe” item (empty, or one that contains a comma, or starts or ends with a space, or is a single brace group), surround it with braces.

Any \texttt{n}-type token list is a valid comma list input for \texttt{l3clist} functions, which will split the token list at every comma and process the items as described above. On the other hand, \texttt{N}-type functions expect comma list variables, which are particular token list variables in which this processing of items (and removal of blank items) has already
occurred. Because comma list variables are token list variables, expanding them once yields their items separated by commas, and \texttt{l3tl} functions such as \texttt{tl\_show:N} can be applied to them. (These functions often have \texttt{l3clist} analogues, which should be preferred.)

Almost all operations on comma lists are noticeably slower than those on sequences so converting the data to sequences using \texttt{\seq\_set\_from\_clist:Nn} (see \texttt{l3seq}) may be advisable if speed is important. The exception is that \texttt{\clist\_if\_in:NnTF} and \texttt{\clist\_-remove\_duplicates:N} may be faster than their sequence analogues for large lists. However, these functions work slowly for “unsafe” items that must be braced, and may produce errors when their argument contains \{,\} or \# (assuming the usual \TeX\ category codes apply). The sequence data type should thus certainly be preferred to comma lists to store such items.

### 22.1 Creating and initialising comma lists

\begin{itemize}
  \item \texttt{\clist\_new:N}\<\comma\ list> \quad \texttt{\clist\_new:c}\<\comma\ list>\n
  \begin{itemize}
    \item \texttt{\clist\_new:N}\<\comma\ list> \quad \texttt{\clist\_new:c}\<\comma\ list>
    \end{itemize}

  \begin{itemize}
    \item Creates a new \(<\ comma\ list>\) or raises an error if the name is already taken. The declaration is global. The \(<\ comma\ list>\) initially contains no items.
    \end{itemize}

  \begin{itemize}
    \item \texttt{\clist\_const:Nn}\<\clist\ var> \{\comma\ list\}
    \end{itemize}

  \begin{itemize}
    \item Creates a new constant \(<\clist\ var>\) or raises an error if the name is already taken. The value of the \(<\clist\ var>\) is set globally to the \(<\ comma\ list>\).
    \end{itemize}

  \begin{itemize}
    \item \texttt{\clist\_clear:N\<\comma\ list>}
    \end{itemize}

  \begin{itemize}
    \item Clears all items from the \(<\ comma\ list>\).
    \end{itemize}

  \begin{itemize}
    \item \texttt{\clist\_clear\_new:N}\<\comma\ list>
    \end{itemize}

  \begin{itemize}
    \item Ensures that the \(<\ comma\ list>\) exists globally by applying \texttt{\clist\_new:N} if necessary, then applies \texttt{\clist\_clear\_new:N} to leave the list empty.
    \end{itemize}

  \begin{itemize}
    \item \texttt{\clist\_set\_eq:NN}\<\comma\ list_1>\<\comma\ list_2>
    \end{itemize}

  \begin{itemize}
    \item Sets the content of \(<\ comma\ list_1>\) equal to that of \(<\ comma\ list_2>\). To set a token list variable equal to a comma list variable, use \texttt{\tl\_set\_eq:NN}. Conversely, setting a comma list variable to a token list is unadvisable unless one checks space-trimming and related issues.
    \end{itemize}

  \begin{itemize}
    \item \texttt{\clist\_set\_from\_seq:NN}\<\comma\ list>\<\sequence>
    \end{itemize}

  \begin{itemize}
    \item Converts the data in the \(<\sequence>\) into a \(<\ comma\ list>\): the original \(<\sequence>\) is unchanged. Items which contain either spaces or commas are surrounded by braces.
    \end{itemize}
\end{itemize}
\clist_concat:NNN \clist_concat:ccc \clist_gconcat:NNN \clist_gconcat:ccc

Concatenates the content of \langle comma list\rangle_2 and \langle comma list\rangle_3 together and saves the result in \langle comma list\rangle. The items in \langle comma list\rangle_2 are placed at the left side of the new comma list.

\clist_concat:NNN \clist_concat:ccc \clist_gconcat:NNN \clist_gconcat:ccc

Tests whether the \langle comma list\rangle is currently defined. This does not check that the \langle comma list\rangle really is a comma list.

\clist_if_exist_p:N \clist_if_exist:NTF \langle comma list\rangle \{ \langle true code\rangle \} \{ \langle false code\rangle \}

22.2 Adding data to comma lists

\clist_set:Nn \langle comma list\rangle { \langle item\rangle_1, \ldots, \langle item\rangle_n }

Sets \langle comma list\rangle to contain the \langle items\rangle, removing any previous content from the variable. Blank items are omitted, spaces are removed from both sides of each item, then a set of braces is removed if the resulting space-trimmed item is braced. To store some \langle tokens\rangle as a single \langle item\rangle even if the \langle tokens\rangle contain commas or spaces, add a set of braces: \clist_set:Nn \langle comma list\rangle { \{ \langle tokens\rangle \} }.

\clist_set:Nn \langle comma list\rangle { \langle item\rangle_1, \ldots, \langle item\rangle_n }

\clist_if_exist_p:N \clist_if_exist:NTF \langle comma list\rangle \{ \langle true code\rangle \} \{ \langle false code\rangle \}

Updated: 2011-09-06

\clist_put_left:Nn \langle comma list\rangle { \langle item\rangle_1, \ldots, \langle item\rangle_n }

Appends the \langle items\rangle to the left of the \langle comma list\rangle. Blank items are omitted, spaces are removed from both sides of each item, then a set of braces is removed if the resulting space-trimmed item is braced. To append some \langle tokens\rangle as a single \langle item\rangle even if the \langle tokens\rangle contain commas or spaces, add a set of braces: \clist_put_left:Nn \langle comma list\rangle { \{ \langle tokens\rangle \} }.

\clist_put_left:Nn \langle comma list\rangle { \langle item\rangle_1, \ldots, \langle item\rangle_n }

\clist_if_exist_p:N \clist_if_exist:NTF \langle comma list\rangle \{ \langle true code\rangle \} \{ \langle false code\rangle \}

Updated: 2011-09-05

\clist_put_right:Nn \langle comma list\rangle { \langle item\rangle_1, \ldots, \langle item\rangle_n }

Appends the \langle items\rangle to the right of the \langle comma list\rangle. Blank items are omitted, spaces are removed from both sides of each item, then a set of braces is removed if the resulting space-trimmed item is braced. To append some \langle tokens\rangle as a single \langle item\rangle even if the \langle tokens\rangle contain commas or spaces, add a set of braces: \clist_put_right:Nn \langle comma list\rangle { \{ \langle tokens\rangle \} }.

\clist_put_right:Nn \langle comma list\rangle { \langle item\rangle_1, \ldots, \langle item\rangle_n }

\clist_if_exist_p:N \clist_if_exist:NTF \langle comma list\rangle \{ \langle true code\rangle \} \{ \langle false code\rangle \}

Updated: 2011-09-05
22.3 Modifying comma lists

While comma lists are normally used as ordered lists, it may be necessary to modify the content. The functions here may be used to update comma lists, while retaining the order of the unaffected entries.

```
\clist_remove_duplicates:N \clist_remove_duplicates:N \langle comma list \rangle
\clist_remove_duplicates:c \clist_gremove_duplicates:N \clist_gremove_duplicates:c
```

Removes duplicate items from the \textit{(comma list)}, leaving the left most copy of each item in the \textit{(comma list)}. The \textit{(item)} comparison takes place on a token basis, as for \texttt{\tl_if_eq:nnTF}.

\textbf{\LaTeX{}hackers note:} This function iterates through every item in the \textit{(comma list)} and does a comparison with the \textit{items} already checked. It is therefore relatively slow with large comma lists. Furthermore, it may fail if any of the items in the \textit{(comma list)} contains \{, \}, or \# (assuming the usual \LaTeX{} category codes apply).

```
\clist_remove_all:Nn \clist_remove_all:Nn \langle comma list \rangle \{ \langle item \rangle \}
\clist_remove_all:cn \clist_gremove_all:cn
```

Removes every occurrence of \textit{(item)} from the \textit{(comma list)}. The \textit{(item)} comparison takes place on a token basis, as for \texttt{\tl_if_eq:nnTF}.

\textbf{\LaTeX{}hackers note:} The function may fail if the \textit{(item)} contains \{, \}, or \# (assuming the usual \LaTeX{} category codes apply).

```
\clist_reverse:N \clist_reverse:c \clist_greverse:N \clist_greverse:c
```

Reverses the order of items stored in the \textit{(comma list)}.

```
\clist_reverse:n \langle comma list \rangle
```

Leaves the items in the \textit{(comma list)} in the input stream in reverse order. Contrarily to other what is done for other \texttt{n}-type \textit{(comma list)} arguments, braces and spaces are preserved by this process.

\textbf{\LaTeX{}hackers note:} The result is returned within \texttt{\unexpanded{}}, which means that the comma list does not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.
\clist_sort:Nn  ⟨clist var⟩ ⟨(comparison code)⟩

Sorts the items in the ⟨clist var⟩ according to the ⟨comparison code⟩, and assigns the result to ⟨clist var⟩. The details of sorting comparison are described in Section 6.1.

22.4 Comma list conditionals

\clist_if_empty_p:N  ⟨comma list⟩
\clist_if_empty:p:c  ⟨comma list⟩ ⟨(true code)⟩ ⟨(false code)⟩
\clist_if_empty:NTF  ⟨comma list⟩ ⟨(true code)⟩ ⟨(false code)⟩
\clist_if_empty:c  ⟨comma list⟩ ⟨(true code)⟩ ⟨(false code)⟩

Tests if the ⟨comma list⟩ is empty (containing no items).

\clist_if_empty_p:n  ⟨comma list⟩
\clist_if_empty:nTF ⟨comma list⟩ ⟨(true code)⟩ ⟨(false code)⟩
\clist_if_empty:n  ⟨comma list⟩ ⟨(true code)⟩ ⟨(false code)⟩
\clist_if_empty:nTF ⟨comma list⟩ ⟨(true code)⟩ ⟨(false code)⟩

Tests if the ⟨comma list⟩ is empty (containing no items). The rules for space trimming are as for other \text{n}-type comma-list functions, hence the comma list {\-,\-,\-,} (without outer braces) is empty, while {\-,\{{},\}} (without outer braces) contains one element, which happens to be empty: the comma-list is not empty.

\clist_if_in:NnTF ⟨comma list⟩ ⟨(item)⟩ ⟨(true code)⟩ ⟨(false code)⟩
\clist_if_in:TFF ⟨comma list⟩ ⟨(item)⟩ ⟨(true code)⟩ ⟨(false code)⟩
\clist_if_in:nnTF ⟨comma list⟩ ⟨(item)⟩ ⟨(true code)⟩ ⟨(false code)⟩
\clist_if_in:nnTFF ⟨comma list⟩ ⟨(item)⟩ ⟨(true code)⟩ ⟨(false code)⟩

Tests if the ⟨item⟩ is present in the ⟨comma list⟩. In the case of an \text{n}-type ⟨comma list⟩, the usual rules of space trimming and brace stripping apply. Hence,

\[
\text{\clist_if_in:nnTF \{a, \{b\}-, \{b\}, c\} \{b\} \{true\} \{false\}}
\]

yields \text{true}.

\text{TeXhackers note:} The function may fail if the ⟨item⟩ contains {, }, or # (assuming the usual \TeX\ category codes apply).

22.5 Mapping over comma lists

The functions described in this section apply a specified function to each item of a comma list. All mappings are done at the current group level, i.e. any local assignments made by the ⟨function⟩ or ⟨code⟩ discussed below remain in effect after the loop.

When the comma list is given explicitly, as an \text{n}-type argument, spaces are trimmed around each item. If the result of trimming spaces is empty, the item is ignored. Otherwise, if the item is surrounded by braces, one set is removed, and the result is passed to the mapped function. Thus, if the comma list that is being mapped is \{a\}, \{b\}, \{c\}, then the arguments passed to the mapped function are ‘a’, ‘\{b\}’, an empty argument, and ‘c’.

175
When the comma list is given as an N-type argument, spaces have already been trimmed on input, and items are simply stripped of one set of braces if any. This case is more efficient than using n-type comma lists.

\clist_map_function:NN ⟨comma list⟩ ⟨function⟩
Applies ⟨function⟩ to every ⟨item⟩ stored in the ⟨comma list⟩. The ⟨function⟩ receives one argument for each iteration. The ⟨items⟩ are returned from left to right. The function \clist_map_inline:Nn is in general more efficient than \clist_map_function:NN.

\clist_map_inline:Nn ⟨comma list⟩ {⟨inline function⟩}
Applies ⟨inline function⟩ to every ⟨item⟩ stored within the ⟨comma list⟩. The ⟨inline function⟩ should consist of code which receives the ⟨item⟩ as #1. The ⟨items⟩ are returned from left to right.

\clist_map_variable:NNn ⟨comma list⟩ ⟨variable⟩ {⟨code⟩}
Stores each ⟨item⟩ of the ⟨comma list⟩ in turn in the ⟨token list⟩ ⟨variable⟩ and applies the ⟨code⟩. The ⟨code⟩ will usually make use of the ⟨variable⟩, but this is not enforced. The assignments to the ⟨variable⟩ are local. Its value after the loop is the last ⟨item⟩ in the ⟨comma list⟩, or its original value if there were no ⟨item⟩. The ⟨items⟩ are returned from left to right.

\clist_map_tokens:Nn ⟨clist var⟩ {⟨code⟩}
\clist_map_tokens:nn {⟨comma list⟩} {⟨code⟩}
Calls ⟨code⟩ {⟨item⟩} for every ⟨item⟩ stored in the ⟨comma list⟩. The ⟨code⟩ receives each ⟨item⟩ as a trailing brace group. If the ⟨code⟩ consists of a single function this is equivalent to \clist_map_function:N.

\clist_map_break:
Used to terminate a \clist_map_... function before all entries in the ⟨comma list⟩ have been processed. This normally takes place within a conditional statement, for example

\clist_map_inline:Nn \l_my_clist
{\str_if_eq:nnTF { #1 } { bingo }
{ \clist_map_break: }
% Do something useful}

Use outside of a \clist_map_... scenario leads to low level \TeX errors.

\TeX hackers note: When the mapping is broken, additional tokens may be inserted before further items are taken from the input stream. This depends on the design of the mapping function.
\clist_map_break:n \clist_map_break:n \langle code \rangle

Used to terminate a \clist_map_... function before all entries in the \langle comma list \rangle have been processed, inserting the \langle code \rangle after the mapping has ended. This normally takes place within a conditional statement, for example

\clist_map_inline:Nn \l_my_clist
{\str_if_eq:nnTF { #1 } { bingo }{ \clist_map_break:n { \langle code \rangle } }%
% Do something useful}

Use outside of a \clist_map_... scenario leads to low level \TeX errors.

\TeXhackers note: When the mapping is broken, additional tokens may be inserted before the \langle code \rangle is inserted into the input stream. This depends on the design of the mapping function.

\clist_count:N \clist_count:N \clist_count:n

Leaves the number of items in the \langle comma list \rangle in the input stream as an \langle integer denotation \rangle. The total number of items in a \langle comma list \rangle includes those which are duplicates, i.e. every item in a \langle comma list \rangle is counted.

\clist_count:N \clist_count:N \clist_count:n \clist_count:c

New: 2012-07-13

\clist_use:Nnnn \clist_use:Nnnn \clist_use:cnnn

\clist_use:Nnnn \langle list var \rangle \langle separator between two \rangle \langle separator between more than two \rangle \langle separator between final two \rangle

Places the contents of the \langle list var \rangle in the input stream, with the appropriate \langle separator \rangle between the items. Namely, if the comma list has more than two items, the \langle separator between more than two \rangle is placed between each pair of items except the last, for which the \langle separator between final two \rangle is used. If the comma list has exactly two items, then they are placed in the input stream separated by the \langle separator between two \rangle. If the comma list has a single item, it is placed in the input stream, and a comma list with no items produces no output. An error is raised if the variable does not exist or if it is invalid.

For example,

\clist_set:Nn \l_tmpa_clist { a , b , , c , {de} , f }
\clist_use:Nnnn \l_tmpa_clist { -and- } { , - } { , - and - }

inserts “a, b, c, de, and f” in the input stream. The first separator argument is not used in this case because the comma list has more than 2 items.

\TeXhackers note: The result is returned within the \unexpanded primitive (\exp_not:n), which means that the \langle items \rangle do not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.

22.6 Using the content of comma lists directly

\clist_use:Nnnn \clist_use:Nnnn

\clist_use:Nnnn \langle list var \rangle \langle separator between two \rangle \langle separator between more than two \rangle \langle separator between final two \rangle

New: 2013-05-26

\clist_use:Nnnn \clist_use:Nnnn

Places the contents of the \langle list var \rangle in the input stream, with the appropriate \langle separator \rangle between the items. Namely, if the comma list has more than two items, the \langle separator between more than two \rangle is placed between each pair of items except the last, for which the \langle separator between final two \rangle is used. If the comma list has exactly two items, then they are placed in the input stream separated by the \langle separator between two \rangle. If the comma list has a single item, it is placed in the input stream, and a comma list with no items produces no output. An error is raised if the variable does not exist or if it is invalid.

For example,

\clist_set:Nn \l_tempt_clist { a , b , , c , {de} , f }
\clist_use:Nnnn \l_tempt_clist { -and- } { , - } { , - and - }

inserts “a, b, c, de, and f” in the input stream. The first separator argument is not used in this case because the comma list has more than 2 items.

\TeXhackers note: The result is returned within the \unexpanded primitive (\exp_not:n), which means that the \langle items \rangle do not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.
\clist_use:Nn \clist_use:cn
\clist_use:nnnn \clist_use:nn
\clist_get:NN \clist_get:cN \clist_get:NN \clist_get:NN \clist_get:cN \clist_get:NN
\clist_pop:NN \clist_pop:cN

\clist_use:Nn \clist_use:cn
\clist_set:Nn \l_tmpa_clist { a , b , , c , \{de\} , f }
\clist_use:Nn \l_tmpa_clist { ~and~ }

\textbf{TEXhackers note:} The result is returned within the \texttt{\unexpanded} primitive \texttt{(\exp_not:n)}, which means that the \langle items \rangle do not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.

\clist_get:NN \l_tmpa_clist { a , b , , c , \{de\} , f }
\clist_get:NN \l_tmpa_clist { -and- }

\textbf{TEXhackers note:} The result is returned within the \texttt{\unexpanded} primitive \texttt{(\exp_not:n)}, which means that the \langle items \rangle do not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.

\section*{22.7 Comma lists as stacks}

Comma lists can be used as stacks, where data is pushed to and popped from the top of the comma list. (The left of a comma list is the top, for performance reasons.) The stack functions for comma lists are not intended to be mixed with the general ordered data functions detailed in the previous section: a comma list should either be used as an ordered data type or as a stack, but not in both ways.

\clist_pop:NN \clist_pop:cN

\textbf{TEXhackers note:} The result is returned within the \texttt{\unexpanded} primitive \texttt{(\exp_not:n)}, which means that the \langle items \rangle do not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.
\clist_gpop:NN \clist_gpop:cN
Pops the left-most item from a \textit{comma list} into the \textit{token list variable}, i.e. removes the item from the comma list and stores it in the \textit{token list variable}. The \textit{comma list} is modified globally, while the assignment of the \textit{token list variable} is local.

\clist_gpop:NNTF \clist_gpop:cNTF
\text{New: 2012-05-14}
If the \textit{comma list} is empty, leaves the \textit{false code} in the input stream. The value of the \textit{token list variable} is not defined in this case and should not be relied upon. If the \textit{comma list} is non-empty, pops the top item from the \textit{comma list} in the \textit{token list variable}, i.e. removes the item from the \textit{comma list}. Both the \textit{comma list} and the \textit{token list variable} are assigned locally.

\clist_gpush:Nn \clist_gpush:NN \clist_gpush:cN
\text{New: 2014-07-17}

\clist_item:Nn \clist_item:cn \clist_item:nn
\text{New: 2014-07-17}

22.8 Using a single item

\clist_item:Nn \clist_item:cn \clist_item:nn
\text{New: 2014-07-17}

Adding the \textit{items} to the top of the \textit{comma list}. Spaces are removed from both sides of each item as for any n-type comma list.

\clist_item:Nn \clist_item:cn \clist_item:nn
\text{New: 2014-07-17}

Indexing items in the \textit{comma list} from 1 at the top (left), this function evaluates the \textit{integer expression} and leaves the appropriate item from the comma list in the input stream. If the \textit{integer expression} is negative, indexing occurs from the bottom (right) of the comma list. When the \textit{integer expression} is larger than the number of items in the \textit{comma list} (as calculated by \clist_count:N) then the function expands to nothing.

\textbf{TeXhacker note:} The result is returned within the \texttt{\unexpanded} primitive (\texttt{\exp_not:n}), which means that the \textit{item} does not expand further when appearing in an x-type or e-type argument expansion.
\clist_rand_item:N \clist_rand_item:c \clist_rand_item:n

Selects a pseudo-random item of the \langle comma list \rangle. If the \langle comma list \rangle has no item, the result is empty.

\texttt{\LaTeX}hackers note: The result is returned within the \texttt{\unexpanded} primitive (\texttt{\exp_not:n}), which means that the \texttt{\item} does not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.

## 22.9 Viewing comma lists

\clist_show:N \clist_show:c

Displays the entries in the \langle comma list \rangle in the terminal.

\clist_show:n \clist_show:c

Displays the entries in the comma list in the terminal.

\clist_log:N \clist_log:c

\clist_log:n \clist_log:c

Writes the entries in the \langle comma list \rangle in the log file. See also \clist_show:N which displays the result in the terminal.

Writes the entries in the comma list in the log file. See also \clist_show:n which displays the result in the terminal.

## 22.10 Constant and scratch comma lists

\c_empty_clist

Constant that is always empty.

\l_tmpa_clist \l_tmpb_clist

Scratch comma lists for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_clist \g_tmpb_clist

Scratch comma lists for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
Chapter 23

The \texttt{l3token} package

Token manipulation

This module deals with tokens. Now this is perhaps not the most precise description so let’s try with a better description: When programming in \TeX, it is often desirable to know just what a certain token is: is it a control sequence or something else. Similarly one often needs to know if a control sequence is expandable or not, a macro or a primitive, how many arguments it takes etc. Another thing of great importance (especially when it comes to document commands) is looking ahead in the token stream to see if a certain character is present and maybe even remove it or disregard other tokens while scanning. This module provides functions for both and as such has two primary function categories: \token\_ for anything that deals with tokens and \peek\_ for looking ahead in the token stream.

Most functions we describe here can be used on control sequences, as those are tokens as well.

It is important to distinguish two aspects of a token: its “shape” (for lack of a better word), which affects the matching of delimited arguments and the comparison of token lists containing this token, and its “meaning”, which affects whether the token expands or what operation it performs. One can have tokens of different shapes with the same meaning, but not the converse.

For instance, \texttt{\if:w}, \texttt{\if\charcode:w}, and \texttt{\tex_if:D} are three names for the same internal operation of \TeX, namely the primitive testing the next two characters for equality of their character code. They have the same meaning hence behave identically in many situations. However, \TeX distinguishes them when searching for a delimited argument. Namely, the example function \texttt{\show_until_if:w} defined below takes everything until \texttt{\if:w} as an argument, despite the presence of other copies of \texttt{\if:w} under different names.

\begin{verbatim}
\cs_new:Npn \show_until_if:w #1 \if:w { \tl_show:n {#1} }
\show_until_if:w \tex_if:D \if\charcode:w \if:w
\end{verbatim}

A list of all possible shapes and a list of all possible meanings are given in section 23.7.
23.1 Creating character tokens

\char_set_active_eq:NN \char \langle \text{function} \rangle

Sets the behaviour of the \langle char \rangle in situations where it is active (category code 13) to be equivalent to that of the \langle function \rangle. The category code of the \langle char \rangle is unchanged by this process. The \langle function \rangle may itself be an active character.

\char_set_active_eq:Nc
\char_gset_active_eq:NN
\char_gset_active_eq:Nc

Updated: 2015-11-12

\char_set_active_eq:nN \{\langle \text{integer expression} \rangle\} \langle \text{function} \rangle

Sets the behaviour of the \langle char \rangle which has character code as given by the \langle integer expression \rangle in situations where it is active (category code 13) to be equivalent to that of the \langle function \rangle. The category code of the \langle char \rangle is unchanged by this process. The \langle function \rangle may itself be an active character.

\char_gset_active_eq:nN
\char_gset_active_eq:nc

New: 2015-11-12

\char_generate:nn \{\langle \text{charcode} \rangle\} \{\langle \text{catcode} \rangle\}

Generates a character token of the given \langle charcode \rangle and \langle catcode \rangle (both of which may be integer expressions). The \langle catcode \rangle may be one of

- 1 (begin group)
- 2 (end group)
- 3 (math toggle)
- 4 (alignment)
- 6 (parameter)
- 7 (math superscript)
- 8 (math subscript)
- 10 (space)
- 11 (letter)
- 12 (other)
- 13 (active)

and other values raise an error. The \langle charcode \rangle may be any one valid for the engine in use, except that for \langle catcode \rangle 10, \langle charcode \rangle 0 is not allowed. Active characters cannot be generated in older versions of X\TeX. Another way to build token lists with unusual category codes is \regex_replace:nnN {.*} \{\langle replacement \rangle\} {\tl \var}.

\textbf{\TeXhackers note}: Exactly two expansions are needed to produce the character.
\char_lowercase:N \* \char_uppercase:N \* \char_titlecase:N \* \char_foldcase:N \* \char_str_lowercase:N \* \char_str_uppercase:N \* \char_str_titlecase:N \* \char_str_foldcase:N \* \char_set_catcode_letter:N \langle character \rangle
\char_set_catcode_escape:N \langle character \rangle
\char_set_catcode_group_begin:N \langle character \rangle
\char_set_catcode_group_end:N \langle character \rangle
\char_set_catcode_math_toggle:N \langle character \rangle
\char_set_catcode_alignment:N \langle character \rangle
\char_set_catcode_end_line:N \langle character \rangle
\char_set_catcode_parameter:N \langle character \rangle
\char_set_catcode_math_superscript:N \langle character \rangle
\char_set_catcode_math_subscript:N \langle character \rangle
\char_set_catcode_ignore:N \langle character \rangle
\char_set_catcode_space:N \langle character \rangle
\char_set_catcode_letter:N \langle character \rangle
\char_set_catcode_other:N \langle character \rangle
\char_set_catcode_active:N \langle character \rangle
\char_set_catcode_comment:N \langle character \rangle
\char_set_catcode_invalid:N \langle character \rangle

Sets the category code of the \langle character \rangle to that indicated in the function name. Depending on the current category code of the \langle token \rangle the escape token may also be needed:

\char_set_catcode_other:N \% \langle token \rangle

The assignment is local.

23.2 Manipulating and interrogating character tokens

Token list containing one character with category code 12, ("other"), and character code 32 (space).

\c_catcode_other_space_tl

Rev: 2020-01-09

Rev: 2011-09-05

New: 2020-01-09

New: 2011-09-05

New: 2015-11-11

Updated: 2015-11-11

Converts the \langle char \rangle to the equivalent case-changed character as detailed by the function name (see \str_foldcase:n and \text_titlecase:n for details of these terms). The case mapping is carried out with no context-dependence (cf. \text_uppercase:n, etc.) The str versions always generate "other" (category code 12) characters, whilst the standard versions generate characters with the category code of the \langle char \rangle (i.e. only the character code changes).
Sets the category code of the \textit{character} which has character code as given by the \textit{integer expression}. This version can be used to set up characters which cannot otherwise be given (cf. the \textit{N}-type variants). The assignment is local.

These functions set the category code of the \textit{character} which has character code as given by the \textit{integer expression}. The first \textit{integer expression} is the character code and the second is the category code to apply. The setting applies within the current \TeX\ group. In general, the symbolic functions \texttt{\char_set_catcode\_\texttt{type}} should be preferred, but there are cases where these lower-level functions may be useful.

Expands to the current category code of the \textit{character} with character code given by the \textit{integer expression}.

Displays the current category code of the \textit{character} with character code given by the \textit{integer expression} on the terminal.

Sets up the behaviour of the \textit{character} when found inside \texttt{\text_lowercase:n}, such that \textit{character}_1 will be converted into \textit{character}_2. The two \textit{characters} may be specified using an \textit{integer expression} for the character code concerned. This may include the \TeX\ \texttt{\texttt{\char\_\texttt{character}}} method for converting a single character into its character code:

\begin{verbatim}
\char_set_lccode:nn { \texttt{\char\_\texttt{A}} } { \texttt{\char\_\texttt{a}} } % Standard behaviour
\char_set_lccode:nn { \texttt{\char\_\texttt{A}} } { \texttt{\char\_\texttt{A} + 32} }
\char_set_lccode:nn { 50 } { 60 }
\end{verbatim}

The setting applies within the current \TeX\ group.
\char_value_lccode:n \char_value_lccode:n \{(integer expression)\}
Expands to the current lower case code of the \langle character \rangle with character code given by the \langle integer expression \rangle .

\char_show_value_lccode:n \char_show_value_lccode:n \{(integer expression)\}
Displays the current lower case code of the \langle character \rangle with character code given by the \langle integer expression \rangle on the terminal.

\char_set_uccode:nn \char_set_uccode:nn \{(intexpr)\} \{(intexpr)\}
Sets up the behaviour of the \langle character \rangle when found inside \text_uppercase:n , such that \langle character_1 \rangle will be converted into \langle character_2 \rangle . The two \langle characters \rangle may be specified using an \langle integer expression \rangle for the character code concerned. This may include the \TeX \ '\langle character \rangle method for converting a single character into its character code:
\begin{verbatim}
\char_set_uccode:nn \{ \textbackslash a \} \{ \textbackslash A \} \ % Standard behaviour
\char_set_uccode:nn \{ \textbackslash A \} \{ \textbackslash A - 32 \}
\char_set_uccode:nn \{ 60 \} \{ 50 \}
\end{verbatim}
The setting applies within the current \TeX group.

\char_value_uccode:n \char_value_uccode:n \{(integer expression)\}
Expands to the current upper case code of the \langle character \rangle with character code given by the \langle integer expression \rangle .

\char_show_value_uccode:n \char_show_value_uccode:n \{(integer expression)\}
Displays the current upper case code of the \langle character \rangle with character code given by the \langle integer expression \rangle on the terminal.

\char_set_mathcode:nn \char_set_mathcode:nn \{(intexpr)\} \{(intexpr)\}
This function sets up the math code of \langle character \rangle . The \langle character \rangle is specified as an \langle integer expression \rangle which will be used as the character code of the relevant character. The setting applies within the current \TeX group.

\char_value_mathcode:n \char_value_mathcode:n \{(integer expression)\}
Expands to the current math code of the \langle character \rangle with character code given by the \langle integer expression \rangle .

\char_show_value_mathcode:n \char_show_value_mathcode:n \{(integer expression)\}
Displays the current math code of the \langle character \rangle with character code given by the \langle integer expression \rangle on the terminal.

\char_set_sfcode:nn \char_set_sfcode:nn \{(intexpr)\} \{(intexpr)\}
This function sets up the space factor for the \langle character \rangle . The \langle character \rangle is specified as an \langle integer expression \rangle which will be used as the character code of the relevant character. The setting applies within the current \TeX group.
\char_value_sfcode:n \star \char_show_value_sfcode:n

Expands to the current space factor for the \langle character \rangle with character code given by the \langle integer expression \rangle.

\char_value.sfcode:n \star \char_show.sfcode:n

Displays the current space factor for the \langle character \rangle with character code given by the \langle integer expression \rangle on the terminal.

\l_char_active_seq

Used to track which tokens may require special handling at the document level as they are (or have been at some point) of category \langle active \rangle (catcode 13). Each entry in the sequence consists of a single escaped token, for example \texttt{\~.} Active tokens should be added to the sequence when they are defined for general document use.

\l_char_special_seq

Used to track which tokens will require special handling when working with verbatim-like material at the document level as they are not of categories \langle letter \rangle (catcode 11) or \langle other \rangle (catcode 12). Each entry in the sequence consists of a single escaped token, for example \texttt{\textbackslash } for the backslash or \texttt{\{ } for an opening brace. Escaped tokens should be added to the sequence when they are defined for general document use.

23.3 Generic tokens

These are implicit tokens which have the category code described by their name. They are used internally for test purposes but are also available to the programmer for other uses.

\c_group_begin_token
\c_group_end_token
\c_math_toggle_token
\c_alignment_token
\c_parameter_token
\c_math_superscript_token
\c_math_subscript_token
\c_space_token

These are implicit tokens which have the category code described by their name. They are used internally for test purposes and should not be used other than for category code tests.

\c_catcode_letter_token
\c_catcode_other_token

A token list containing an active token. This is used internally for test purposes and should not be used other than in appropriately-constructed category code tests.
23.4 Converting tokens

\token_to_meaning:N \token_to_meaning:c 

Inserts the current meaning of the \langle token \rangle into the input stream as a series of characters of category code 12 (other). This is the primitive \TeX description of the \langle token \rangle, thus for example both functions defined by \cs_set_nopar:Npn and token list variables defined using \tl_new:N are described as macros.

\TeXhackers note: This is the \TeX primitive \meaning. The \langle token \rangle can thus be an explicit space tokens or an explicit begin-group or end-group character token \{ or \} when normal \TeX category codes apply) even though these are not valid N-type arguments.

\token_to_str:N \token_to_str:c 

Converts the given \langle token \rangle into a series of characters with category code 12 (other). If the \langle token \rangle is a control sequence, this will start with the current escape character with category code 12 (the escape character is part of the \langle token \rangle). This function requires only a single expansion.

\TeXhackers note: \token_to_str:N is the \TeX primitive \string renamed. The \langle token \rangle can thus be an explicit space tokens or an explicit begin-group or end-group character token \{ or \} when normal \TeX category codes apply) even though these are not valid N-type arguments.

23.5 Token conditionals

\token_if_group_begin_p:N \token_if_group_begin:NTF 

Tests if \langle token \rangle has the category code of a begin group token \{ when normal \TeX category codes are in force). Note that an explicit begin group token cannot be tested in this way, as it is not a valid N-type argument.

\token_if_group_end_p:N \token_if_group_end:NTF 

Tests if \langle token \rangle has the category code of an end group token \} when normal \TeX category codes are in force). Note that an explicit end group token cannot be tested in this way, as it is not a valid N-type argument.

\token_if_math_toggle_p:N \token_if_math_toggle:NTF 

Tests if \langle token \rangle has the category code of a math shift token \$ when normal \TeX category codes are in force).
Tests if \langle \text{token} \rangle has the category code of an alignment token (& when normal \TeX{} category codes are in force).

Tests if \langle \text{token} \rangle has the category code of a macro parameter token (# when normal \TeX{} category codes are in force).

Tests if \langle \text{token} \rangle has the category code of a superscript token (^ when normal \TeX{} category codes are in force).

Tests if \langle \text{token} \rangle has the category code of a subscript token (_ when normal \TeX{} category codes are in force).

Tests if \langle \text{token} \rangle has the category code of a space token. Note that an explicit space token with character code 32 cannot be tested in this way, as it is not a valid N-type argument.

Tests if \langle \text{token} \rangle has the category code of a letter token.

Tests if \langle \text{token} \rangle has the category code of an “other” token.

Tests if \langle \text{token} \rangle has the category code of an active character.

Tests if the two \langle \text{tokens} \rangle have the same category code.

Tests if the two \langle \text{tokens} \rangle have the same character code.
\token_if_eq_meaning_p:NN * \token_if_eq_meaning_p:NN ⟨token1⟩ ⟨token2⟩
\token_if_eq_meaning:NNTF * \token_if_eq_meaning:NNTF ⟨token1⟩ ⟨token2⟩ {⟨true code⟩} {⟨false code⟩}

Tests if the two ⟨tokens⟩ have the same meaning when expanded.

\token_if_macro_p:N * \token_if_macro_p:N ⟨token⟩
\token_if_macro:NTF * \token_if_macro:NTF ⟨token⟩ {⟨true code⟩} {⟨false code⟩}

Tests if the ⟨token⟩ is a \TeX{} macro.

\token_if_cs_p:N * \token_if_cs_p:N ⟨token⟩
\token_if_cs:NTF * \token_if_cs:NTF ⟨token⟩ {⟨true code⟩} {⟨false code⟩}

Tests if the ⟨token⟩ is a control sequence.

\token_if_expandable_p:N * \token_if_expandable_p:N ⟨token⟩
\token_if_expandable:NTF * \token_if_expandable:NTF ⟨token⟩ {⟨true code⟩} {⟨false code⟩}

Tests if the ⟨token⟩ is expandable. This test returns ⟨false⟩ for an undefined token.

\token_if_long_macro_p:N * \token_if_long_macro_p:N ⟨token⟩
\token_if_long_macro:NTF * \token_if_long_macro:NTF ⟨token⟩ {⟨true code⟩} {⟨false code⟩}

Tests if the ⟨token⟩ is a long macro.

\token_if_protected_macro_p:N * \token_if_protected_macro_p:N ⟨token⟩
\token_if_protected_macro:NTF * \token_if_protected_macro:NTF ⟨token⟩ {⟨true code⟩} {⟨false code⟩}

Tests if the ⟨token⟩ is a protected macro: for a macro which is both protected and long this returns false.

\token_if_protected_long_macro_p:N * \token_if_protected_long_macro_p:N ⟨token⟩
\token_if_protected_long_macro:NTF * \token_if_protected_long_macro:NTF ⟨token⟩ {⟨true code⟩} {⟨false code⟩}

Tests if the ⟨token⟩ is a protected long macro.

\token_if_chardef_p:N * \token_if_chardef_p:N ⟨token⟩
\token_if_chardef:NTF * \token_if_chardef:NTF ⟨token⟩ {⟨true code⟩} {⟨false code⟩}

Tests if the ⟨token⟩ is defined to be a chardef.

\TeXhackers note: Booleans, boxes and small integer constants are implemented as \chardefs.

\token_if_mathchardef_p:N * \token_if_mathchardef_p:N ⟨token⟩
\token_if_mathchardef:NTF * \token_if_mathchardef:NTF ⟨token⟩ {⟨true code⟩} {⟨false code⟩}

Tests if the ⟨token⟩ is defined to be a mathchardef.
Tests if the \langle token \rangle is defined to be a font selection command.

Tests if the \langle token \rangle is defined to be a dimension register.

Tests if the \langle token \rangle is defined to be an integer register.

\TeX hackers note: Constant integers may be implemented as integer registers, \texttt{\chardef}s, or \texttt{\mathchardef}s depending on their value.

Tests if the \langle token \rangle is defined to be a muskip register.

Tests if the \langle token \rangle is defined to be a skip register.

Tests if the \langle token \rangle is defined to be a toks register (not used by \LaTeX{}3).

Tests if the \langle token \rangle is an engine primitive. In Lua\TeX{} this includes primitive-like commands defined using \texttt{\token.set_lua}.
This function compares the \textit{test token} in turn with each of the \textit{token cases}. If the two are equal (as described for \token_if_eq_catcode:NNTF, \token_if_eq_charcode:NNTF and \token_if_eq_meaning:NNTF, respectively) then the associated \textit{code} is left in the input stream and other cases are discarded. If any of the cases are matched, the \textit{true code} is also inserted into the input stream (after the code for the appropriate case), while if none match then the \textit{false code} is inserted. The functions \token_case_catcode:Nn, \token_case_charcode:Nn, and \token_case_meaning:Nn, which do nothing if there is no match, are also available.

### 23.6 Peeking ahead at the next token

There is often a need to look ahead at the next token in the input stream while leaving it in place. This is handled using the “peek” functions. The generic \peek_after:Nw is provided along with a family of predefined tests for common cases. As peeking ahead does not skip spaces the predefined tests include both a space-respecting and space-skipping version. In addition, using \peek_analysis_map_inline:n, one can map through the following tokens in the input stream and repeatedly perform some tests.

\begin{verbatim}
\peek_after:Nw \peek_after:Nw \peek_gafter:Nw
\end{verbatim}

Locally sets the test variable \l_peek_token equal to \textit{token} (as an implicit token, not as a token list), and then expands the \textit{function}. The \textit{token} remains in the input stream as the next item after the \textit{function}. The \textit{token} here may be \l, \{ or \} (assuming normal \TeX category codes), i.e. it is not necessarily the next argument which would be grabbed by a normal function.

Globally sets the test variable \g_peek_token equal to \textit{token} (as an implicit token, not as a token list), and then expands the \textit{function}. The \textit{token} remains in the input stream as the next item after the \textit{function}. The \textit{token} here may be \l, \{ or \} (assuming normal \TeX category codes), i.e. it is not necessarily the next argument which would be grabbed by a normal function.

Token set by \peek_after:Nw and available for testing as described above.

Token set by \peek_gafter:Nw and available for testing as described above.
\peek_catcode:NTF (test token) \{\text{true code}\} \{\text{false code}\}

Tests if the next \textit{token} in the input stream has the same category code as the \textit{test token} (as defined by the test \texttt{\token_if_eq_catcode:NTF}). Spaces are respected by the test and the \textit{token} is left in the input stream after the \textit{true code} or \textit{false code} (as appropriate to the result of the test).

\peek_catcode_remove:NTF (test token) \{\text{true code}\} \{\text{false code}\}

Tests if the next \textit{token} in the input stream has the same category code as the \textit{test token} (as defined by the test \texttt{\token_if_eq_catcode:NTF}). Spaces are respected by the test and the \textit{token} is removed from the input stream if the test is true. The function then places either the \textit{true code} or \textit{false code} in the input stream (as appropriate to the result of the test).

\peek_charcode:NTF (test token) \{\text{true code}\} \{\text{false code}\}

Tests if the next \textit{token} in the input stream has the same character code as the \textit{test token} (as defined by the test \texttt{\token_if_eq_charcode:NTF}). Spaces are respected by the test and the \textit{token} is left in the input stream after the \textit{true code} or \textit{false code} (as appropriate to the result of the test).

\peek_charcode_remove:NTF (test token) \{\text{true code}\} \{\text{false code}\}

Tests if the next \textit{token} in the input stream has the same character code as the \textit{test token} (as defined by the test \texttt{\token_if_eq_charcode:NTF}). Spaces are respected by the test and the \textit{token} is removed from the input stream if the test is true. The function then places either the \textit{true code} or \textit{false code} in the input stream (as appropriate to the result of the test).

\peek_meaning:NTF (test token) \{\text{true code}\} \{\text{false code}\}

Tests if the next \textit{token} in the input stream has the same meaning as the \textit{test token} (as defined by the test \texttt{\token_if_eq_meaning:NTF}). Spaces are respected by the test and the \textit{token} is left in the input stream after the \textit{true code} or \textit{false code} (as appropriate to the result of the test).

\peek_meaning_remove:NTF (test token) \{\text{true code}\} \{\text{false code}\}

Tests if the next \textit{token} in the input stream has the same meaning as the \textit{test token} (as defined by the test \texttt{\token_if_eq_meaning:NTF}). Spaces are respected by the test and the \textit{token} is removed from the input stream if the test is true. The function then places either the \textit{true code} or \textit{false code} in the input stream (as appropriate to the result of the test).

\peek_remove_spaces:n \{\text{code}\}

Peeks ahead and detects if the following token is a space (category code 10 and character code 32). If so, removes the token and checks the next token. Once a non-space token is found, the \textit{code} will be inserted into the input stream. Typically this will contain a \peek operation, but this is not required.

Updated: 2012-12-20

Updated: 2011-07-02

Updated: 2018-10-01
\peek_remove_filler:n\peek_remove_filler:n \{\langle code\rangle\}

Peeks ahead and detect if the following token is a space (category code 10) or has meaning equal to \scan_stop:. If so, removes the token and checks the next token. If neither of these cases apply, expands the next token using \_type expansion, then checks the resulting leading token in the same way. If after expansion the next token is neither of the two test cases, the \langle code\rangle will be inserted into the input stream. Typically this will contain a \peek operation, but this is not required.

\TeXhackers note: This is essentially a macro-based implementation of how \TeX\ handles the search for a left brace after for example \everypar, except that any non-expandable token cleanly ends the \langle filler\rangle (i.e. it does not lead to a \TeX error).

In contrast to \TeX's filler removal, a construct \exp_not:N \foo will be treated in the same way as \foo.

\peek_N_type:TF\peek_N_type:TF \{\langle true\ code\rangle\} \{\langle false\ code\rangle\}

Tests if the next \langle token\rangle in the input stream can be safely grabbed as an \texttt{N}-type argument. The test is \langle false\rangle if the next \langle token\rangle is either an explicit or implicit begin-group or end-group token (with any character code), or an explicit or implicit space character (with character code 32 and category code 10), or an outer token (never used in \EMX) and \langle true\rangle in all other cases. Note that a \langle true\rangle result ensures that the next \langle token\rangle is a valid \texttt{N}-type argument. However, if the next \langle token\rangle is for instance \cs_space_token, the test takes the \langle false\rangle branch, even though the next \langle token\rangle is in fact a valid \texttt{N}-type argument. The \langle token\rangle is left in the input stream after the \langle true\ code\rangle or \langle false\ code\rangle (as appropriate to the result of the test).
Repeatably removes one \textit{token} from the input stream and applies the \textit{inline function} to it, until \texttt{	extbackslash peek_analysis_map_break:} is called. The \textit{inline function} receives three arguments for each \textit{token} in the input stream:

- \textit{tokens}, which both o-expand and x-expand to the \textit{token}. The detailed form of \textit{tokens} may change in later releases.
- \textit{char code}, a decimal representation of the character code of the \textit{token}, −1 if it is a control sequence.
- \textit{catcode}, a capital hexadecimal digit which denotes the category code of the \textit{token} (0: control sequence, 1: begin-group, 2: end-group, 3: math shift, 4: alignment tab, 6: parameter, 7: superscript, 8: subscript, A: space, B: letter, C: other, D: active). This can be converted to an integer by writing \texttt{\textit{catcode}}.

These arguments are the same as for \texttt{\textbackslash tl_analysis_map_inline:nn} defined in \texttt{l3tl-analysis}. The \textit{char code} and \textit{catcode} do not take the meaning of a control sequence or active character into account: for instance, upon encountering the token \texttt{\c_group_begin_token} in the input stream, \texttt{\textbackslash peek_analysis_map_inline:n} calls the \textit{inline function} with \texttt{#1} being \texttt{\exp_not:n \{ \c_group_begin_token \}} (with the current implementation), \texttt{#2} being \texttt{-1}, and \texttt{#3} being \texttt{0}, as for any other control sequence. In contrast, upon encountering an explicit begin-group token \texttt{\{}, the \textit{inline function} is called with arguments \texttt{\exp_after:wN \{ \iffalse: \fi: }} \texttt{123} and \texttt{1}.

The mapping is done at the current group level, \textit{i.e.} any local assignments made by the \textit{inline function} remain in effect after the loop. Within the code, \texttt{\l_peek_token} is set equal (as a token, not a token list) to the token under consideration.

Tests if the \textit{tokens} that follow in the input stream match the \textit{regular expression}. Any \textit{tokens} that have been read are left in the input stream after the \textit{true code} or \textit{false code} (as appropriate to the result of the test). See \texttt{l3regex} for documentation of the syntax of regular expressions. The \textit{regular expression} is implicitly anchored at the start, so for instance \texttt{\peek_regex:nTF \{ \regex \} \{ \true_code \} \{ \false_code \}} will only inspect the first token \texttt{a} even though the first branch \texttt{abc} of the alternative is preferred in functions such as \texttt{\peek_regex_remove_once:n}. This may have an effect on tokenization if the input stream has not yet been tokenized and category codes are changed.
Tests if the \langle tokens\rangle that follow in the input stream match the \langle regex\rangle. If the test is true, the \langle tokens\rangle are removed from the input stream and the \langle true code\rangle is inserted, while if the test is false, the \langle false code\rangle is inserted followed by the \langle tokens\rangle that were originally in the input stream. See l3regex for documentation of the syntax of regular expressions. The \langle regular expression\rangle is implicitly anchored at the start, so for instance \texttt{\textbackslash peek_regex_remove_once:nTF \{ a \}} is essentially equivalent to \texttt{\textbackslash peek_charcode_remove:NTF a}.

\textbf{TeXhackers note:} Implicit character tokens are correctly considered by \texttt{\textbackslash peek_regex_remove_once:nTF} as control sequences, while functions that inspect individual tokens (for instance \texttt{\textbackslash peek_charcode:NTF}) only take into account their meaning.

If the \langle tokens\rangle that follow in the input stream match the \langle regex\rangle, replaces them according to the \langle replacement\rangle as for \texttt{\textbackslash regex_replace_once:nnN}, and leaves the result in the input stream, after the \langle true code\rangle. Otherwise, leaves \langle false code\rangle followed by the \langle tokens\rangle that were originally in the input stream, with no modifications. See l3regex for documentation of the syntax of regular expressions and of the \langle replacement\rangle: for instance \texttt{\textbackslash 0} in the \langle replacement\rangle is replaced by the tokens that were matched in the input stream. The \langle regular expression\rangle is implicitly anchored at the start. In contrast to \texttt{\textbackslash regex_replace_once:nnN}, no error arises if the \langle replacement\rangle leads to an unbalanced token list: the tokens are inserted into the input stream without issue.

\textbf{TeXhackers note:} Implicit character tokens are correctly considered by \texttt{\textbackslash peek_regex_replace_once:nnTF} as control sequences, while functions that inspect individual tokens (for instance \texttt{\textbackslash peek_charcode:NTF}) only take into account their meaning.

### 23.7 Description of all possible tokens

Let us end by reviewing every case that a given token can fall into. This section is quite technical and some details are only meant for completeness. We distinguish the meaning of the token, which controls the expansion of the token and its effect on \TeX’s state, and its shape, which is used when comparing token lists such as for delimited arguments. Two tokens of the same shape must have the same meaning, but the converse does not hold.

A token has one of the following shapes.

- A control sequence, characterized by the sequence of characters that constitute its name: for instance, \texttt{\textbackslash use:n} is a five-letter control sequence.
• An active character token, characterized by its character code (between 0 and 1114111 for LuaTeX and XeTeX and less for other engines) and category code 13.

• A character token, characterized by its character code and category code (one of 1, 2, 3, 4, 6, 7, 8, 10, 11 or 12 whose meaning is described below).

There are also a few internal tokens. The following list may be incomplete in some engines.

• Expanding \textfont results in a token that looks identical to the command that was used to select the current font (such as \textit{em}) but it differs from it in shape.

• A “frozen” \relax, which differs from the primitive in shape (but has the same meaning), is inserted when the closing \fi of a conditional is encountered before the conditional is evaluated.

• Expanding \noexpand \langle \text{token} \rangle (when the \langle \text{token} \rangle is expandable) results in an internal token, displayed (temporarily) as \notexpanded: \langle \text{token} \rangle, whose shape coincides with the \langle \text{token} \rangle and whose meaning differs from \relax.

• An \outerendtemplate: can be encountered when peeking ahead at the next token; this expands to another internal token, end of alignment template.

• Tricky programming might access a frozen \endwrite.

• Some frozen tokens can only be accessed in interactive sessions: \cr, \right, \endgroup, \fi, \inaccessible.

• In LuaTeX, there is also the strange case of “bytes” \text{1100xy} where x, y are any two lowercase hexadecimal digits, so that the hexadecimal number ranges from \text{110000}=1114112 to \text{1100ff}=1114367. These are used to output individual bytes to files, rather than UTF-8. For the purposes of token comparisons they behave like non-expandable primitive control sequences (\textit{not characters}) whose \meaning is the \textit{character}, followed by the given byte. If this byte is in the range \text{80–ff} this gives an “invalid utf-8 sequence” error: applying \token_to_str:N or \token_to_meaning:N to these tokens is unsafe. Unfortunately, they don’t seem to be detectable safely by any means except perhaps Lua code.

The meaning of a (non-active) character token is fixed by its category code (and character code) and cannot be changed. We call these tokens \textit{explicit} character tokens. Category codes that a character token can have are listed below by giving a sample output of the \TeX\ primitive \meaning, together with their \LaTeX\ names and most common example:

1 begin-group character (\textit{group_begin}, often \{),

2 end-group character (\textit{group_end}, often \}),

3 math shift character (\textit{math_toggle}, often \$),

4 alignment tab character (\textit{alignment}, often \&),

6 macro parameter character (\textit{parameter}, often \#),

7 superscript character (\textit{math_superscript}, often \^),
8 subscript character (\texttt{math\_subscript}, often \_),
10 blank space (\texttt{space}, often character code 32),
11 the letter (\texttt{letter}, such as A),
12 the character (\texttt{other}, such as 0).

Category code 13 (\texttt{active}) is discussed below. Input characters can also have several other category codes which do not lead to character tokens for later processing: 0 (\texttt{escape}), 5 (\texttt{end\_line}), 9 (\texttt{ignore}), 14 (\texttt{comment}), and 15 (\texttt{invalid}).

The meaning of a control sequence or active character can be identical to that of any character token listed above (with any character code), and we call such tokens \textit{implicit} character tokens. The meaning is otherwise in the following list:

- a macro, used in \LaTeX3 for most functions and some variables (\texttt{tl}, \texttt{fp}, \texttt{seq}, ...),
- a primitive such as \texttt{\def} or \texttt{\topmark}, used in \LaTeX3 for some functions,
- a register such as \texttt{\count123}, used in \LaTeX3 for the implementation of some variables (\texttt{int}, \texttt{dim}, ...),
- a constant integer such as \texttt{\char"56} or \texttt{\mathchar"121},
- a font selection command,
- undefined.

Macros can be \texttt{\protected} or not, \texttt{\long} or not (the opposite of what \LaTeX3 calls \texttt{nopar}), and \texttt{\outer} or not (unused in \LaTeX3). Their \texttt{meaning} takes the form

\begin{verbatim}
(prefix) macro:(argument)->(replacement)
\end{verbatim}

where \texttt{(prefix)} is among \texttt{\protected\long\outer}, \texttt{(argument)} describes parameters that the macro expects, such as \texttt{#1#2#3}, and \texttt{(replacement)} describes how the parameters are manipulated, such as \texttt{\int\_eval\_n{#2+#1*#3}}.

Now is perhaps a good time to mention some subtleties relating to tokens with category code 10 (\texttt{space}). Any input character with this category code (normally, space and tab characters) becomes a normal space, with character code 32 and category code 10.

When a macro takes an undelimited argument, explicit space characters (with character code 32 and category code 10) are ignored. If the following token is an explicit character token with category code 1 (\texttt{begin\_group}) and an arbitrary character code, then \TeX scans ahead to obtain an equal number of explicit character tokens with category code 1 (\texttt{begin\_group}) and 2 (\texttt{end\_group}), and the resulting list of tokens (with outer braces removed) becomes the argument. Otherwise, a single token is taken as the argument for the macro: we call such single tokens “N-type”, as they are suitable to be used as an argument for a function with the signature :\texttt{N}.

When a macro takes a delimited argument \TeX scans ahead until finding the delimiter (outside any pairs of begin-group/end-group explicit characters), and the resulting list of tokens (with outer braces removed) becomes the argument. Note that explicit space characters at the start of the argument are \textit{not} ignored in this case (and they prevent brace-stripping).
Chapter 24

The l3prop package

Property lists

\texttt{l3prop} implements a “property list” data type, which contain an unordered list of entries each of which consists of a \texttt{(key)} and an associated \texttt{(value)}. The \texttt{(key)} and \texttt{(value)} may both be any \texttt{(balanced text)}, the \texttt{(key)} is processed using \texttt{\tl_to_str:n}, meaning that category codes are ignored. It is possible to map functions to property lists such that the function is applied to every key–value pair within the list.

Each entry in a property list must have a unique \texttt{(key)}: if an entry is added to a property list which already contains the \texttt{(key)} then the new entry overwrites the existing one. The \texttt{(keys)} are compared on a string basis, using the same method as \texttt{\str_if_-eq:nn}.

Property lists are intended for storing key-based information for use within code. This is in contrast to key–value lists, which are a form of \textit{input} parsed by the \texttt{l3keys} module.

24.1 Creating and initialising property lists

\begin{itemize}
  \item \texttt{\prop_new:N} \texttt{(property list)}
  \item \texttt{\prop_new:c} \texttt{(property list)}
  \item \texttt{\prop_clear:N} \texttt{(property list)}
  \item \texttt{\prop_clear:c} \texttt{(property list)}
  \item \texttt{\prop_gclear:N} \texttt{(property list)}
  \item \texttt{\prop_gclear:c} \texttt{(property list)}
\end{itemize}

\texttt{\prop_new:N} \texttt{(property list)}

Creates a new \texttt{(property list)} or raises an error if the name is already taken. The declaration is global. The \texttt{(property list)} initially contains no entries.

\texttt{\prop_clear:N} \texttt{(property list)}

Clears all entries from the \texttt{(property list)}.

\texttt{\prop_gclear:N} \texttt{(property list)}

Ensures that the \texttt{(property list)} exists globally by applying \texttt{\prop_new:N} if necessary, then applies \texttt{\prop_(g)clear:N} to leave the list empty.
\prop_set_eq:NN \langle property list \rangle \langle property list \rangle
Sets the content of \langle property list \rangle equal to that of \langle property list \rangle.

\prop_set_eq:NN \prop_gset_eq:NN \langle \langle cN | Nc | cc \rangle \rangle
\prop_set_from_keyval:Nn \langle prop var \rangle \{ \langle \langle key1 \rangle \rangle = \langle \langle value1 \rangle \rangle , \langle \langle key2 \rangle \rangle = \langle \langle value2 \rangle \rangle , \ldots \}
Sets \langle prop var \rangle to contain key–value pairs given in the second argument. If duplicate keys appear only the last of the values is kept.

Spaces are trimmed around every \langle key \rangle and every \langle value \rangle, and if the result of trimming spaces consists of a single brace group then a set of outer braces is removed. This enables both the \langle key \rangle and the \langle value \rangle to contain spaces, commas or equal signs. The \langle key \rangle is then processed by \tl_to_str:n. This function correctly detects the = and , signs provided they have the standard category code 12 or they are active.

\prop_set_from_keyval:Nn \prop_gset_from_keyval:Nn \prop_const_from_keyval:Nn \prop_const_from_keyval:cn
New: 2017-11-28
Updated: 2021-11-07
\prop_const_from_keyval:Nn \prop_const_from_keyval:cn
Creates a new constant \langle prop var \rangle or raises an error if the name is already taken. The \langle prop var \rangle is set globally to contain key–value pairs given in the second argument, processed in the way described for \prop_set_from_keyval:Nn. If duplicate keys appear only the last of the values is kept. This function correctly detects the = and , signs provided they have the standard category code 12 or they are active.

Notice that in contrast to most keyval lists (e.g. those in l3keys), each key here must be followed with an = sign.

24.2 Adding and updating property list entries

\prop_put:Nnn \prop_gput:Nnn \langle \langle property list \rangle \rangle \{ \langle \langle key \rangle \rangle \} \{ \langle \langle value \rangle \rangle \}
\prop_gput:Nnn \langle \langle property list \rangle \rangle \{ \langle \langle key \rangle \rangle \} \{ \langle \langle value \rangle \rangle \}
Adds an entry to the \langle property list \rangle which may be accessed using the \langle key \rangle and which has \langle value \rangle. If the \langle key \rangle is already present in the \langle property list \rangle, the existing entry is overwritten by the new \langle value \rangle. Both the \langle key \rangle and \langle value \rangle may contain any \langle balanced text \rangle. The \langle key \rangle is stored after processing with \tl_to_str:n, meaning that category codes are ignored.
\input{content/24.3-Recovering-values-from-property-lists}

24.3 Recovering values from property lists

\begin{verbatim}
\prop_put:NNn \prop_put:Nnn \prop_gput:Nnn
\prop_get:NNn \prop_get:Nnn \prop_pop:NNn
\prop_put_from_keyval:Nn \prop_put_from_keyval:cn
\prop_gput_from_keyval:Nn \prop_gput_from_keyval:cn
\end{verbatim}

If the \textit{key} is present in the \textit{property list} then no action is taken. Otherwise, a new entry is added as described for \prop_put:Nnn.

\begin{verbatim}
\prop_put_if_new:NNn \prop_put_if_new:cnn \prop_gput_if_new:NNn \prop_gput_if_new:cnn
\end{verbatim}

If the \textit{key} is present in the \textit{property list} then no action is taken. Otherwise, a new entry is added as described for \prop_put:Nnn.

\begin{verbatim}
\prop_concat:NNN \prop_concat:ccc \prop_gconcat:NNN \prop_gconcat:ccc
\end{verbatim}

 Combines the key–value pairs of \textit{prop var\_2} and \textit{prop var\_3}, and saves the result in \textit{prop var\_1}. If a key appears in both \textit{prop var\_2} and \textit{prop var\_3} then the last value, namely the value in \textit{prop var\_3} is kept.

\begin{verbatim}
\prop_put_from_keyval:Nn \prop_put_from_keyval:cn \prop_gput_from_keyval:Nn \prop_gput_from_keyval:cn
\end{verbatim}

Updated: 2021-11-07

Recover the \textit{value} stored with \textit{key} from the \textit{property list}, and places this in the \textit{token list variable}. If the \textit{key} is not found in the \textit{property list} then the \textit{token list variable} is set to the special marker \texttt{\string q\_no\_value}. The \textit{token list variable} is set within the current \TeX{} group. See also \prop_get:NnNTF.

\begin{verbatim}
\prop_get:NNn \prop_get:Nnn \prop_pop:NNn \prop_pop:Nnn
\end{verbatim}

Recover the \textit{value} stored with \textit{key} from the \textit{property list}, and places this in the \textit{token list variable}. If the \textit{key} is not found in the \textit{property list} then the \textit{token list variable} is set to the special marker \texttt{\string q\_no\_value}. The \textit{key} and \textit{value} are then deleted from the property list. Both assignments are local. See also \prop_pop:NnNTF.
\prop_gpop: \{property list\} \{key\} \{tl var\}
Recovers the \{value\} stored with \{key\} from the \{property list\}, and places this in the \{token list variable\}. If the \{key\} is not found in the \{property list\} then the \{token list variable\} is set to the special marker \texttt{\q_no\_value}. The \{key\} and \{value\} are then deleted from the property list. The \{property list\} is modified globally, while the assignment of the \{token list variable\} is local. See also \prop_gpop:NTF.

\prop_item: \{property list\} \{key\}
Expands to the \{value\} corresponding to the \{key\} in the \{property list\}. If the \{key\} is missing, this has an empty expansion.

\TeXhackers\ note: This function is slower than the non-expandable analogue \prop_-get: \{property list\}. The result is returned within the \texttt{\unexpanded} primitive \texttt{\exp_not:n}, which means that the \{value\} does not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.

\prop_count: \{property list\}
Leaves the number of key–value pairs in the \{property list\} in the input stream as an \{integer denotation\}.

\prop_to_keyval: \{property list\}
Expands to the \{property list\} in a key–value notation. Keep in mind that a \{property list\} is unordered, while key–value interfaces don’t necessarily are, so this can’t be used for arbitrary interfaces.

\TeXhackers\ note: The result is returned within the \texttt{\unexpanded} primitive \texttt{\exp_not:n}, which means that the key–value list does not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion. It also needs exactly two steps of expansion.

24.4 Modifying property lists

\prop_remove: \{property list\} \{key\}
Removes the entry listed under \{key\} from the \{property list\}. If the \{key\} is not found in the \{property list\} no change occurs, \textit{i.e} there is no need to test for the existence of a key before deleting it.

24.5 Property list conditionals

\prop_if_exist_p: \{property list\}
\prop_if_exist:NTF \{property list\} \{true code\} \{false code\}
Tests whether the \{property list\} is currently defined. This does not check that the \{property list\} really is a property list variable.
Tests if the \(\textit{property list}\) is empty (containing no entries).

\textbf{T\kern-1pt\TeX\kern-0.6pt Xhackers note:} This function iterates through every key–value pair in the \(\textit{property list}\) and is therefore slower than using the non-expandable \propget{\textit{property list}}{\textit{key}}{\textit{true code}}{\textit{false code}}.

## 24.6 Recovering values from property lists with branching

The functions in this section combine tests for the presence of a key in a property list with recovery of the associated value. This makes them useful for cases where different cases follow dependent on the presence or absence of a key in a property list. They offer increased readability and performance over separate testing and recovery phases.

If the \(\textit{key}\) is not present in the \(\textit{property list}\), leaves the \(\textit{false code}\) in the input stream. The value of the \(\textit{token list variable}\) is not defined in this case and should not be relied upon. If the \(\textit{key}\) is present in the \(\textit{property list}\), stores the corresponding \(\textit{value}\) in the \(\textit{token list variable}\) without removing it from the \(\textit{property list}\), then leaves the \(\textit{true code}\) in the input stream. The \(\textit{token list variable}\) is assigned locally.
If the \texttt{\langle key\rangle} is not present in the \texttt{\langle property list\rangle}, leaves the \texttt{\langle false code\rangle} in the input stream. The value of the \texttt{\langle token list variable\rangle} is not defined in this case and should not be relied upon. If the \texttt{\langle key\rangle} is present in the \texttt{\langle property list\rangle}, pops the corresponding \texttt{\langle value\rangle} in the \texttt{\langle token list variable\rangle}, i.e. removes the item from the \texttt{\langle property list\rangle}. The \texttt{\langle property list\rangle} is modified globally, while the \texttt{\langle token list variable\rangle} is assigned locally.

### 24.7 Mapping over property lists

All mappings are done at the current group level, \textit{i.e.} any local assignments made by the \texttt{\langle function\rangle} or \texttt{\langle code\rangle} discussed below remain in effect after the loop.

\begin{verbatim}
\prop_map_function:NN \prop_map_function:CN
\end{verbatim}

Applies \texttt{\langle function\rangle} to every \texttt{\langle entry\rangle} stored in the \texttt{\langle property list\rangle}. The \texttt{\langle function\rangle} receives two arguments for each iteration: the \texttt{\langle key\rangle} and associated \texttt{\langle value\rangle}. The order in which \texttt{\langle entries\rangle} are returned is not defined and should not be relied upon. To pass further arguments to the \texttt{\langle function\rangle}, see \texttt{\prop_map_tokens:Nn}.

\begin{verbatim}
\prop_map_inline:NN \prop_map_inline:CN
\end{verbatim}

Applies \texttt{\langle inline function\rangle} to every \texttt{\langle entry\rangle} stored within the \texttt{\langle property list\rangle}. The \texttt{\langle inline function\rangle} should consist of code which receives the \texttt{\langle key\rangle} as \#1 and the \texttt{\langle value\rangle} as \#2. The order in which \texttt{\langle entries\rangle} are returned is not defined and should not be relied upon.

\begin{verbatim}
\prop_map_tokens:NN \prop_map_tokens:CN
\end{verbatim}

Analogue of \texttt{\prop_map_function:NN} which maps several tokens instead of a single function. The \texttt{\langle code\rangle} receives each key–value pair in the \texttt{\langle property list\rangle} as two trailing brace groups. For instance,

\begin{verbatim}
\prop_map_tokens:NN \l_my_prop \str_if_eq:nnT \mykey \str_if_eq:nnT \mykey \end{verbatim}

expands to the value corresponding to \texttt{\mykey}: for each pair in \texttt{\l_my_prop} the function \texttt{\str_if_eq:nnT} receives \texttt{\mykey}, the \texttt{\langle key\rangle} and the \texttt{\langle value\rangle} as its three arguments. For that specific task, \texttt{\prop_item:Nn} is faster.
\prop_map_break: \prop_map_break:n

Used to terminate a \prop_map... function before all entries in the \textit{property list} have been processed. This normally takes place within a conditional statement, for example

\prop_map_inline:Nn \l_my_prop
\{
\str_if_eq:nnTF { #1 } { bingo } \{
\prop_map_break: \}
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\prop_log:N \prop_log:c

Writes the entries in the \prop_list in the log file.

24.9 Scratch property lists

\l_tmpa_prop \l_tmpb_prop
New: 2012-06-23

Scratch property lists for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_prop \g_tmpb_prop
New: 2012-06-23

Scratch property lists for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

24.10 Constants

\c_empty_prop

A permanently-empty property list used for internal comparisons.
Chapter 25

The \texttt{l3skip} package
Dimensions and skips

\LaTeX{} provides two general length variables: \texttt{dim} and \texttt{skip}. Lengths stored as \texttt{dim} variables have a fixed length, whereas \texttt{skip} lengths have a rubber (stretch/shrink) component. In addition, the \texttt{muskip} type is available for use in math mode: this is a special form of \texttt{skip} where the lengths involved are determined by the current math font (in \texttt{mu}). There are common features in the creation and setting of length variables, but for clarity the functions are grouped by variable type.

25.1 Creating and initialising \texttt{dim} variables

\begin{itemize}
\item \texttt{\dim_new:N} \texttt{\dim_new:c}
\item \texttt{\dim_const:Nn} \texttt{\dim_const:cn}
\item \texttt{\dim_zero:N} \texttt{\dim_zero:c} \texttt{\dim_gzero:N} \texttt{\dim_gzero:c}
\end{itemize}

New: 2012-03-05

\begin{itemize}
\item \texttt{\dim_new:N \langle\textit{dimension}\rangle} \texttt{\dim_new:c}
\item \texttt{\dim_const:Nn \langle\textit{dimension}\rangle} \texttt{\langle\textit{dimension expression}\rangle} \texttt{\dim_const:cn}
\item \texttt{\dim_zero:N} \texttt{\dim_zero:c} \texttt{\dim_gzero:N} \texttt{\dim_gzero:c}
\end{itemize}

New: 2012-01-07

\begin{itemize}
\item Creates a new \texttt{\langle\textit{dimension}\rangle} or raises an error if the name is already taken. The declaration is globally initialised to 0 pt.
\item Creates a new constant \texttt{\langle\textit{dimension}\rangle} \texttt{\langle\textit{dimension expression}\rangle} or raises an error if the name is already taken. The value of the \texttt{\langle\textit{dimension}\rangle} is set globally to the \texttt{\langle\textit{dimension expression}\rangle}.
\item Sets \texttt{\langle\textit{dimension}\rangle} to 0 pt.
\item Ensures that the \texttt{\langle\textit{dimension}\rangle} exists globally by applying \texttt{\dim_new:N} if necessary, then applies \texttt{\dim\_\{g\}zero:N} to leave the \texttt{\langle\textit{dimension}\rangle} set to zero.
\end{itemize}
Tests whether the \textit{dimension} is currently defined. This does not check that the \textit{dimension} really is a dimension variable.

### 25.2 Setting dim variables

\begin{verbatim}
\dim_add:Nn \dim_add:cn \dim_gadd:Nn \dim_gadd:cn
\end{verbatim}

\texttt{\dim_add:Nn \{dimension\} \{dimension expression\}}

Adds the result of the \textit{dimension expression} to the current content of the \textit{dimension}.

\begin{verbatim}
\dim_set:Nn \dim_set:cn \dim_gset:Nn \dim_gset:cn
\end{verbatim}

\texttt{\dim_set:Nn \{dimension\} \{dimension expression\}}

Sets \textit{dimension} to the value of \textit{dimension expression}, which must evaluate to a length with units.

\begin{verbatim}
\dim_set_eq:NN \dim_set_eq: \{dimension\} \{dimension\}
\dim_gset_eq:NN \dim_gset_eq: \{dimension\} \{dimension\}
\end{verbatim}

\texttt{\dim_set_eq:NN \{dimension\}_1 \{dimension\}_2}

Sets the content of \textit{dimension}_1 equal to that of \textit{dimension}_2.

\begin{verbatim}
\dim_sub:Nn \dim_sub:cn \dim_gsub:Nn \dim_gsub:cn
\end{verbatim}

\texttt{\dim_sub:Nn \{dimension\} \{dimension expression\}}

Subtracts the result of the \textit{dimension expression} from the current content of the \textit{dimension}.

### 25.3 Utilities for dimension calculations

\begin{verbatim}
\dim_abs:n \dim_abs:cn \dim_gadd:n \dim_gadd:cn
\end{verbatim}

\texttt{\dim_abs:n \{dimexpr\}}

Converts the \textit{dimexpr} to its absolute value, leaving the result in the input stream as a \textit{dimension denotation}.

\begin{verbatim}
\dim_max:nn \dim_max: \{dimexpr_1\} \{dimexpr_2\}
\dim_min:nn \dim_min: \{dimexpr_1\} \{dimexpr_2\}
\end{verbatim}

\texttt{\dim_max:nn \{dimexpr_1\} \{dimexpr_2\}} \texttt{\dim_min:nn \{dimexpr_1\} \{dimexpr_2\}}

Evaluates the two \textit{dimension expressions} and leaves either the maximum or minimum value in the input stream as appropriate, as a \textit{dimension denotation}.
\dim_ratio:nn

Parses the two \textit{dimension expressions} and converts the ratio of the two to a form suitable for use inside a \textit{dimension expression}. This ratio is then left in the input stream, allowing syntax such as

\begin{verbatim}
\dim_set:Nn \l_my_dim
  { 10 \text{ pt} \times \dim_ratio:nn { 5 \text{ pt} } { 10 \text{ pt} } }
\end{verbatim}

The output of \texttt{\dim_ratio:nn} on full expansion is a ratio expression between two integers, with all distances converted to scaled points. Thus

\begin{verbatim}
\tl_set:Nx \l_my_tl { \dim_ratio:nn { 5 \text{ pt} } { 10 \text{ pt} } }
\tl_show:N \l_my_tl
\end{verbatim}

displays 327680/655360 on the terminal.

\section*{25.4 Dimension expression conditionals}

\dim_compare_p:nNn
\dim_compare:nNnTF

This function first evaluates each of the \textit{dimension expressions} as described for \texttt{\dim_eval:n}. The two results are then compared using the \textit{relation}:

\begin{verbatim}
  Equal =
  Greater than >
  Less than <
\end{verbatim}

This function is less flexible than \texttt{\dim_compare:nTF} but around 5 times faster.
This function evaluates the \( \text{(dimension expressions)} \) as described for \texttt{\dim_eval:n} and compares consecutive result using the corresponding \( \text{(relations)} \), namely it compares \( \langle \text{dimexpr}_1 \rangle \text{ and } \langle \text{dimexpr}_2 \rangle \) using the \( \langle \text{relation}_1 \rangle \), then \( \langle \text{dimexpr}_2 \rangle \text{ and } \langle \text{dimexpr}_3 \rangle \) using the \( \langle \text{relation}_2 \rangle \), until finally comparing \( \langle \text{dimexpr}_N \rangle \text{ and } \langle \text{dimexpr}_{N+1} \rangle \) using the \( \langle \text{relation}_N \rangle \). The test yields \texttt{true} if all comparisons are \texttt{true}. Each \( \text{(dimension expression)} \) is evaluated only once, and the evaluation is lazy, in the sense that if one comparison is \texttt{false}, then no other \( \text{(dimension expression)} \) is evaluated and no other comparison is performed. The \( \text{(relations)} \) can be any of the following:

- Equal \( = \) or \( == \)
- Greater than or equal to \( >= \)
- Greater than \( > \)
- Less than or equal to \( <= \)
- Less than \( < \)
- Not equal \( != \)

This function is more flexible than \texttt{\dim_compare:nNnTF} but around 5 times slower.
This function evaluates the \textit{(test dimension expression)} and compares this in turn to each of the \textit{(dimension expression cases)}. If the two are equal then the associated \textit{(code)} is left in the input stream and other cases are discarded. If any of the cases are matched, the \textit{(true code)} is also inserted into the input stream (after the code for the appropriate case), while if none match then the \textit{(false code)} is inserted. The function \texttt{\dim_case:nn}, which does nothing if there is no match, is also available. For example

\begin{verbatim}
\dim_set:Nn \l_tmpa_dim { 5 pt }
\dim_case:nnF
 { 2 \l_tmpa_dim }
 { 5 pt } { Small }{ 4 pt + 6 pt } { Medium }
 { - 10 pt } { Negative }
 { No idea! }
\end{verbatim}

leaves “Medium” in the input stream.

### 25.5 Dimension expression loops

\begin{verbatim}
\dim_do_until:nNnn {{dimexpr}} {relation} {{dimexpr}} {\langle code\rangle}
\end{verbatim}

Places the \textit{(code)} in the input stream for \TeX{} to process, and then evaluates the relationship between the two \textit{(dimension expressions)} as described for \texttt{\dim_compare:nNnTF}. If the test is \texttt{false} then the \textit{(code)} is inserted into the input stream again and a loop occurs until the \textit{(relation)} is \texttt{true}.

\begin{verbatim}
\dim_do_while:nNnn {{dimexpr}} {relation} {{dimexpr}} {\langle code\rangle}
\end{verbatim}

Places the \textit{(code)} in the input stream for \TeX{} to process, and then evaluates the relationship between the two \textit{(dimension expressions)} as described for \texttt{\dim_compare:nNnTF}. If the test is \texttt{true} then the \textit{(code)} is inserted into the input stream again and a loop occurs until the \textit{(relation)} is \texttt{false}.

\begin{verbatim}
\dim_until_do:nNnn {{dimexpr}} {relation} {{dimexpr}} {\langle code\rangle}
\end{verbatim}

Evaluates the relationship between the two \textit{(dimension expressions)} as described for \texttt{\dim_compare:nNnTF}, and then places the \textit{(code)} in the input stream if the \textit{(relation)} is \texttt{false}. After the \textit{(code)} has been processed by \TeX{} the test is repeated, and a loop occurs until the test is \texttt{true}.

210
25.6 Dimension step functions

\dim_step_function:nnnN \{initial value\} \{step\} \{final value\} \{function\}

This function first evaluates the \{initial value\}, \{step\} and \{final value\}, all of which should be dimension expressions. The \{function\} is then placed in front of each \{value\} from the \{initial value\} to the \{final value\} in turn (using \{step\} between each \{value\}). The \{step\} must be non-zero. If the \{step\} is positive, the loop stops when the \{value\} becomes larger than the \{final value\}. If the \{step\} is negative, the loop stops when the \{value\} becomes smaller than the \{final value\}. The \{function\} should absorb one argument.

\dim_step_inline:nnn \{initial value\} \{step\} \{final value\} \{code\}

This function first evaluates the \{initial value\}, \{step\} and \{final value\}, all of which should be dimension expressions. Then for each \{value\} from the \{initial value\} to the \{final value\} in turn (using \{step\} between each \{value\}), the \{code\} is inserted into the input stream with \#1 replaced by the current \{value\}. Thus the \{code\} should define a function of one argument (\#1).
This function first evaluates the ⟨initial value⟩, ⟨step⟩ and ⟨final value⟩, all of which should be dimension expressions. Then for each ⟨value⟩ from the ⟨initial value⟩ to the ⟨final value⟩ in turn (using ⟨step⟩ between each ⟨value⟩), the ⟨code⟩ is inserted into the input stream, with the ⟨tl var⟩ defined as the current ⟨value⟩. Thus the ⟨code⟩ should make use of the ⟨tl var⟩.

### 25.7 Using dim expressions and variables

\dim_eval:n \{⟨dimension expression⟩\}

Evaluates the ⟨dimension expression⟩, expanding any dimensions and token list variables within the ⟨expression⟩ to their content (without requiring \dim_use:N/\tl_use:N) and applying the standard mathematical rules. The result of the calculation is left in the input stream as a ⟨dimension denotation⟩ after two expansions. This is expressed in points (pt), and requires suitable termination if used in a \TeX-style assignment as it is not an ⟨internal dimension⟩.

\dim_sign:n \{⟨dimexpr⟩\}

Evaluates the ⟨dimexpr⟩ then leaves 1 or 0 or −1 in the input stream according to the sign of the result.

\dim_use:N \{⟨dimension⟩\}

Recovers the content of a ⟨dimension⟩ and places it directly in the input stream. An error is raised if the variable does not exist or if it is invalid. Can be omitted in places where a ⟨dimension⟩ is required (such as in the argument of \dim_eval:n).

**\TeXhackers note:** \dim_use:N is the \TeX primitive \the: this is one of several \LaTeX names for this primitive.

\dim_to_decimal:n \{⟨dimexpr⟩\}

Evaluates the ⟨dimension expression⟩, and leaves the result, expressed in points (pt) in the input stream, with no units. The result is rounded by \TeX to four or five decimal places. If the decimal part of the result is zero, it is omitted, together with the decimal marker.

For example

\dim_to_decimal:n \{ 1bp \}

leaves 1.00374 in the input stream, i.e. the magnitude of one “big point” when converted to (\TeX) points.
\texttt{\textbackslash dim\_to\_decimal\_in\_bp:n} \ *
\texttt{\textbackslash dim\_to\_decimal\_in\_bp:n \{\textbackslash dimexpr\}}

Evaluates the \textit{(dimension expression)}, and leaves the result, expressed in big points (bp) in the input stream, \textit{with no units}. The result is rounded by \TeX to four or five decimal places. If the decimal part of the result is zero, it is omitted, together with the decimal marker.

For example
\begin{verbatim}
\texttt{\textbackslash dim\_to\_decimal\_in\_bp:n \{ \textnormal{1pt} \}}
\end{verbatim}
leaves 0.99628 in the input stream, \textit{i.e.} the magnitude of one (\TeX) point when converted to big points.

\texttt{\textbackslash dim\_to\_decimal\_in\_sp:n} \ *
\texttt{\textbackslash dim\_to\_decimal\_in\_sp:n \{\textbackslash dimexpr\}}

Evaluates the \textit{(dimension expression)}, and leaves the result, expressed in scaled points (sp) in the input stream, \textit{with no units}. The result is necessarily an integer.

\texttt{\textbackslash dim\_to\_decimal\_in\_unit:nn} \ *
\texttt{\textbackslash dim\_to\_decimal\_in\_unit:nn \{\textbackslash dimexpr\_1\} \{\textbackslash dimexpr\_2\}}

Evaluates the \textit{(dimension expressions)}, and leaves the value of \textit{\{\textbackslash dimexpr\_1\}}, expressed in a unit given by \textit{\{\textbackslash dimexpr\_2\}}, in the input stream. The result is a decimal number, rounded by \TeX to four or five decimal places. If the decimal part of the result is zero, it is omitted, together with the decimal marker.

For example
\begin{verbatim}
\texttt{\textbackslash dim\_to\_decimal\_in\_unit:nn \{ \textnormal{1bp} \} \{ \textnormal{1mm} \}}
\end{verbatim}
leaves 0.35277 in the input stream, \textit{i.e.} the magnitude of one big point when converted to millimetres.

Note that this function is not optimised for any particular output and as such may give different results to \texttt{\textbackslash dim\_to\_decimal\_in\_bp:n} or \texttt{\textbackslash dim\_to\_decimal\_in\_sp:n}. In particular, the latter is able to take a wider range of input values as it is not limited by the ability to calculate a ratio using \texttt{\varepsilon}\TeX primitives, which is required internally by \texttt{\textbackslash dim\_to\_decimal\_in\_unit:nn}.

\texttt{\textbackslash dim\_to\_fp:n} \ *
\texttt{\textbackslash dim\_to\_fp:n \{\textbackslash dimexpr\}}

Expands to an internal floating point number equal to the value of the \textit{\{\textbackslash dimexpr\}} in pt. Since dimension expressions are evaluated much faster than their floating point equivalent, \texttt{\textbackslash dim\_to\_fp:n} can be used to speed up parts of a computation where a low precision and a smaller range are acceptable.

### 25.8 Viewing dim variables

\texttt{\textbackslash dim\_show:N} \ *
\texttt{\textbackslash dim\_show:N \{\textbackslash dimension\}}

Displays the value of the \textit{\{\textbackslash dimension\}} on the terminal.
\dim_show:n \dim_show:n \langle \text{dimension expression} \rangle

Displays the result of evaluating the \langle \text{dimension expression} \rangle on the terminal.

\dim_log:N \dim_log:c \dim_log:n \dim_log:n

\dim_log:N \langle \text{dimension} \rangle

Writes the value of the \langle \text{dimension} \rangle in the log file.

\dim_log:n \langle \text{dimension expression} \rangle

Writes the result of evaluating the \langle \text{dimension expression} \rangle in the log file.

25.9 Constant dimensions

\c_max_dim

The maximum value that can be stored as a dimension. This can also be used as a component of a skip.

\c_zero_dim

A zero length as a dimension. This can also be used as a component of a skip.

25.10 Scratch dimensions

\l_tmpa_dim \l_tmpb_dim

Scratch dimension for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_dim \g_tmpb_dim

Scratch dimension for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

25.11 Creating and initialising skip variables

\skip_new:N \skip_new:c

\skip_new:N \langle \text{skip} \rangle

Creates a new \langle \text{skip} \rangle or raises an error if the name is already taken. The declaration is global. The \langle \text{skip} \rangle is initially equal to 0 pt.
\skip_const:Nn \skip_const:cn
New: 2012-03-05
\skip_const:Nn (\skip expression)

Creates a new constant \langle \skip \rangle or raises an error if the name is already taken. The value of the \langle \skip \rangle is set globally to the \langle \skip expression \rangle.

\skip_zero:N \skip_zero:c
\skip_gzero:N \skip_gzero:c
\skip_zero_new:N \skip_zero_new:c
\skip_gzero_new:N \skip_gzero_new:c
New: 2012-01-07
\skip_zero:N (\skip)
\skip_zero:c
\skip_gzero:N (\skip)
\skip_gzero:c
\skip_zero_new:N (\skip)
\skip_zero_new:c
\skip_gzero_new:N (\skip)
\skip_gzero_new:c
Ensures that the \langle \skip \rangle exists globally by applying \skip_new:N if necessary, then applies \skip_(g)zero:N to leave the \langle \skip \rangle set to zero.

\skip_if_exist_p:N \skip_if_exist_p:c \star
\skip_if_exist:NTF \skip_if_exist:TN \star \skip_if_exist:N (\skip) \star \skip_if_exist:c (\skip)
Updated: 2011-03-03
\skip_if_exist_p:N (\skip)
\skip_if_exist_p:c \star
\skip_if_exist:NTF \skip_if_exist:TN \star
\skip_if_exist:N (\skip) \star \skip_if_exist:c (\skip)

Tests whether the \langle \skip \rangle is currently defined. This does not check that the \langle \skip \rangle really is a skip variable.

25.12 Setting skip variables

\skip_add:Nn \skip_add:cn
\skip_gadd:Nn \skip_gadd:cn
Updated: 2011-10-22
\skip_add:Nn (\skip) \{\langle \skip expression \rangle\}
\skip_gadd:Nn (\skip)
\skip_gadd:cn

Adds the result of the \langle \skip expression \rangle to the current content of the \langle \skip \rangle.

\skip_set:Nn \skip_set:cn
\skip_gset:Nn \skip_gset:cn
Updated: 2011-10-22
\skip_set:Nn (\skip) \{\langle \skip expression \rangle\}
\skip_gset:Nn (\skip)
\skip_gset:cn

Sets \langle \skip \rangle to the value of \langle \skip expression \rangle, which must evaluate to a length with units and may include a rubber component (for example 1 cm plus 0.5 cm).

\skip_set_eq:NN \skip_set_eq:(cN|Nc|cc)
\skip_gset_eq:NN \skip_gset_eq:(cN|Nc|cc)
\skip_set_eq:NN (\skip_1) (\skip_2)
\skip_gset_eq:NN (\skip_1) (\skip_2)

Sets the content of \langle \skip_1 \rangle equal to that of \langle \skip_2 \rangle.

\skip_sub:Nn \skip_sub:cn
\skip_gsub:Nn \skip_gsub:cn
Updated: 2011-10-22
\skip_sub:Nn (\skip) \{\langle \skip expression \rangle\}
\skip_gsub:Nn (\skip)
\skip_gsub:cn

Subtracts the result of the \langle \skip expression \rangle from the current content of the \langle \skip \rangle.
25.13 Skip expression conditionals

\skip_if_eq_p:nn \skip_if_eq:nnTF

\skip_if_eq_p:nn \langle \text{skipexpr}_1 \rangle \{ \text{skipexpr}_2 \}
\skip_if_eq:nnTF \{ \text{skipexpr}_1 \} \{ \text{skipexpr}_2 \}
\{ \langle \text{true code} \rangle \} \{ \langle \text{false code} \rangle \}

This function first evaluates each of the \langle \text{skip expressions} \rangle as described for \texttt{\skip_eval:n}. The two results are then compared for exact equality, \textit{i.e.} both the fixed and rubber components must be the same for the test to be true.

\skip_if_finite_p:n \skip_if_finite:nTF

\skip_if_finite_p:n \langle \text{skipexpr} \rangle
\skip_if_finite:nTF \{ \text{skipexpr} \} \{ \langle \text{true code} \rangle \} \{ \langle \text{false code} \rangle \}

Evaluates the \langle \text{skip expression} \rangle as described for \texttt{\skip_eval:n}, and then tests if all of its components are finite.

25.14 Using skip expressions and variables

\skip_eval:n

\skip_eval:n \langle \text{skip expression} \rangle

Evaluates the \langle \text{skip expression} \rangle, expanding any skips and token list variables within the \langle \text{expression} \rangle to their content (without requiring \texttt{\skip_use:N/\tl_use:N}) and applying the standard mathematical rules. The result of the calculation is left in the input stream as a \langle glue denotation \rangle after two expansions. This is expressed in points (pt), and requires suitable termination if used in a \TeX-style assignment as it is \textit{not} an \langle internal glue \rangle.

\skip_use:N \skip_use:c

\skip_use:N \langle \text{skip} \rangle
\skip_use:c

Recovering the content of a \langle \text{skip} \rangle and places it directly in the input stream. An error is raised if the variable does not exist or if it is invalid. Can be omitted in places where a \langle \text{dimension} \rangle or \langle \text{skip} \rangle is required (such as in the argument of \texttt{\skip_eval:n}).

\TeXhackers note: \texttt{\skip_use:N} is the \TeX primitive \texttt{\the}: this is one of several \LaTeX names for this primitive.

25.15 Viewing skip variables

\skip_show:N \skip_show:c \skip_show:n

\skip_show:N \langle \text{skip} \rangle
\skip_show:c

Displays the value of the \langle \text{skip} \rangle on the terminal.

\skip_show:n \langle \text{skip expression} \rangle

Displays the result of evaluating the \langle \text{skip expression} \rangle on the terminal.
\skip_log:N \skip_log:c
New: 2014-08-22
Updated: 2015-08-03

\skip_log:n
New: 2014-08-22
Updated: 2015-08-07

Writes the value of the \langle \textit{skip} \rangle in the log file.

Writes the result of evaluating the \langle \textit{skip expression} \rangle in the log file.

### 25.16 Constant skips

The maximum value that can be stored as a skip (equal to \texttt{\c_max_dim} in length), with no stretch nor shrink component.

\texttt{\c_zero_skip} Updated: 2012-11-01

A zero length as a skip, with no stretch nor shrink component.

### 25.17 Scratch skips

\texttt{\l_tmpa_skip} \texttt{\l_tmpb_skip}

Scratch skip for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX{}-3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\texttt{\g_tmpa_skip} \texttt{\g_tmpb_skip}

Scratch skip for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX{}-3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

### 25.18 Inserting skips into the output

\texttt{\skip_horizontal:N \skip_horizontal:c}
\texttt{\skip_horizontal:n \{\textit{skipexpr}\}}

Updated: 2011-10-22

Inserts a horizontal \langle \textit{skip} \rangle into the current list. The argument can also be a \langle \texttt{dim} \rangle.

\textit{\TeX} hackers note: \texttt{\skip_horizontal:N} is the \TeX primitive \texttt{\hspace} renamed.
\skip_vertical:N \skip_vertical:c \skip_vertical:n

Updated: 2011-10-22

\skip_vertical:N \skip_vertical:c \skip_vertical:n

Inserts a vertical \choose{skip} into the current list. The argument can also be a \choose{dim}.

\TeXhacksnote{\skip_vertical:N is the \TeX primitive \vskip renamed.}

25.19 Creating and initialising muskip variables

\muskip_new:N \muskip_new:c

Creates a new \choose{muskip} or raises an error if the name is already taken. The declaration
is global. The \choose{muskip} is initially equal to 0 mu.

\muskip_new:N \muskip_new:c

\muskip_const:NN \muskip_const:cn

Creates a new constant \choose{muskip} or raises an error if the name is already taken. The
value of the \choose{muskip} is set globally to the \choose{muskip expression}.

\skip_zero:N \skip_zero:c \skip_zero:n \skip_zero:c

\muskip_zero:N \muskip_zero:c \muskip_gzero:N \muskip_gzero:c

\muskip_zero_new:N \muskip_zero_new:c \muskip_gzero_new:N \muskip_gzero_new:c

\muskip_if_exist_p:N \muskip_if_exist_p:c \muskip_if_exist:NTF \muskip_if_exist:cn

Tests whether the \choose{muskip} is currently defined. This does not check that the \choose{muskip}
really is a muskip variable.

New: 2012-03-03

25.20 Setting muskip variables

\muskip_add:Nn \muskip_add:cn \muskip_gadd:Nn \muskip_gadd:cn

\muskip_add:Nn \muskip_add:cn

Adds the result of the \choose{muskip expression} to the current content of the \choose{muskip}. 218
\texttt{\textbackslash \text{muskip}\_set:Nn} \texttt{(muskip) \{muskip expression\}}

Sets (muskip) to the value of (muskip expression), which must evaluate to a math length with units and may include a rubber component (for example 1 \text{mu} plus 0.5 \text{mu}).

\texttt{\textbackslash \text{muskip}\_set:cn} \texttt{\textbackslash \text{muskip}\_gset:Nn} \texttt{\textbackslash \text{muskip}\_gset:cn}

Updated: 2011-10-22

\texttt{\textbackslash \text{muskip}\_set\_eq:NN} \texttt{(muskip}_1\texttt{) (muskip}_2\texttt{)}

Sets the content of (muskip}_1\texttt{) equal to that of (muskip}_2\texttt{).}

\texttt{\textbackslash \text{muskip}\_sub:Nn} \texttt{(muskip) \{muskip expression\}}

Subtracts the result of the (muskip expression) from the current content of the (muskip).

\texttt{\textbackslash \text{muskip}\_eval:n \star}

\texttt{\textbackslash \text{muskip}\_eval:n \{muskip expression\}}

Evaluates the (muskip expression), expanding any skips and token list variables within the (expression) to their content (without requiring \texttt{\textbackslash \text{muskip}\_use:N/\texttt{\textbackslash tl}\_use:N}) and applying the standard mathematical rules. The result of the calculation is left in the input stream as a (muglue denotation) after two expansions. This is expressed in \text{mu}, and requires suitable termination if used in a \LaTeX-style assignment as it is not an (internal muglue).

\texttt{\textbackslash \text{muskip}\_use:N \star}

\texttt{\textbackslash \text{muskip}\_use:N (muskip)}

Recovers the content of a (skip) and places it directly in the input stream. An error is raised if the variable does not exist or if it is invalid. Can be omitted in places where a (dimension) is required (such as in the argument of \texttt{\textbackslash \text{muskip}\_eval:n}).

\textbf{\LaTeXhacksnote:} \texttt{\textbackslash \text{muskip}\_use:N} is the \LaTeX primitive \texttt{\textbackslash \the:} this is one of several \LaTeXX names for this primitive.

\texttt{\textbackslash \text{muskip}\_show:N \star}

\texttt{\textbackslash \text{muskip}\_show:N (muskip)}

Displays the value of the (muskip) on the terminal.

\section{Using muskip expressions and variables}

\texttt{\textbackslash \text{muskip}\_set_eq:NN} \texttt{(muskip}_1\texttt{) (muskip}_2\texttt{)}

\section{Viewing muskip variables}

Updated: 2015-08-03
25.23 Constant muskips

\c_max_muskip

The maximum value that can be stored as a muskip, with no stretch nor shrink component.

\c_zero_muskip

A zero length as a muskip, with no stretch nor shrink component.

25.24 Scratch muskips

\l_tmpa_muskip \l_tmpb_muskip

Scratch muskip for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_muskip \g_tmpb_muskip

Scratch muskip for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

25.25 Primitive conditional

\if_dim:w \star
\if_dim:w ⟨dimen⟩ ⟨relation⟩ ⟨dimen⟩
⟨true code⟩
\else:
⟨false⟩
\fi:

Compare two dimensions. The ⟨relation⟩ is one of <, = or > with category code 12.

\TeXhackers note: This is the \TeX primitive \ifdim.
Chapter 26

The \texttt{l3keys} package

Key–value interfaces

The key–value method is a popular system for creating large numbers of settings for controlling function or package behaviour. The system normally results in input of the form

\begin{verbatim}
\MyModuleSetup{
  key-one = value one,
  key-two = value two
}
\end{verbatim}

or

\begin{verbatim}
\MyModuleMacro[
  key-one = value one,
  key-two = value two
]{argument}
\end{verbatim}

for the user.

The high level functions here are intended as a method to create key–value controls. Keys are themselves created using a key–value interface, minimising the number of functions and arguments required. Each key is created by setting one or more \textit{properties} of the key:

\begin{verbatim}
\keys_define:nn { mymodule }
  {
    key-one .code:n  = code including parameter \#1,
    key-two .tl_set:N = \l_mymodule_store_tl
  }
\end{verbatim}

These values can then be set as with other key–value approaches:

\begin{verbatim}
\keys_set:nn { mymodule }
  {
    key-one = value one,
    key-two = value two
  }
\end{verbatim}
As illustrated, keys are created inside a \textit{(module)}: a set of related keys, typically those for a single \LaTeX \textit{2e} package. See Section for suggestions on how to divide large numbers of keys for a single module.

At a document level, \texttt{\keys_set:nn} is used within a document function, for example

\begin{verbatim}
\DeclareDocumentCommand \MyModuleSetup { m }{ \keys_set:nn { mymodule } { #1 } }
\DeclareDocumentCommand \MyModuleMacro { o m }{\group_begin:
\keys_set:nn { mymodule } { #1 }
% Main code for \MyModuleMacro
\group_end:}
\end{verbatim}

Key names may contain any tokens, as they are handled internally using \texttt{\tl_to_str:n}. As discussed in section 26.2, it is suggested that the character / is reserved for sub-division of keys into logical groups. Functions and variables are not expanded when creating key names, and so

\begin{verbatim}
\tl_set:Nn \l_mymodule_tmp_tl { key }
\keys_define:nn { mymodule }
{ \l_mymodule_tmp_tl.code:n = code }
\end{verbatim}

creates a key called \texttt{\l_mymodule_tmp_tl}, and not one called key.

\section{Creating keys}

\keys_define:nn \keys_define:nn \{\texttt{\module}\} \{\\\texttt{\keyval-list}\}
\keys_define:nn \keys_define:nn \{\texttt{\module}\} \{\\texttt{\keyval-list}\}
Parses the \texttt{\keyval-list} and defines the keys listed there for \texttt{\module}. The \texttt{\module} name is treated as a string. In practice the \texttt{\module} should be chosen to be unique to the module in question (unless deliberately adding keys to an existing module).

The \texttt{\keyval-list} should consist of one or more key names along with an associated key property. The properties of a key determine how it acts. The individual properties are described in the following text; a typical use of \texttt{\keys_define:nn} might read

\begin{verbatim}
\keys_define:nn { mymodule }
{ 
  keyname .code:n = Some-code-using-#1,
  keyname .value_required:n = true
}
\end{verbatim}

where the properties of the key begin from the , after the key name.

The various properties available take either no arguments at all, or require one or more arguments. This is indicated in the name of the property using an argument specification. In the following discussion, each property is illustrated attached to an arbitrary \texttt{\key}, which when used may be supplied with a \texttt{\value}. All key \textit{definitions} are local.
Key properties are applied in the reading order and so the ordering is significant. Key properties which define “actions”, such as `.code:n`, `.tl_set:N`, etc., override one another. Some other properties are mutually exclusive, notably `.value_required:n` and `.value_forbidden:n`, and so they replace one another. However, properties covering non-exclusive behaviours may be given in any order. Thus for example the following definitions are equivalent.

\keys_define:nn { mymodule }
{  
  keyname .code:n = Some-code-using-#1,
  keyname .value_required:n = true
}
\keys_define:nn { mymodule }
{  
  keyname .value_required:n = true,
  keyname .code:n = Some-code-using-#1
}

Note that with the exception of the special `.undefine:` property, all key properties define the key within the current \TeX{} scope.

\bool_set:N  ⟨key⟩.bool_set:N = ⟨boolean variable⟩
\bool_set:c  ⟨key⟩.bool_set:c
\bool_gset:N  ⟨key⟩.bool_gset:N
\bool_gset:c  ⟨key⟩.bool_gset:c

Defines ⟨key⟩ to set ⟨boolean variable⟩ to ⟨value⟩ (which must be either “true” or “false”). If the variable does not exist, it will be created globally at the point that the key is set up.

\bool_set_inverse:N  ⟨key⟩.bool_set_inverse:N = ⟨boolean variable⟩
\bool_set_inverse:c  ⟨key⟩.bool_set_inverse:c
\bool_gset_inverse:N  ⟨key⟩.bool_gset_inverse:N
\bool_gset_inverse:c  ⟨key⟩.bool_gset_inverse:c

Defines ⟨key⟩ to set ⟨boolean variable⟩ to the logical inverse of ⟨value⟩ (which must be either “true” or “false”). If the ⟨boolean variable⟩ does not exist, it will be created globally at the point that the key is set up.

\choice:  ⟨key⟩.choice:

Sets ⟨key⟩ to act as a choice key. Each valid choice for ⟨key⟩ must then be created, as discussed in section 26.3.

\choices:nn  ⟨key⟩.choices:nn = {⟨choices⟩} ⟨{code}⟩
\choices:(Vn/on/xn)

Sets ⟨key⟩ to act as a choice key, and defines a series ⟨choices⟩ which are implemented using the ⟨code⟩. Inside ⟨code⟩, \l_keys_choice_tl will be the name of the choice made, and \l_keys_choice_int will be the position of the choice in the list of ⟨choices⟩ (indexed from 1). Choices are discussed in detail in section 26.3.

\clist_set:N  ⟨key⟩.clist_set:N = ⟨comma list variable⟩
\clist_set:c  ⟨key⟩.clist_set:c
\clist_gset:N  ⟨key⟩.clist_gset:N
\clist_gset:c  ⟨key⟩.clist_gset:c

Revised: 2011-09-11

Defines ⟨key⟩ to set ⟨comma list variable⟩ to ⟨value⟩. Spaces around commas and empty items will be stripped. If the variable does not exist, it is created globally at the point that the key is set up.
\input{chapters/ch02_keydef}

\begin{quote}
\texttt{\langle key\rangle .code:n = \{\langle code\rangle\} }
\end{quote}

Stores the \langle code\rangle for execution when \langle key\rangle is used. The \langle code\rangle can include one parameter (\#1), which will be the \langle value\rangle given for the \langle key\rangle.

\begin{quote}
\texttt{\langle key\rangle .cs_set:Np = \langle control sequence\rangle \{arg. spec.\} }
\end{quote}

Defines \langle key\rangle to set \langle control sequence\rangle to have \langle arg. spec.\rangle and replacement text \langle value\rangle.

\begin{quote}
\texttt{\langle key\rangle .default:n = \{\langle default\rangle\} }
\end{quote}

Creates a \langle default\rangle value for \langle key\rangle, which is used if no value is given. This will be used if only the key name is given, but not if a blank \langle value\rangle is given:

\begin{verbatim}
\keys_define:nn { mymodule } 
{ 
  key .code:n   = Hello-#1, 
  key .default:n = World
}
\keys_set:nn { mymodule }
{ 
  key = Fred, % Prints 'Hello Fred'
  key,     % Prints 'Hello World'
  key = ,   % Prints 'Hello '
}
\end{verbatim}

The default does not affect keys where values are required or forbidden. Thus a required value cannot be supplied by a default value, and giving a default value for a key which cannot take a value does not trigger an error.

\begin{quote}
\texttt{\langle key\rangle .dim_set:N = \{\langle dimension\rangle\} }
\end{quote}

Defines \langle key\rangle to set \langle dimension\rangle to \langle value\rangle (which must a dimension expression). If the variable does not exist, it is created globally at the point that the key is set up. The key will require a value at point-of-use unless a default is set.

\begin{quote}
\texttt{\langle key\rangle .fp_set:N = \{\langle floating point\rangle\} }
\end{quote}

Defines \langle key\rangle to set \langle floating point\rangle to \langle value\rangle (which must a floating point expression). If the variable does not exist, it is created globally at the point that the key is set up. The key will require a value at point-of-use unless a default is set.

\begin{quote}
\texttt{\langle key\rangle .groups:n = \{\langle groups\rangle\} }
\end{quote}

Defines \langle key\rangle as belonging to the \langle groups\rangle declared. Groups provide a “secondary axis” for selectively setting keys, and are described in Section 26.7.
Specifies that the \langle key \rangle path should inherit the keys listed as any of the \langle parents \rangle (a comma list), which can be a module or a subgroup. For example, after setting
\begin{verbatim}
\keys_define:nn { foo } { test .code:n = \tl_show:n {#1} }
\keys_define:nn { } { bar .inherit:n = foo }
\end{verbatim}
setting
\begin{verbatim}
\keys_set:nn { bar } { test = a }
\end{verbatim}
will be equivalent to
\begin{verbatim}
\keys_set:nn { foo } { test = a }
\end{verbatim}
Inheritance applies at point of use, not at definition, thus keys may be added to the \langle parent \rangle after the use of .\inherit:n and will be active. If more than one \langle parent \rangle is specified, the presence of the \langle key \rangle will be tested for each in turn, with the first successful hit taking priority.

\begin{verbatim}
\keys_define:nn { foo } { test .code:n = \tl_show:n {#1} }
\keys_define:nn { } { bar .inherit:n = foo }
\end{verbatim}

\begin{verbatim}
\keys_set:nn { foo } { test = a }
\end{verbatim}

\begin{verbatim}
\keys_set:nn { bar } { test = a }
\end{verbatim}

Updated: 2013-07-10

\begin{verbatim}
\keys_define:nn { foo } { test .code:n = \tl_show:n {#1} }
\keys_define:nn { } { bar .inherit:n = foo }
\end{verbatim}

Updated: 2020-01-17
Sets \langle key \rangle to act as a multiple choice key. Each valid choice for \langle key \rangle must then be created, as discussed in section 26.3.

\langle key \rangle \cdot \text{multichoices:nn} \{ \langle choices \rangle \} \{ \langle code \rangle \}

Sets \langle key \rangle to act as a multiple choice key, and defines a series \langle choices \rangle which are implemented using the \langle code \rangle. Inside \langle code \rangle, \text{l_keys_choice_tl} will be the name of the choice made, and \text{l_keys_choice_int} will be the position of the choice in the list of \langle choices \rangle (indexed from 1). Choices are discussed in detail in section 26.3.

\langle key \rangle \cdot \text{musrkip_set:N} = \langle muskip \rangle

Defines \langle key \rangle to set \langle muskip \rangle to \langle value \rangle (which must be a muskip expression). If the variable does not exist, it is created globally at the point that the key is set up. The key will require a value at point-of-use unless a default is set.

\langle key \rangle \cdot \text{prop_put:N} = \langle property list \rangle

Defines \langle key \rangle to put the \langle value \rangle onto the \langle property list \rangle stored under the \langle key \rangle. If the variable does not exist, it is created globally at the point that the key is set up.

\langle key \rangle \cdot \text{skip_set:N} = \langle skip \rangle

Defines \langle key \rangle to set \langle skip \rangle to \langle value \rangle (which must be a skip expression). If the variable does not exist, it is created globally at the point that the key is set up. The key will require a value at point-of-use unless a default is set.

\langle key \rangle \cdot \text{str_set:N} = \langle string variable \rangle

Defines \langle key \rangle to set \langle string variable \rangle to \langle value \rangle. If the variable does not exist, it is created globally at the point that the key is set up.

\langle key \rangle \cdot \text{str_set:x:N} = \langle string variable \rangle

Defines \langle key \rangle to set \langle string variable \rangle to \langle value \rangle, which will be subjected to an x-type expansion (i.e. using \text{str_set:Nx}). If the variable does not exist, it is created globally at the point that the key is set up.

\langle key \rangle \cdot \text{tl_set:N} = \langle token list variable \rangle

Defines \langle key \rangle to set \langle token list variable \rangle to \langle value \rangle. If the variable does not exist, it is created globally at the point that the key is set up.
26.2 Sub-dividing keys

When creating large numbers of keys, it may be desirable to divide them into several sub-groups for a given module. This can be achieved either by adding a sub-division to the module name:

\keys_define:nn { mymodule / subgroup }
{ key .code:n = code }

or to the key name:

\keys_define:nn { mymodule }
{ subgroup / key .code:n = code }

As illustrated, the best choice of token for sub-dividing keys in this way is /. This is because of the method that is used to represent keys internally. Both of the above code fragments set the same key, which has full name mymodule/subgroup/key.

As illustrated in the next section, this subdivision is particularly relevant to making multiple choices.

26.3 Choice and multiple choice keys

The l3keys system supports two types of choice key, in which a series of pre-defined input values are linked to varying implementations. Choice keys are usually created so that the various values are mutually-exclusive: only one can apply at any one time. “Multiple” choice keys are also supported: these allow a selection of values to be chosen at the same time.

Mutually-exclusive choices are created by setting the .choice: property:

\keys_define:nn { mymodule }
{ key .choice: }

\keys_define:nn { mymodule / subgroup }
{ key .code:n = code }
For keys which are set up as choices, the valid choices are generated by creating sub-keys of the choice key. This can be carried out in two ways.

In many cases, choices execute similar code which is dependant only on the name of the choice or the position of the choice in the list of all possibilities. Here, the keys can share the same code, and can be rapidly created using the `.choices:nn` property.

\begin{verbatim}
\keys_define:nn { mymodule }
{
  key .choices:nn =
  { choice-a, choice-b, choice-c }
  { You\-gave\-choice-'\\tl_use:N \l_keys_choice_tl',-
    which\-is\-in\-position-\int_use:N \l_keys_choice_int \c_space_tl
    in\-the\-list.
  }
}
\end{verbatim}

The index \l_keys_choice_int in the list of choices starts at 1.

\texttt{l_keys_choice_int} \texttt{l_keys_choice_tl}

Inside the code block for a choice generated using `.choices:nn`, the variables \l_keys_choice_tl and \l_keys_choice_int are available to indicate the name of the current choice, and its position in the comma list. The position is indexed from 1. Note that, as with standard key code generated using `.code:n`, the value passed to the key (i.e. the choice name) is also available as \#1.

On the other hand, it is sometimes useful to create choices which use entirely different code from one another. This can be achieved by setting the `.choice:` property of a key, then manually defining sub-keys.

\begin{verbatim}
\keys_define:nn { mymodule }
{
  key .choice:,
  key / choice-a .code:n = code-a,
  key / choice-b .code:n = code-b,
  key / choice-c .code:n = code-c,
}
\end{verbatim}

It is possible to mix the two methods, but manually-created choices should \textit{not} use \l_keys_choice_tl or \l_keys_choice_int. These variables do not have defined behaviour when used outside of code created using `.choices:nn` (i.e. anything might happen).

It is possible to allow choice keys to take values which have not previously been defined by adding code for the special `unknown` choice. The general behavior of the `unknown` key is described in Section 26.6. A typical example in the case of a choice would be to issue a custom error message:

\begin{verbatim}
\keys_define:nn { mymodule }
{
  key .choice:,
  key / choice-a .code:n = code-a,
  key / choice-b .code:n = code-b,
  key / choice-c .code:n = code-c,
}
\end{verbatim}

228
Multiple choices are created in a very similar manner to mutually-exclusive choices, using the properties `.multichoice:` and `.multichoices:nn`. As with mutually exclusive choices, multiple choices are define as sub-keys. Thus both

\keys_define:nn { mymodule } { 
  key .multichoices:nn = 
  { choice-a, choice-b, choice-c } 
  { You gave choice `	l_use:N \l_keys_choice_tl`, which is in position \int_use:N \l_keys_choice_int \c_space_tl in-the-list. } 
}

and

\keys_define:nn { mymodule } { 
  key .multichoice:, 
  key / choice-a .code:n = code-a, 
  key / choice-b .code:n = code-b, 
  key / choice-c .code:n = code-c, 
}

are valid.

When a multiple choice key is set

\keys_set:nn { mymodule } { 
  key = { a , b , c } % 'key' defined as a multiple choice 
}

each choice is applied in turn, equivalent to a clist mapping or to applying each value individually:

\keys_set:nn { mymodule } { 
  key = a , 
  key = b , 
  key = c , 
}

Thus each separate choice will have passed to it the \l_keys_choice_tl and \l_keys_choice_int in exactly the same way as described for `.choices:nn`. 

229
26.4 Key usage scope

Some keys will be used as settings which have a strictly limited scope of usage. Some
will be only available once, others will only be valid until typesetting begins. To allow
formats to support this in a structured way, \l3keys allows this information to be specified
using the .usage:n property.

\usage:n
\langle key \rangle .usage:n = \langle scope \rangle

Defines the \langle key \rangle to have usage within the \langle scope \rangle, which should be one of general,
preamble or load.

\l_keys_usage_load_prop
\l_keys_usage_preamble_prop

\l3keys itself does not attempt to redefine keys based on the usage scope. Rather, this
information is made available with these two property lists. These hold an entry for each
module (prefix); the value of each entry is a comma-separated list of the usage-restricted
key(s).

26.5 Setting keys

\keys_set:nn
\keys_set:(nVnvno)

Parses the \langle keyval list \rangle, and sets those keys which are defined for \langle module \rangle. The behaviour
on finding an unknown key can be set by defining a special unknown key: this is illustrated
later.

\l_keys_key_str
\l_keys_path_str
\l_keys_value_tl

For each key processed, information of the full path of the key, the name of the key and
the value of the key is available within three token list variables. These may be used
within the code of the key.

The value is everything after the =, which may be empty if no value was given. This
is stored in \l_keys_value_tl, and is not processed in any way by \keys_set:nn.

The path of the key is a “full” description of the key, and is unique for each key. It
consists of the module and full key name, thus for example

\keys_set:nn { mymodule } { key-a = some-value }

has path mymodule/key-a while

\keys_set:nn { mymodule } { subset / key-a = some-value }

has path mymodule/subset/key-a. This information is stored in \l_keys_path_str.

The name of the key is the part of the path after the last /, and thus is not unique.
In the preceding examples, both keys have name key-a despite having different paths.
This information is stored in \l_keys_key_str.
26.6 Handling of unknown keys

If a key has not previously been defined (is unknown), \keys_set:nn looks for a special unknown key for the same module, and if this is not defined raises an error indicating that the key name was unknown. This mechanism can be used for example to issue custom error texts.

\keys_define:nn { mymodule }
{
  unknown .code:n =
    You-tried-to-set-key-'\l_keys_key_str'-to-'#1'.
}

These functions set keys which are known for the ⟨module⟩, and simply ignore other keys. The \keys_set_known:nn function parses the ⟨keyval list⟩, and sets those keys which are defined for ⟨module⟩. Any keys which are unknown are not processed further by the parser. In addition, \keys_set_known:nnN stores the key–value pairs in the ⟨tl⟩ in comma-separated form (i.e. an edited version of the ⟨keyval list⟩). When a ⟨root⟩ is given (\keys_set_known:nnnN), the key–value entries are returned relative to this point in the key tree. When it is absent, only the key name and value are provided. The correct list is returned by nested calls.

26.7 Selective key setting

In some cases it may be useful to be able to select only some keys for setting, even though these keys have the same path. For example, with a set of keys defined using

\keys_define:nn { mymodule }
{
  key-one .code:n  = \my_func:n {#1} ,
  key-two .tl_set:N = \l_my_a_tl ,
  key-three .tl_set:N = \l_my_b_tl ,
  key-four .fp_set:N = \l_my_a_fp ,
}

the use of \keys_set:nn attempts to set all four keys. However, in some contexts it may only be sensible to set some keys, or to control the order of setting. To do this, keys may be assigned to groups: arbitrary sets which are independent of the key tree. Thus modifying the example to read
\keys define:nn { mymodule }
{
  key-one .code:n = \my_func:n {#1} ,
  key-one .groups:n = { first } ,
  key-two .tl_set:N = \l_my_a_tl ,
  key-two .groups:n = { first } ,
  key-three .tl_set:N = \l_my_b_tl ,
  key-three .groups:n = { second } ,
  key-four .fp_set:N = \l_my_a_fp ,
}

assigns key-one and key-two to group first, key-three to group second, while
key-four is not assigned to a group.

Selective key setting may be achieved either by selecting one or more groups to be
made “active”, or by marking one or more groups to be ignored in key setting.

\keys_set_filter:nnn
\keys_set_filter:nnnN
\keys_set_filter:nnnnN
New: 2013-07-14
Updated: 2019-01-29

Activates key filtering in an “opt-out” sense: keys assigned to any of the ⟨groups⟩ specified
are ignored. The ⟨groups⟩ are given as a comma-separated list. Unknown keys are not
assigned to any group and are thus always set. The key–value pairs for each key which
is filtered out are stored in the ⟨tl⟩ in a comma-separated form (i.e. an edited version of
the ⟨keyval list⟩). The \keys_set_filter:nnn version skips this stage.

Use of \keys_set_filter:nnnN can be nested, with the correct residual ⟨keyval
list⟩ returned at each stage. In the version which takes a ⟨root⟩ argument, the key list is
returned relative to that point in the key tree. In the cases without a ⟨root⟩ argument,
only the key names and values are returned.

\keys_set_groups:nnn
\keys_set_groups:nnnN
New: 2013-07-14
Updated: 2017-05-27

Activates key filtering in an “opt-in” sense: only keys assigned to one or more of the
⟨groups⟩ specified are set. The ⟨groups⟩ are given as a comma-separated list. Unknown
keys are not assigned to any group and are thus never set.

\keys_precompile:nnN
New: 2022-03-09
Updated: 2022-03-09

Parses the ⟨keyval list⟩ as for \keys_set:nn, placing the resulting code for those which
set variables or functions into the ⟨tl⟩. Thus this function “precompiles” the keyval list
into a set of results which can be applied rapidly.

26.8 Digesting keys
26.9 Utility functions for keys

```
\keys_if_exist:nn \{\langle\text{module}\rangle\} \{\langle\text{key}\rangle\}
\keys_if_exist:nnTF \{\langle\text{module}\rangle\} \{\langle\text{key}\rangle\} \{\langle\text{true code}\rangle\} \{\langle\text{false code}\rangle\}
```

Tests if the \langle\text{key}\rangle exists for \langle\text{module}\rangle, i.e. if any code has been defined for \langle\text{key}\rangle.

```
\keys_if_choice_exist:nn \{\langle\text{module}\rangle\} \{\langle\text{key}\rangle\} \{\langle\text{choice}\rangle\}
\keys_if_choice_exist:nnTF \{\langle\text{module}\rangle\} \{\langle\text{key}\rangle\} \{\langle\text{choice}\rangle\} \{\langle\text{true code}\rangle\} \{\langle\text{false code}\rangle\}
```

Tests if the \langle\text{choice}\rangle is defined for the \langle\text{key}\rangle within the \langle\text{module}\rangle, i.e. if any code has been defined for \langle\text{key}\rangle/\langle\text{choice}\rangle. The test is false if the \langle\text{key}\rangle itself is not defined.

```
\keys_show:nn \{\langle\text{module}\rangle\} \{\langle\text{key}\rangle\}
```

Displays in the terminal the information associated to the \langle\text{key}\rangle for a \langle\text{module}\rangle, including the function which is used to actually implement it.

```
\keys_log:nn \{\langle\text{module}\rangle\} \{\langle\text{key}\rangle\}
```

Writes in the log file the information associated to the \langle\text{key}\rangle for a \langle\text{module}\rangle. See also \keys_show:nn which displays the result in the terminal.

26.10 Low-level interface for parsing key–val lists

To re-cap from earlier, a key–value list is input of the form

```
KeyOne = ValueOne , 
KeyTwo = ValueTwo , 
KeyThree
```

where each key–value pair is separated by a comma from the rest of the list, and each
key–value pair does not necessarily contain an equals sign or a value! Processing this type of input correctly requires a number of careful steps, to correctly account for braces, spaces and the category codes of separators.

While the functions described earlier are used as a high-level interface for processing such input, in special circumstances you may wish to use a lower-level approach. The low-level parsing system converts a \langle\text{key–value list}\rangle into \langle\text{keys}\rangle and associated \langle\text{values}\rangle. After the parsing phase is completed, the resulting keys and values (or keys alone) are available for further processing. This processing is not carried out by the low-level parser itself, and so the parser requires the names of two functions along with the key–value list. One function is needed to process key–value pairs (it receives two arguments), and a second function is required for keys given without any value (it is called with a single argument).

The parser does not double # tokens or expand any input. Active tokens = and , appearing at the outer level of braces are converted to category “other” (12) so that the
The \keywordvalparse command does not "miss" any due to category code changes. Spaces are removed from the ends of the keys and values. Keys and values which are given in braces have exactly one set removed (after space trimming), thus

\verbatimverbatim{key = \{value here\},}

and

\verbatimverbatim{key = value here,}

are treated identically.

\verbatimverbatim{\keywordvalparse\{\texttt{code_1}\}\{\texttt{code_2}\}\{\texttt{key-value list}\}}

parses the \texttt{(key-value list)} into a series of \texttt{(keys)} and associated \texttt{(values)}, or keys alone (if no \texttt{(value)} was given). \texttt{(code_1)} receives each \texttt{(key)} (with no \texttt{(value)}) as a trailing brace group, whereas \texttt{(code_2)} is appended by two brace groups, the \texttt{(key)} and \texttt{(value)}. The order of the \texttt{(keys)} in the \texttt{(key-value list)} is preserved. Thus

\verbatimverbatim{\keywordvalparse
\begin{verbatim}
\{ \use_none:nn \{ code 1 \}\}
\{ \use_none:nnn \{ code 2 \}\}
\{ key1 = value1 , key2 = value2 , key3 = , key4 \}
\end{verbatim}
}

is converted into an input stream

\verbatimverbatim{\use_none:nn \{ code 2 \} \{ key1 \} \{ value1 \}}
\verbatimverbatim{\use_none:nn \{ code 2 \} \{ key2 \} \{ value2 \}}
\verbatimverbatim{\use_none:nn \{ code 2 \} \{ key3 \} \{ \} }
\verbatimverbatim{\use_none:nn \{ code 1 \} \{ key4 \}}

Note that there is a difference between an empty value (an equals sign followed by nothing) and a missing value (no equals sign at all). Spaces are trimmed from the ends of the \texttt{(key)} and \texttt{(value)}, then an outer set of braces is removed from the \texttt{(key)} and \texttt{(value)} as part of the processing. If you need exactly the output shown above, you’ll need to either \texttt{x-type} or \texttt{e-type} expand the function.

\TeXhackersnote: The result of each list element is returned within \texttt{\exp_not:n}, which means that the converted input stream does not expand further when appearing in an \texttt{x-type} or \texttt{e-type} argument expansion.
\keyval_parse:NNn \keyval_parse:NNn \{function\} \{function\} \{\langle key–value list\rangle\}

Parses the \langle key–value list\rangle into a series of \langle keys\rangle and associated \langle values\rangle, or keys alone (if no \langle value\rangle was given). \langle function_1\rangle should take one argument, while \langle function_2\rangle should absorb two arguments. After \keyval_parse:NNn has parsed the \langle key–value list\rangle, \langle function_1\rangle is used to process keys given with no value and \langle function_2\rangle is used to process keys given with a value. The order of the \langle keys\rangle in the \langle key–value list\rangle is preserved. Thus

\keyval_parse:NNn \function:n \function:nn
{ key1 = value1 , key2 = value2, key3 = , key4 }

is converted into an input stream

\function:nn { key1 } { value1 }
\function:nn { key2 } { value2 }
\function:nn { key3 } { }
\function:n { key4 }

Note that there is a difference between an empty value (an equals sign followed by nothing) and a missing value (no equals sign at all). Spaces are trimmed from the ends of the \langle key\rangle and \langle value\rangle, then one outer set of braces is removed from the \langle key\rangle and \langle value\rangle as part of the processing.

This shares the implementation of \keyval_parse:nnn, the difference is only semantically.

\TeXhackers note: The result is returned within \exp_not:n, which means that the converted input stream does not expand further when appearing in an x-type or e-type argument expansion.
Chapter 27

The \texttt{l3intarray} package: fast
global integer arrays

27.1 \texttt{l3intarray} documentation

For applications requiring heavy use of integers, this module provides arrays which can
be accessed in constant time (contrast \texttt{l3seq}, where access time is linear). These arrays
have several important features

- The size of the array is fixed and must be given at point of initialisation
- The absolute value of each entry has maximum $2^{30} - 1$ (i.e. one power lower than
  the usual \texttt{c\_max\_int} ceiling of $2^{31} - 1$)

The use of \texttt{intarray} data is therefore recommended for cases where the need for fast
access is of paramount importance.

\begin{verbatim}
\intarray_new:Nn \langle intarray var \rangle \{ \langle size \rangle \}
\end{verbatim}

Evaluates the integer expression \( \langle size \rangle \) and allocates an \langle integer array variable \rangle with that
number of (zero) entries. The variable name should start with \texttt{\g_} because assignments
are always global.

\begin{verbatim}
\intarray_count:N \langle intarray var \rangle
\end{verbatim}

Expands to the number of entries in the \langle integer array variable \rangle. Contrarily to \texttt{\seq\_count:N}
this is performed in constant time.

\begin{verbatim}
\intarray_gset:Nnn \langle intarray var \rangle \{ \langle position \rangle \} \{ \langle value \rangle \}
\end{verbatim}

Stores the result of evaluating the integer expression \( \langle value \rangle \) into the \langle integer array variable \rangle at the (integer expression) \langle position \rangle. If the \langle position \rangle is not between 1 and
the \texttt{\intarray_count:N}, or the \langle value \rangle’s absolute value is bigger than $2^{30} - 1$, an error
occurs. Assignments are always global.

236
\texttt{\intarray\_const\_from\_clist:Nn} \hspace{1em} \texttt{\intarray\_const\_from\_clist:cn}

\begin{itemize}
\item \texttt{\intarray\_const\_from\_clist:Nn} \langle \text{intarray var} \rangle \langle \text{intexpr clist} \rangle
\item \texttt{\intarray\_const\_from\_clist:cn}
\end{itemize}
New: 2018-05-04

Creates a new constant \textit{(integer array variable)} or raises an error if the name is already taken. The \textit{(integer array variable)} is set (globally) to contain as its items the results of evaluating each \textit{(integer expression)} in the \textit{(comma list)}.

\texttt{\intarray\_gzero:N} \hspace{1em} \texttt{\intarray\_gzero:c}

\begin{itemize}
\item \texttt{\intarray\_gzero:N} \langle \text{intarray var} \rangle
\item \texttt{\intarray\_gzero:c}
\end{itemize}
New: 2018-05-04

Sets all entries of the \textit{(integer array variable)} to zero. Assignments are always global.

\texttt{\intarray\_item:Nn} \hspace{1em} \texttt{\intarray\_item:cn}

\begin{itemize}
\item \texttt{\intarray\_item:Nn} \langle \text{intarray var} \rangle \{\langle \text{position} \rangle\}
\item \texttt{\intarray\_item:cn}
\end{itemize}
New: 2018-03-29

Expands to the integer entry stored at the \textit{(integer expression)} \textit{(position)} in the \textit{(integer array variable)}. If the \textit{(position)} is not between 1 and the \texttt{\intarray\_count:N}, an error occurs.

\texttt{\intarray\_rand\_item:N} \hspace{1em} \texttt{\intarray\_rand\_item:c}

\begin{itemize}
\item \texttt{\intarray\_rand\_item:N} \langle \text{intarray var} \rangle
\item \texttt{\intarray\_rand\_item:c}
\end{itemize}
New: 2018-05-05

Selects a pseudo-random item of the \textit{(integer array)}. If the \textit{(integer array)} is empty, produce an error.

\texttt{\intarray\_show:N} \hspace{1em} \texttt{\intarray\_show:c}
\texttt{\intarray\_log:N} \hspace{1em} \texttt{\intarray\_log:c}

\begin{itemize}
\item \texttt{\intarray\_show:N} \langle \text{intarray var} \rangle
\item \texttt{\intarray\_log:N} \langle \text{intarray var} \rangle
\item \texttt{\intarray\_log:c}
\end{itemize}
New: 2018-05-04

Displays the items in the \textit{(integer array variable)} in the terminal or writes them in the log file.

### 27.1.1 Implementation notes

It is a wrapper around the \texttt{\fontdimen} primitive, used to store arrays of integers (with a restricted range: absolute value at most $2^{30} - 1$). In contrast to \texttt{l3seq} sequences the access to individual entries is done in constant time rather than linear time, but only integers can be stored. More precisely, the primitive \texttt{\fontdimen} stores dimensions but the \texttt{l3intarray} package transparently converts these from/to integers. Assignments are always global.

While Lua\TeX{}’s memory is extensible, other engines can “only” deal with a bit less than $4 \times 10^9$ entries in all \texttt{\fontdimen} arrays combined (with default \TeX{} Live settings).
A decimal floating point number is one which is stored as a significand and a separate exponent. The module implements expandably a wide set of arithmetic, trigonometric, and other operations on decimal floating point numbers, to be used within floating point expressions. Floating point expressions support the following operations with their usual precedence.

- Basic arithmetic: addition $x + y$, subtraction $x - y$, multiplication $x * y$, division $x/y$, square root $\sqrt{x}$, and parentheses.
- Comparison operators: $x < y$, $x <= y$, $x > y$, $x! = y$ etc.
- Boolean logic: sign $\text{sign} x$, negation $!x$, conjunction $x && y$, disjunction $x || y$, ternary operator $x?y:z$.
- Exponentials: $\exp x$, $\ln x$, $x^y$, $\log_b x$.
- Integer factorial: $\text{fact} x$.
- Trigonometry: $\sin x$, $\cos x$, $\tan x$, $\cot x$, $\sec x$, $\csc x$ expecting their arguments in radians, and $\text{sind} x$, $\text{cosd} x$, $\text{tand} x$, $\text{cotd} x$, $\text{secd} x$, $\text{csed} x$ expecting their arguments in degrees.
- Inverse trigonometric functions: $\text{asin} x$, $\text{acos} x$, $\text{atan} x$, $\text{acot} x$, $\text{asec} x$, $\text{acsc} x$ giving a result in radians, and $\text{asind} x$, $\text{acosd} x$, $\text{atand} x$, $\text{acotd} x$, $\text{asecd} x$, $\text{acsed} x$ giving a result in degrees.

(not yet) Hyperbolic functions and their inverse functions: $\sinh x$, $\cosh x$, $\tanh x$, $\coth x$, $\sech x$, $\csch$, and $\text{asinh} x$, $\text{acosh} x$, $\text{atanh} x$, $\text{acoth} x$, $\text{asech} x$, $\text{acsch} x$.

- Extrema: $\text{max}(x_1, x_2, \ldots)$, $\text{min}(x_1, x_2, \ldots)$, $\text{abs}(x)$.
- Rounding functions, controlled by two optional values, $n$ (number of places, 0 by default) and $t$ (behavior on a tie, $\text{nan}$ by default):
  - $\text{trunc}(x, n)$ rounds towards zero,
  - $\text{floor}(x, n)$ rounds towards $-\infty$, 

238
ceil($x,n$) rounds towards $+\infty$,
round($x,n,t$) rounds to the closest value, with ties rounded to an even value
by default, towards zero if $t = 0$, towards $+\infty$ if $t > 0$ and towards $-\infty$ if
$t < 0$.

And (not yet) modulo, and “quantize”.

- Random numbers: $\text{rand}()$, $\text{randint}(m,n)$.
- Constants: $\pi$, $\text{deg}$ (one degree in radians).
- Dimensions, automatically expressed in points, e.g., $\text{pc}$ is 12.
- Automatic conversion (no need for \langle type\rangle\_use:N) of integer, dimension, and skip
variables to floating point numbers, expressing dimensions in points and ignoring
the stretch and shrink components of skips.
- Tuples: $(x_1,\ldots,x_n)$ that can be stored in variables, added together, multiplied or
divided by a floating point number, and nested.

Floating point numbers can be given either explicitly (in a form such as $1.234\times10^{-34}$, or
$-0.0001$), or as a stored floating point variable, which is automatically replaced by its
current value. A “floating point” is a floating point number or a tuple thereof. See
section 28.9.1 for a description of what a floating point is, section 28.9.2 for details about
how an expression is parsed, and section 28.9.3 to know what the various operations do.
Some operations may raise exceptions (error messages), described in section 28.7.

An example of use could be the following.

\LaTeX{} can now compute: $\frac{\sin (3.5)}{2} + 2\cdot 10^{-3}$
= \ExplSyntaxOn \fp_to_decimal:n {sin(3.5)/2 + 2e-3} \ExplSyntaxOff.

The operation round can be used to limit the result’s precision. Adding +0 avoids the
possibly undesirable output $-0$, replacing it by $+0$. However, the l3fp module is mostly
meant as an underlying tool for higher-level commands. For example, one could provide
a function to typeset nicely the result of floating point computations.

\documentclass{article}
\usepackage{xparse, siunitx}
\ExplSyntaxOn
\NewDocumentCommand { \calcnum } { m } { \num { \fp_to_scientific:n {#1} } }
\ExplSyntaxOff
\begin{document}
\calcnum { 2 \pi * \sin ( 2.3 ^ 5 ) }
\end{document}

See the documentation of siunitx for various options of \num.
28.1 Creating and initialising floating point variables

\fp_new:N \langle fp\ var \rangle

Creates a new \langle fp\ var \rangle or raises an error if the name is already taken. The declaration is global. The \langle fp\ var \rangle is initially +0.

\fp_const:Nn \langle fp\ var \rangle \{ \langle floating point expression \rangle \}

Creates a new constant \langle fp\ var \rangle or raises an error if the name is already taken. The \langle fp\ var \rangle is set globally equal to the result of evaluating the \langle floating point expression \rangle.

\fp_zero:N \langle fp\ var \rangle

Sets the \langle fp\ var \rangle to +0.

\fp_zero_new:N \langle fp\ var \rangle

Ensures that the \langle fp\ var \rangle exists globally by applying \fp_new:N if necessary, then applies \fp_(g)zero:N to leave the \langle fp\ var \rangle set to +0.

28.2 Setting floating point variables

\fp_set:Nn \langle fp\ var \rangle \{ \langle floating point expression \rangle \}

Sets \langle fp\ var \rangle equal to the result of computing the \langle floating point expression \rangle.

\fp_set_eq:NN \langle fp\ var_1 \rangle \langle fp\ var_2 \rangle

Sets the floating point variable \langle fp\ var_1 \rangle equal to the current value of \langle fp\ var_2 \rangle.

\fp_add:Nn \langle fp\ var \rangle \{ \langle floating point expression \rangle \}

Adds the result of computing the \langle floating point expression \rangle to the \langle fp\ var \rangle. This also applies if \langle fp\ var \rangle and \langle floating point expression \rangle evaluate to tuples of the same size.
\fp_sub:Nn \fp_sub:cn \fp_gsub:Nn \fp_gsub:cn

\fp_sub:Nn \fp_var \{\textit{floating point expression}\}
Subtracts the result of computing the \textit{floating point expression} from the \textit{fp var}. This also applies if \textit{fp var} and \textit{floating point expression} evaluate to tuples of the same size.

\fp_eval:n \fp_eval:cn
\fp_gsub:Nn \fp_gsub:cn

\fp_eval:n \{\textit{floating point expression}\}
Evaluates the \textit{floating point expression} and expresses the result as a decimal number with no exponent. Leading or trailing zeros may be inserted to compensate for the exponent. Non-significant trailing zeros are trimmed, and integers are expressed without a decimal separator. The values $\pm \infty$ and \texttt{nan} trigger an “invalid operation” exception. For a tuple, each item is converted using \fp_eval:n and they are combined as $(\langle fp_1 \rangle, \langle fp_2 \rangle, \ldots, \langle fp_n \rangle)$ if $n > 1$ and $(\langle fp_1 \rangle)$ or $()$ for fewer items. This function is identical to \fp_to_decimal:n.

\fp_sign:n \fp_sign:cn

\fp_sign:n \{\textit{fpexpr}\}
Evaluates the \textit{fpexpr} and leaves its sign in the input stream using \fp_eval:n $\{\text{sign}(\langle \text{result} \rangle)\}$: $+1$ for positive numbers and for $+\infty$, $-1$ for negative numbers and for $-\infty$, $\pm 0$ for $\pm 0$. If the operand is a tuple or is \texttt{nan}, then “invalid operation” occurs and the result is 0.

\fp_to_decimal:N \fp_to_decimal:c \fp_to_decimal:n

\fp_to_decimal:N \fp_var \fp_to_decimal:n \{\textit{floating point expression}\}
Evaluates the \textit{floating point expression} and expresses the result as a decimal number with no exponent. Leading or trailing zeros may be inserted to compensate for the exponent. Non-significant trailing zeros are trimmed, and integers are expressed without a decimal separator. The values $\pm \infty$ and \texttt{nan} trigger an “invalid operation” exception. For a tuple, each item is converted using \fp_to_decimal:n and they are combined as $(\langle fp_1 \rangle, \langle fp_2 \rangle, \ldots, \langle fp_n \rangle)$ if $n > 1$ and $(\langle fp_1 \rangle)$ or $()$ for fewer items.

\fp_to_dim:N \fp_to_dim:c \fp_to_dim:n

\fp_to_dim:N \fp_var \fp_to_dim:n \{\textit{floating point expression}\}
Evaluates the \textit{floating point expression} and expresses the result as a dimension (in pt) suitable for use in dimension expressions. The output is identical to \fp_to_decimal:n, with an additional trailing \texttt{pt} (both letter tokens). In particular, the result may be outside the range $[-2^{14} + 2^{−17}, 2^{14} − 2^{−17}]$ of valid \TeX\ dimensions, leading to overflow errors if used as a dimension. Tuples, as well as the values $\pm \infty$ and \texttt{nan}, trigger an “invalid operation” exception.

\fp_to_int:N \fp_to_int:c \fp_to_int:n

\fp_to_int:N \fp_var \fp_to_int:n \{\textit{floating point expression}\}
Evaluates the \textit{floating point expression}, and rounds the result to the closest integer, rounding exact ties to an even integer. The result may be outside the range $[-2^{31} + 1, 2^{31} − 1]$ of valid \TeX\ integers, leading to overflow errors if used in an integer expression. Tuples, as well as the values $\pm \infty$ and \texttt{nan}, trigger an “invalid operation” exception.

28.3 Using floating points
\texttt{\textbackslash fp\_to\_scientific:N} (*)\texttt{\textbackslash fp\_to\_scientific:c} (*)\texttt{\textbackslash fp\_to\_scientific:n} (*)

Revised: 2012-05-08
Updated: 2016-03-22

\texttt{\textbackslash fp\_to\_scientific:N} (fp var)\texttt{\textbackslash fp\_to\_scientific:c} \texttt{\textbackslash fp\_to\_scientific:n} \{\texttt{floating point expression}\}

Evaluates the \texttt{\textbackslash{}floating point expression} and expresses the result in scientific notation: 
\begin{align*}
\text{optional -}\langle \text{digit}\rangle .\langle 15\text{ digits}\rangle \text{e}(\text{optional sign})\langle \text{exponent}\rangle
\end{align*}

The leading \langle \text{digit}\rangle is non-zero except in the case of ±0. The values ±∞ and \texttt{nan} trigger an “invalid operation” exception. Normal category codes apply: thus the \texttt{e} is category code 11 (a letter). For a tuple, each item is converted using \texttt{\textbackslash fp\_to\_scientific:n} and they are combined as \((\langle fp_1\rangle,\langle fp_2\rangle,\ldots,\langle fp_n\rangle)\) if \(n > 1\) and \((\langle fp_1\rangle,)\) or () for fewer items.

\texttt{\textbackslash fp\_to\_tl:N} (*)\texttt{\textbackslash fp\_to\_tl:c} (*)\texttt{\textbackslash fp\_to\_tl:n} (*)

Updated: 2016-03-22

\texttt{\textbackslash fp\_to\_tl:N} (fp var)\texttt{\textbackslash fp\_to\_tl:c} \texttt{\textbackslash fp\_to\_tl:n} \{\texttt{floating point expression}\}

Evaluates the \texttt{(floating point expression)} and expresses the result in (almost) the shortest possible form. Numbers in the ranges \((0, 10^{-3})\) and \([10^{16}, \infty)\) are expressed in scientific notation with trailing zeros trimmed and no decimal separator when there is a single significant digit (this differs from \texttt{\textbackslash fp\_to\_scientific:n}). Numbers in the range \([10^{-3}, 10^{16})\) are expressed in a decimal notation without exponent, with trailing zeros trimmed, and no decimal separator for integer values (see \texttt{\textbackslash fp\_to\_decimal:n}). Negative numbers start with −. The special values ±0, ±∞ and \texttt{nan} are rendered as 0, -0, \texttt{inf}, -\texttt{inf}, and \texttt{nan} respectively. Normal category codes apply and thus \texttt{inf} or \texttt{nan}, if produced, are made up of letters. For a tuple, each item is converted using \texttt{\textbackslash fp\_to\_tl:n} and they are combined as \((\langle fp_1\rangle,\langle fp_2\rangle,\ldots,\langle fp_n\rangle)\) if \(n > 1\) and \((\langle fp_1\rangle,)\) or () for fewer items. This function is identical to \texttt{\textbackslash fp\_to\_decimal:n}.

\texttt{\textbackslash fp\_use:N} (*)\texttt{\textbackslash fp\_use:c} (*)

Updated: 2012-07-08

\texttt{\textbackslash fp\_use:N} (fp var)\texttt{\textbackslash fp\_use:c} \texttt{\textbackslash fp\_use:N} \{\texttt{floating point expression}\}

Inserts the value of the \texttt{\textbackslash{}floating point expression} into the input stream as a decimal number with no exponent. Leading or trailing zeros may be inserted to compensate for the exponent. Non-significant trailing zeros are trimmed. Integers are expressed without a decimal separator. The values ±∞ and \texttt{nan} trigger an “invalid operation” exception. For a tuple, each item is converted using \texttt{\textbackslash fp\_to\_decimal:n} and they are combined as \((\langle fp_1\rangle,\langle fp_2\rangle,\ldots,\langle fp_n\rangle)\) if \(n > 1\) and \((\langle fp_1\rangle,)\) or () for fewer items. This function is identical to \texttt{\textbackslash fp\_to\_decimal:n}.

\texttt{\textbackslash fp\_if\_exist\_p:N} (*)\texttt{\textbackslash fp\_if\_exist\_p:c} (*)\texttt{\textbackslash fp\_if\_exist\_NTF} (*)\texttt{\textbackslash fp\_if\_exist\_CTF} (*)

Updated: 2012-05-08

\texttt{\textbackslash fp\_if\_exist\_p:N} (fp var)\texttt{\textbackslash fp\_if\_exist\_p:c} \texttt{\textbackslash fp\_if\_exist\_NTF} \texttt{\textbackslash fp\_if\_exist\_CTF} \{\texttt{true code}\} \{\texttt{false code}\}

Tests whether the \texttt{\textbackslash{}floating point expression} is currently defined. This does not check that the \texttt{\textbackslash{}floating point expression} really is a floating point variable.

\section{28.4 Floating point conditionals}

\texttt{\textbackslash fp\_if\_exist\_p:N} (fp var)\texttt{\textbackslash fp\_if\_exist\_p:c} \texttt{\textbackslash fp\_if\_exist\_NTF} \texttt{\textbackslash fp\_if\_exist\_CTF} \{\texttt{true code}\} \{\texttt{false code}\}

Tests whether the \texttt{\textbackslash{}floating point expression} is currently defined. This does not check that the \texttt{\textbackslash{}floating point expression} really is a floating point variable.
\fp_compare_p:nNn \fp_compare_p:nNn {\texttt{\textbackslash fpexpr}_1} \texttt{\langle relation \rangle} \{\texttt{\textbackslash fpexpr}_2\}
\fp_compare:nNnTF {\texttt{\textbackslash fpexpr}_1} \texttt{\langle relation \rangle} {\texttt{\textbackslash fpexpr}_2} {\texttt{\textbackslash true code}} {\texttt{\textbackslash false code}}

Compares the \texttt{\textbackslash fpexpr}_1 and the \texttt{\textbackslash fpexpr}_2, and returns \texttt{true} if the \texttt{\langle relation \rangle} is obeyed. Two floating points \texttt{x} and \texttt{y} may obey four mutually exclusive relations: \texttt{x < y}, \texttt{x = y}, \texttt{x > y}, or \texttt{\textbackslash x\textbackslash ?\textbackslash y} (“not ordered”). The last case occurs exactly if one or both operands is \texttt{nan} or is a tuple, unless they are equal tuples. Note that a \texttt{nan} is distinct from any value, even another \texttt{nan}, hence \texttt{x = x} is not true for a \texttt{nan}. To test if a value is \texttt{nan}, compare it to an arbitrary number with the “not ordered” relation.

\fp_compare:nNnTF { <value> } \texttt{\textbackslash ?} { 0 }
{ } % <value> is nan
{ } % <value> is not nan

Tuples are equal if they have the same number of items and items compare equal (in particular there must be no \texttt{nan}). At present any other comparison with tuples yields \texttt{\textbackslash ?} (not ordered). This is experimental.

This function is less flexible than \texttt{\textbackslash fp_compare:nTF} but slightly faster. It is provided for consistency with \texttt{\textbackslash int_compare:nNnTF} and \texttt{\textbackslash dim_compare:nNnTF}.

243
\texttt{\textbackslash fp\_compare\_p:n} \hspace{1cm} \texttt{\textbackslash fp\_compare\_n:n}  
\{
  \langle \text{fpexpr}_1 \rangle \langle \text{relation}_1 \rangle 
  \ldots 
  \langle \text{fpexpr}_N \rangle \langle \text{relation}_N \rangle 
  \langle \text{fpexpr}_{N+1} \rangle 
\}
\texttt{\textbackslash fp\_compare\_nTF}  
\{
  \langle \text{fpexpr}_1 \rangle \langle \text{relation}_1 \rangle 
  \ldots 
  \langle \text{fpexpr}_N \rangle \langle \text{relation}_N \rangle 
  \langle \text{fpexpr}_{N+1} \rangle 
\}
\{\langle \text{true code} \rangle\} \{\langle \text{false code} \rangle\}

Evaluates the \textit{(floating point expressions)} as described for \texttt{\textbackslash fp\_eval:n} and compares consecutive result using the corresponding \langle \text{relation} \rangle, namely it compares \langle \text{fpexpr}_1 \rangle and \langle \text{fpexpr}_2 \rangle using the \langle \text{relation}_1 \rangle, then \langle \text{fpexpr}_2 \rangle and \langle \text{fpexpr}_3 \rangle using the \langle \text{relation}_2 \rangle, until finally comparing \langle \text{fpexpr}_N \rangle and \langle \text{fpexpr}_{N+1} \rangle using the \langle \text{relation}_N \rangle. The test yields \texttt{true} if all comparisons are \texttt{true}. Each \langle \text{floating point expression} \rangle is evaluated only once. Contrarily to \texttt{\textbackslash int\_compare:nTF}, all \langle \text{floating point expressions} \rangle are computed, even if one comparison is \texttt{false}. Two floating points \(x\) and \(y\) may obey four mutually exclusive relations: \(x<y\), \(x=y\), \(x>y\), or \(x\not\approx y\) (“not ordered”). The last case occurs exactly if one or both operands is \texttt{nan} or is a tuple, unless they are equal tuples. Each \langle \text{relation} \rangle can be any (non-empty) combination of \(<, =, >, \text{and } \not\approx\), plus an optional leading \texttt{!} (which negates the \langle \text{relation} \rangle), with the restriction that the \langle \text{relation} \rangle may not start with \not\approx, as this symbol has a different meaning (in combination with \texttt{:}) within floating point expressions. The comparison \(x \langle \text{relation} \rangle y\) is then \texttt{true} if the \langle \text{relation} \rangle does not start with \texttt{!} and the actual relation (\(<, =, >, \text{or } \texttt{:}\)) between \(x\) and \(y\) appears within the \langle \text{relation} \rangle, or on the contrary if the \langle \text{relation} \rangle starts with \texttt{!} and the relation between \(x\) and \(y\) does not appear within the \langle \text{relation} \rangle. Common choices of \langle \text{relation} \rangle include \texttt{>=} (greater or equal), \texttt{!=} (not equal), \texttt{!?} or \texttt{<=} (comparable).

This function is more flexible than \texttt{\textbackslash fp\_compare:nNnTF} and only slightly slower.

### 28.5 Floating point expression loops

\texttt{\textbackslash fp\_do\_until:nNn} \hspace{1cm} \texttt{\textbackslash fp\_do\_until:nNn} \{\langle \text{fpexpr}_1 \rangle\} \{\langle \text{relation} \rangle\} \{\langle \text{fpexpr}_2 \rangle\} \{\langle \text{code} \rangle\}

Places the \langle \text{code} \rangle in the input stream for \TeX{} to process, and then evaluates the relationship between the two \langle \text{floating point expressions} \rangle as described for \texttt{\textbackslash fp\_compare:nNnTF}. If the test is \texttt{false} then the \langle \text{code} \rangle is inserted into the input stream again and a loop occurs until the \langle \text{relation} \rangle is \texttt{true}.

\texttt{\textbackslash fp\_do\_while:nNn} \hspace{1cm} \texttt{\textbackslash fp\_do\_while:nNn} \{\langle \text{fpexpr}_1 \rangle\} \{\langle \text{relation} \rangle\} \{\langle \text{fpexpr}_2 \rangle\} \{\langle \text{code} \rangle\}

Places the \langle \text{code} \rangle in the input stream for \TeX{} to process, and then evaluates the relationship between the two \langle \text{floating point expressions} \rangle as described for \texttt{\textbackslash fp\_compare:nNnTF}. If the test is \texttt{true} then the \langle \text{code} \rangle is inserted into the input stream again and a loop occurs until the \langle \text{relation} \rangle is \texttt{false}.
\fp_until_do:nNnn
\fp_compare:nNnTF \fpexpr{1} \relation \fpexpr{2} \{\code\}

Evaluates the relationship between the two \textit{floating point expressions} as described for \fp_compare:nNnTF, and then places the \textit{code} in the input stream if the \textit{relation} is \textit{false}. After the \textit{code} has been processed by \TeX{} the test is repeated, and a loop occurs until the test is \textit{true}.

\fp_until_do:nn
\fp_compare:nTF \fpexpr{1} \relation \fpexpr{2} \{\code\}

Evaluates the relationship between the two \textit{floating point expressions} as described for \fp_compare:nTF, and then places the \textit{code} in the input stream if the \textit{relation} is \textit{false}. After the \textit{code} has been processed by \TeX{} the test is repeated, and a loop occurs until the test is \textit{true}.

\fp_while_do:nNnn
\fp_compare:nNnTF \fpexpr{1} \relation \fpexpr{2} \{\code\}

Evaluates the relationship between the two \textit{floating point expressions} as described for \fp_compare:nNnTF, and then places the \textit{code} in the input stream if the \textit{relation} is \textit{true}. After the \textit{code} has been processed by \TeX{} the test is repeated, and a loop occurs until the test is \textit{false}.

\fp_while_do:nn
\fp_compare:nTF \fpexpr{1} \relation \fpexpr{2} \{\code\}

Evaluates the relationship between the two \textit{floating point expressions} as described for \fp_compare:nTF, and then places the \textit{code} in the input stream if the \textit{relation} is \textit{true}. After the \textit{code} has been processed by \TeX{} the test is repeated, and a loop occurs until the test is \textit{false}.
\fp_step_function:nnnN \star \fp_step_function:nnnc \star

New: 2016-11-21
Updated: 2016-12-06

This function first evaluates the \langle initial value \rangle, \langle step \rangle and \langle final value \rangle, each of which should be a floating point expression evaluating to a floating point number, not a tuple.

The \langle function \rangle is then placed in front of each \langle value \rangle from the \langle initial value \rangle to the \langle final value \rangle in turn (using \langle step \rangle between each \langle value \rangle). The \langle step \rangle must be non-zero. If the \langle step \rangle is positive, the loop stops when the \langle value \rangle becomes larger than the \langle final value \rangle.

If the \langle step \rangle is negative, the loop stops when the \langle value \rangle becomes smaller than the \langle final value \rangle. The \langle function \rangle should absorb one numerical argument. For example
\cs_set:Npn \my_func:n #1 { \[I \text{ saw } #1\] \quad }
\fp_step_function:nnnN { 1.0 } { 0.1 } { 1.5 } \my_func:n
would print
[I saw 1.0] [I saw 1.1] [I saw 1.2] [I saw 1.3] [I saw 1.4] [I saw 1.5]

\TeXhackers note: Due to rounding, it may happen that adding the \langle step \rangle to the \langle value \rangle does not change the \langle value \rangle; such cases give an error, as they would otherwise lead to an infinite loop.

\fp_step_inline:nnnn

New: 2016-11-21
Updated: 2016-12-06

This function first evaluates the \langle initial value \rangle, \langle step \rangle and \langle final value \rangle, all of which should be floating point expressions evaluating to a floating point number, not a tuple.

Then for each \langle value \rangle from the \langle initial value \rangle to the \langle final value \rangle in turn (using \langle step \rangle between each \langle value \rangle), the \langle code \rangle is inserted into the input stream with \#1 replaced by the current \langle value \rangle. Thus the \langle code \rangle should define a function of one argument (\#1).

\fp_step_variable:nnnNn

New: 2017-04-12

This function first evaluates the \langle initial value \rangle, \langle step \rangle and \langle final value \rangle, all of which should be floating point expressions evaluating to a floating point number, not a tuple.

Then for each \langle value \rangle from the \langle initial value \rangle to the \langle final value \rangle in turn (using \langle step \rangle between each \langle value \rangle), the \langle code \rangle is inserted into the input stream, with the \langle tl var \rangle defined as the current \langle value \rangle. Thus the \langle code \rangle should make use of the \langle tl var \rangle.

28.6 Some useful constants, and scratch variables

\c_zero_fp
\c_minus_zero_fp

New: 2012-05-08

Zero, with either sign.

\c_one_fp

New: 2012-05-08

One as an \fp: useful for comparisons in some places.
Infinity, with either sign. These can be input directly in a floating point expression as \texttt{inf} and \texttt{-inf}.

The value of the base of the natural logarithm, \(e = \exp(1)\).

The value of \(\pi\). This can be input directly in a floating point expression as \texttt{pi}.

The value of \(1^\circ\) in radians. Multiply an angle given in degrees by this value to obtain a result in radians. Note that trigonometric functions expecting an argument in radians or in degrees are both available. Within floating point expressions, this can be accessed as \texttt{deg}.

Scratch floating points for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

Scratch floating points for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

### 28.7 Floating point exceptions

The functions defined in this section are experimental, and their functionality may be altered or removed altogether.

“Exceptions” may occur when performing some floating point operations, such as \(0 / 0\), or \(10 ** 1e9999\). The relevant IEEE standard defines 5 types of exceptions, of which we implement 4.

- **Overflow** occurs whenever the result of an operation is too large to be represented as a normal floating point number. This results in \(\pm\infty\).
- **Underflow** occurs whenever the result of an operation is too close to 0 to be represented as a normal floating point number. This results in \(\pm0\).
- **Invalid operation** occurs for operations with no defined outcome, for instance \(0/0\) or \(\sin(\infty)\), and results in a \texttt{nan}. It also occurs for conversion functions whose target type does not have the appropriate infinite or \texttt{nan} value (e.g., \texttt{fp_to_dim:n}).
- **Division by zero** occurs when dividing a non-zero number by 0, or when evaluating functions at poles, e.g., \(\ln(0)\) or \(\cot(0)\). This results in \(\pm\infty\).
(not yet) Inexact occurs whenever the result of a computation is not exact, in other words, almost always. At the moment, this exception is entirely ignored in \LaTeX3.

To each exception we associate a “flag”: `fp_overflow`, `fp_underflow`, `fp_invalid_operation` and `fp_division_by_zero`. The state of these flags can be tested and modified with commands from \l3flag

By default, the “invalid operation” exception triggers an (expandable) error, and raises the corresponding flag. Other exceptions raise the corresponding flag but do not trigger an error. The behaviour when an exception occurs can be modified (using `\fp_trap:nn`) to either produce an error and raise the flag, or only raise the flag, or do nothing at all.

```
\fp_trap:nn {⟨exception⟩} {⟨trap type⟩}
```

All occurrences of the ⟨exception⟩ (overflow, underflow, invalid_operation or division_by_zero) within the current group are treated as ⟨trap type⟩, which can be

- none: the ⟨exception⟩ will be entirely ignored, and leave no trace;
- flag: the ⟨exception⟩ will turn the corresponding flag on when it occurs;
- error: additionally, the ⟨exception⟩ will halt the \TeX run and display some information about the current operation in the terminal.

This function is experimental, and may be altered or removed.

Flags denoting the occurrence of various floating-point exceptions:

```
flag_fp_overflow
flag_fp_underflow
flag_fp_invalid_operation
flag_fp_division_by_zero
```

28.8 Viewing floating points

```
\fp_show:N ⟨fp var⟩
\fp_show:c
\fp_show:n {⟨floating point expression⟩}
```

Evaluates the ⟨floating point expression⟩ and displays the result in the terminal.

```
\fp_log:N ⟨fp var⟩
\fp_log:c
\fp_log:n {⟨floating point expression⟩}
```

Evaluates the ⟨floating point expression⟩ and writes the result in the log file.
28.9 Floating point expressions

28.9.1 Input of floating point numbers

We support four types of floating point numbers:

- $\pm m \cdot 10^n$, a floating point number, with integer $1 \leq m \leq 10^{16}$, and $-10000 \leq n \leq 10000$;
- $\pm 0$, zero, with a given sign;
- $\pm \infty$, infinity, with a given sign;
- $\text{nan}$, is “not a number”, and can be either quiet or signalling (not yet: this distinction is currently unsupported);

Normal floating point numbers are stored in base 10, with up to 16 significant figures.

On input, a normal floating point number consists of:

- $\langle \text{sign} \rangle$: a possibly empty string of + and – characters;
- $\langle \text{significand} \rangle$: a non-empty string of digits together with zero or one dot;
- $\langle \text{exponent} \rangle$: optionally: the character e or E, followed by a possibly empty string of + and – tokens, and a non-empty string of digits.

The sign of the resulting number is + if $\langle \text{sign} \rangle$ contains an even number of –, and – otherwise, hence, an empty $\langle \text{sign} \rangle$ denotes a non-negative input. The stored significand is obtained from $\langle \text{significand} \rangle$ by omitting the decimal separator and leading zeros, and rounding to 16 significant digits, filling with trailing zeros if necessary. In particular, the value stored is exact if the input $\langle \text{significand} \rangle$ has at most 16 digits. The stored $\langle \text{exponent} \rangle$ is obtained by combining the input $\langle \text{exponent} \rangle$ (0 if absent) with a shift depending on the position of the significand and the number of leading zeros.

A special case arises if the resulting $\langle \text{exponent} \rangle$ is either too large or too small for the floating point number to be represented. This results either in an overflow (the number is then replaced by $\pm \infty$), or an underflow (resulting in $\pm 0$).

The result is thus $\pm 0$ if and only if $\langle \text{significand} \rangle$ contains no non-zero digit (i.e., consists only in characters 0, and an optional period), or if there is an underflow. Note that a single dot is currently a valid floating point number, equal to $+0$, but that is not guaranteed to remain true.

The $\langle \text{significand} \rangle$ must be non-empty, so e1 and e-1 are not valid floating point numbers. Note that the latter could be mistaken with the difference of “e” and 1. To avoid confusions, the base of natural logarithms cannot be input as e and should be input as exp(1) or c_e_fp (which is faster).

Special numbers are input as follows:

- inf represents $+\infty$, and can be preceded by any $\langle \text{sign} \rangle$, yielding $\pm \infty$ as appropriate.
- nan represents a (quiet) non-number. It can be preceded by any sign, but that sign is ignored.
- Any unrecognizable string triggers an error, and produces a nan.
- Note that commands such as \infty, \pi, or \sin do not work in floating point expressions. They may silently be interpreted as completely unexpected numbers, because integer constants (allowed in expressions) are commonly stored as mathematical characters.
28.9.2 Precedence of operators

We list here all the operations supported in floating point expressions, in order of decreasing precedence: operations listed earlier bind more tightly than operations listed below them.

- Function calls ($\sin$, $\ln$, etc).
- Binary $\ast\ast$ and $\sim$ (right associative).
- Unary $+$, $-$, $!$.
- Implicit multiplication by juxtaposition ($2\pi$) when neither factor is in parentheses.
- Binary $\ast$ and $/$, implicit multiplication by juxtaposition with parentheses (for instance $3(4+5)$).
- Binary $+$ and $-$.
- Comparisons $\geq$, $\neq$, $<$, $<$?, etc.
- Logical $\text{and}$, denoted by $\&\&$.
- Logical $\text{or}$, denoted by $||$.
- Ternary operator $?:$ (right associative).
- Comma (to build tuples).

The precedence of operations can be overridden using parentheses. In particular, the precedence of juxtaposition implies that

$$\begin{align*}
1/2\pi &= 1/(2\pi), \\
1/2\pi(p + p) &= (2\pi)^{-1}(\pi + \pi) \approx 1, \\
\sin2\pi &= \sin(2\pi)\pi \neq 0, \\
2^{-2}\max(3, 5) &= 2^{-2}\max(3, 5) = 20, \\
1\text{in}/1\text{cm} &= (1\text{in})/(1\text{cm}) = 2.54.
\end{align*}$$

Functions are called on the value of their argument, contrarily to \TeX macros.

28.9.3 Operations

We now present the various operations allowed in floating point expressions, from the lowest precedence to the highest. When used as a truth value, a floating point expression is \text{false} if it is $\pm0$, and \text{true} otherwise, including when it is \text{nan} or a tuple such as $(0, 0)$. Tuples are only supported to some extent by operations that work with truth values ($?\?, 11, \&\&$), by comparisons ($!\equiv$?), and by $+$, $-$, $\ast$, $/$. Unless otherwise specified, providing a tuple as an argument of any other operation yields the “invalid operation” exception and a \text{nan} result.
The ternary operator `?:` results in `⟨operand2⟩` if `⟨operand1⟩` is true (not ±0), and `⟨operand3⟩` if `⟨operand1⟩` is false (±0). All three `⟨operands⟩` are evaluated in all cases; they may be tuples. The operator is right associative, hence

\[
\text{\texttt{\textbackslash fp\textunderscore eval:n}} \{ \langle \text{operand}_1 \rangle \ ? \langle \text{operand}_2 \rangle : \langle \text{operand}_3 \rangle \} \]

first tests whether \(1 + 3 > 4\); since this isn’t true, the branch following : is taken, and \(2 + 4 > 5\) is compared; since this is true, the branch before : is taken, and everything else is (evaluated then) ignored. That allows testing for various cases in a concise manner, with the drawback that all computations are made in all cases.

\[
\text{\texttt{\textbackslash fp\textunderscore eval:n}} \{ \langle \text{operand}_1 \rangle \ || \langle \text{operand}_2 \rangle \} \]

If `⟨operand1⟩` is true (not ±0), use that value, otherwise the value of `⟨operand2⟩`. Both `⟨operands⟩` are evaluated in all cases; they may be tuples. In `⟨operand_1⟩ || ⟨operand_2⟩ || ... || ⟨operand_N⟩`, the first true (nonzero) `⟨operand⟩` is used and if all are zero the last one (±0) is used.

\[
\text{\texttt{\textbackslash fp\textunderscore eval:n}} \{ \langle \text{operand}_1 \rangle \ &\& \langle \text{operand}_2 \rangle \} \]

If `⟨operand1⟩` is false (equal to ±0), use that value, otherwise the value of `⟨operand2⟩`. Both `⟨operands⟩` are evaluated in all cases; they may be tuples. In `⟨operand_1⟩ &\& ⟨operand_2⟩ &\& ... &\& ⟨operand_N⟩`, the first false (±0) `⟨operand⟩` is used and if none is zero the last one is used.

Each `⟨relation⟩` consists of a non-empty string of `<`, `=`, `>`, and `?`, optionally preceded by `!`, and may not start with `?`. This evaluates to +1 if all comparisons `⟨operand_i⟩ ⟨relation_i⟩ ⟨operand_{i+1}⟩` are true, and +0 otherwise. All `⟨operands⟩` are evaluated (once) in all cases. See `\texttt{\textbackslash fp\textunderscore compare:nTF}` for details.

\[
\text{\texttt{\textbackslash fp\textunderscore eval:n}} \{ \langle \text{operand}_1 \rangle + \langle \text{operand}_2 \rangle \} \]

\[
\text{\texttt{\textbackslash fp\textunderscore eval:n}} \{ \langle \text{operand}_1 \rangle - \langle \text{operand}_2 \rangle \} \]

Computes the sum or the difference of its two `⟨operands⟩`. The “invalid operation” exception occurs for ∞ − ∞. “Underflow” and “overflow” occur when appropriate. These operations supports the itemwise addition or subtraction of two tuples, but if they have a different number of items the “invalid operation” exception occurs and the result is `\texttt{\textbackslash math:nan}`.
Computes the product or the ratio of its two \emph{operands}. The “invalid operation” exception occurs for $\infty/\infty$, $0/0$, or $0 \times \infty$. “Division by zero” occurs when dividing a finite non-zero number by $\pm 0$. “Underflow” and “overflow” occur when appropriate. When $\langle \text{operand}_1 \rangle$ is a tuple and $\langle \text{operand}_2 \rangle$ is a floating point number, each item of $\langle \text{operand}_1 \rangle$ is multiplied or divided by $\langle \text{operand}_2 \rangle$. Other combinations yield an “invalid operation” exception and a \texttt{nan} result.

The unary + does nothing, the unary - changes the sign of the $\langle \text{operand} \rangle$ (for a tuple, of all its components), and ! $\langle \text{operand} \rangle$ evaluates to 1 if $\langle \text{operand} \rangle$ is false (is $\pm 0$) and 0 otherwise (this is the not boolean function). Those operations never raise exceptions.

Raises $\langle \text{operand}_1 \rangle$ to the power $\langle \text{operand}_2 \rangle$. This operation is right associative, hence $2 ** 2 ** 3$ equals $2^{2^3} = 256$. If $\langle \text{operand}_1 \rangle$ is negative or $-0$ then: the result’s sign is + if the $\langle \text{operand}_2 \rangle$ is infinite and $(-1)^p$ if the $\langle \text{operand}_2 \rangle$ is $p/5^q$ with $p, q$ integers; the result is +0 if abs($\langle \text{operand}_1 \rangle$)**$\langle \text{operand}_2 \rangle$ evaluates to zero; in other cases the “invalid operation” exception occurs because the sign cannot be determined. “Division by zero” occurs when raising $\pm 0$ to a finite strictly negative power. “Underflow” and “overflow” occur when appropriate. If either operand is a tuple, “invalid operation” occurs.

Computes the absolute value of the $\langle \text{fpexpr} \rangle$. If the operand is a tuple, “invalid operation” occurs. This operation does not raise exceptions in other cases. See also \fp_abs:n.

Computes the exponential of the $\langle \text{fpexpr} \rangle$. “Underflow” and “overflow” occur when appropriate. If the operand is a tuple, “invalid operation” occurs.

Computes the factorial of the $\langle \text{fpexpr} \rangle$. If the $\langle \text{fpexpr} \rangle$ is an integer between $-0$ and 3248 included, the result is finite and correctly rounded. Larger positive integers give $+\infty$ with “overflow”, while fact($+\infty$) = $+\infty$ and fact(nan) = nan with no exception. All other inputs give nan with the “invalid operation” exception.

Computes the natural logarithm of the $\langle \text{fpexpr} \rangle$. Negative numbers have no (real) logarithm, hence the “invalid operation” is raised in that case, including for \ln($-0$). “Division by zero” occurs when evaluating \ln($+0$) = $-\infty$. “Underflow” and “overflow” occur when appropriate. If the operand is a tuple, “invalid operation” occurs.
\texttt{\textbackslash fp\_eval\_n \{ logb ( \textbackslash{fpexpr} ) \}}

Determines the exponent of the \textbackslash{fpexpr}, namely the floor of the base-10 logarithm of its absolute value. “Division by zero” occurs when evaluating \texttt{logb(0)} = \(-\infty\). Other special values are \texttt{logb(±\infty)} = \(+\infty\) and \texttt{logb(nan)} = \texttt{nan}. If the operand is a tuple or is \texttt{nan}, then “invalid operation” occurs and the result is \texttt{nan}.

\texttt{\textbackslash fp\_eval\_n \{ max ( \textbackslash{fpexpr}_1, \textbackslash{fpexpr}_2, \ldots ) \}}
\texttt{\textbackslash fp\_eval\_n \{ min ( \textbackslash{fpexpr}_1, \textbackslash{fpexpr}_2, \ldots ) \}}

Evaluates each \textbackslash{fpexpr} and computes the largest (smallest) of those. If any of the \textbackslash{fpexpr} is a \texttt{nan} or tuple, the result is \texttt{nan}. If any operand is a tuple, “invalid operation” occurs; these operations do not raise exceptions in other cases.

\texttt{\textbackslash fp\_eval\_n \{ round ( \textbackslash{fpexpr} ) \}}
\texttt{\textbackslash fp\_eval\_n \{ round ( \textbackslash{fpexpr}_1, \textbackslash{fpexpr}_2 ) \}}
\texttt{\textbackslash fp\_eval\_n \{ round ( \textbackslash{fpexpr}_1, \textbackslash{fpexpr}_2, \textbackslash{fpexpr}_3 ) \}}

Only \texttt{round} accepts a third argument. Evaluates \texttt{(fpexpr)1 = x} and \texttt{(fpexpr)2 = n} and \texttt{(fpexpr)t = t} then rounds \(x\) to \(n\) places. If \(n\) is an integer, this rounds \(x\) to a multiple of \(10^{-n}\); if \(n = +\infty\), this always yields \(x\); if \(n = -\infty\), this yields one of \(±0\), \(±\infty\), or \texttt{nan}; if \(n = \texttt{nan}\), this yields \texttt{nan}; if \(n\) is neither \(±\infty\) nor an integer, then an “invalid operation” exception is raised. When \texttt{(fpexpr)2} is omitted, \(n = 0\), i.e., \texttt{(fpexpr)1} is rounded to an integer. The rounding direction depends on the function.

- \texttt{round} yields the multiple of \(10^{-n}\) closest to \(x\), with ties \((x\) half-way between two such multiples) rounded as follows. If \(t\) is \texttt{nan} (or not given) the even multiple is chosen (“ties to even”), if \(t = ±0\) the multiple closest to \(0\) is chosen (“ties to zero”), if \(t\) is positive/negative the multiple closest to \(+\infty/-\infty\) is chosen (“ties towards positive/negative infinity”).
- \texttt{floor} yields the largest multiple of \(10^{-n}\) smaller or equal to \(x\) (“round towards negative infinity”);
- \texttt{ceil} yields the smallest multiple of \(10^{-n}\) greater or equal to \(x\) (“round towards positive infinity”);
- \texttt{trunc} yields a multiple of \(10^{-n}\) with the same sign as \(x\) and with the largest absolute value less than that of \(x\) (“round towards zero”).

“Overflow” occurs if \(x\) is finite and the result is infinite (this can only happen if \texttt{(fpexpr)2} < \(-9984\)). If any operand is a tuple, “invalid operation” occurs.

\texttt{\textbackslash fp\_eval\_n \{ sign ( \textbackslash{fpexpr} ) \}}

Evaluates the \textbackslash{fpexpr} and determines its sign: \(+1\) for positive numbers and for \(+\infty\), \(-1\) for negative numbers and for \(-\infty\), \(±0\) for \(±0\), and \texttt{nan} for \texttt{nan}. If the operand is a tuple, “invalid operation” occurs. This operation does not raise exceptions in other cases.
Computes the sine, cosine, tangent, cotangent, cosecant, or secant of the \( \langle \text{fpexpr} \rangle \) given in radians. For arguments given in degrees, see \texttt{sind}, \texttt{cosd}, etc. Note that since \( \pi \) is irrational, \( \sin(8\pi) \) is not quite zero, while its analogue \( \text{sind}(8 \times 180) \) is exactly zero. The trigonometric functions are undefined for an argument of \( \pm \infty \), leading to the “invalid operation” exception. Additionally, evaluating tangent, cotangent, cosecant, or secant at one of their poles leads to a “division by zero” exception. “Underflow” and “overflow” occur when appropriate. If the operand is a tuple, “invalid operation” occurs.

Updated: 2013-11-17

Computes the sine, cosine, tangent, cotangent, cosecant, or secant of the \( \langle \text{fpexpr} \rangle \) given in degrees. For arguments given in radians, see \texttt{sin}, \texttt{cos}, etc. Note that since \( \pi \) is irrational, \( \sin(8\pi) \) is not quite zero, while its analogue \( \text{sind}(8 \times 180) \) is exactly zero. The trigonometric functions are undefined for an argument of \( \pm \infty \), leading to the “invalid operation” exception. Additionally, evaluating tangent, cotangent, cosecant, or secant at one of their poles leads to a “division by zero” exception. “Underflow” and “overflow” occur when appropriate. If the operand is a tuple, “invalid operation” occurs.

New: 2013-11-02

Computes the arcsine, arccosine, arccosecant, or arccosecant of the \( \langle \text{fpexpr} \rangle \) and returns the result in radians, in the range \([-\pi/2, \pi/2]\) for \texttt{asin} and \texttt{acsc} and \([0, \pi]\) for \texttt{acos} and \texttt{asec}. For a result in degrees, use \texttt{asind}, etc. If the argument of \texttt{asin} or \texttt{acos} lies outside the range \([-1, 1]\), or the argument of \texttt{acsc} or \texttt{asec} inside the range \((-1, 1)\), an “invalid operation” exception is raised. “Underflow” and “overflow” occur when appropriate. If the operand is a tuple, “invalid operation” occurs.

New: 2013-11-02

Computes the arcsine, arccosine, arccosecant, or arccosecant of the \( \langle \text{fpexpr} \rangle \) and returns the result in degrees, in the range \([-90, 90]\) for \texttt{asin} and \texttt{acsc} and \([0, 180]\) for \texttt{acos} and \texttt{asec}. For a result in radians, use \texttt{asind}, etc. If the argument of \texttt{asin} or \texttt{acos} lies outside the range \([-1, 1]\), or the argument of \texttt{acsc} or \texttt{asec} inside the range \((-1, 1)\), an “invalid operation” exception is raised. “Underflow” and “overflow” occur when appropriate. If the operand is a tuple, “invalid operation” occurs.

New: 2013-11-02

254
Those functions yield an angle in radians: \texttt{atan} and \texttt{acot} are their analogs in degrees.
The one-argument versions compute the arctangent or arccotangent of the \texttt{fexpr}: arctangent takes values in the range \([-\pi/2, \pi/2]\), and arccotangent in the range \([0, \pi]\). The two-argument arctangent computes the angle in polar coordinates of the point with Cartesian coordinates \((\texttt{fexpr}_2, \texttt{fexpr}_1)\): this is the arctangent of \(\texttt{fexpr}_1/\texttt{fexpr}_2\), possibly shifted by \(\pi\) depending on the signs of \(\texttt{fexpr}_1\) and \(\texttt{fexpr}_2\). The two-argument arccotangent computes the angle in polar coordinates of the point \((\texttt{fexpr}_1, \texttt{fexpr}_2)\), equal to the arccotangent of \(\texttt{fexpr}_1/\texttt{fexpr}_2\), possibly shifted by \(\pi\). Both two-argument functions take values in the wider range \([-\pi, \pi]\). The ratio \(\texttt{fexpr}_1/\texttt{fexpr}_2\) need not be defined for the two-argument arctangent: when both expressions yield \(\pm0\), or when both yield \(\pm\infty\), the resulting angle is one of \(\{\pm\pi/4, \pm3\pi/4\}\) depending on signs. The “underflow” exception can occur. If any operand is a tuple, “invalid operation” occurs.

Those functions yield an angle in degrees: \texttt{atand} and \texttt{acotd} are their analogs in radians. The one-argument versions compute the arctangent or arccotangent of the \texttt{fexpr}: arctangent takes values in the range \([-90, 90]\), and arccotangent in the range \([0, 180]\). The two-argument arctangent computes the angle in polar coordinates of the point with Cartesian coordinates \((\texttt{fexpr}_2, \texttt{fexpr}_1)\): this is the arctangent of \(\texttt{fexpr}_1/\texttt{fexpr}_2\), possibly shifted by 180 depending on the signs of \(\texttt{fexpr}_1\) and \(\texttt{fexpr}_2\). The two-argument arccotangent computes the angle in polar coordinates of the point \((\texttt{fexpr}_1, \texttt{fexpr}_2)\), equal to the arccotangent of \(\texttt{fexpr}_1/\texttt{fexpr}_2\), possibly shifted by 180. Both two-argument functions take values in the wider range \([-180, 180]\]. The ratio \(\texttt{fexpr}_1/\texttt{fexpr}_2\) need not be defined for the two-argument arctangent: when both expressions yield \(\pm0\), or when both yield \(\pm\infty\), the resulting angle is one of \(\{\pm45, \pm135\}\) depending on signs. The “underflow” exception can occur. If any operand is a tuple, “invalid operation” occurs.

Computes the square root of the \texttt{fexpr}. The “invalid operation” is raised when the \texttt{fexpr} is negative or is a tuple; no other exception can occur. Special values yield 
\[
\sqrt{-0} = -0, \quad \sqrt{+0} = +0, \quad \sqrt{-\infty} = +\infty \quad \text{and} \quad \sqrt{\texttt{nan}} = \texttt{nan}.
\]
fp_eval:n { rand() }

Produces a pseudo-random floating-point number (multiple of $10^{-16}$) between 0 included and 1 excluded. This is not available in older versions of \TeX. The random seed can be queried using \sys_rand_seed: and set using \sys_gset_rand_seed:n.

\TeXhacker note: This is based on pseudo-random numbers provided by the engine’s primitive \pdfuniformdeviate in pdf\TeX, \p\TeX, up\TeX and \uniformdeviate in Lua\TeX and \Xe\TeX. The underlying code is based on Metapost, which follows an additive scheme recommended in Section 3.6 of “The Art of Computer Programming, Volume 2”.

While we are more careful than \uniformdeviate to preserve uniformity of the underlying stream of 28-bit pseudo-random integers, these pseudo-random numbers should of course not be relied upon for serious numerical computations nor cryptography.

randint

fp_eval:n { randint( ⟨fpexpr⟩ ) }
fp_eval:n { randint( ⟨fpexpr⟩₁ , ⟨fpexpr⟩₂ ) }

Produces a pseudo-random integer between 1 and ⟨fpexpr⟩ or between ⟨fpexpr⟩₁ and ⟨fpexpr⟩₂ inclusive. The bounds must be integers in the range $(-10^{16}, 10^{16})$ and the first must be smaller or equal to the second. See rand for important comments on how these pseudo-random numbers are generated.

inf The special values $+\infty$, $-\infty$, and nan are represented as inf, -inf and nan (see \c_-inf_fp, \c_minus_inf_fp and \c_nan_fp).

nan

pi The value of $\pi$ (see \c_pi_fp).

deg The value of 1° in radians (see \c_one_degree_fp).
Those units of measurement are equal to their values in pt, namely

<table>
<thead>
<tr>
<th>Unit</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 in</td>
<td>72.27 pt</td>
</tr>
<tr>
<td>1 pt</td>
<td>1 pt</td>
</tr>
<tr>
<td>1 pc</td>
<td>12 pt</td>
</tr>
<tr>
<td>1 cm</td>
<td>$\frac{1}{2.54}$ in = 28.45275590551181 pt</td>
</tr>
<tr>
<td>1 mm</td>
<td>$\frac{1}{25.4}$ in = 2.845275590551181 pt</td>
</tr>
<tr>
<td>1 dd</td>
<td>0.376065 mm = 1.07000856496063 pt</td>
</tr>
<tr>
<td>1 cc</td>
<td>12 dd = 12.84010277952756 pt</td>
</tr>
<tr>
<td>1 nd</td>
<td>0.375 mm = 1.066978346456693 pt</td>
</tr>
<tr>
<td>1 nc</td>
<td>12 nd = 12.80374015748031 pt</td>
</tr>
<tr>
<td>1 bp</td>
<td>$\frac{1}{72}$ in = 1.00375 pt</td>
</tr>
<tr>
<td>1 sp</td>
<td>$2^{-16}$ pt = 1.52587890625 × 10^{-5} pt.</td>
</tr>
</tbody>
</table>

The values of the (font-dependent) units em and ex are gathered from \TeX when the surrounding floating point expression is evaluated.

### \fp_abs:n

\fp_abs:n \{⟨floating point expression⟩\}

Evaluates the ⟨floating point expression⟩ as described for \fp_eval:n and leaves the absolute value of the result in the input stream. If the argument is ±∞, nan or a tuple, “invalid operation” occurs. Within floating point expressions, abs() can be used; it accepts ±∞ and nan as arguments.

### \fp_max:nn \fp_min:nn

\fp_max:nn \{⟨fp expression 1⟩\} \{⟨fp expression 2⟩\}

\fp_min:nn \{⟨fp expression 1⟩\} \{⟨fp expression 2⟩\}

Evaluates the (floating point expressions) as described for \fp_eval:n and leaves the resulting larger (max) or smaller (min) value in the input stream. If the argument is a tuple, “invalid operation” occurs, but no other case raises exceptions. Within floating point expressions, max() and min() can be used.

## 28.10 Disclaimer and roadmap

The package may break down if the escape character is among 0123456789_+, or if it receives a \TeX primitive conditional affected by \exp_not:N.

The following need to be done. I’ll try to time-order the items.

- Function to count items in a tuple (and to determine if something is a tuple).
- Decide what exponent range to consider.
• Support signalling nan.

• Modulo and remainder, and rounding function quantize (and its friends analogous to trunc, ceil, floor).

• \fp_format::nn {\fpexpr} {\langle format \rangle}, but what should \langle format \rangle be? More general pretty printing?

• Add and, or, xor? Perhaps under the names all, any, and xor?

• Add log(x,b) for logarithm of x in base b.

• hypot (Euclidean length). Cartesian-to-polar transform.

• Hyperbolic functions cosh, sinh, tanh.

• Inverse hyperbolics.

• Base conversion, input such as 0xAB.CDEF.

• Factorial (not with !), gamma function.

• Improve coefficients of the sin and tan series.

• Treat upper and lower case letters identically in identifiers, and ignore underscores.

• Add an array(1,2,3) and i=complex(0,1).

• Provide an experimental map function? Perhaps easier to implement if it is a single character, @sin(1,2)?

• Provide an isnan function analogue of \fp_if_nan:nTF?

• Support keyword arguments?

Pgfmath also provides box-measurements (depth, height, width), but boxes are not possible expandably.

Bugs, and tests to add.

• Check that functions are monotonic when they should.

• Add exceptions to ?:, !==>?, &&, ||, and !.

• Logarithms of numbers very close to 1 are inaccurate.

• When rounding towards −∞, \dim_to_fp:n {0pt} should return −0, not +0.

• The result of (±0) + (±0), of x + (−x), and of (−x) + x should depend on the rounding mode.

• 0e9999999999 gives a TpX “number too large” error.

• Subnormals are not implemented.

Possible optimizations/improvements.

• Document that l3trial/l3fp-types introduces tools for adding new types.

• In subsection 28.9.1, write a grammar.
• It would be nice if the parse auxiliaries for each operation were set up in the corresponding module, rather than centralizing in l3fp-parse.

• Some functions should get an _o ending to indicate that they expand after their result.

• More care should be given to distinguish expandable/restricted expandable (auxiliary and internal) functions.

• The code for the ternary set of functions is ugly.

• There are many - missing in the doc to avoid bad line-breaks.

• The algorithm for computing the logarithm of the significand could be made to use a 5 terms Taylor series instead of 10 terms by taking \( c = \frac{2000}{(\lfloor 200x \rfloor + 1)} \in [10, 95] \) instead of \( c \in [1, 10] \). Also, it would then be possible to simplify the computation of \( t \). However, we would then have to hard-code the logarithms of 44 small integers instead of 9.

• Improve notations in the explanations of the division algorithm (l3fp-basics).

• Understand and document \texttt{\_\_fp\_basics\_pack\_weird\_low:NNNNw} and \texttt{\_\_fp\_-\_basics\_pack\_weird\_high:NNNNNNNNNw} better. Move the other basics\_pack auxiliaries to l3fp-aux under a better name.

• Find out if underflow can really occur for trigonometric functions, and redoc as appropriate.

• Add bibliography. Some of Kahan’s articles, some previous \TeX\ fp packages, the international standards,…

• Also take into account the “inexact” exception?

• Support multi-character prefix operators (e.g., @/ or whatever)?
Chapter 29

The l3fparray package: fast global floating point arrays

29.1 l3fparray documentation

For applications requiring heavy use of floating points, this module provides arrays which can be accessed in constant time (contrast l3seq, where access time is linear). The interface is very close to that of l3intarray. The size of the array is fixed and must be given at point of initialisation.

\fparray_new:Nn \fparray_new:Nn \langle fparray var \rangle \{\langle size\rangle\}

Evaluates the integer expression \langle size\rangle and allocates an \langle floating point array variable\rangle with that number of (zero) entries. The variable name should start with \g_ because assignments are always global.

\fparray_count:N \fparray_count:N \langle fparray var \rangle

Expands to the number of entries in the \langle floating point array variable\rangle. This is performed in constant time.

\fparray_gset:Nnn \fparray_gset:Nnn \langle fparray var \rangle \{\langle position\rangle\} \{\langle value\rangle\}

Stores the result of evaluating the floating point expression \langle value\rangle into the \langle floating point array variable\rangle at the (integer expression) \langle position\rangle. If the \langle position\rangle is not between 1 and the \fparray_count:N, an error occurs. Assignments are always global.

\fparray_gzero:N \fparray_gzero:N \langle fparray var \rangle

Sets all entries of the \langle floating point array variable\rangle to +0. Assignments are always global.

\fparray_item:Nn \fparray_item:Nn \langle fparray var \rangle \{\langle position\rangle\}

Applies \fp_use:N or \fp_to_tl:N (respectively) to the floating point entry stored at the (integer expression) \langle position\rangle in the \langle floating point array variable\rangle. If the \langle position\rangle is not between 1 and the \fparray_count:N, an error occurs.
Chapter 30

The l3cctab package
Category code tables

A category code table enables rapid switching of all category codes in one operation. For LuaLaTeX, this is possible over the entire Unicode range. For other engines, only the 8-bit range (0–255) is covered by such tables.

30.1 Creating and initialising category code tables

\cctab_new:N \cctab_new:c

Creates a new \textit{category code table} variable or raises an error if the name is already taken. The declaration is global. The \textit{category code table} is initialised with the codes as used by ini\TeX{}.

\cctab_const:Nn \cctab_const:cn

Creates a new \textit{category code table}, applies (in a group) the \textit{category code set up} on top of ini\TeX{} settings, then saves them globally as a constant table. The \textit{category code set up} can include a call to \cctab_select:N.

\cctab_gset:Nn \cctab_gset:cn

Starting from the ini\TeX{} category codes, applies (in a group) the \textit{category code set up}, then saves them globally in the \textit{category code table}. The \textit{category code set up} can include a call to \cctab_select:N.

30.2 Using category code tables

\cctab_begin:N \cctab_begin:c

Switches locally the category codes in force to those stored in the \textit{category code table}. The prevailing codes before the function is called are added to a stack, for use with \cctab_end:. This function does not start a \TeX{} group.
\cctab_end:

Ends the scope of a \langle category code table \rangle started using \cctab_begin:N, returning the codes to those in force before the matching \cctab_begin:N was used. This must be used within the same \TeX{} group (and at the same \TeX{} group level) as the matching \cctab_begin:N.

\cctab_select:N \cctab_select:c

Selects the \langle category code table \rangle for the scope of the current group. This is in particular useful in the \langle setup \rangle arguments of \tl_set_rescan:Nnn, \tl_rescan:nn, \cctab_const:Nn, and \cctab_gset:Nn.

\cctab_item:Nn \cctab_item:cn ⋆

\cctab_item:Nn ⋆ \cctab_item:cn ⋆

\cctab_if_exist_p:N \cctab_if_exist:p:c \cctab_if_exist:NTF \cctab_if_exist:cTF

\cctab_if_exist_p:N ⋆ \cctab_if_exist_p:c ⋆ \cctab_if_exist:NTF ⋆ \cctab_if_exist:cTF ⋆

Tests whether the \langle category code table \rangle is currently defined. This does not check that the \langle category code table \rangle really is a category code table.

### 30.3 Category code table conditionals

\c_code_cctab

Category code table for the expl3 code environment; this does not include \&, which is retained as an “other” character.

\c_document_cctab

Category code table for a standard \LaTeX{} document, as set by the \LaTeX{} kernel. In particular, the upper-half of the 8-bit range will be set to “active” with pdf\TeX{} only. No babel shorthands will be activated.

\c_initex_cctab

Category code table as set up by ini\TeX{}.

\c_other_cctab

Category code table where all characters have category code 12 (other).

\c_str_cctab

Category code table where all characters have category code 12 (other) with the exception of spaces, which have category code 10 (space).

262
Part V

Text manipulation
Chapter 31

The l3unicode package: Unicode support functions

This module provides Unicode-specific functions along with loading data from a range of Unicode Consortium files. At present, it provides no public functions.
Chapter 32

The l3text package: text processing

This module deals with manipulation of (formatted) text; such material is comprised of a restricted set of token list content. The functions provided here concern conversion of textual content for example in case changing, generation of bookmarks and extraction to tags. All of the major functions operate by expansion. Begin-group and end-group tokens in the ⟨text⟩ are normalized and become { and }, respectively.

32.1 Expanding text

\text\_expand\_n \text\_expand\_n \text\_expand\_n \text\_expand\_n ⟨text⟩

Takes user input ⟨text⟩ and expands the content. Protected commands (typically formatting) are left in place, and no processing takes place of math mode material (as delimited by pairs given in \l_text\_math\_delims\_tl or as the argument to commands listed in \l_text\_math\_arg\_tl). Commands which are neither engine- nor \LaTeX protected are expanded exhaustively. Any commands listed in \l_text\_expand\_exclude\_tl, \l_text\_accents\_tl and \l_text\_letterlike\_tl are excluded from expansion.

\text\_declare\_expand\_equivalent\_Nn \text\_declare\_expand\_equivalent\_Nn ⟨cmd⟩ \{⟨replacement⟩\}

\text\_declare\_expand\_equivalent\_cn

Declares that the ⟨replacement⟩ tokens should be used whenever the ⟨cmd⟩ (a single token) is encountered. The ⟨replacement⟩ tokens should be expandable.
32.2 Case changing

\text_lowercasen \star \text_uppercasen \{(tokens)\}
\text_uppercasen \star \text_uppercasen \{(language)\} \{(tokens)\}
\text_titlecasen \star \text_titlecasen \{(language)\} \{(tokens)\}
\text_titlecase_firstn \star \text_titlecasen \star \text_lowercasen \star \text_uppercasen \star \text_titlecasen \star \text_titlecase_firstn \star

Takes user input (text) first applies \text_expand, then transforms the case of character tokens as specified by the function name. The category code of letters are not changed by this process (at least where they can be represented by the engine as a single token: 8-bit engines may require active characters).

Upper- and lowercase have the obvious meanings. Titlecasing may be regarded informally as converting the first character of the (tokens) to uppercase and the rest to lowercase. However, the process is more complex than this as there are some situations where a single lowercase character maps to a special form, for example ij in Dutch which becomes IJ. The titlecase_first variant does not attempt any case changing at all after the first letter has been processed.

Importantly, notice that these functions are intended for working with user text for typesetting. For case changing programmatic data see the \lstr module and discussion there of \str_lowercase, \str_uppercase and \str_foldcase.

Case changing does not take place within math mode material so for example

\text_uppercasen \{ Some-text-$y = mx + c$-with\{Braces\} \}

becomes

SOME TEXT \$y = mx + c$ WITH \{BRACES\}

The arguments of commands listed in \text_case_exclude_arg_tl are excluded from case changing; the latter are entirely non-textual content (such as labels).

As is generally true for expl3, these functions are designed to work with Unicode input only. As such, UTF-8 input is assumed for all engines. When used with Xe\TeX{} or Lua\TeX{} a full range of Unicode transformations are enabled. Specifically, the standard mappings here follow those defined by the Unicode Consortium in \texttt{UnicodeData.txt} and \texttt{SpecialCasing.txt}. In the case of 8-bit engines, mappings are provided for characters which can be represented in output typeset using the T1, T2 and LGR font encodings. Thus for example ä can be case-changed using pdf\TeX. For \TeX{} only the ASCII range is covered as the engine treats input outside of this range as east Asian.

Language-sensitive conversions are enabled using the (language) argument, and follow Unicode Consortium guidelines. Currently, the languages recognised for special handling are as follows.

- Azeri and Turkish (az and tr). The case pairs I/i-dotless and I-dot/i are activated for these languages. The combining dot mark is removed when lowercasing I-dot and introduced when upper casing i-dotless.

- German (de-xeszett). An alternative mapping for German in which the lowercase Eszett maps to a großes Eszett. Since there is a T1 slot for the großes Eszett in T1, this tailoring is available with pdf\TeX{} as well as in the Unicode \TeX{} engines.
• Greek (el). Removes accents from Greek letters when uppercasing; titlecasing leaves accents in place. (At present this is implemented only for Unicode engines.) A variant el-xiota is available which retains the ypogegrammeni (subscript muted iota) when uppercasing: the standard version converts these to a following capital iota.

• Lithuanian (lt). The lowercase letters i and j should retain a dot above when the accents grave, acute or tilde are present. This is implemented for lowercasing of the relevant uppercase letters both when input as single Unicode codepoints and when using combining accents. The combining dot is removed when uppercasing in these cases. Note that only the accents used in Lithuanian are covered: the behaviour of other accents are not modified.

• Dutch (nl). Capitalisation of ij at the beginning of titlecased input produces IJ rather than Ij. The output retains two separate letters, thus this transformation is available using pdfTeX.

For titlecasing, note that there are two functions available. The function \textitlecase:n applies (broadly) uppercasing to the first letter of the input, then lowercasing to the remainder. In contrast, \textitlecase_first:n only carries out the uppercasing operation, and leaves the balance of the input unchanged. Determining whether non-letter characters at the start of text should switch from upper- to lowercasing is controllable. When \ltextitlecase_check_letter_bool is true, characters which are not letters (category code 11) are left unchanged and “skipped”: the first letter is uppercased. (With 8-bit engines, this is extended to active characters which form part of a multi-byte letter codepoint.) When \ltextitlecase_check_letter_bool is false, the first character is uppercased, and the rest lowercased, irrespective of the nature of the character.

32.3 Removing formatting from text

\text_purify:n \{text\} \text_purify:n \{text\}

Takes user input \langle text \rangle and expands as described for \text_expand:n, then removes all functions from the resulting text. Math mode material (as delimited by pairs given in \l_text_math_delims_tl or as the argument to commands listed in \l_text_math_arg_tl) is left contained in a pair of $ delimiters. Non-expandable functions present in the \langle text \rangle must either have a defined equivalent (see \text_declare_purify_equivalent:Nn) or will be removed from the result. Implicit tokens are converted to their explicit equivalent.

\text_declare_purify_equivalent:Nn \text_declare_purify_equivalent:Nx \text_declare_purify_equivalent:Nn \{cmd\} \{replacement\} \text_declare_purify_equivalent:Nx

Declares that the \langle replacement \rangle tokens should be used whenever the \langle cmd \rangle (a single token) is encountered. The \langle replacement \rangle tokens should be expandable.
32.4 Control variables

\l_text_accents_tl  Lists commands which represent accents, and which are left unchanged by expansion. (Defined only for the \LaTeX2ε package.)

\l_text_letterlike_tl  Lists commands which represent letters; these are left unchanged by expansion. (Defined only for the \LaTeX2ε package.)

\l_text_math_arg_tl  Lists commands present in the \text{⟨text⟩} where the argument of the command should be treated as math mode material. The treatment here is similar to \l_text_math_delims_tl but for a command rather than paired delimiters.

\l_text_math_delims_tl  Lists pairs of tokens which delimit (in-line) math mode content; such content may be excluded from processing.

\l_text_case_exclude_arg_tl  Lists commands which are excluded from case changing. This protection includes everything up to and including their first braced argument.

\l_text_expand_exclude_tl  Lists commands which are excluded from expansion. This protection includes everything up to and including their first braced argument.

\l_text_titlecase_check_letter_bool  Controls how the start of titlecasing is handled: when \text{true}, the first \text{letter} in text is considered. The standard setting is \text{true}.
Part VI

Typesetting
Chapter 33
The l3box package

Boxes

Box variables contain typeset material that can be inserted on the page or in other boxes. Their contents cannot be converted back to lists of tokens. There are three kinds of box operations: horizontal mode denoted with prefix \hbox_, vertical mode with prefix \vbox_, and the generic operations working in both modes with prefix \box_. For instance, a new box variable containing the words “Hello, world!” (in a horizontal box) can be obtained by the following code.

\box_new:N \l_hello_box
\hbox_set:Nn \l_hello_box { Hello, ~ world! }

The argument is typeset inside a TEX group so that any variables assigned during the construction of this box restores its value afterwards.

Box variables from l3box are compatible with those of \LaTeX{} and plain TEX and can be used interchangeably. The l3box commands to construct boxes, such as \hbox:n or \hbox_set:Nn, are “color-safe”, meaning that

\hbox:n { \color_select:n { blue } Hello, } ~ world!

will result in “Hello,” taking the color blue, but “world!” remaining with the prevailing color outside the box.

33.1 Creating and initialising boxes

\box_new:N \box_new:c
\box_clear:N \box_clear:c \box_gclear:N \box_gclear:c

\box_new:N \box_new:c

Creates a new \langle box \rangle or raises an error if the name is already taken. The declaration is global. The \langle box \rangle is initially void.

\box_clear:N \box_clear:c \box_gclear:N \box_gclear:c

Clears the content of the \langle box \rangle by setting the box equal to \c_empty_box.
\box_clear_new:N  
\box_clear_new:c  
\box_gclear_new:N  
\box_gclear_new:c  

\box_set_eq:NN  
\box_set_eq:cnNc  
\box_gset_eq:NN  
\box_gset_eq:cnNc  

\box_if_exist_p:N  
\box_if_exist:NTF  
\box_if_exist:pNc  
\box_if_exist:cNTF  
\box_if_exist:cP  

\box_use:N  
\box_use:c  

\box_move_right:nn  
\box_move_left:nn  

\box_move_up:nn  
\box_move_down:nn  

33.2 Using boxes

\box_use:N  
\box_use:c  

Inserts the current content of the \textit{box} onto the current list for typesetting. An error is raised if the variable does not exist or if it is invalid.

\textbf{\LaTeX{}hackers note:} This is the \LaTeX{} primitive \texttt{\copy{}}.

\box_move_right:nn  
\box_move_left:nn  

This function operates in vertical mode, and inserts the material specified by the \textit{box function} such that its reference point is displaced horizontally by the given \texttt{(dimexpr)} from the reference point for typesetting, to the right or left as appropriate. The \textit{box function} should be a box operation such as \texttt{\box_use:N \box} or a “raw” box specification such as \texttt{\vbox:n \{ \texttt{xyz} \}}.

\box_move_up:nn  
\box_move_down:nn  

This function operates in horizontal mode, and inserts the material specified by the \textit{box function} such that its reference point is displaced vertically by the given \texttt{(dimexpr)} from the reference point for typesetting, up or down as appropriate. The \textit{box function} should be a box operation such as \texttt{\box_use:N \box} or a “raw” box specification such as \texttt{\vbox:n \{ \texttt{xyz} \}}.

New: 2012-03-03

271
### 33.3 Measuring and setting box dimensions

- **\box_dp:N**  
  **\box_dp:c**  
  Calculates the depth (below the baseline) of the \(\langle box\rangle\) in a form suitable for use in a \langle dimension expression \rangle.

  **\TeXhackers note:** This is the \TeX primitive \dp.

- **\box_ht:N**  
  **\box_ht:c**  
  Calculates the height (above the baseline) of the \(\langle box\rangle\) in a form suitable for use in a \langle dimension expression \rangle.

  **\TeXhackers note:** This is the \TeX primitive \ht.

- **\box_wd:N**  
  **\box_wd:c**  
  Calculates the width of the \(\langle box\rangle\) in a form suitable for use in a \langle dimension expression \rangle.

  **\TeXhackers note:** This is the \TeX primitive \wd.

- **\box_ht_plus_dp:N**  
  **\box_ht_plus_dp:c**  
  Calculates the total vertical size (height plus depth) of the \(\langle box\rangle\) in a form suitable for use in a \langle dimension expression \rangle.

- **\box_set_dp:Nn**  
  **\box_set_dp:cn**  
  **\box_gset_dp:Nn**  
  **\box_gset_dp:cn**  
  Set the depth (below the baseline) of the \(\langle box\rangle\) to the value of the \(\{\langle dimension expression \rangle\}\).

  **\TeXhackers note:** This is the \TeX primitive \dp.

- **\box_set_ht:Nn**  
  **\box_set_ht:cn**  
  **\box_gset_ht:Nn**  
  **\box_gset_ht:cn**  
  Set the height (above the baseline) of the \(\langle box\rangle\) to the value of the \(\{\langle dimension expression \rangle\}\).

- **\box_set_wd:Nn**  
  **\box_set_wd:cn**  
  **\box_gset_wd:Nn**  
  **\box_gset_wd:cn**  
  Set the width of the \(\langle box\rangle\) to the value of the \(\{\langle dimension expression \rangle\}\).
33.4 Box conditionals

\box_if_empty_p:N \box_if_empty_p:c \box_if_empty:NTF \box_if_empty:cTF

Tests if \langle box \rangle is a empty (equal to \c_empty_box).

\box_if_horizontal_p:N \box_if_horizontal_p:c \box_if_horizontal:NTF \box_if_horizontal:cTF

Tests if \langle box \rangle is a horizontal box.

\box_if_vertical_p:N \box_if_vertical_p:c \box_if_vertical:NTF \box_if_vertical:cTF

Tests if \langle box \rangle is a vertical box.

33.5 The last box inserted

\box_set_to_last:N \box_set_to_last:c \box_gset_to_last:N \box_gset_to_last:c

Sets the \langle box \rangle equal to the last item (box) added to the current partial list, removing the item from the list at the same time. When applied to the main vertical list, the \langle box \rangle is always void as it is not possible to recover the last added item.

33.6 Constant boxes

\c_empty_box

This is a permanently empty box, which is neither set as horizontal nor vertical.\TeXhackers note: At the \TeX level this is a void box.

33.7 Scratch boxes

\l_tmpa_box \l_tmpb_box

Scratch boxes for local assignment. These are never used by the kernel code, and so are safe for use with any \EM\TeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.

\g_tmpa_box \g_tmpb_box

Scratch boxes for global assignment. These are never used by the kernel code, and so are safe for use with any \EM\TeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
33.8 Viewing box contents

\box_show:N\box_show:c
Updated: 2012-05-11
\box_show:N (box)
Shows full details of the content of the (box) in the terminal.

\box_show:Nnn\box_show:cn
Updated: 2012-05-11
\box_show:Nnn (box) \{\intexpr_1\} \{\intexpr_2\}
Display the contents of (box) in the terminal, showing the first \intexpr_1 items of the box, and descending into \intexpr_2 group levels.

\box_log:N\box_log:c
Updated: 2012-05-11
\box_log:N (box)
Writes full details of the content of the (box) to the log.

\box_log:Nnn\box_log:cn
Updated: 2012-05-11
\box_log:Nnn (box) \{\intexpr_1\} \{\intexpr_2\}
Writes the contents of (box) to the log, showing the first \intexpr_1 items of the box, and descending into \intexpr_2 group levels.

33.9 Boxes and color

All \LaTeX3 boxes are “color safe”: a color set inside the box stops applying after the end of the box has occurred.

33.10 Horizontal mode boxes

\hbox:n\hbox_set:NN\hbox_gset:NN\hbox_to_zero:n
Updated: 2017-04-05
\hbox:n \{\contents\}
Typesets the \contents into a horizontal box of natural width and then includes this box in the current list for typesetting.

\hbox_to_wd:nn\hbox_to_wd:cn\hbox_gset:NN\hbox_gset:cn
Updated: 2017-04-05
\hbox_to_wd:nn \{\dimexpr\} \{\contents\}
Typesets the \contents into a horizontal box of width \dimexpr and then includes this box in the current list for typesetting.

\hbox_to_zero:n\hbox_to_zero:cn
Updated: 2017-04-05
\hbox_to_zero:n \{\contents\}
Typesets the \contents into a horizontal box of zero width and then includes this box in the current list for typesetting.

\hbox_set:Nn\hbox_set:cn\hbox_gset:Nn\hbox_gset:cn
Updated: 2017-04-05
\hbox_set:Nn \{\contents\}
Typesets the \contents at natural width and then stores the result inside the (box).
\hbox_set:WNw \hbox_set:WCw \hbox_set_end: \hbox_unpack:N
Unpacks the content of the horizontal \textit{box}, retaining any stretching or shrinking applied when the \textit{box} was set.

\textbf{TeXhackers note}: This is the \TeX{} primitive \texttt{\unhcopy}.  

\section*{33.11 Vertical mode boxes}

Vertical boxes inherit their baseline from their contents. The standard case is that the baseline of the box is at the same position as that of the last item added to the box. This means that the box has no depth unless the last item added to it had depth. As a result most vertical boxes have a large height value and small or zero depth. The exception are
_top boxes, where the reference point is that of the first item added. These tend to have a large depth and small height, although the latter is typically non-zero.

\vbox:n \{(contents)\}

Typesets the \(\langle \text{contents} \rangle\) into a vertical box of natural height and includes this box in the current list for typesetting.

\vbox_top:n \{(contents)\}

Typesets the \(\langle \text{contents} \rangle\) into a vertical box of natural height and includes this box in the current list for typesetting. The baseline of the box is equal to that of the first item added to the box.

\vbox_to_ht:nn \{(dimexpr)\} \{(contents)\}

Typesets the \(\langle \text{contents} \rangle\) into a vertical box of height \(\langle \text{dimexpr} \rangle\) and then includes this box in the current list for typesetting.

\vbox_to_zero:n \{(contents)\}

Typesets the \(\langle \text{contents} \rangle\) into a vertical box of zero height and then includes this box in the current list for typesetting.

\vbox_set:Nn \langle box \rangle \{(contents)\}

Typesets the \(\langle \text{contents} \rangle\) at natural height and then stores the result inside the \(\langle \text{box} \rangle\).

\vbox_set_top:Nn \langle box \rangle \{(contents)\}

Typesets the \(\langle \text{contents} \rangle\) at natural height and then stores the result inside the \(\langle \text{box} \rangle\). The baseline of the box is equal to that of the first item added to the box.

\vbox_set_to_ht:NNn \langle box \rangle \{(dimexpr)\} \{(contents)\}

Typesets the \(\langle \text{contents} \rangle\) to the height given by the \(\langle \text{dimexpr} \rangle\) and then stores the result inside the \(\langle \text{box} \rangle\).

\vbox_set:Nw \langle box \rangle \langle \text{contents} \rangle \vbox_set_end:

Typesets the \(\langle \text{contents} \rangle\) at natural height and then stores the result inside the \(\langle \text{box} \rangle\). In contrast to \vbox_set:Nn this function does not absorb the argument when finding the \(\langle \text{content} \rangle\), and so can be used in circumstances where the \(\langle \text{content} \rangle\) may not be a simple argument.
\vbox_set_to_ht:Nnw \vbox_set_to_ht:cnw \vbox_gset_to_ht:Nnw \vbox_gset_to_ht:cnw

\vbox_set_split_to_ht:NNn \vbox_set_split_to_ht:(cNn|Ncn|ccn) \vbox_gset_split_to_ht:NNn \vbox_gset_split_to_ht:(cNn|Ncn|ccn)

\vbox_unpack:N \vbox_unpack:c

33.12 Using boxes efficiently

The functions above for using box contents work in exactly the same way as for any other expl3 variable. However, for efficiency reasons, it is also useful to have functions which drop box contents on use. When a box is dropped, the box becomes empty at the group level where the box was originally set rather than necessarily at the current group level. For example, with

\hbox_set:Nn \l_tmpa_box { A } \group_begin: \hbox_set:Nn \l_tmpa_box { B } \group_begin: \box_use_drop:N \l_tmpa_box \group_end: \box_show:N \l_tmpa_box \group_end: \box_show:N \l_tmpa_box

the first use of \box_show:N will show an entirely cleared (void) box, and the second will show the letter A in the box.

These functions should be preferred when the content of the box is no longer required after use. Note that due to the unusual scoping behaviour of drop functions they may be applied to both local and global boxes: the latter will naturally be set and thus cleared at a global level.
\box_use_drop:N \box_use_drop:c

Inserts the current content of the \texttt{<box>} onto the current list for typesetting then drops the box content. An error is raised if the variable does not exist or if it is invalid. This function may be applied to local or global boxes.

\textbf{\TeX}hackers note: This is the \texttt{\box} primitive.

\box_set_eq_drop:NN \box_set_eq_drop:(ch|nc|cc)

Sets the content of \texttt{<box>_1} equal to that of \texttt{<box>_2}, then drops \texttt{<box>_2}.

\textbf{\TeX}hackers note: New: 2019-01-17

\box_gset_eq_drop:NN \box_gset_eq_drop:(ch|nc|cc)

Sets the content of \texttt{<box>_1} globally equal to that of \texttt{<box>_2}, then drops \texttt{<box>_2}.

\textbf{\TeX}hackers note: New: 2019-01-17

\hbox_unpack_drop:N \hbox_unpack_drop:c

Unpacks the content of the horizontal \texttt{<box>}, retaining any stretching or shrinking applied when the \texttt{<box>} was set. The original \texttt{<box>} is then dropped.

\textbf{\TeX}hackers note: This is the \TeX{} primitive \texttt{\unhbox}.

\vbox_unpack_drop:N \vbox_unpack_drop:c

Unpacks the content of the vertical \texttt{<box>}, retaining any stretching or shrinking applied when the \texttt{<box>} was set. The original \texttt{<box>} is then dropped.

\textbf{\TeX}hackers note: This is the \TeX{} primitive \texttt{\unvbox}.

\section{Affine transformations}

Affine transformations are changes which (informally) preserve straight lines. Simple translations are affine transformations, but are better handled in \TeX{} by doing the translation first, then inserting an unmodified box. On the other hand, rotation and resizing of boxed material can best be handled by modifying boxes. These transformations are described here.
Resizes the ⟨box⟩ to fit within the given ⟨x-size⟩ (horizontally) and ⟨y-size⟩ (vertically); both of the sizes are dimension expressions. The ⟨y-size⟩ is the height only: it does not include any depth. The updated ⟨box⟩ is an hbox, irrespective of the nature of the ⟨box⟩ before the resizing is applied. The final size of the ⟨box⟩ is the smaller of {⟨x-size⟩} and {⟨y-size⟩}, i.e. the result fits within the dimensions specified. Negative sizes cause the material in the ⟨box⟩ to be reversed in direction, but the reference point of the ⟨box⟩ is unchanged. Thus a negative ⟨y-size⟩ results in the ⟨box⟩ having a depth dependent on the height of the original and vice versa.

Resizes the ⟨box⟩ to fit within the given ⟨x-size⟩ (horizontally) and ⟨y-size⟩ (vertically); both of the sizes are dimension expressions. The ⟨y-size⟩ is the total vertical size (height plus depth). The updated ⟨box⟩ is an hbox, irrespective of the nature of the ⟨box⟩ before the resizing is applied. The final size of the ⟨box⟩ is the smaller of {⟨x-size⟩} and {⟨y-size⟩}, i.e. the result fits within the dimensions specified. Negative sizes cause the material in the ⟨box⟩ to be reversed in direction, but the reference point of the ⟨box⟩ is unchanged. Thus a negative ⟨y-size⟩ results in the ⟨box⟩ having a depth dependent on the height of the original and vice versa.

Resizes the ⟨box⟩ to ⟨y-size⟩ (vertically), scaling the horizontal size by the same amount; ⟨y-size⟩ is a dimension expression. The ⟨y-size⟩ is the height only: it does not include any depth. The updated ⟨box⟩ is an hbox, irrespective of the nature of the ⟨box⟩ before the resizing is applied. A negative ⟨y-size⟩ causes the material in the ⟨box⟩ to be reversed in direction, but the reference point of the ⟨box⟩ is unchanged. Thus a negative ⟨y-size⟩ results in the ⟨box⟩ having a depth dependent on the height of the original and vice versa.
Resizes the `⟨box⟩` to `⟨y-size⟩` (vertically), scaling the horizontal size by the same amount; `⟨y-size⟩` is a dimension expression. The `⟨y-size⟩` is the total vertical size (height plus depth). The updated `⟨box⟩` is an `hbox`, irrespective of the nature of the `⟨box⟩` before the resizing is applied. A negative `⟨y-size⟩` causes the material in the `⟨box⟩` to be reversed in direction, but the reference point of the `⟨box⟩` is unchanged. Thus a negative `⟨y-size⟩` results in the `⟨box⟩` having a depth dependent on the height of the original and *vice versa*.

Resizes the `⟨box⟩` to `⟨x-size⟩` (horizontally), scaling the vertical size by the same amount; `⟨x-size⟩` is a dimension expression. The updated `⟨box⟩` is an `hbox`, irrespective of the nature of the `⟨box⟩` before the resizing is applied. A negative `⟨x-size⟩` causes the material in the `⟨box⟩` to be reversed in direction, but the reference point of the `⟨box⟩` is unchanged. Thus a negative `⟨x-size⟩` results in the `⟨box⟩` having a depth dependent on the height of the original and *vice versa*.

Resizes the `⟨box⟩` to `⟨x-size⟩` (horizontally) and `⟨y-size⟩` (vertically); both of the sizes are dimension expressions. The `⟨y-size⟩` is the height only and does not include any depth. The updated `⟨box⟩` is an `hbox`, irrespective of the nature of the `⟨box⟩` before the resizing is applied. Negative sizes cause the material in the `⟨box⟩` to be reversed in direction, but the reference point of the `⟨box⟩` is unchanged. Thus a negative `⟨y-size⟩` results in the `⟨box⟩` having a depth dependent on the height of the original and *vice versa*.

Resizes the `⟨box⟩` to `⟨x-size⟩` (horizontally) and `⟨y-size⟩` (vertically); both of the sizes are dimension expressions. The `⟨y-size⟩` is the total vertical size (height plus depth). The updated `⟨box⟩` is an `hbox`, irrespective of the nature of the `⟨box⟩` before the resizing is applied. Negative sizes cause the material in the `⟨box⟩` to be reversed in direction, but the reference point of the `⟨box⟩` is unchanged. Thus a negative `⟨y-size⟩` results in the `⟨box⟩` having a depth dependent on the height of the original and *vice versa*.
\box_rotate: Nn \box_rotate: cn
\box_grotate: Nn \box_grotate: cn
Updated: 2019-01-22

Rotates the \langle box \rangle by \langle angle \rangle (in degrees) anti-clockwise about its reference point. The reference point of the updated box is moved horizontally such that it is at the left side of the smallest rectangle enclosing the rotated material. The updated \langle box \rangle is an hbox, irrespective of the nature of the \langle box \rangle before the rotation is applied.

\box_scale: Nnn \box_scale: cn
\box_gscale: Nnn \box_gscale: cn
Updated: 2019-01-22

Scales the \langle box \rangle by factors \langle x-scale \rangle and \langle y-scale \rangle in the horizontal and vertical directions, respectively (both scales are integer expressions). The updated \langle box \rangle is an hbox, irrespective of the nature of the \langle box \rangle before the scaling is applied. Negative scalings cause the material in the \langle box \rangle to be reversed in direction, but the reference point of the \langle box \rangle is unchanged. Thus a negative \langle y-scale \rangle results in the \langle box \rangle having a depth dependent on the height of the original and vice versa.

33.14 Primitive box conditionals

\if_hbox: N * \if_hbox: N \langle box \rangle
\true code
\else:
\false code
\fi:
Tests is \langle box \rangle is a horizontal box.

\TeXhackers note: This is the \TeX primitive \ifhbox.

\if_vbox: N * \if_vbox: N \langle box \rangle
\true code
\else:
\false code
\fi:
Tests is \langle box \rangle is a vertical box.

\TeXhackers note: This is the \TeX primitive \ifvbox.

\if_box_empty: N * \if_box_empty: N \langle box \rangle
\true code
\else:
\false code
\fi:
Tests is \langle box \rangle is an empty (void) box.

\TeXhackers note: This is the \TeX primitive \ifvoid.
Chapter 34

The l3coffins package
Coffin code layer

The material in this module provides the low-level support system for coffins. For details about the design concept of a coffin, see the xcoffins module (in the l3experimental bundle).

34.1 Creating and initialising coffins

\coffin_new:N \coffin_new:c
New: 2011-08-17

Creates a new \texttt{coffin} or raises an error if the name is already taken. The declaration is global. The \texttt{coffin} is initially empty.

\coffin_clear:N \coffin_clear:c
\coffin_gclear:N \coffin_gclear:c
New: 2011-08-17
Updated: 2019-01-21

Clears the content of the \texttt{coffin}.

\coffin_set_eq:NN \coffin_gset_eq:NN
\coffin_set_eq:(Nc|N|cc) \coffin_gset_eq:(Nc|N|cc)
New: 2011-08-17
Updated: 2019-01-21

Sets both the content and poles of \texttt{coffin} equal to those of \texttt{coffin}.

\coffin_if_exist_p:N \coffin_if_exist_p:c \coffin_if_exist:NTF \coffin_if_exist:CTF
New: 2012-06-20

Tests whether the \texttt{coffin} is currently defined.
34.2 Setting coffin content and poles

\coffin_set:Nn \coffin_set:cn \coffin_gset:Nn \coffin_gset:cn

New: 2011-08-17
Updated: 2019-01-21

\coffin_set:Nw \coffin_set:cn \coffin_set_end:
\coffin_gset:Nw \coffin_gset:cn

New: 2011-09-10
Updated: 2019-01-21

\vcoffin_set:Nnn \vcoffin_set:cn \vcoffin_gset:Nnn \vcoffin_gset:cn

New: 2011-08-17
Updated: 2019-01-21

\vcoffin_set:Nnw \vcoffin_set:cnw \vcoffin_set_end:
\vcoffin_gset:Nnw \vcoffin_gset:cnw

New: 2011-09-10
Updated: 2019-01-21

\coffin_set_horizontal_pole:Nnn \coffin_set_horizontal_pole:cn \coffin_gset_horizontal_pole:Nnn \coffin_gset_horizontal_pole:cn

New: 2012-07-20
Updated: 2019-01-21

Sets the \textit{pole} to run horizontally through the \textit{coffin}. The \textit{pole} is placed at the \textit{offset} from the baseline of the \textit{coffin}. The \textit{offset} should be given as a dimension expression.

\coffin_set:Nn \langle\textit{coffin}\rangle \{\langle\textit{material}\rangle\}

Typesets the \textit{material} in horizontal mode, storing the result in the \textit{coffin}. The standard poles for the \textit{coffin} are then set up based on the size of the typeset material.

\coffin_set:Nw \langle\textit{coffin}\rangle \langle\textit{material}\rangle \coffin_set_end:

Typesets the \textit{material} in horizontal mode, storing the result in the \textit{coffin}. The standard poles for the \textit{coffin} are then set up based on the size of the typeset material. These functions are useful for setting the entire contents of an environment in a coffin.

\vcoffin_set:Nnn \langle\textit{coffin}\rangle \{\langle\textit{width}\rangle\} \{\langle\textit{material}\rangle\}

Typesets the \textit{material} in vertical mode constrained to the given \textit{width} and stores the result in the \textit{coffin}. The standard poles for the \textit{coffin} are then set up based on the size of the typeset material.

\vcoffin_set:Nnw \langle\textit{coffin}\rangle \{\langle\textit{width}\rangle\} \{\langle\textit{material}\rangle\} \vcoffin_set_end:

Typesets the \textit{material} in vertical mode constrained to the given \textit{width} and stores the result in the \textit{coffin}. The standard poles for the \textit{coffin} are then set up based on the size of the typeset material. These functions are useful for setting the entire contents of an environment in a coffin.
Sets the ⟨pole⟩ to run vertically through the ⟨coffin⟩. The ⟨pole⟩ is placed at the ⟨offset⟩ from the left-hand edge of the bounding box of the ⟨coffin⟩. The ⟨offset⟩ should be given as a dimension expression.

### 34.3 Coffin affine transformations

Resized the ⟨coffin⟩ to ⟨width⟩ and ⟨total-height⟩, both of which should be given as dimension expressions.

Rotates the ⟨coffin⟩ by the given ⟨angle⟩ (given in degrees counter-clockwise). This process rotates both the coffin content and poles. Multiple rotations do not result in the bounding box of the coffin growing unnecessarily.

Scales the ⟨coffin⟩ by factors ⟨x-scale⟩ and ⟨y-scale⟩ in the horizontal and vertical directions, respectively. The two scale factors should be given as real numbers.

### 34.4 Joining and using coffins

This function attaches ⟨coffin2⟩ to ⟨coffin1⟩ such that the bounding box of ⟨coffin1⟩ is not altered, i.e. ⟨coffin2⟩ can protrude outside of the bounding box of the coffin. The alignment is carried out by first calculating ⟨handle1⟩, the point of intersection of ⟨coffin1-pole1⟩ and ⟨coffin1-pole2⟩, and ⟨handle2⟩, the point of intersection of ⟨coffin2-pole1⟩ and ⟨coffin2-pole2⟩. ⟨coffin2⟩ is then attached to ⟨coffin1⟩ such that the relationship between ⟨handle1⟩ and ⟨handle2⟩ is described by the ⟨x-offset⟩ and ⟨y-offset⟩. The two offsets should be given as dimension expressions.
This function joins \langle coffin \rangle 1 to \langle coffin \rangle 2 such that the bounding box of \langle coffin \rangle 1 may expand. The new bounding box covers the area containing the bounding boxes of the two original coffins. The alignment is carried out by first calculating \langle handle \rangle 1, the point of intersection of \langle coffin-pole \rangle 1 and \langle coffin-pole \rangle 2, and \langle handle \rangle 2, the point of intersection of \langle coffin2-pole \rangle 1 and \langle coffin2-pole \rangle 2. \langle coffin \rangle 2 is then attached to \langle coffin \rangle 1 such that the relationship between \langle handle \rangle 1 and \langle handle \rangle 2 is described by the \langle x-offset \rangle and \langle y-offset \rangle. The two offsets should be given as dimension expressions.

Typesetting is carried out by first calculating \langle handle \rangle, the point of intersection of \langle pole \rangle 1 and \langle pole \rangle 2. The coffin is then typeset in horizontal mode such that the relationship between the current reference point in the document and the \langle handle \rangle is described by the \langle x-offset \rangle and \langle y-offset \rangle. The two offsets should be given as dimension expressions. Typesetting a coffin is therefore analogous to carrying out an alignment where the “parent” coffin is the current insertion point.

### 34.5 Measuring coffins

\begin{align*}
\text{\textbackslash coffin_dp:N} & \quad \text{\textbackslash coffin_dp:c} \\
\text{\textbackslash coffin_dp:N} & \quad \text{\textbackslash coffin_dp:c} \\
\text{\textbackslash coffin_ht:N} & \quad \text{\textbackslash coffin_ht:c} \\
\text{\textbackslash coffin_wd:N} & \quad \text{\textbackslash coffin_wd:c} \\
\end{align*}

Calculates the depth (below the baseline) of the \langle coffin \rangle in a form suitable for use in a \langle dimension expression \rangle.

Calculates the height (above the baseline) of the \langle coffin \rangle in a form suitable for use in a \langle dimension expression \rangle.

Calculates the width of the \langle coffin \rangle in a form suitable for use in a \langle dimension expression \rangle.

### 34.6 Coffin diagnostics

\begin{align*}
\text{\textbackslash coffin_display_handles:N} & \quad \text{\textbackslash coffin_display_handles:N} \\
\text{\textbackslash coffin_display_handles:N} & \quad \text{\textbackslash coffin_display_handles:N} \\
\end{align*}

This function first calculates the intersections between all of the \langle poles \rangle of the \langle coffin \rangle to give a set of \langle handles \rangle. It then prints the \langle coffin \rangle at the current location in the source, with the position of the \langle handles \rangle marked on the coffin. The \langle handles \rangle are labelled as part of this process: the locations of the \langle handles \rangle and the labels are both printed in the \langle color \rangle specified.
This function first calculates the ⟨handle⟩ for the ⟨coffin⟩ as defined by the intersection of ⟨pole1⟩ and ⟨pole2⟩. It then marks the position of the ⟨handle⟩ on the ⟨coffin⟩. The ⟨handle⟩ are labelled as part of this process: the location of the ⟨handle⟩ and the label are both printed in the ⟨color⟩ specified.

This function shows the structural information about the ⟨coffin⟩ in the terminal. The width, height and depth of the typeset material are given, along with the location of all of the poles of the coffin.

Notice that the poles of a coffin are defined by four values: the $x$ and $y$ co-ordinates of a point that the pole passes through and the $x$- and $y$-components of a vector denoting the direction of the pole. It is the ratio between the later, rather than the absolute values, which determines the direction of the pole.

This function writes the structural information about the ⟨coffin⟩ in the log file. See also \coffin_show_structure:N which displays the result in the terminal.

Shows full details of poles and contents of the ⟨coffin⟩ in the terminal or log file. See \coffin_show_structure:N and \box_show:N to show separately the pole structure and the contents.

Shows poles and contents of the ⟨coffin⟩ in the terminal or log file, showing the first ⟨intexpr1⟩ items in the coffin, and descending into ⟨intexpr2⟩ group levels. See \coffin_show_structure:N and \box_show:N to show separately the pole structure and the contents.

34.7 Constants and variables

A permanently empty coffin.

Scratch coffins for local assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX3-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
Scratch coffins for global assignment. These are never used by the kernel code, and so are safe for use with any \LaTeX\-defined function. However, they may be overwritten by other non-kernel code and so should only be used for short-term storage.
Chapter 35

The l3color package

Color support

35.1 Color in boxes

Controlling the color of text in boxes requires a small number of control functions, so that the boxed material uses the color at the point where it is set, rather than where it is used.

\color_group_begin:
\color_group_end:

New: 2011-09-03

Creates a color group: one used to “trap” color settings. This grouping is built in to for example \hbox_set:Nn.

\color_ensure_current:

New: 2011-09-03

Ensures that material inside a box uses the foreground color at the point where the box is set, rather than that in force when the box is used. This function should usually be used within a \color_group_begin: ... \color_group_end: group.

35.2 Color models

A color model is a way to represent sets of colors. Different models are particularly suitable for different output methods, e.g. screen or print. Parameter-based models can describe a very large number of unique colors, and have a varying number of axes which define a color space. In contrast, various proprietary models are available which define spot colors (more formally separations).

Core models are used to pass color information to output; these are “native” to l3color. Core models use real numbers in the range \([0,1]\) to represent values. The core models supported here are

- gray Grayscale color, with a single axis running from 0 (fully black) to 1 (fully white)
- rgb Red-green-blue color, with three axes, one for each of the components
• **cmyk** Cyan-magenta-yellow-black color, with four axes, one for each of the components.

There are also interface models: these are convenient for users but have to be manipulated before storing/passing to the backend. Interface models are primarily integer-based: see below for more detail. The supported interface models are

- **Gray** Grayscale color, with a single axis running from 0 (fully black) to 15 (fully white)
- **hsb** Hue-saturation-brightness color, with three axes, all real values in the range [0, 1] for hue saturation and brightness
- **Hsb** Hue-saturation-brightness color, with three axes, integer in the range [0, 360] for hue, real values in the range [0, 1] for saturation and brightness
- **HSB** Hue-saturation-brightness color, with three axes, integers in the range [0, 240] for hue, saturation and brightness
- **HTML** HTML format representation of RGB color given as a single six-digit hexadecimal number
- **RGB** Red-green-blue color, with three axes, one for each of the components, values as integers from 0 to 255
- **wave** Light wavelength, a real number in the range 380 to 780 (nanometres)

All interface models are internally stored as **rgb**.

To allow parsing of data from **xcolor**, any leading model up the first : will be discarded; the approach of selecting an internal form for data is not used in **l3color**.

Additional models may be created to allow mixing of separation colors with each other or with those from other models. See Section 35.9 for more detail of color support for additional models.

When color is selected by model, the **values** given are specified as a comma-separated list. The length of the list will therefore be determined by the detail of the model involved.

Color models (and interconversion) are complex, and more details are given in the manual to the **l3tex2e xcolor** package and in the *PostScript Language Reference Manual*, published by Addison–Wesley.

### 35.3 Color expressions

In addition to allowing specification of color by model and values, **l3color** also supports color expressions. These are created by combining one or more color names, with the amount of each specified as a percentage. The latter is given between ! symbols in the expression. Thus for example

```
red!50!green
```

is a mixture of 50% red and 50% green. A trailing percentage is interpreted as implicitly followed by **white**, and so

```
red!25
```
specifies 25% red mixed with 75% white.

Where the models for the mixed colors are different, the model of the first color is used. Thus

\texttt{red!50!cyan}

will result in a color specification using the \texttt{rgb} model, made up of 50% red and 50% of cyan \textit{expressed in rgb}. This may be important as color model interconversion is not exact.

The one exception to the above is where the first model in an expression is \texttt{gray}. In this case, the order of mixing is “swapped” internally, so that for example

\texttt{black!50!red}

has the same result as

\texttt{red!50!black}

(the predefined colors \texttt{black} and \texttt{white} use the \texttt{gray} model).

Where more than two colors are mixed in an expression, evaluation takes place in a stepwise fashion. Thus in

\texttt{cyan!50!magenta!10!yellow}

the sub-expression

\texttt{cyan!50!magenta}

is first evaluated to give an intermediate color specification, before the second step

\texttt{<intermediate>!10!yellow}

where \texttt{<intermediate>} represents this transitory calculated value.

Within a color expression, \texttt{.} may be used to represent the color active for typesetting (the current color). This allows for example

\texttt{.!50}

to mean a mixture of 50% of current color with white.

(Color expressions supported here are a subset of those provided by the \LaTeX\ \texttt{xcolor} package. At present, only such features as are clearly useful have been added here.)

### 35.4 Named colors

Color names are stored in a single namespace, which makes them accessible as part of color expressions. Whilst they are not reserved in a technical sense, the names \texttt{black}, \texttt{white}, \texttt{red}, \texttt{green}, \texttt{blue}, \texttt{cyan}, \texttt{magenta} and \texttt{yellow} have special meaning and should not be redefined. Color names should be made up of letters, numbers and spaces only: other characters are reserved for use in color expressions. In particular, \texttt{.} represents the current color at the start of a color expression.

\begin{verbatim}
\color_set:nn \color_set:nn \{name\} \{(color expression)\}
\end{verbatim}

Evaluates the \texttt{(color expression)} and stores the resulting color specification as the \texttt{(name)}.
Stores the color specification equivalent to the \langle model(s) \rangle and \langle values \rangle as the \langle name \rangle.

Copies the color specification in \langle name2 \rangle to \langle name1 \rangle. The special name \texttt{.} may be used to represent the current color, allowing it to be saved to a name.

Displays the color specification stored in the \langle name \rangle on the terminal or log file.

35.5 Selecting specification colors

General selection of color is safe when split across pages: a stack is used to ensure that the correct color is re-selected on the new page.

These commands set the current color (\texttt{.}): other more specialised functions such as fill and stroke selectors do \textit{not} adjust this value.

Parses the \langle color expression \rangle and then activates the resulting color specification for typeset material.

Activates the color specification equivalent to the \langle model(s) \rangle and \langle value(s) \rangle for typeset material.

When this is set to a non-empty value, colors will be converted to the specified model when they are selected. Note that included images and similar are not influenced by this setting.

35.6 Colors for fills and strokes

Colors for drawing operations and so forth are split into strokes and fills (the latter may also be referred to as non-stroke color). The fill color is used for text under normal circumstances. Depending on the backend, stroke color may use a \textit{stack}, in which case it exhibits the same page breaking behavior as general color. However, \texttt{dvips/dvisvgm} do not support this, and so color will need to be contained within a scope, such as \texttt{\draw_begin:/\draw_end:}.

Parses the \langle color expression \rangle and then activates the resulting color specification for filling or stroking.
\color_fill:nn \color_fill {\langle model(s)\rangle\}\{\langle value(s)\rangle\}
Activates the color specification equivalent to the \langle model(s)\rangle and \langle value(s)\rangle for filling or stroking.

\color_stroke:nn
When using dvips, this PostScript variables hold the stroke color.

35.6.1 Coloring math mode material
Coloring math mode material using \color_select:nn(n) has some restrictions and often leads to spacing issues and/or poor input syntax. Avoiding generating \mathord atoms whilst coloring only those parts of the input which are required needs careful handling. The functionality here covers this important use case.

\color_math:nn \color_math {\langle color expression\rangle}\{\langle content\rangle\}
\color_math:nnn \color_math {\langle model(s)\rangle}\{\langle value(s)\rangle\}\{\langle content\rangle\}
Works as for \color_select:n(n) but applies color only to the math mode \langle content\rangle. The function does not generate a group and the \langle content\rangle therefore retains its math atom states. Sub/superscripts are also properly handled.

\l_color_math_active_tl
New: 2022-01-26
This list controls which tokens are considered as math active and should therefore be replaced by their definition during searching for sub/superscripts.

35.7 Multiple color models
When selecting or setting a color with an explicit model, it is possible to give values for more than one model at one time. This is particularly useful where automated conversion between models does not give the desired outcome. To do this, the list of models and list of values are both subdivided using / characters (as for the similar function in xcolor).

For example, to save a color with explicit cmyk and rgb values, one could use

\color_set:nnn { foo } { cmyk / rgb }
\{ 0.1 , 0.2 , 0.3 , 0.4 / 0.1, 0.2 , 0.3 \}

The manually-specified conversion will be used in preference to automated calculation whenever the model(s) listed are used: both in expressions and when a fixed model is active.

Similarly, the same syntax can be applied to directly selecting a color.

\color_select:nn { cmyk / rgb }
\{ 0.1 , 0.2 , 0.3 , 0.4 / 0.1, 0.2 , 0.3 \}

Again, this list is used when a fixed model is active: the first entry is used unless there is a fixed model matching one of the other entries.
35.8 Exporting color specifications

The major use of color expressions is in setting typesetting output, but there are other places in which some form of color information is required. These may need data in a different format or using a different model to the internal representation. Thus a set of functions are available to export colors in different formats.

Valid export targets are

- **backend** Two brace groups: the first containing the model, the second containing space-separated values appropriate for the model; this is the format required by backend functions of expl3
- **comma-sep-cmyk** Comma-separated cyan-magenta-yellow-black values
- **comma-sep-rgb** Comma-separated red-green-blue values suitable for use as a PDF annotation color
- **HTML** Uppercase two-digit hexadecimal values, expressing a red-green-blue color; the digits are not separated
- **space-sep-cmyk** Space-separated cyan-magenta-yellow-black values
- **space-sep-rgb** Space-separated red-green-blue values suitable for use as a PDF annotation color

\[\texttt{\textbackslash color\textunderscore export:nnN} \{(color\ expression)\} \{(format)\} \{(tl)\}\]

 Parses the \(\text{color expression}\) as described earlier, then converts to the \(\text{format}\) specified and assigns the data to the \(\text{tl}\).

\[\texttt{\textbackslash color\textunderscore export:nnnN} \{(model)\} \{(value(s))\} \{(format)\} \{(tl)\}\]

 Expresses the combination of \(\text{model}\) and \(\text{value(s)}\) in an internal representation, then converts to the \(\text{format}\) specified and assigns the data to the \(\text{tl}\).

35.9 Creating new color models

Additional color models are required to support specialist workflows, for example those involving separations (see https://helpx.adobe.com/indesign/using/spot-process-colors.html for details of the use of separations in print). Color models may be split into families; for the standard device-based color models (DeviceCMYK, DeviceRGB, DeviceGray), these are synonymous. This is not generally the case: see the PDF reference for more details. (Note that l3color uses the shorter names cmyk, etc.)
\color_model_new:nnn \color_model_new:nnn \{\langle\text{model}\rangle\} \{\langle\text{family}\rangle\} \{\langle\text{params}\rangle\}\\ Creates a new \langle\text{model}\rangle which is derived from the color model \langle\text{family}\rangle. The latter should be one of\\ \begin{itemize}\\ \item DeviceN\\ \item ICCBased\\ \item Separation\\ \end{itemize}\\ (The \langle\text{family}\rangle may be given in mixed case as-in the PDF reference: internally, case of these strings is folded.) Depending on the \langle\text{family}\rangle, one or more \langle\text{params}\rangle are mandatory or optional.\\ For a \text{Separation} space, there are three \text{compulsory} keys.\\ \begin{itemize}\\ \item \text{name} The name of the Separation, for example the formal name of a spot color ink. Such a \langle\text{name}\rangle may contain spaces, etc., which are not permitted in the \langle\text{model}\rangle.\\ \item \text{alternative-model} An alternative device colorspace, one of cmyk, rgb, gray or CIELAB. The three parameter-based models work as described above; see below for details of CIELAB colors.\\ \item \text{alternative-values} A comma-separated list of values appropriate to the \text{alternative-model}. This information is used by the PDF application if the \text{Separation} is not available.\\ \end{itemize}\\ CIELAB color separations are created using the \text{alternative-model} = CIELAB setting. These colors must also have an \text{illuminant} key, one of a, c, e, d50, d55, d65 or d75. The \text{alternative-values} in this case are the three parameters \(L^*\), \(a^*\) and \(b^*\) of the CIELAB model. Full details of this device-independent color approach are given in the documentation to the \text{colorspace} package.\\ CIELAB colors cannot be converted into other device-dependent color spaces, and as such, mixing can only occur if colors set up using the CIELAB model are also given with an alternative parameter-based model. If that is not the case, \text{3color} will fallback to using black as the colorant in any mixing.\\ For a \text{DeviceN} space, there is one \text{compulsory} key.\\ \begin{itemize}\\ \item \text{names} The names of the components of the \text{DeviceN} space. Each should be either the \langle\text{name}\rangle of a \text{Separation} model, a process color name (cyan, etc.) or the special name none.\\ \end{itemize}\\ For a \text{ICCBased} space, there is one \text{compulsory} key.\\ \begin{itemize}\\ \item \text{file} The name of the file containing the profile.\\ \end{itemize}\\ 35.9.1 Color profiles\\ Color profiles are used to ensure color accuracy by linking to collaboration. Applying a profile can be used to standardise color which is otherwise device-dependence.\\ \color_profile_apply:nn \color_profile_apply:nn \{\langle\text{profile}\rangle\} \{\langle\text{model}\rangle\}\\ This function applies a \langle\text{profile}\rangle to one of the device \langle\text{models}\rangle. The profile will then apply to all color of the selected \langle\text{model}\rangle. The \langle\text{profile}\rangle should specify an ICC profile file. The \langle\text{model}\rangle has to be one the standard device models: cmyk, gray or rgb.
Chapter 36

The \texttt{l3pdf} package

Core PDF support

36.1 Objects

\texttt{\textbackslash pdf\_object\_new:nn} {⟨object⟩} {⟨type⟩}

Declares ⟨object⟩ as a PDF object of ⟨type⟩, which should be one of

\begin{itemize}
  \item array
  \item dict
  \item fstream
  \item stream
\end{itemize}

The object may be referenced from this point on, and written later using \texttt{\textbackslash pdf\_object\_write:nn}.

\texttt{\textbackslash pdf\_object\_if\_exist\_p:n} {⟨object⟩}

\texttt{\textbackslash pdf\_object\_if\_exist:nTF} {⟨object⟩}

Tests whether an object with name {⟨object⟩} has been defined.

\texttt{\textbackslash pdf\_object\_write:nn} {⟨object⟩} {⟨content⟩}

Writes the ⟨content⟩ as content of the ⟨object⟩. Depending on the ⟨type⟩ declared for the object, the format required for the ⟨data⟩ will vary

\begin{description}
  \item [array] A space-separated list of values
  \item [dict] Key-value pairs in the form /⟨key⟩ ⟨value⟩
  \item [fstream] Two brace groups: ⟨file name⟩ and ⟨file content⟩
  \item [stream] Two brace groups: ⟨attributes (dictionary)⟩ and ⟨stream contents⟩
\end{description}
\pdf_object_ref:n  * \pdf_object_ref:n \{object\}

Inserts the appropriate information to reference the \langle object \rangle in for example page resource allocation.

\pdf_object_unnamed_write:nn \pdf_object_unnamed_write:nn \{type\} \{content\}
\pdf_object_unnamed_write:nx

New: 2021-02-10

Writes the \langle content \rangle as content of an anonymous object. Depending on the \langle type \rangle, the format required for the \langle data \rangle will vary.

array A space-separated list of values

dict Key–value pairs in the form /\langle key \rangle /\langle value \rangle

fstream Two brace groups: \langle attributes (dictionary) \rangle and \langle file name \rangle

stream Two brace groups: \langle attributes (dictionary) \rangle and \langle stream contents \rangle

\pdf_object_ref_last:  * \pdf_object_ref_last:

New: 2021-02-10

Inserts the appropriate information to reference the last \langle object \rangle created. This is particularly useful for anonymous objects.

\pdf_pageobject_ref:n  * \pdf_pageobject_ref:n \{pageobject\}

New: 2021-02-10

 Inserts the appropriate information to reference the \langle pageobject \rangle.

### 36.2 Version

\pdf_version_compare_p:Nn  * \pdf_version_compare_p:NnTF \{comparator\} \{\langle version \rangle\} \{(true code)\} \{(false code)\}

New: 2021-02-10

Compares the version of the PDF being created with the \langle version \rangle string specified, using the \langle comparator \rangle. Either the \langle true code \rangle or \langle false code \rangle will be left in the output stream.

\pdf_version_gset:n \pdf_version_min_gset:n

New: 2021-02-10

Sets the \langle version \rangle of the PDF being created. The \texttt{min} version will not alter the output version unless it is currently lower than the \langle version \rangle requested.

This function may only be used up to the point where the PDF file is initialised. With \texttt{dvips} it sets \texttt{pdf_version_major:} and \texttt{pdf_version_minor:} and allows to compare the values with \texttt{pdf_version_compare:Nn}, but the PDF version itself still has to be set with the command line option \texttt{-dCompatibilityLevel} of \texttt{ps2pdf}.

\pdf_version:  * \pdf_version_major:  * \pdf_version_minor:  *

New: 2021-02-10

Expands to the currently-active PDF version.
### 36.3 Compression

\pdf_uncompress: Disables any compression of the PDF, where possible.

This function may only be used up to the point where the PDF file is initialised.

### 36.4 Destinations

Destinations are the places a link jumped too. Unlike the name may suggest they don’t described an exact location in the PDF. Instead a destination contains a reference to a page along with an instruction how to display this page. The normally used “XYZ top left zoom” for example instructs the viewer to show the page with the given zoom and the top left corner at the top left coordinates—which then gives the impression that there is an anchor at this position.

If an instruction takes a coordinate, it is calculated by the following commands relative to the location the command is issued. So to get a specific coordinate one has to move the command to the right place.

\pdf_destination:nn \{\text{name}\} \{\text{type or integer}\}

This creates a destination. \{(\text{type or integer})\} can be one of \text{fit, fith, fitv, fitb, fitbh, fitbv, fitr, xyz} or an integer representing a scale factor in percent. \text{fitr} here gives only a lightweight version of /FitR: The backend code defines fitr so that it will with pdftex and LuaTeX use the coordinates of the surrounding box, with dvips and dvipdfmx it falls back to fit. For full control use \pdf_destination:nnnn.

The keywords match to the PDF names as described in the following tabular.

<table>
<thead>
<tr>
<th>Keyword</th>
<th>PDF</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>fit</td>
<td>/Fit</td>
<td>Fits the page to the window</td>
</tr>
<tr>
<td>fith</td>
<td>/FitH \text{top}</td>
<td>Fits the width of the page to the window</td>
</tr>
<tr>
<td>fitv</td>
<td>/FitV \text{left}</td>
<td>Fits the height of the page to the window</td>
</tr>
<tr>
<td>fitb</td>
<td>/FitB</td>
<td>Fits the page bounding box to the window</td>
</tr>
<tr>
<td>fitbh</td>
<td>/FitBH \text{top}</td>
<td>Fits the width of the page bounding box to the window.</td>
</tr>
<tr>
<td>fitbv</td>
<td>/FitBV \text{left}</td>
<td>Fits the height of the page bounding box to the window.</td>
</tr>
<tr>
<td>fitr</td>
<td>/FitR \text{left bottom right top}</td>
<td>Fits the rectangle specified by the four coordinates to the window (see above for the restrictions)</td>
</tr>
<tr>
<td>xyz</td>
<td>/XYZ \text{left top null}</td>
<td>Sets a coordinate but doesn’t change the zoom.</td>
</tr>
</tbody>
</table>
| \{(\text{integer})\} | /XYZ \text{left top zoom} | Sets a coordinate and a zoom meaning \{(\text{integer})\}\%.
\pdf_destination:nmmm \pdf_destination:nmmm \{\text{name}\} \{\text{width}\} \{\text{height}\} \{\text{depth}\}

This creates a destination with /FitR type with the given dimensions relative to the current location. The destination is in a box of size zero, but it doesn’t switch to horizontal mode.
Part VII
Additions and removals
Chapter 37

The \texttt{l3candidates} package
Experimental additions to \texttt{l3kernel}

37.1 Important notice

This module provides a space in which functions can be added to \texttt{l3kernel} (expl3) while still being experimental.

As such, the functions here may not remain in their current form, or indeed at all, in \texttt{l3kernel} in the future.

In contrast to the material in \texttt{l3experimental}, the functions here are all \textit{small} additions to the kernel. We encourage programmers to test them out and report back on the \texttt{LaTeX-L} mailing list.

Thus, if you intend to use any of these functions from the candidate module in a public package offered to others for productive use (e.g., being placed on CTAN) please consider the following points carefully:

- Be prepared that your public packages might require updating when such functions are being finalized.
- Consider informing us that you use a particular function in your public package, e.g., by discussing this on the \texttt{LaTeX-L} mailing list. This way it becomes easier to coordinate any updates necessary without issues for the users of your package.
- Discussing and understanding use cases for a particular addition or concept also helps to ensure that we provide the right interfaces in the final version so please give us feedback if you consider a certain candidate function useful (or not).

We only add functions in this space if we consider them being serious candidates for a final inclusion into the kernel. However, real use sometimes leads to better ideas, so functions from this module are \textbf{not necessarily stable} and we may have to adjust them!
37.2 Additions to \texttt{l3box}

\texttt{\textbackslash box\_clip:N} \texttt{\langle box \rangle}

Clips the \langle box \rangle in the output so that only material inside the bounding box is displayed in the output. The updated \langle box \rangle is an hbox, irrespective of the nature of the \langle box \rangle before the clipping is applied.

**These functions require the \LaTeX{}3 native drivers: they do not work with the \LaTeX{}2\epsilon graphics drivers!**

\textbf{\textsc{\LaTeX}hackers note:} Clipping is implemented by the driver, and as such the full content of the box is placed in the output file. Thus clipping does not remove any information from the raw output, and hidden material can therefore be viewed by direct examination of the file.

\texttt{\textbackslash box\_clip:c}
\texttt{\textbackslash box\_gclip:N}
\texttt{\textbackslash box\_gclip:c}

\texttt{\textbackslash box\_set\_trim:Nnnnn \langle box \rangle \{\langle left \rangle \} \{\langle bottom \rangle \} \{\langle right \rangle \} \{\langle top \rangle \}}

Adjusts the bounding box of the \langle box \rangle \langle left \rangle is removed from the left-hand edge of the bounding box, \langle right \rangle from the right-hand edge and so forth. All adjustments are \langle dimension expressions \rangle. Material outside of the bounding box is still displayed in the output unless \texttt{\textbackslash box\_clip:N} is subsequently applied. The updated \langle box \rangle is an hbox, irrespective of the nature of the \langle box \rangle before the trim operation is applied. The behavior of the operation where the trims requested is greater than the size of the box is undefined.

\texttt{\textbackslash box\_set\_viewport:Nnnnn}
\texttt{\textbackslash box\_set\_viewport:ccnnnn}
\texttt{\textbackslash box\_gset\_viewport:Nnnnn}
\texttt{\textbackslash box\_gset\_viewport:ccnnnn}

\texttt{\textbackslash exp\_args\_generate:n \{(variant argument specifiers)\}}

Defines \texttt{\exp\_args:N(variant)} functions for each \langle variant \rangle given in the comma list \{\langle variant argument specifiers \rangle\}. Each \langle variant \rangle should consist of the letters N, c, n, V, v, o, f, e, x, p and the resulting function is protected if the letter x appears in the \langle variant \rangle. This is only useful for cases where \texttt{\cs\_generate\_variant:Nn} is not applicable.

37.3 Additions to \texttt{l3expan}

\texttt{\exp\_args\_generate:n \{(variant argument specifiers)\}}

Evaluates the \langle fpexpr \rangle and tests whether the result is exactly nan. The test returns false for any other result, even a tuple containing nan.

\texttt{\textbackslash fp\_if\_nan:p:n * \textbackslash fp\_if\_nan:nTF *}

Updated: 2019–02–08
37.5 Additions to \texttt{l3file}

\\texttt{\textbackslash iow\_allow\_break:} \hfill \texttt{\textbackslash iow\_allow\_break:}

\hspace{1cm} \texttt{Nov: 2018-12-29}

In the first argument of \texttt{\textbackslash iow\_wrap:nN} (for instance in messages), inserts a break-point that allows a line break. In other words this is a zero-width breaking space.

\\texttt{\textbackslash ior\_get\_term:nN} \hfill \texttt{\textbackslash ior\_str\_get\_term:nN}

\hspace{1cm} \texttt{Nov: 2019-03-23}

\texttt{\textbackslash ior\_get\_term:nN \langle prompt \rangle \langle token\ list\ variable \rangle}

Function that reads one or more lines (until an equal number of left and right braces are found) from the terminal and stores the result locally in the \langle token list variable \rangle. Tokenization occurs as described for \texttt{\textbackslash ior\_get:NN} or \texttt{\textbackslash ior\_str\_get:NN}, respectively. When the \langle prompt \rangle is empty, \TeX will wait for input without any other indication: typically the programmer will have provided a suitable text using e.g. \texttt{\textbackslash iow\_term:n}.

Where the \langle prompt \rangle is given, it will appear in the terminal followed by an =, e.g.

\texttt{prompt=}

\hspace{1cm} \texttt{\textbackslash ior\_shell\_open:Nn}

\hspace{1cm} \texttt{Nov: 2019-05-08}

\texttt{\textbackslash ior\_shell\_open:Nn \langle stream \rangle \{\langle shell\ command \rangle\}}

Opens the pseudo-file created by the output of the \langle shell command \rangle for reading using \langle stream \rangle as the control sequence for access. If the \langle stream \rangle was already open it is closed before the new operation begins. The \langle stream \rangle is available for access immediately and will remain allocated to \langle shell command \rangle until a \texttt{\textbackslash ior\_close:N} instruction is given or the \TeX run ends. If piped system calls are disabled an error is raised.

For details of handling of the \langle shell command \rangle, see \texttt{\textbackslash sys\_get\_shell:mmN}.T.

37.6 Additions to \texttt{l3flag}

\texttt{\textbackslash flag\_raise\_if\_clear:n \star}

\hspace{1cm} \texttt{\textbackslash flag\_raise\_if\_clear:n \langle flag\ name \rangle}

Ensures the \langle flag \rangle is raised by making its height at least 1, locally.

37.7 Additions to \texttt{l3intarray}

\\texttt{\textbackslash intarray\_gset\_rand:Nnn} \hfill \texttt{\textbackslash intarray\_gset\_rand:cn}

\hspace{1cm} \texttt{\textbackslash intarray\_gset\_rand:Nn} \hfill \texttt{\textbackslash intarray\_gset\_rand:cn}

\hspace{1cm} \texttt{Nov: 2018-05-05}

\texttt{\textbackslash intarray\_gset\_rand:Nnn \langle intarray\ var \rangle \{\langle minimum \rangle\} \{\langle maximum \rangle\}}

\texttt{\textbackslash intarray\_gset\_rand:Nn \langle intarray\ var \rangle \{\langle maximum \rangle\}}

Evaluates the integer expressions \langle minimum \rangle and \langle maximum \rangle then sets each entry (independently) of the \langle integer array variable \rangle to a pseudo-random number between the two (with bounds included). If the absolute value of either bound is bigger than $2^{30} - 1$, an error occurs. Entries are generated in the same way as repeated calls to \texttt{\textbackslash int\_rand:nn} or \texttt{\textbackslash int\_rand:n} respectively, in particular for the second function the \langle minimum \rangle is 1. Assignments are always global. This is not available in older versions of \TeX.

\\texttt{\textbackslash intarray\_to\_clist:N \star}

\hspace{1cm} \texttt{\textbackslash intarray\_to\_clist:N \langle intarray\ var \rangle}

\hspace{1cm} \texttt{Nov: 2018-05-04}

Converts the \langle intarray \rangle to integer denotations separated by commas. All tokens have category code other. If the \langle intarray \rangle has no entry the result is empty; otherwise the result has one fewer comma than the number of items.

302
### 37.8 Additions to l3msg

\texttt{\msg\_show\_eval:NN} (\texttt{function}) \{\texttt{expression}\}

Shows or logs the \texttt{expression} (turned into a string), an equal sign, and the result of applying the \texttt{function} to the \{\texttt{expression}\} (with f-expansion). For instance, if the \texttt{function} is \texttt{\int\_eval:n} and the \texttt{expression} is \texttt{1+2} then this logs \texttt{1+2=3}.  

\begin{verbatim}
\msg\_show\_item:nn  *  \seq\_map\_function:NN \seq  \msg\_show\_item:nn
\msg\_show\_item\_unbraced:nn  *  \prop\_map\_function:NN \prop  \msg\_show\_item:nn
\msg\_show\_item:nn
\msg\_show\_item\_unbraced:nn
\end{verbatim}

Rev: 2017-12-04

Used in the text of messages for \texttt{\msg\_show:nnxxx} to show or log a list of items or key-value pairs. The one-argument functions are used for sequences, clist or token lists and the others for property lists. These functions turn their arguments to strings.

### 37.9 Additions to l3prg

\texttt{\bool\_set\_inverse:N} (\texttt{boolean})

Toggles the \texttt{boolean} from \texttt{true} to \texttt{false} and conversely: sets it to the inverse of its current value.
\bool_case_true:n * \bool_case_true:nTF
\bool_case_true:nTF * { {\boolexpr{case_1}} {\langle\langle\{\text{code case_1}\}\rangle}}
\bool_case_false:n * { {\boolexpr{case_2}} {\langle\langle\{\text{code case_2}\}\rangle}}
\bool_case_false:nTF * ...
{ {\boolexpr{case_n}} {\langle\langle\{\text{code case_n}\}\rangle}}
\langle\langle\{\text{true code}\}\rangle}
\langle\langle\{\text{false code}\}\rangle}

Evaluates in turn each of the \langle boolean expression cases \rangle until the first one that evaluates to true or to false, for \bool_case_true:n and \bool_case_false:n, respectively. The \langle code \rangle associated to this first case is left in the input stream, followed by the \langle true code \rangle, and other cases are discarded. If none of the cases match then only the \langle false code \rangle is inserted. The functions \bool_case_true:n and \bool_case_false:n, which do nothing if there is no match, are also available. For example

\bool_case_true:nF
{ \dim_compare_p:n { \l__mypkg_wd_dim <= 10pt } }
{ Fits }
{ \int_compare_p:n { \l__mypkg_total_int >= 10 } }
{ Many }
{ \l__mypkg_special_bool }
{ Special }
{ No idea! }

leaves “Fits” or “Many” or “Special” or “No idea!” in the input stream, in a way similar to some other language’s “if ... elseif ... elseif ... else ...”.

37.10 Additions to l3prop

\prop_rand_key_value:N * \prop_rand_key_value:N \prop var
\prop_rand_key_value:c * \prop_rand_key_value:c \prop_var

Selects a pseudo-random key–value pair from the \langle property list \rangle and returns \{\langle key\rangle\} and \{\langle value\rangle\}. If the \langle property list \rangle is empty the result is empty. This is not available in older versions of \TeX{}.\TeX{}hackers note: The result is returned within the \unexpanded primitive \exp_not:n, which means that the \langle value\rangle does not expand further when appearing in an x-type or e-type argument expansion.
37.11 Additions to \l3seq

\seq_mapthread_function:NNN \→ \seq_mapthread_function:NNN \langle seq1 \rangle \langle seq2 \rangle \langle function \rangle

Applies \langle function \rangle to every pair of items \langle seq1-item \rangle–\langle seq2-item \rangle from the two sequences, returning items from both sequences from left to right. The \langle function \rangle receives two \textit{n}-type arguments for each iteration. The mapping terminates when the end of either sequence is reached \textit{(i.e.} whichever sequence has fewer items determines how many iterations occur).

\seq_set_filter:NNn \seq_gset_filter:NNn

\seq_set_filter:NNn \langle sequence1 \rangle \langle sequence2 \rangle \langle \textit{inline} \ boolexpr \rangle

Evaluates the \langle \textit{inline} boolexpr \rangle for every \langle item \rangle stored within the \langle sequence2 \rangle. The \langle \textit{inline} boolexpr \rangle receives the \langle item \rangle as \#1. The sequence of all \langle items \rangle for which the \langle \textit{inline} boolexpr \rangle evaluated to \texttt{true} is assigned to \langle sequence1 \rangle.

\textbf{\TeX}Hackers note: Contrarily to other mapping functions, \texttt{\seq_map_break} cannot be used in this function, and would lead to low-level \TeX{} errors.

\seq_set_from_function:NnN \seq_gset_from_function:NnN

\seq_set_from_function:NnN \langle seq var \rangle \langle \textit{loop code} \rangle \langle function \rangle

Sets the \langle seq var \rangle equal to a sequence whose items are obtained by \textit{x}-expanding \langle \textit{loop code} \rangle \langle function \rangle. This expansion must result in successive calls to the \langle function \rangle with no nonexpandable tokens in between. More precisely the \langle function \rangle is replaced by a wrapper function that inserts the appropriate separators between items in the sequence. The \langle \textit{loop code} \rangle must be expandable; it can be for example \texttt{\tl_map_function:NN} \langle tl var \rangle or \texttt{\clist_map_function:nN} \langle clist \rangle or \texttt{\int_step_function:nnnN} \langle \textit{initial value} \rangle \langle \textit{step} \rangle \langle \textit{final value} \rangle.

\seq_set_from_inline_x:Nnn \seq_gset_from_inline_x:Nnn

\seq_set_from_inline_x:Nnn \langle seq var \rangle \langle \textit{loop code} \rangle \langle \textit{inline code} \rangle

Sets the \langle seq var \rangle equal to a sequence whose items are obtained by \textit{x}-expanding \langle \textit{loop code} \rangle applied to a \langle \textit{function} \rangle derived from the \langle \textit{inline code} \rangle. A \langle \textit{function} \rangle is defined, that takes one argument, \textit{x}-expands the \langle \textit{inline code} \rangle with that argument as \#1, then adds appropriate separators to turn the result into an item of the sequence. The \textit{x}-expansion of \langle \textit{loop code} \rangle \langle \textit{function} \rangle must result in successive calls to the \langle \textit{function} \rangle with no nonexpandable tokens in between. The \langle \textit{loop code} \rangle must be expandable; it can be for example \texttt{\tl_map_function:NN} \langle tl var \rangle or \texttt{\clist_map_function:nN} \langle clist \rangle or \texttt{\int_step_function:nnnN} \langle \textit{initial value} \rangle \langle \textit{step} \rangle \langle \textit{final value} \rangle, but not the analogous “\textit{inline}” mappings.
Removes the item of \langle sequence \rangle at the position given by evaluating the \langle integer expression \rangle and replaces it by \langle item \rangle. Items are indexed from 1 on the left/top of the \langle sequence \rangle, or from \(-1\) on the right/bottom. If the \langle integer expression \rangle is zero or is larger (in absolute value) than the number of items in the sequence, the \langle sequence \rangle is not modified. In these cases, \seq_set_item:Nnn \safeverb raises an error while \seq_set_item:NnnTF runs the \langle false code \rangle. In cases where the assignment was successful, \langle true code \rangle is run afterwards.

Removes the \langle item \rangle at position \langle integer expression \rangle in the \langle sequence \rangle, and places it in the \langle token list variable \rangle. Items are indexed from 1 on the left/top of the \langle sequence \rangle, or from \(-1\) on the right/bottom. If the position is zero or is larger (in absolute value) than the number of items in the sequence, the \langle seq var \rangle is not modified, the \langle token list \rangle is set to the special marker \texttt{\textless no_value}, and the \langle false code \rangle is left in the input stream; otherwise the \langle true code \rangle is. The \langle token list \rangle assignment is local while the \langle sequence \rangle is assigned locally for \texttt{pop} or globally for \texttt{gpop} functions.

### 37.12 Additions to l3sys

The version string of the current engine, in the same form as given in the banner issued when running a job. For pdf\TeX{} and Lua\TeX{} this is of the form
\[
\langle major \rangle.\langle minor \rangle.\langle revision \rangle
\]
For X\TeX{}, the form is
\[
\langle major \rangle.\langle minor \rangle
\]
For \p\TeX{} and up\TeX{}, only releases since \TeX{} Live 2018 make the data available, and the form is more complex, as it comprises the \p\TeX{} version, the up\TeX{} version and the e-\p\TeX{} version.

\[
p\langle major \rangle.\langle minor \rangle.\langle revision \rangle-u\langle major \rangle.\langle minor \rangle-\langle e\TeX{} \rangle
\]
where the \texttt{u} part is only present for up\TeX{}.

Tests if the engine has a pseudo-random number generator. Currently this is the case in pdf\TeX{}, Lua\TeX{}, \p\TeX{}, up\TeX{} and recent releases of X\TeX{}.
37.13 Additions to l3tl

\tl_range_braced:Nnn \tl_range_braced:cnn \tl_range_braced:nnn \tl_range_unbraced:Nnn \tl_range_unbraced:cnn \tl_range_unbraced:nnn

Leaves in the input stream the items from the \langle start index \rangle to the \langle end index \rangle inclusive, using the same indexing as \tl_range:nnn. Spaces are ignored. Regardless of whether items appear with or without braces in the \langle token list \rangle, the \tl_range_braced:nnn function wraps each item in braces, while \tl_range_unbraced:nnn does not (overall it removes an outer set of braces). For instance,

\iw_term:x \{ \tl_range_braced:nnn \{ abcd-\{e\}\}f \} \{ 2 \} \{ 5 \} \}
\iw_term:x \{ \tl_range_braced:nnn \{ abcd-\{e\}\}f \} \{ -4 \} \{ -1 \} \}
\iw_term:x \{ \tl_range_braced:nnn \{ abcd-\{e\}\}f \} \{ -2 \} \{ -1 \} \}
\iw_term:x \{ \tl_range_braced:nnn \{ abcd-\{e\}\}f \} \{ 0 \} \{ -1 \} \}

prints \{b}\{c\}d\{e\}\}, \{c\}d\{e\}\{f\}, \{e\}\{f\}, and an empty line to the terminal, while

\iw_term:x \{ \tl_range_unbraced:nnn \{ abcd-\{e\}\}f \} \{ 2 \} \{ 5 \} \}
\iw_term:x \{ \tl_range_unbraced:nnn \{ abcd-\{e\}\}f \} \{ -4 \} \{ -1 \} \}
\iw_term:x \{ \tl_range_unbraced:nnn \{ abcd-\{e\}\}f \} \{ -2 \} \{ -1 \} \}
\iw_term:x \{ \tl_range_unbraced:nnn \{ abcd-\{e\}\}f \} \{ 0 \} \{ -1 \} \}

prints bcde{} , cde{}f , e{}f , and an empty line to the terminal. Because braces are removed, the result of \tl_range_unbraced:nnn may have a different number of items as for \tl_range:nnn or \tl_range_braced:nnn. In cases where preserving spaces is important, consider the slower function \tl_range:nnn.

\textit{TeXhackers note:} The result is returned within the \texttt{\unexpanded} primitive (\texttt{\exp_not:n}), which means that the \langle item \rangle does not expand further when appearing in an \texttt{x}-type or \texttt{e}-type argument expansion.

\tl_build_begin:N \tl_build_gbegin:N
Revised: 2018-04-01
\tl_build_clear:N \tl_build_gclear:N
Revised: 2018-04-01

Clears the \langle tl var \rangle and sets it up to support other \tl_build\_\ldots functions, which allow accumulating large numbers of tokens piece by piece much more efficiently than standard \l3tl functions. Until \tl_build_end:N \langle tl var \rangle is called, applying any function from \l3tl other than \tl_build\_\ldots will lead to incorrect results. The \texttt{begin} and \texttt{gbegin} functions must be used for local and global \langle tl var \rangle respectively.

Clears the \langle tl var \rangle and sets it up to support other \tl_build\_\ldots functions. The \texttt{clear} and \texttt{gclear} functions must be used for local and global \langle tl var \rangle respectively.
\tl_build_put_left:Nn \tl_build_put_left:Nx \tl_build_gput_left:Nn \tl_build_gput_left:Nx \tl_build_put_right:Nn \tl_build_put_right:Nx \tl_build_gput_right:Nn \tl_build_gput_right:Nx

Rev: 2018-04-01

\tl_build_get:NN \tl_build_get:NN ⟨tl var⟩ ⟨tl var⟩

Stores the contents of the ⟨tl var1⟩ in the ⟨tl var2⟩. The ⟨tl var1⟩ must have been set up with \tl_build_begin:N or \tl_build_gbegin:N. The ⟨tl var2⟩ is a “normal” token list variable, assigned locally using \tl_set:Nn.

Rev: 2018-04-01

\tl_build_end:N \tl_build_gend:N

\tl_build_end:N ⟨tl var⟩ \tl_build_gend:N

Gets the contents of ⟨tl var⟩ and stores that into the ⟨tl var⟩ using \tl_set:Nn or \tl_gset:Nn. The ⟨tl var⟩ must have been set up with \tl_build_begin:N or \tl_build_gbegin:N. The end and gend functions must be used for local and global ⟨tl var⟩ respectively. These functions completely remove the setup code that enabled ⟨tl var⟩ to be used for other \tl_build_... functions.

37.14 Additions to \l3token

\c_catcode_active_space_tl

Rev: 2017-08-07

\char_to_utfviii_bytes:n \char_to_utfviii_bytes:n ⟨codepoint⟩

Converting the (Unicode) ⟨codepoint⟩ to UTF-8 bytes. The expansion of this function comprises four brace groups, each of which will contain a hexadecimal value: the appropriate byte. As UTF-8 is a variable-length, one or more of the groups may be empty: the bytes read in the logical order, such that a two-byte codepoint will have groups #1 and #2 filled and #3 and #4 empty.

Rev: 2020-01-09

\char_to_nfd:N \char_to_nfd:N ⟨char⟩

Converting the ⟨char⟩ to the Unicode Normalization Form Canonical Decomposition. The category code of the generated character is the same as the ⟨char⟩. With 8-bit engines, no change is made to the character.

Rev: 2020-01-02

308
Collects and removes tokens from the input stream until finding a token that does not match the `<test token>` (as defined by the test `\token_if_eq_catcode:NNTF` or `\token_if_eq_charcode:NNTF` or `\token_if_eq_meaning:NNTF`). The collected tokens are passed to the `<inline code>` as #1. When begin-group or end-group tokens (usually `{ or `}`) are collected they are replaced by implicit `\c_group_begin_token` and `\c_group_end_token`, and when spaces (including `\c_space_token`) are collected they are replaced by explicit spaces.

For example the following code prints “Hello” to the terminal and leave “, world!” in the input stream.

```
\peek_catcode_collect_inline:N \iow_term:n {#1} Hello,~world!
```

Another example is that the following code tests if the next token is *, ignoring intervening spaces, but putting them back using #1 if there is no *.

```
\peek_meaning_collect_inline:N \c_space_token
{ \peek_charcode:NTF * \iow_term:n {#1} \iow_term:n {no-star #1} }
```
Part VIII
Implementation
Chapter 38

\texttt{l3bootstrap} implementation

\section*{38.1 \LaTeX{}-specific code}

Depending on the versions available, the \LaTeX{} format may not have the raw \texttt{\textbackslash Umath} primitive names available. We fix that globally: it should cause no issues. Older \LaTeX{} versions do not have a pre-built table of the primitive names here so sort one out ourselves. These end up globally-defined but at that is true with a newer format anyway and as they all start \texttt{\textbackslash U} this should be reasonably safe.

\begin{verbatim}
\begingroup
\expandafter\ifx\csname directlua\endcsname\relax
\else
\directlua{%
local i
local t = { }
for _,i in pairs(tex.extraprimitives("luatex")) do
  if string.match(i,"^U") then
    if not string.match(i,"^Uchar$") then $%
      table.insert(t,i)
    end
  end
end
\endgroup
\end{verbatim}

\section*{38.2 The \texttt{\textbackslash pdfstrcmp} primitive in \texttt{Xe\TeX{}}} 

Only \texttt{pdf\TeX{}} has a primitive called \texttt{pdfstrcmp}. The \texttt{Xe\TeX{}} version is just \texttt{\textbackslash strcmp}, so there is some shuffling to do. As this is still a real primitive, using the \texttt{pdf\TeX{}} name is “safe”.

\begin{verbatim}
\begingroup\expandafter\expandafter\expandafter\endgroup
\expandafter\ifx\csname pdfstrcmp\endcsname\relax
\let pdfstrcmp strcmp
\endgroup
\end{verbatim}
38.3 Loading support Lua code

When LuaTeX is used there are various pieces of Lua code which need to be loaded. The code itself is defined in \texttt{l3luatex} and is extracted into a separate file. Thus here the task is to load the Lua code both now and (if required) at the start of each job.

```latex
\begingroup\expandafter\expandafter\expandafter\endgroup
\expandafter\ifx\csname directlua\endcsname\relax
\else
\ifnum\luatexversion<110
\else
\input{ltluatex}%
\fi
\begingroup\expandafter\expandafter\expandafter\endgroup
\expandafter\ifx\csname newluabytecode\endcsname\relax
\else
\newluabytecode\@expl@luadata@bytecode
\fi
\directlua{require("expl3")}%
\fi
\fi
```

As the user might be making a custom format, no assumption is made about matching package mode with only loading the Lua code once. Instead, a query to Lua reveals what mode is in operation.

```latex
\ifnum 0\%\directlua{
if status.ini_version then
tex.write("1")
end
}>0 \%
\everyjob\expandafter{\the\expandafter\everyjob\csname\detokenize{lua_now:n}\endcsname{require("expl3")}}%\fi
\fi
```

38.4 Engine requirements

The code currently requires \textsc{e}-\TeX and functionality equivalent to \texttt{pdfstrcmp}, and also driver and Unicode character support. This is available in a reasonably-wide range of engines.

For Lua\TeX, we require at least Lua 5.3 and the \texttt{token.set_lua} function. This is available at least since Lua\TeX 1.10.

```latex
\begingroup
\def\next{\endgroup}%
\def\ShortText{Required primitives not found}%
\fi
```

312
The L3 programming layer requires the e-TeX primitives and additional functionality as described in the README file. These are available in the engines pdfTeX v1.40.0, XeTeX v0.99992, LuaTeX v1.10, e-(u)pTeX mid-2012, or later.

38.5 Extending allocators

The ability to extend \TeX’s allocation routine to allow for e-\TeX has been around since 1997 in the etex package. Loading this support is delayed until here as we are now sure
that the \-\eTeX\ extensions and \texttt{\pdffstrcmp} or equivalent are available. Thus there is no danger of an “uncontrolled” error if the engine requirements are not met.

For \LaTeX{} 2\textepsilon\ we need to make sure that the extended pool is being used: \texttt{expl3} uses a lot of registers. For formats from 2015 onward there is nothing to do as this is automatic. For older formats, the \texttt{etex} package needs to be loaded to do the job. In that case, some inserts are reserved also as these have to be from the standard pool. Note that \texttt{\reserveinserts} is \texttt{\outer} and so is accessed here by csname. In earlier versions, loading \texttt{etex} was done directly and so \texttt{\reserveinserts} appeared in the code: this then required a \texttt{\relax} after \texttt{\RequirePackage} to prevent an error with “unsafe” definitions as seen for example with \texttt{capoptions}. The optional loading here is done using a group and \texttt{\ifx} test as we are not quite in the position to have a single name for \texttt{\pdffstrcmp} just yet.

\begin{verbatim}
\begingroup
\def\@tempa{LaTeX2e}\
\def\next{}\
\ifx\fmtname\@tempa
\expandafter\ifx\csname extrafloats\endcsname\relax
\def\next{\RequirePackage{etex}\csname reserveinserts\endcsname{32}}
\fi
\fi
\expandafter\endgroup
\next
\end{verbatim}

38.6 The \LaTeX{}3 code environment

The code environment is now set up.

\texttt{\ExplSyntaxOff} Before changing any category codes, in package mode we need to save the situation before loading. Note the set up here means that once applied \texttt{\ExplSyntaxOff} becomes a “do nothing” command until \texttt{\ExplSyntaxOn} is used.

\begin{verbatim}
\protectededef\ExplSyntaxOff
\protectededef\noexpand\ExplSyntaxOff{}
\catcode 9 = \the\catcode 9\relax
\catcode 32 = \the\catcode 32\relax
\catcode 34 = \the\catcode 34\relax
\catcode 38 = \the\catcode 38\relax
\catcode 58 = \the\catcode 58\relax
\catcode 94 = \the\catcode 94\relax
\catcode 95 = \the\catcode 95\relax
\catcode 124 = \the\catcode 124\relax
\catcode 126 = \the\catcode 126\relax
\endlinechar = \the\endlinechar\relax
\chardef\csname\detokenize{l__kernel_expl_bool}\endcsname = 0\relax
\end{verbatim}

(End definition for \texttt{\ExplSyntaxOff}. This function is documented on page 9.)
The code environment is now set up.
\catcode 9 = 9\relax
\catcode 32 = 9\relax
\catcode 34 = 12\relax
\catcode 38 = 4\relax
\catcode 58 = 11\relax
\catcode 94 = 7\relax
\catcode 95 = 11\relax
\catcode 124 = 12\relax
\catcode 126 = 10\relax
\endlinechar = 32\relax
\l__kernel_expl_bool
The status for code syntax: this is on at present.
\chardef\l__kernel_expl_bool = 1\relax
(End definition for \l__kernel_expl_bool.)
\ExplSyntaxOn
The idea here is that multiple \ExplSyntaxOn calls are not going to mess up category codes, and that multiple calls to \ExplSyntaxOff are also not wasting time. Applying \ExplSyntaxOn alters the definition of \ExplSyntaxOff and so in package mode this function should not be used until after the end of the loading process!
\protected \def \ExplSyntaxOn
{ \bool_if:NF \l__kernel_expl_bool
{ \cs_set_protected:Npx \ExplSyntaxOff
{ \char_set_catcode:nn { 9 } { \char_value_catcode:n { 9 } }
\char_set_catcode:nn { 32 } { \char_value_catcode:n { 32 } }
\char_set_catcode:nn { 34 } { \char_value_catcode:n { 34 } }
\char_set_catcode:nn { 38 } { \char_value_catcode:n { 38 } }
\char_set_catcode:nn { 58 } { \char_value_catcode:n { 58 } }
\char_set_catcode:nn { 94 } { \char_value_catcode:n { 94 } }
\char_set_catcode:nn { 95 } { \char_value_catcode:n { 95 } }
\char_set_catcode:nn { 124 } { \char_value_catcode:n { 124 } }
\char_set_catcode:nn { 126 } { \char_value_catcode:n { 126 } }
\tex_endlinechar:D = \tex_the:D \tex_endlinechar:D \scan_stop:
\bool_set_false:N \l__kernel_expl_bool
\cs_set_protected:Npx \ExplSyntaxOff { }
\}
}
\char_set_catcode:nn { 9 } { \char_value_catcode:n { 9 } }
\char_set_catcode:nn { 32 } { \char_value_catcode:n { 32 } }
\char_set_catcode:nn { 34 } { \char_value_catcode:n { 34 } }
\char_set_catcode:nn { 38 } { \char_value_catcode:n { 38 } }
\char_set_catcode:nn { 58 } { \char_value_catcode:n { 58 } }
\char_set_catcode:nn { 94 } { \char_value_catcode:n { 94 } }
\char_set_catcode:nn { 95 } { \char_value_catcode:n { 95 } }
\char_set_catcode:nn { 124 } { \char_value_catcode:n { 124 } }
\char_set_catcode:nn { 126 } { \char_value_catcode:n { 126 } }
\tex_endlinechar:D = \tex_the:D \tex_endlinechar:D \scan_stop:
\bool_set_false:N \l__kernel_expl_bool
\cs_set_protected:Npx \ExplSyntaxOff { }
\char_set_catcode:nn { 9 } { \char_value_catcode:n { 9 } }
\char_set_catcode:nn { 32 } { \char_value_catcode:n { 32 } }
\char_set_catcode:nn { 34 } { \char_value_catcode:n { 34 } }
\char_set_catcode:nn { 38 } { \char_value_catcode:n { 38 } }
\char_set_catcode:nn { 58 } { \char_value_catcode:n { 58 } }
\char_set_catcode:nn { 94 } { \char_value_catcode:n { 94 } }
\char_set_catcode:nn { 95 } { \char_value_catcode:n { 95 } }
\char_set_catcode:nn { 124 } { \char_value_catcode:n { 124 } }
\char_set_catcode:nn { 126 } { \char_value_catcode:n { 126 } }
\tex_endlinechar:D = 32 \scan_stop:
\bool_set_true:N \l__kernel_expl_bool
\l__kernel_expl_bool
(End definition for \ExplSyntaxOn. This function is documented on page 9.)

178 ⟨/package⟩
Chapter 39

\texttt{l3names} implementation

The prefix here is \texttt{kernel}. A few places need \texttt{@@} to be left as is; this is obtained as \texttt{@@@}. The code here simply renames all of the primitives to new, internal, names. The \texttt{\let} primitive is renamed by hand first as it is essential for the entire process to follow. This also uses \texttt{\global}, as that way we avoid leaving an unneeded csname in the hash table.

Everything is inside a (rather long) group, which keeps \texttt{\_\_kernel\_primitive:NN} trapped.

\begin{verbatim}
\def \__kernel_primitive:NN #1#2{ \tex_global:D \tex_let:D #2 #1 }
\end{verbatim}

\begin{verbatim}
\long \def \__kernel_primitive:NN #1#2{ \tex_global:D \tex_let:D #2 #1 }
\end{verbatim}

To allow extracting “just the names”, a bit of DocStrip fiddling.

In the current incarnation of this package, all \TeX\ primitives are given a new name of the form \texttt{\textunderscore oldname:D}. But first three special cases which have symbolic original names. These are given modified new names, so that they may be entered without catcode tricks.

\begin{verbatim}
\__kernel_primitive:NN \tex_space:D
\__kernel_primitive:NN \textunderscore \tex_italiccorrection:D
\__kernel_primitive:NN \textunderscore \tex_hyphen:D
\end{verbatim}

Now all the other primitives.

\begin{verbatim}
\__kernel_primitive:NN \textunderscore \above:D
\__kernel_primitive:NN \textunderscore \abovedisplayshortskip:D
\__kernel_primitive:NN \textunderscore \abovedisplayshortskip:D
\__kernel_primitive:NN \textunderscore \abovedisplayskip:D
\__kernel_primitive:NN \textunderscore \abovedisplayskip:D
\__kernel_primitive:NN \textunderscore \abovewithdelims:D
\__kernel_primitive:NN \textunderscore \abovewithdelims:D
\end{verbatim}
Primitives introduced by \TeX.

\begin{itemize}
    \item \_kernel_primitive:NN \toks \tex_toks:D
    \item \_kernel_primitive:NN \toksdef \tex_toksdef:D
    \item \_kernel_primitive:NN \tolerance \tex_tolerance:D
    \item \_kernel_primitive:NN \topmark \tex_topmark:D
    \item \_kernel_primitive:NN \topskip \tex_topskip:D
    \item \_kernel_primitive:NN \tracingcommands \tex_tracingcommands:D
    \item \_kernel_primitive:NN \tracinglostchars \tex_tracinglostchars:D
    \item \_kernel_primitive:NN \tracingmacros \tex_tracingmacros:D
    \item \_kernel_primitive:NN \tracingonline \tex_tracingonline:D
    \item \_kernel_primitive:NN \tracingoutput \tex_tracingoutput:D
    \item \_kernel_primitive:NN \tracingpages \tex_tracingpages:D
    \item \_kernel_primitive:NN \tracingparagraphs \tex_tracingparagraphs:D
    \item \_kernel_primitive:NN \tracingrestores \tex_tracingrestores:D
    \item \_kernel_primitive:NN \tracingstats \tex_tracingstats:D
    \item \_kernel_primitive:NN \uccode \tex_uccode:D
    \item \_kernel_primitive:NN \uchyph \tex_uchyph:D
    \item \_kernel_primitive:NN \underline \tex_underline:D
    \item \_kernel_primitive:NN \unhbox \tex_unhbox:D
    \item \_kernel_primitive:NN \unhcopy \tex_unhcopy:D
    \item \_kernel_primitive:NN \unkern \tex_unkern:D
    \item \_kernel_primitive:NN \unpenalty \tex_unpenalty:D
    \item \_kernel_primitive:NN \unskip \tex_unskip:D
    \item \_kernel_primitive:NN \unvbox \tex_unvbox:D
    \item \_kernel_primitive:NN \unvcopy \tex_unvcopy:D
    \item \_kernel_primitive:NN \uppercase \tex_uppercase:D
    \item \_kernel_primitive:NN \vadjust \tex_vadjust:D
    \item \_kernel_primitive:NN \valign \tex_valign:D
    \item \_kernel_primitive:NN \vbadness \tex_vbadness:D
    \item \_kernel_primitive:NN \vbox \tex_vbox:D
    \item \_kernel_primitive:NN \vcenter \tex_vcenter:D
    \item \_kernel_primitive:NN \vfil \tex_vfil:D
    \item \_kernel_primitive:NN \vfill \tex_vfill:D
    \item \_kernel_primitive:NN \vfilneg \tex_vfilneg:D
    \item \_kernel_primitive:NN \vfuzz \tex_vfuzz:D
    \item \_kernel_primitive:NN \voffset \tex_voffset:D
    \item \_kernel_primitive:NN \vrule \tex_vrule:D
    \item \_kernel_primitive:NN \vsize \tex_vsize:D
    \item \_kernel_primitive:NN \vskip \tex_vskip:D
    \item \_kernel_primitive:NN \vtop \tex_vtop:D
    \item \_kernel_primitive:NN \wd \tex_wd:D
    \item \_kernel_primitive:NN \widowpenalty \tex_widowpenalty:D
    \item \_kernel_primitive:NN \write \tex_write:D
    \item \_kernel_primitive:NN \xdef \tex_xdef:D
    \item \_kernel_primitive:NN \xleaders \tex_xleaders:D
    \item \_kernel_primitive:NN \xspaceskip \tex_xspaceskip:D
    \item \_kernel_primitive:NN \year \tex_year:D
    \item \_kernel_primitive:NN \beginL \tex_beginL:D
    \item \_kernel_primitive:NN \beginR \tex_beginR:D
    \item \_kernel_primitive:NN \botmarks \tex_botmarks:D
    \item \_kernel_primitive:NN \clubpenalties \tex_clubpenalties:D
    \item \_kernel_primitive:NN \currentgrouplevel \tex_currentgrouplevel:D
\end{itemize}
Post-ε-TeX primitives do not always end up with the same name in all engines, if indeed they are available cross-engine anyway. We therefore take the approach of preferring the shortest name that makes sense. First, we deal with the primitives introduced by pdfTeX which directly relate to PDF output: these are copied with the names unchanged.
These are not related to PDF output and either already appear in other engines without the `\pdf` prefix, or might reasonably do so at some future stage. We therefore drop the leading `pdf` here.
The version primitives are not related to PDF mode but are pdfTEX-specific, so again are carried forward unchanged.

These ones appear in pdfTEX but don’t have pdf in the name at all: no decisions to make.

Post pdfTEX primitive availability gets more complex. Both XeLaTeX and LuaLaTeX have varying names for some primitives from pdfTeX. Particularly for LuaLaTeX tracking all of that would be hard. Instead, we now check that we only save primitives if they actually exist.

Some pdfTeX primitives are handled here because they got dropped in LuaLaTeX but the corresponding internal names are emulated later. The Lua code is already loaded at this point, so we shouldn’t overwrite them.
XeLaTeX-specific primitives. Note that XeLaTeX’s strcmp is handled earlier and is “rolled up” into pdfstrcmp. A few cross-compatibility names which lack the pdf of the original are handled later.
Primitives from pdfTeX that \LaTeX\ renames: also helps with Lua\TeX.  

Primitives from Lua\TeX, some of which have been ported back to \LaTeX.  

\begin{verbatim}
\__kernel_primitive:NN \creationdate \tex_creationdate:D
\__kernel_primitive:NN \elapsedtime \tex_elapsedtime:D
\__kernel_primitive:NN \filedump \tex_filedump:D
\__kernel_primitive:NN \filemoddate \tex_filemoddate:D
\__kernel_primitive:NN \filesize \tex_filesize:D
\__kernel_primitive:NN \mdfivesum \tex_mdfivesum:D
\__kernel_primitive:NN \realtime \tex_realtime:D
\__kernel_primitive:NN \resettimer \tex_resettimer:D
\__kernel_primitive:NN \shellescape \tex_shellescape:D
\__kernel_primitive:NN \alignmark \tex_alignmark:D
\__kernel_primitive:NN \aligntab \tex_aligntab:D
\__kernel_primitive:NN \attribute \tex_attribute:D
\__kernel_primitive:NN \attributedef \tex_attributedef:D
\__kernel_primitive:NN \automaticdiscretionary \tex_automaticdiscretionary:D
\__kernel_primitive:NN \automatichyphenmode \tex_automatichyphenmode:D
\__kernel_primitive:NN \automatichyphenpenalty \tex_automatichyphenpenalty:D
\__kernel_primitive:NN \begincsname \tex_begincsname:D
\__kernel_primitive:NN \bodydir \tex_bodydir:D
\__kernel_primitive:NN \bodydirection \tex_bodydirection:D
\__kernel_primitive:NN \boxdir \tex_boxdir:D
\__kernel_primitive:NN \boxdirection \tex_boxdirection:D
\__kernel_primitive:NN \breakafterdirmode \tex_breakafterdirmode:D
\__kernel_primitive:NN \catcodetable \tex_catcodetable:D
\__kernel_primitive:NN \clearmarks \tex_clearmarks:D
\__kernel_primitive:NN \crampeddisplaystyle \tex_crampeddisplaystyle:D
\__kernel_primitive:NN \crampedscriptscriptstyle \tex_crampedscriptscriptstyle:D
\__kernel_primitive:NN \crampedscriptstyle \tex_crampedscriptstyle:D
\__kernel_primitive:NN \crampedtextstyle \tex_crampedtextstyle:D
\__kernel_primitive:NN \csstring \tex_csstring:D
\__kernel_primitive:NN \discretionaryligaturemode \tex_discretionaryligaturemode:D
\__kernel_primitive:NN \directlua \tex_directlua:D
\__kernel_primitive:NN \dviextension \tex_dviextension:D
\__kernel_primitive:NN \dvifeedback \tex_dvifeedback:D
\end{verbatim}

329
\_kernel_primitive:NN \dvivariable \tex_dvivariable:D
\_kernel_primitive:NN \eTeXglueshrinkorder \tex_eTeXglueshrinkorder:D
\_kernel_primitive:NN \eTeXgluestretchorder \tex_eTeXgluestretchorder:D
\_kernel_primitive:NN \etoksapp \tex_etoksapp:D
\_kernel_primitive:NN \etokspre \tex_etokspre:D
\_kernel_primitive:NN \exceptionpenalty \tex_exceptionpenalty:D
\_kernel_primitive:NN \explicithyphenpenalty \tex_explicithyphenpenalty:D
\_kernel_primitive:NN \expanded \tex_expanded:D
\_kernel_primitive:NN \explicitdiscretionary \tex_explicitdiscretionary:D
\_kernel_primitive:NN \firstvalidlanguage \tex_firstvalidlanguage:D
\_kernel_primitive:NN \fontid \tex_fontid:D
\_kernel_primitive:NN \formatname \tex_formatname:D
\_kernel_primitive:NN \hjcode \tex_hjcode:D
\_kernel_primitive:NN \hpack \tex_hpack:D
\_kernel_primitive:NN \hyphenationbounds \tex_hyphenationbounds:D
\_kernel_primitive:NN \hyphenationmin \tex_hyphenationmin:D
\_kernel_primitive:NN \hyphenpenaltymode \tex_hyphenpenaltymode:D
\_kernel_primitive:NN \gleaders \tex_gleaders:D
\_kernel_primitive:NN \ifcondition \tex_ifcondition:D
\_kernel_primitive:NN \immediateassigned \tex_immediateassigned:D
\_kernel_primitive:NN \immediateassignment \tex_immediateassignment:D
\_kernel_primitive:NN \initcatcodetable \tex_initcatcodetable:D
\_kernel_primitive:NN \lastnamedcs \tex_lastnamedcs:D
\_kernel_primitive:NN \latelua \tex_latelua:D
\_kernel_primitive:NN \lateluafunction \tex_lateluafunction:D
\_kernel_primitive:NN \leftghost \tex_leftghost:D
\_kernel_primitive:NN \letcharcode \tex_letcharcode:D
\_kernel_primitive:NN \linedir \tex_linedir:D
\_kernel_primitive:NN \linedirection \tex_linedirection:D
\_kernel_primitive:NN \localbrokenpenalty \tex_localbrokenpenalty:D
\_kernel_primitive:NN \localinterlinepenalty \tex_localinterlinepenalty:D
\_kernel_primitive:NN \luabytecode \tex_luabytecode:D
\_kernel_primitive:NN \luabytecodecall \tex_luabytecodecall:D
\_kernel_primitive:NN \luacopyinputnodes \tex_luacopyinputnodes:D
\_kernel_primitive:NN \luadef \tex_luadef:D
\_kernel_primitive:NN \localeleftbox \tex_localeleftbox:D
\_kernel_primitive:NN \locallinebreakpenalty \tex_locallinebreakpenalty:D
\_kernel_primitive:NN \luacrashcode \tex_luacrashcode:D
\_kernel_primitive:NN \luaescapestring \tex_luaescapestring:D
\_kernel_primitive:NN \luafunction \tex_luafunction:D
\_kernel_primitive:NN \luafunctioncall \tex_luafunctioncall:D
\_kernel_primitive:NN \luatexbanner \tex_luatexbanner:D
\_kernel_primitive:NN \luatexversion \tex_luatexversion:D
\_kernel_primitive:NN \mathdefaultsmode \tex_mathdefaultsmode:D
\_kernel_primitive:NN \mathdelimitersmode \tex_mathdelimitersmode:D
\_kernel_primitive:NN \mathdir \tex_mathdir:D
\_kernel_primitive:NN \mathdirection \tex_mathdirection:D
\_kernel_primitive:NN \mathdisplayskipmode \tex_mathdisplayskipmode:D
\_kernel_primitive:NN \matheqdirmode \tex_matheqdirmode:D
\_kernel_primitive:NN \matheqnogapstep \tex_matheqnogapstep:D
\_kernel_primitive:NN \mathoption \tex_mathoption:D
\_kernel_primitive:NN \mathpenaltiesmode \tex_mathpenaltiesmode:D
\_kernel_primitive:NN \mathrulesfam \tex_mathrulesfam:D
\_kernel_primitive:NN \mathdefaultsmode \tex_mathdefaultsmode:D
\_kernel_primitive:NN \mathdelimitersmode \tex_mathdelimitersmode:D
\_kernel_primitive:NN \mathdir \tex_mathdir:D
\_kernel_primitive:NN \mathdirection \tex_mathdirection:D
\_kernel_primitive:NN \mathdisplayskipmode \tex_mathdisplayskipmode:D
\_kernel_primitive:NN \matheqdirmode \tex_matheqdirmode:D
\_kernel_primitive:NN \matheqnogapstep \tex_matheqnogapstep:D
\_kernel_primitive:NN \mathoption \tex_mathoption:D
\_kernel_primitive:NN \mathpenaltiesmode \tex_mathpenaltiesmode:D
\_kernel_primitive:NN \mathrulesfam \tex_mathrulesfam:D
Primitives from pdFTeX that LuaFTeX renames.

\_\_kernel\_primitive:NN \adjustspacing \_\_\_kernel\_primitive:tex_adjustspacing
\_\_kernel\_primitive:NN \copyfont \_\_\_kernel\_primitive:tex_copyfont
\_\_kernel\_primitive:NN \draftmode \_\_\_kernel\_primitive:tex_draftmode
\_\_kernel\_primitive:NN \expandafter \_\_\_kernel\_primitive:tex_expandafter
\_\_kernel\_primitive:NN \ifaddfont \_\_\_kernel\_primitive:tex_ifaddfont
\_\_kernel\_primitive:NN \ifabedim \_\_\_kernel\_primitive:tex_ifabedim
\_\_kernel\_primitive:NN \ifabanum \_\_\_kernel\_primitive:tex_ifabanum

331
The set of Unicode math primitives were introduced by \LaTeXe\ and \LuaTeX\ in a somewhat complex fashion: a few first as \Xe\TeX..., which were then renamed with \LuaTeX\ having a lot more. These names now all start \U... and mainly \Umath...
Primitives from µE
c

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE

Primitives from µE
Primitives from upTEX.

\__kernel_primitive:NN \iftdir \tex_iftdir:D
\__kernel_primitive:NN \ifybox \tex_ifybox:D
\__kernel_primitive:NN \ifydir \tex_ifydir:D
\__kernel_primitive:NN \inhibitglue \tex_inhibitglue:D
\__kernel_primitive:NN \inhibitxspcode \tex_inhibitxspcode:D
\__kernel_primitive:NN \jcharwidowpenalty \tex_jcharwidowpenalty:D
\__kernel_primitive:NN \jfam \tex_jfam:D
\__kernel_primitive:NN \jfont \tex_jfont:D
\__kernel_primitive:NN \jis \tex_jis:D
\__kernel_primitive:NN \kanjiskip \tex_kanjiskip:D
\__kernel_primitive:NN \kansuji \tex_kansuji:D
\__kernel_primitive:NN \kansujichar \tex_kansujichar:D
\__kernel_primitive:NN \kcatcode \tex_kcatcode:D
\__kernel_primitive:NN \kuten \tex_kuten:D
\__kernel_primitive:NN \lastnodechar \tex_lastnodechar:D
\__kernel_primitive:NN \lastnodefont \tex_lastnodefont:D
\__kernel_primitive:NN \lastnodesubtype \tex_lastnodesubtype:D
\__kernel_primitive:NN \noautospacing \tex_noautospacing:D
\__kernel_primitive:NN \noautoxspacing \tex_noautoxspacing:D
\__kernel_primitive:NN \pagefistretch \tex_pagefistretch:D
\__kernel_primitive:NN \postbreakpenalty \tex_postbreakpenalty:D
\__kernel_primitive:NN \ptexlineendmode \tex_lineendmode:D
\__kernel_primitive:NN \ptexminorversion \tex_ptexminorversion:D
\__kernel_primitive:NN \ptexrevision \tex_ptexrevision:D
\__kernel_primitive:NN \readpapersizespecial \tex_readpapersizespecial:D
\__kernel_primitive:NN \scriptbaselineshiftfactor \tex_scriptbaselineshiftfactor:D
\__kernel_primitive:NN \scriptscriptbaselineshiftfactor \tex_scriptscriptbaselineshiftfactor:D
\__kernel_primitive:NN \showmode \tex_showmode:D
\__kernel_primitive:NN \sjis \tex_sjis:D
\__kernel_primitive:NN \tate \tex_tate:D
\__kernel_primitive:NN \tbaselineshift \tex_tbaselineshift:D
\__kernel_primitive:NN \textbaselineshiftfactor \tex_textbaselineshiftfactor:D
\__kernel_primitive:NN \tfont \tex_tfont:D
\__kernel_primitive:NN \toucs \tex_toucs:D
\__kernel_primitive:NN \ucs \tex_ucs:D
\__kernel_primitive:NN \xkanjiskip \tex_xkanjiskip:D
\__kernel_primitive:NN \xspcode \tex_xspcode:D
\__kernel_primitive:NN \ybaselineshift \tex_ybaselineshift:D
\__kernel_primitive:NN \yoko \tex_yoko:D
\__kernel_primitive:NN \vfi \tex_vfi:D

Primitives from upTeX.

\__kernel_primitive:NN \currentcjktoken \tex_currentcjktoken:D
\__kernel_primitive:NN \disablecjktoken \tex_disablecjktoken:D
\__kernel_primitive:NN \enablecjktoken \tex_enablecjktoken:D
\__kernel_primitive:NN \forcecjktoken \tex_forcecjktoken:D
\__kernel_primitive:NN \kchar \tex_kchar:D
\__kernel_primitive:NN \kchardef \tex_kchardef:D
\__kernel_primitive:NN \kuten \tex_kuten:D
\__kernel_primitive:NN \uptexrevision \tex_uptexrevision:D

336
Omega primitives provided by \TeX (listed separately mainly to allow understanding of their source).

Newer cross-engine primitives.

End of the “just the names” part of the source.

The job is done: close the group (using the primitive renamed!).

\LaTeXe moves a few primitives, so these are sorted out. In newer versions of \LaTeXe, \texttt{expl3} is loaded rather early, so only some primitives are already renamed, so we need two tests here. At the beginning of the \LaTeXe format, the primitives \texttt{\end} and \texttt{\input} are renamed, and only later on the other ones.

If \texttt{\@@@@hyph} is defined, we are loading \texttt{expl3} in a pre-2020/10/01 release of \LaTeXe, so a few other primitives have to be tested as well.

The \texttt{\shipout} primitive is particularly tricky as a number of packages want to hook in here. First, we see if a sufficiently-new kernel has saved a copy: if it has, just use that. Otherwise, we need to check each of the possible packages/classes that might move it: here, we are looking for those which do not delay action to the \texttt{\AtBeginDocument} hook. (We cannot use \texttt{\primitive} as that doesn’t allow us to make a direct copy of the primitive itself.) As we know that \LaTeXe is in use, we use it’s \texttt{\@tfor} loop here.
Some tidying up is needed for \(\text{pdf})\text{tracingfonts}. Newer Lua\TeX{} has this simply as \texttt{tracingfonts}, but that is overwritten by the \LaTeX{}2\varepsilon kernel. So any spurious definition has to be removed, then the real version saved either from the pdf\TeX{} name or from Lua\TeX{}. In the latter case, we leave \texttt{@@tracingfonts} available: this might be useful and almost all \LaTeX{}2\varepsilon users will have expl3 loaded by fontspec. (We follow the usual kernel convention that \texttt{@@} is used for saved primitives.)

That is also true for the Lua\TeX{} primitives under \LaTeX{}2\varepsilon (depending on the format-building date). There are a few primitives that get the right names anyway so are missing here!
Which also covers those slightly odd ones.

Only pdfTeX and LuaTeX define \pdfmapfile and \pdfmapline: Tidy up the fact that some format-building processes leave a couple of questionable decisions about that!
A few packages do unfortunate things to date-related primitives.

Up to v0.80, LuaTeX defines the pdfTeX version data: rather confusing. Removing them means that \text{pdfTeXversion} is a marker for pdfTeX alone: useful in engine-dependent code later.

cslatex moves a couple of primitives which we recover here; as there is no other marker, we can only work by looking for the names.
a few more primitives are moved: they are implemented using some Lua code in the current ConTeXt.

\text_ifdefined:D \normalend  \normalend
\text_let:D \tex_end:D \normaleveryjob
\text_let:D \tex_input:D \normalinput
\text_let:D \tex_language:D \normallanguage
\text_let:D \tex_mathop:D \normalmathop
\text_let:D \tex_month:D \normalmonth
\text_let:D \tex_outer:D \normalouter
\text_let:D \tex_over:D \normalvcenter
\text_let:D \tex_vcenter:D \normalvcenter
\text_let:D \tex_unexpanded:D \normalunexpanded
\text_let:D \tex_expanded:D \normalexpanded
\tex_fi:D
\text_ifdefined:D \normalitaliccorrection
\text_let:D \tex_hoffset:D \normalhoffset
\text_let:D \tex_italiccorrection:D \normalitaliccorrection
\text_let:D \tex_voffset:D \normalvoffset
\text_let:D \tex_showtokens:D \normalshowtokens
\text_let:D \tex_bodydir:D \spac_directions_normal_body_dir
\text_let:D \tex_pagedir:D \spac_directions_normal_page_dir
\tex_fi:D
\text_ifdefined:D \normalleft
\text_let:D \tex_left:D \normalleft
\text_let:D \tex_middle:D \normalmiddle
\text_let:D \tex_right:D \normalright
\tex_fi:D
\langle \text \rangle

In Lua\TeX, we additionally emulate some primitives using Lua code.

\text_strcmp:D Compare two strings, expanding to 0 if they are equal, \(-1\) if the first one is smaller and 1 if the second one is smaller. Here “smaller” refers to codepoint order which does not correspond to the user expected order for most non-ASCII strings.

local minus_tok = token_new(string.byte‘-’, 12)
local zero_tok = token_new(string.byte‘0’, 12)
local one_tok = token_new(string.byte‘1’, 12)
lucmd(’\text_strcmp:D’, function()
local first = scan_string()
local second = scan_string()
if first < second then
  put_next(minus_tok, one_tok)
else
  put_next(first == second and zero_tok or one_tok)
end
end)

(End definition for \text_strcmp:D. This function is documented on page ??.)

\text_Ucharcat:D Creating arbitrary chars using \text.cprint. The alternative approach using token.new(...) is about 10% slower but needed to create arbitrary space tokens.
local sprint = tex.sprint
local cprint = tex.cprint
luacmd('\text_Ucharcat:\text_D', function()
  local charcode = scan_int()
  local categcode = scan_int()
  if categcode == 10 then
    sprint(token_new(charcode, 10))
  else
    cprint(charcode, utf8_char(charcode))
  end
end, 'global')

(End definition for \text_Ucharcat:\text_D. This function is documented on page ??.)

\text_filesize:\text_D
Wrap the function from \text_ltxutils\text{}.  
luacmd('\text_filesize:\text_D', function()
  local size = filesize(scan_string())
  if size then write(size) end
end, 'global')

(End definition for \text_filesize:\text_D. This function is documented on page ??.)

\text_mdfivesum:\text_D
There are two cases: Either hash a file or a string. Both are already implemented in \text_l3luatex\text{} or built-in.
luacmd('\text_mdfivesum:\text_D', function()
  local hash
  if scan_keyword"file" then
    hash = filemd5sum(scan_string())
  else
    hash = md5_HEX(scan_string())
  end
  if hash then write(hash) end
end, 'global')

(End definition for \text_mdfivesum:\text_D. This function is documented on page ??.)

\text_filemoddate:\text_D
A primitive for getting the modification date of a file.
luacmd('\text_filemoddate:\text_D', function()
  local date = filemoddate(scan_string())
  if date then write(date) end
end, 'global')

(End definition for \text_filemoddate:\text_D. This function is documented on page ??.)

\text_filedump:\text_D
An emulated primitive for getting a hexdump from a (partial) file. The length has a default of 0. This is consistent with pdf\textTeX\text{}, but it effectively makes the primitive useless without an explicit length. Therefore we allow the keyword whole to be used instead of a length, indicating that the whole remaining file should be read.
luacmd('\text_filedump:\text_D', function()
  local offset = scan_keyword"offset" and scan_int() or nil
  local length = scan_keyword"length" and scan_int()
  or not scan_keyword"whole" and 0 or nil
  local data = filedump(scan_string(), offset, length)
  if data then write(data) end
end, 'global')
(End definition for \text{\texttt{filedump}}:D. This function is documented on page ??.)

1426 ⟨/lua⟩

1427 ⟨/package⟩
Chapter 40

l3kernel-functions:
kernel-reserved functions

40.1 Internal kernel functions

\__kernel_chk_cs_exist:N \__kernel_chk_cs_exist:N \{cs\}
This function is only created if debugging is enabled. It checks that \(cs\) exists according to the criteria for \cs_if_exist_p:N\, and if not raises a kernel-level error.

\__kernel_chk_defined:NT
\__kernel_chk_defined:NT \{variable\} \{(true code)\}
If \{variable\} is not defined (according to \cs_if_exist:NTF\), this triggers an error, otherwise the \{true code\} is run.

\__kernel_chk_expr:nNnN \{\{expr\}\} \{eval\} \{(convert)\} \{caller\}
This function is only created if debugging is enabled. By default it is equivalent to \use_i:nnnn. When expression checking is enabled, it leaves in the input stream the result of \tex_the:D \{eval\} \{expr\} \tex_relax:D after checking that no token was left over. If any token was not taken as part of the expression, there is an error message displaying the result of the evaluation as well as the \{caller\}. For instance \{eval\} can be \__int_eval:w and \{caller\} can be \int_eval:n or \int_set:Nn. The argument \{convert\} is empty except for mu expressions where it is \tex_mutoglue:D, used for internal purposes.

\__kernel_chk_tl_type:NnnT \__kernel_chk_tl_type:NnnT \{control sequence\} \{(specific type)\}
\{(reconstruction)\} \{(true code)\}
Helper to test that the \{control sequence\} is a variable of the given \{specific type\} of token list. Produces suitable error messages if the \{control sequence\} does not exist, or if it is not a token list variable at all, or if the \{control sequence\} differs from the result of x-expanding \{reconstruction\}. If all of these tests succeed then the \{true code\} is run.
\__kernel_cs_parm_from_arg_count:nnF \__kernel_cs_parm_from_arg_count:nnF \{\texttt{follow-on}\}\ {\{\texttt{args}\}\} \{\texttt{false code}\}

Evaluates the number of \langle \texttt{args} \rangle and leaves the \langle \texttt{follow-on} \rangle code followed by a brace group containing the required number of primitive parameter markers (#1, etc.). If the number of \langle \texttt{args} \rangle is outside the range [0, 9], the \langle \texttt{false code} \rangle is inserted instead of the \langle \texttt{follow-on} \rangle.

\__kernel_dependency_version_check:Nn \__kernel_dependency_version_check:Nn \{\langle \texttt{date} \rangle\}\ {\{\texttt{file}\}\}
\__kernel_dependency_version_check:nn \__kernel_dependency_version_check:nn \{\langle \texttt{date} \rangle\}\ {\{\texttt{file}\}\}

Checks if the loaded version of the expl3 kernel is at least \langle \texttt{date} \rangle, required by \langle \texttt{file}\rangle. If the kernel date is older than \langle \texttt{date} \rangle, the loading of \langle \texttt{file} \rangle is aborted and an error is raised.

\__kernel_deprecation_code:nn \__kernel_deprecation_code:nn \{\langle \texttt{error code} \rangle\}\ {\{\texttt{working code}\}\}

Stores both an \langle \texttt{error} \rangle and \langle \texttt{working} \rangle definition for given material such that they can be exchanged by \texttt{\debug_on:} and \texttt{\debug_off:}.

\__kernel_exp_not:w ∗ \__kernel_exp_not:w \{\langle \texttt{expandable tokens} \rangle\} \{\langle \texttt{content} \rangle\}

Carries out expansion on the \langle \texttt{expandable tokens} \rangle before preventing further expansion of the \langle \texttt{content} \rangle as for \texttt{\exp_not:n}. Typically, the \langle \texttt{expandable tokens} \rangle will alter the nature of the \langle \texttt{content} \rangle, i.e. allow it to be generated in some way.

\l__kernel_expl_bool
A boolean which records the current code syntax status: \texttt{true} if currently inside a code environment. This variable should only be set by \texttt{\ExplSyntaxOn/\ExplSyntaxOff}.

(End definition for \l__kernel_expl_bool.)

\c__kernel_expl_date_tl
A token list containing the release date of the l3kernel preloaded in \LaTeX{} used to check if dependencies match.

(End definition for \c__kernel_expl_date_tl.)

\__kernel_file_missing:n \__kernel_file_missing:n \{\langle \texttt{name} \rangle\}

Expands the \langle \texttt{name} \rangle as per \__kernel_file_name_sanitize:n then produces an error message indicating that this file was not found.

\__kernel_file_name_sanitize:n ∗ \__kernel_file_name_sanitize:n \{\langle \texttt{name} \rangle\}

Expands the file name using a \csname-based approach, and relies on active characters (for example from UTF-8 characters) being properly set up to expand to a expansion-safe version using \texttt{\ifcsname}. This is less conservative than the token-by-token approach used before, but it is much faster.

\__kernel_file_input_push:n \__kernel_file_input_push:n \{\langle \texttt{name} \rangle\}
\__kernel_file_input_pop: \__kernel_file_input_pop:

Used to push and pop data from the internal file stack: needed only in package mode, where interfacing with the \LaTeX{} kernel is necessary.
\__kernel_int_add:nnn \__kernel_int_add:nnn \{\{integer_1\}\} \{\{integer_2\}\} \{\{integer_3\}\}

Expands to the result of adding the three \{\{integers\}\} (which must be suitable input for \int_eval:w), avoiding intermediate overflow. Overflow occurs only if the overall result is outside \([-2^{31}+1, 2^{31}−1]\). The \{\{integers\}\} may be of the form \int_eval:w \ldots \text{\scan_stop}: but may be evaluated more than once.

\__kernel_intarray_gset:Nnn \__kernel_intarray_gset:Nnn \{intarray\var\} \{\{index\}\} \{\{value\}\}

New: 2018-03-31

Faster version of \intarray_gset:Nnn. Stores the \{\{value\}\} into the \{\{integer array\} variable\} at the \{\{position\}\}. The \{\{index\}\} and \{\{value\}\} must be suitable for a direct assignment to a \TeX{} count register, for instance expanding to an integer denotation or obtained through the primitive \numexpr (which may be un-terminated). No bound checking is performed: the caller is responsible for ensuring that the \{\{position\}\} is between 1 and the \intarray_count:N, and the \{\{value\}\}'s absolute value is at most \(2^{30}−1\). Assignments are always global.

\__kernel_intarray_item:Nn \__kernel_intarray_item:Nn \{intarray\var\} \{\{index\}\}

New: 2018-03-31

Faster version of \intarray_item:Nn. Expands to the integer entry stored at the \{\{index\}\} in the \{\{integer array\} variable\}. The \{\{index\}\} must be suitable for a direct assignment to a \TeX{} count register and must be between 1 and the \intarray_count:N, lest a low-level \TeX{} error occur.

\__kernel_intarray_range_to_clist:Nnn \__kernel_intarray_range_to_clist:Nnn \{intarray\var\} \{\{start\ index\}\} \{\{end\ index\}\}

New: 2020-07-12

Converts to integer denotations separated by commas the entries of the \{\{intarray\}\} from positions \{\{start\ index\}\} to \{\{end\ index\}\} included. The \{\{start\ index\}\} and \{\{end\ index\}\} must be suitable for a direct assignment to a \TeX{} count register, must be between 1 and the \intarray_count:N, and be suitably ordered. All tokens have category code other.

\__kernel_intarray_gset_range_from_clist:Nnn \__kernel_intarray_gset_range_from_clist:Nnn \{intarray\var\} \{\{start\ index\}\} \{\{integer\ comma\ list\}\}

New: 2020-07-12

Stores the entries of the \{\{clist\}\} as entries of the \{\{intarray\var\}\} starting from the \{\{start\ index\}\}, upwards. This is done without any bound checking. The \{\{start\ index\}\} and all entries of the \{\{integer comma list\}\} (which do not undergo space trimming and brace stripping as in normal clist mappings) must be suitable for a direct assignment to a \TeX{} count register. An empty entry may stop the loop.

\__kernel_ior_open:Nn \__kernel_ior_open:No

\{\{stream\}\} \{\{file\ name\}\}

This function has identical syntax to the public version. However, is does not take precautions against active characters in the \{\{file\ name\}\}, and it does not attempt to add a \{\{path\}\} to the \{\{file\ name\}\}: it is therefore intended to be used by higher-level functions which have already fully expanded the \{\{file\ name\}\} and which need to perform multiple open or close operations. See for example the implementation of \file_get_full-name:nN,
\_\_kernel\_iow\_with:NNn  \_\_kernel\_iow\_with:NNn \(\text{integer}\) \{\(\text{value}\)\} \{\(\text{code}\)\}

If the \(\text{integer}\) is equal to the \(\text{value}\) then this function simply runs the \(\text{code}\). Otherwise it saves the current value of the \(\text{integer}\), sets it to the \(\text{value}\), runs the \(\text{code}\), and restores the \(\text{integer}\) to its former value. This is used to ensure that the \texttt{\_newlinechar} is 10 when writing to a stream, which lets \texttt{\_iow\_newline} work, and that \texttt{\_errorcontextlines} is \(-1\) when displaying a message.

\_\_kernel\_kern:n  \_\_kernel\_kern:n \{\(\text{length}\)\}

Inserts a kern of the specified \(\text{length}\), a dimension expression.

(End definition for \_\_kernel\_kern:n.)

\_\_kernel\_prg\_map\_int  \_\_kernel\_prg\_map\_int

This integer is used by non-expandable mapping functions to track the level of nesting in force. The functions \(\texttt{\langle type\rangle\_map\_1:w}, \texttt{\langle type\rangle\_map\_2:w}, \text{etc.}\), labelled by \texttt{\_\_kernel\_prg\_map\_int} hold functions to be mapped over various list datatypes in inline and variable mappings.

(End definition for \_\_kernel\_prg\_map\_int.)
\__kernel_quark_new_test:N \__kernel_quark_new_test:N \langle name \rangle:(arg spec)

Defines a quark-test function \langle name \rangle:(arg spec) which tests if its argument is \q__-\langle namespace\rangle_{recursion\_tail}, then acts accordingly, as described below for each possible (arg spec).

The \langle namespace\rangle is determined as the first (nonempty) \_ delimited word in \langle name \rangle and is used internally in the definition of auxiliaries. The function \__kernel_quark_new_test:N does not define the \q__\langle namespace\rangle_{recursion\_tail} and \q__\langle namespace\rangle_{recursion\_stop} quarks. They should be manually defined with \quark\_new:N.

There are 6 different types of quark-test functions. Which one is defined depends on the (arg spec), which must be one of the options listed now. Four of them are modeled after \quark_if_recursion_tail:(N|n) and \quark_if_recursion_tail_\_do:(N|n)n.

\n defines \langle name \rangle:n such that it checks if #1 contains only \q__\langle namespace\rangle_{recursion\_tail}, and if so consumes all tokens up to \q__\langle namespace\rangle_{recursion\_stop} (c.f. \quark_if_recursion_tail_\_stop:n).

\nn defines \langle name \rangle:nn such that it checks if #1 contains only \q__\langle namespace\rangle_{recursion\_tail}, and if so consumes all tokens up to \q__\langle namespace\rangle_{recursion\_stop}, then executes the code #2 after that (c.f. \quark_if_recursion_tail_\_stop_\_do:nn).

\N defines \langle name \rangle:N such that it checks if #1 is \q__\langle namespace\rangle_{recursion\_tail}, and if so consumes all tokens up to \q__\langle namespace\rangle_{recursion\_stop} (c.f. \quark_if_recursion_tail_stop:N).

\nn defines \langle name \rangle:nn such that it checks if #1 is \q__\langle namespace\rangle_{recursion\_tail}, and if so consumes all tokens up to \q__\langle namespace\rangle_{recursion\_stop}, then executes the code #2 after that (c.f. \quark_if_recursion_tail_stop_\_do:nn).

The last two are modeled after \quark_if_recursion_tail_break:(n|N)N, and in those cases the quark \q__\langle namespace\rangle_{recursion\_stop} is not used (and thus needs not be defined).

\nN defines \langle name \rangle:nN such that it checks if #1 contains only \q__\langle namespace\rangle_{recursion\_tail}, and if so uses the \langle type\rangle_{map\_break}: function #2.

\NN defines \langle name \rangle:NN such that it checks if #1 is \q__\langle namespace\rangle_{recursion\_tail}, and if so uses the \langle type\rangle_{map\_break}: function #2.

Any other signature, as well as a function without signature are errors, and in such case the definition is aborted.
\_\_kernel_quark_new_conditional:Nn \_\_kernel_quark_new_conditional:Nn \_\_\langle namespace\rangle_quark_if\_\langle name\rangle:\langle arg spec\rangle \{\langle conditions\rangle\}

Defines a collection of quark conditionals that test if their argument is the quark \q\_\_\langle namespace\rangle\_\langle name\rangle and perform suitable actions. The \langle conditions\rangle are a comma-separated list of one or more of p, T, F, and TF, and one conditional is defined for each \langle condition\rangle in the list, as described for \texttt{\prg_new_conditional:Nn}. The conditionals are defined using \texttt{\prg_new_conditional:Nn}, so that their name is obtained by adding p, T, F, or TF to the base name \_\_\langle namespace\rangle\_quark_if\_\langle name\rangle:\langle arg spec\rangle.

The first argument of \_\_kernel_quark_new_conditional:Nn must contain \texttt{\quark_if} and :, as these markers are used to determine the \langle name\rangle of the quark \q\_\_\langle namespace\rangle\_\langle name\rangle to be tested. This quark should be manually defined with \texttt{\quark_new:N}, as \_\_kernel_quark_new_conditional:Nn does \textit{not} define it.

The function \_\_kernel_quark_new_conditional:Nn can define 2 different types of quark conditionals. Which one is defined depends on the \langle arg spec\rangle, which \textit{must} be one of the following options, modeled after \texttt{\quark_if_nil:(N|n)(TF)}.

\texttt{n} defines \_\_\langle namespace\rangle\_quark_if\_\langle name\rangle:n(TF) such that it checks if \#1 contains only q\_\_\langle namespace\rangle\_\langle name\rangle, and executes the proper conditional branch.

\texttt{N} defines \_\_\langle namespace\rangle\_quark_if\_\langle name\rangle:N(TF) such that it checks if \#1 is q\_\_\langle namespace\rangle\_\langle name\rangle, and executes the proper conditional branch.

Any other signature, as well as a function without signature are errors, and in such case the definition is aborted.

\c__kernel_randint_max_int Maximal allowed argument to \_\_kernel_randint:n. Equal to $2^{17} - 1$.

(End definition for \c__kernel_randint_max_int)

\_\_kernel_randint:n \_\_kernel_randint:n \{\langle max\rangle\}

Used in an integer expression this gives a pseudo-random number between 1 and \langle max\rangle included. One must have \langle max\rangle \leq 2^{17} - 1. The \langle max\rangle must be suitable for \texttt{\int_value:w} (and any \texttt{\int_eval:w} must be terminated by \texttt{\scan_stop:} or equivalent).

\_\_kernel_randint:nn \_\_kernel_randint:nn \{\langle min\rangle\} \{\langle max\rangle\}

Used in an integer expression this gives a pseudo-random number between \langle min\rangle and \langle max\rangle included. The \langle min\rangle and \langle max\rangle must be suitable for \texttt{\int_value:w} (and any \texttt{\int_eval:w} must be terminated by \texttt{\scan_stop:} or equivalent). For small ranges $R = \langle max\rangle - \langle min\rangle + 1 \leq 2^{17} - 1$, \langle min\rangle - 1 + \_\_kernel_randint:n\{R\} is faster.

\_\_kernel_register_show:N \_\_kernel_register_show:N \_\_kernel_register_show:N \langle register\rangle

Used to show the contents of a \texttt{\TeX} register at the terminal, formatted such that internal parts of the mechanism are not visible.

\_\_kernel_register_log:N \_\_kernel_register_log:N \_\_kernel_register_log:N \langle register\rangle

Used to write the contents of a \texttt{\TeX} register to the log file in a form similar to \_\_kernel_register_show:N.

349
\_\text{\texttt{\_\_kernel\_str\_to\_other}}:n \quad \_\text{\texttt{\_\_kernel\_str\_to\_other}}:n \ (\langle \text{token list} \rangle)

Converts the \langle \text{token list} \rangle to a \langle \text{other string} \rangle, where spaces have category code “other”. This function can be \texttt{f}-expanded without fear of losing a leading space, since spaces do not have category code 10 in its result. It takes a time quadratic in the character count of the string.

\_\text{\texttt{\_\_kernel\_str\_to\_other\_fast}}:n \quad \_\text{\texttt{\_\_kernel\_str\_to\_other\_fast}}:n \ (\langle \text{token list} \rangle)

Same behaviour \_\text{\texttt{\_\_kernel\_str\_to\_other}}:n but only restricted-expandable. It takes a time linear in the character count of the string.

\_\text{\texttt{\_\_kernel\_tl\_to\_str}}:w \quad \_\text{\texttt{\_\_kernel\_tl\_to\_str}}:w \ (\langle \text{expandable tokens} \rangle) \ (\langle \text{tokens} \rangle)

Carries out expansion on the \langle \text{expandable tokens} \rangle before conversion of the \langle \text{tokens} \rangle to a string as describe for \_\text{\texttt{\_\_tl\_to\_str}}:n. Typically, the \langle \text{expandable tokens} \rangle will alter the nature of the \langle \text{tokens} \rangle, i.e. allow it to be generated in some way. This function requires only a single expansion.

\_\text{\texttt{\_\_kernel\_tl\_set}}:Nx \quad \_\text{\texttt{\_\_kernel\_tl\_set}}:Nx \ (\text{tl var}) \ (\langle \text{tokens} \rangle)
\_\text{\texttt{\_\_kernel\_tl\_gset}}:Nx

Fully expands \langle \text{tokens} \rangle and assigns the result to \langle \text{tl var} \rangle. \langle \text{tokens} \rangle must be given in braces and there must be no token between \langle \text{tl var} \rangle and \langle \text{tokens} \rangle.

\section{40.2 Kernel backend functions}

These functions are required to pass information to the backend. The nature of these means that they are defined only when the relevant backend is in use.

\_\text{\texttt{\_\_kernel\_backend\_literal}}:n \quad \_\text{\texttt{\_\_kernel\_backend\_literal}}:n \ (\langle \text{content} \rangle)
\_\text{\texttt{\_\_kernel\_backend\_literal}}:x
\_\text{\texttt{\_\_kernel\_backend\_literal}}:nx

Adds the \langle \text{content} \rangle literally to the current vertical list as a whatsit. The nature of the \langle \text{content} \rangle will depend on the backend in use.

\_\text{\texttt{\_\_kernel\_backend\_literal\_postscript}}:n \quad \_\text{\texttt{\_\_kernel\_backend\_literal\_postscript}}:n \ (\langle \text{PostScript} \rangle)
\_\text{\texttt{\_\_kernel\_backend\_literal\_postscript}}:x

Adds the \langle \text{PostScript} \rangle literally to the current vertical list as a whatsit. No positioning is applied.

\_\text{\texttt{\_\_kernel\_backend\_literal\_pdf}}:n \quad \_\text{\texttt{\_\_kernel\_backend\_literal\_pdf}}:n \ (\langle \text{PDF instructions} \rangle)
\_\text{\texttt{\_\_kernel\_backend\_literal\_pdf}}:x

Adds the \langle \text{PDF instructions} \rangle literally to the current vertical list as a whatsit. No positioning is applied.

\_\text{\texttt{\_\_kernel\_backend\_literal\_svg}}:n \quad \_\text{\texttt{\_\_kernel\_backend\_literal\_svg}}:n \ (\langle \text{SVG instructions} \rangle)
\_\text{\texttt{\_\_kernel\_backend\_literal\_svg}}:x

Adds the \langle \text{SVG instructions} \rangle literally to the current vertical list as a whatsit. No positioning is applied.
\_\_kernel_backend_postscript:n \_\_kernel_backend_postscript:n \{\langle PostScript \rangle\}
\_\_kernel_backend_postscript:x

Adds the \langle PostScript \rangle to the current vertical list as a whatsit. The PostScript reference point is adjusted to match the current position. The PostScript is inserted inside a SDict begin/end pair.

\_\_kernel_backend_align_begin: \_\_kernel_backend_align_begin:
\_\_kernel_backend_align_end: \_\_kernel_backend_align_end:
\langle PostScript literals \rangle
\_\_kernel_backend_align_end:

Arranges to align the PostScript and DVI current positions and scales.

\_\_kernel_backend_scope_begin: \_\_kernel_backend_scope_begin:
\_\_kernel_backend_scope_end: \_\_kernel_backend_scope_end:
\langle content \rangle
\_\_kernel_backend_scope_end:

Creates a scope for instructions at the backend level.

\_\_kernel_backend_matrix:n \_\_kernel_backend_matrix:n \{\langle matrix \rangle\}
\_\_kernel_backend_matrix:x

Applies the \langle matrix \rangle to the current transformation matrix.

\_\_kernel_backend_header_bool

Specifies whether to write headers for the backend.

\_\_kernel_color_stack_int

The color stack used in pdf\TeX\ and Lua\TeX\ for the main color.
Chapter 41

l3basics implementation

41.1 Renaming some \TeX{} primitives (again)

Having given all the \TeX{} primitives a consistent name, we need to give sensible names to the ones we actually want to use. These will be defined as needed in the appropriate modules, but we do a few now, just to get started.\footnote{This renaming gets expensive in terms of csname usage, an alternative scheme would be to just use the \texttt{\textbackslash tex\ldots:} name in the cases where no good alternative exists.}

\begin{verbatim}
\iftrue: \iffalse: \or: \else: \fi:
\reverse_if:N \if:w \if_charcode:w \if_catcode:w \if_meaning:w
\tex_let:D \if_true: \tex_iftrue:D \tex_let:D \if_false: \tex_iffalse:D
\tex_let:D \or: \tex_or:D \tex_let:D \else: \tex_else:D \tex_let:D \fi: \tex_fi:D
\tex_let:D \reverse_if:N \tex_unless:D \tex_let:D \if:w \tex_if:D
\tex_let:D \if_charcode:w \tex_if:D \tex_let:D \if_catcode:w \tex_ifcat:D
\tex_let:D \if_meaning:w \tex_ifx:D \tex_let:D \if_bool:N \tex_ifodd:D
\end{verbatim}

Then some conditionals.

\begin{verbatim}
\if_mode_math: \if_mode_horizontal: \if_mode_vertical: \if_mode_inner:
\tex_let:D \if_mode_math: \tex_ifmmode:D \tex_let:D \if_mode_horizontal: \tex_ifhmode:D
\tex_let:D \if_mode_vertical: \tex_ifvmode:D \tex_let:D \if_mode_inner: \tex_ifinner:D
\end{verbatim}

\text{TEX} lets us detect some of its modes.

\begin{verbatim}
\if_cs_exist:N \if_cs_exist:w \cs:w \cs_end:
\tex_let:D \if_cs_exist:N \tex_ifdefined:D \tex_let:D \if_cs_exist:w \tex_ifcsname:D
\cs:w \tex_csname:D \tex_let:D \cs_end: \tex_endcsname:D
\end{verbatim}

Building csnames and testing if control sequences exist.
The five \exp\_ functions are used in the \texttt{l3expan} module where they are described.

\begin{verbatim}
\texttt{\exp_after:wN} \texttt{\exp_not:N} \texttt{\exp_not:n} \texttt{\exp:w} \texttt{\exp:romannumeral:D}
\end{verbatim}

(End definition for \exp\_after:wN, \exp\_not:N, and \exp\_not:n. These functions are documented on page 28.)

\begin{verbatim}
\texttt{\token_to_meaning:N} \texttt{\cs_meaning:N} \texttt{\tl_to_str:n} \texttt{\token_to_str:N} \texttt{\__kernel_tl_to_str:w}
\end{verbatim}

(End definition for \token_to_meaning:N and \cs_meaning:N. These functions are documented on page 38.)

\begin{verbatim}
\texttt{\scan_stop:} \texttt{\group_begin:} \texttt{\group_end:}
\end{verbatim}

(End definition for \scan\_stop:, \group\_begin:, and \group\_end:. These functions are documented on page 13.)

\begin{verbatim}
\texttt{\if_int_compare:w} \texttt{\__int_to_roman:w}
\end{verbatim}

(End definition for \if\_int\_compare:w and \__int\_to\_roman:w. This function is documented on page 167.)

\begin{verbatim}
\texttt{\group_insert_after:N}
\end{verbatim}

(End definition for \group\_insert\_after:N. This function is documented on page 14.)

\begin{verbatim}
\texttt{\exp_args:Nc} \texttt{\exp_args:cc}
\end{verbatim}

(End definition for \exp\_args:Nc and \exp\_args:cc. These functions are documented on page 34.)

353
A small number of variants defined by hand. Some of the necessary functions (\use_{i:nn}, \use_{ii:nn}, and \exp_args:NNc) are not defined at that point yet, but will be defined before those variants are used. The \cs_meaning:c command must check for an undefined control sequence to avoid defining it mistakenly.

\tex_def:D \token_to_str:c \{ \exp_args:Nc \token_to_str:N \} \cs_meaning:c #1
{ \if_cs_exist:w #1 \cs_end:
  \exp_after:wN \use_i:nn
\else:
  \exp_after:wN \use_ii:nn
\fi:
{ \exp_args:Nc \cs_meaning:N {#1} }
{ \tl_to_str:n {undefined} }
}
\tex_let:D \token_to_meaning:c = \cs_meaning:c

(End definition for \token_to_meaning:N. This function is documented on page 187.)

\c_zero_int
We need the constant \c_zero_int which is used by some functions in the l3alloc module. The rest are defined in the l3int module – at least for the ones that can be defined with \tex_chardef:D or \tex_mathchardef:D. For other constants the l3int module is required but it can't be used until the allocation has been set up properly!

\tex_chardef:D \c_zero_int = 0 ~
\tex_ifdefined:D \tex_luatexversion:D
\tex_chardef:D \c_max_register_int = 65 535 ~
\tex_else:D
\tex_ifdefined:D \tex_omathchardef:D
\tex_omathchardef:D \c_max_register_int = 65535 ~
\tex_else:D
\tex_mathchardef:D \c_max_register_int = 32767 ~
\tex_fi:D
\tex_fi:D

(End definition for \c_zero_int. This variable is documented on page 166.)

\c_max_register_int
This is here as this particular integer is needed both in package mode and to bootstrap l3alloc, and is documented in l3int. LuaTeX and those which contain parts of the Omega extensions have more registers available than \eTeX.

\tex_ifdefined:D \tex_luatexversion:D
\tex_chardef:D \c_max_register_int = 65 535 ~
\tex_else:D
\tex_ifdefined:D \tex_omathchardef:D
\tex_omathchardef:D \c_max_register_int = 65535 ~
\tex_else:D
\tex_mathchardef:D \c_max_register_int = 32767 ~
\tex_fi:D
\tex_fi:D

(End definition for \c_max_register_int. This variable is documented on page 166.)

41.3 Defining functions
We start by providing functions for the typical definition functions. First the local ones.

All assignment functions in \LaTeX3 should be naturally protected; after all, the \TeX primitives for assignments are and it can be a cause of problems if others aren't.

\tex_def:D \cs_set_nopar:Npn \cs_set_nopar:Npx \cs_set:Npn \cs_set:Npx
\cs_set_protected_nopar:Npn \cs_set_protected:Npx
\cs_set_protected:Nn
\cs_set_protected:Nx

354
1493 \text_protected:D \text_long:D \text_def:D \cs_set:Npn
1494 { \text_long:D \text_def:D }
1495 \text_protected:D \text_long:D \text_def:D \cs_set:Npx
1496 { \text_long:D \text_def:D }
1497 \text_protected:D \text_long:D \text_def:D \cs_set_protected_nopar:Npn
1498 { \text_protected:D \text_def:D }
1499 \text_protected:D \text_long:D \text_def:D \cs_set_protected_nopar:Npx
1500 { \text_protected:D \text_def:D }
1501 \text_protected:D \text_long:D \text_def:D \cs_set_protected:Npn
1502 { \text_protected:D \text_long:D \text_def:D }
1503 \text_protected:D \text_long:D \text_def:D \cs_set_protected:Npx
1504 { \text_protected:D \text_long:D \text_def:D }

(End definition for \cs_set_nopar:Npn and others. These functions are documented on page 16.)

\cssetnopar:Npn \cssetnopar:Npx \csset:Npn \csset:Npx \csset_protected:Npn \csset_protected:Npx

Global versions of the above functions.

\csset_nopar:Npx \csset_nopar:Npx \csset:Npx \csset:Npx

(End definition for \csset_nopar:Npn and others. These functions are documented on page 16.)

41.4 Selecting tokens

\l__exp_internal_tl Scratch token list variable for \l3exp, used by \use:x, used in defining conditionals. We don’t use \tl methods because \l3basics is loaded earlier.

\csset_nopar:Npn \l__exp_internal_tl \l__exp_internal_tl

(End definition for \l__exp_internal_tl.)

\use:c This macro grabs its argument and returns a csname from it.

\csset:Npn \use:c #1 \cs:w #1 \csend:

(End definition for \use:c. This function is documented on page 20.)

\use:x Fully expands its argument and passes it to the input stream. Uses the reserved \l__-exp_internal_tl which we’ve set up above.

\csset_protected:Npn \use:x #1
{ \csset_nopar:Npx \l__exp_internal_tl \l__exp_internal_tl }
\use:e

In non-Lua\TeX\ engines older than 2019, \expanded is emulated.
\begin{verbatim}
\cs_set:Npn \use:e #1 { \tex_expanded:D {#1} }
\tex_ifdefined:D \tex_expanded:D \tex_else:D
\cs_set:Npn \use:e #1 { \exp_args:Ne \use:n {#1} }
\tex_fi:D
\end{verbatim}

(End definition for \use:e. This function is documented on page 25.)

\use:n, \use:nn, \use:nnn, \use:nnnn

These macros grab their arguments and return them back to the input (with outer braces removed).
\begin{verbatim}
\cs_set:Npn \use:n #1 {#1}
\cs_set:Npn \use:nn #1#2 {#1#2}
\cs_set:Npn \use:nnn #1#2#3 {#1#2#3}
\cs_set:Npn \use:nnnn #1#2#3#4 {#1#2#3#4}
\end{verbatim}

(End definition for \use:n and others. These functions are documented on page 23.)

\use_i:nn, \use_ii:nn, \use_iii:nn, \use_i_ii:nn, \use_i:nnn, \use_ii:nnn, \use_iii:nnn, \use_i_ii:nnn

The equivalent to \LaTeX\2e\’s \texttt{@firstoftwo} and \texttt{@secondoftwo}.
\begin{verbatim}
\cs_set:Npn \use_i:nn #1#2 {#1}
\cs_set:Npn \use_ii:nn #1#2 {#2}
\cs_set:Npn \use_i_ii:nn #1#2 {#1#2}
\cs_set:Npn \use_i:nnn #1#2#3 {#1}
\cs_set:Npn \use_ii:nnn #1#2#3 {#2}
\cs_set:Npn \use_iii:nnn #1#2#3 {#3}
\cs_set:Npn \use_i_ii:nnn #1#2#3 {#1#2}
\end{verbatim}

(End definition for \use_i:nn and \use_ii:nn. These functions are documented on page 24.)

\use_i:nnn, \use_ii:nnn, \use_iii:nnn, \use_iv:nnnn

We also need something for picking up arguments from a longer list.
\begin{verbatim}
\cs_set:Npn \use_i:nnn #1#2#3 {#1}
\cs_set:Npn \use_ii:nnn #1#2#3 {#2}
\cs_set:Npn \use_iii:nnn #1#2#3 {#3}
\cs_set:Npn \use_i_ii:nnn #1#2#3 {#1#2}
\cs_set:Npn \use_i:nnnn #1#2#3#4 {#1}
\cs_set:Npn \use_ii:nnnn #1#2#3#4 {#2}
\cs_set:Npn \use_iii:nnnn #1#2#3#4 {#3}
\cs_set:Npn \use_iv:nnnn #1#2#3#4 {#4}
\end{verbatim}

(End definition for \use_i:nnn and others. These functions are documented on page 24.)

\use_i_ii:nnn
\begin{verbatim}
\cs_set:Npn \use_i_ii:nn #1#2 {#2 #1}
\end{verbatim}

(End definition for \use_i_ii:nn. This function is documented on page 24.)

\use_none_delimit_by_q_nil:w, \use_none_delimit_by_q_stop:w, \use_none_delimit_by_q_recursion_stop:w

Functions that gobble everything until they see either \texttt{\q_nil}, \texttt{\q_stop}, or \texttt{\q_-recursion_stop}, respectively.
\begin{verbatim}
\cs_set:Npn \use_none_delimit_by_q_nil:w #1 \q_nil { }
\cs_set:Npn \use_none_delimit_by_q_stop:w #1 \q_stop { }
\cs_set:Npn \use_none_delimit_by_q_recursion_stop:w #1 \q_recursion_stop { }
\end{verbatim}

(End definition for \use_none_delimit_by_q_nil:w, \use_none_delimit_by_q_stop:w, and \use_none_delimit_by_q_recursion_stop:w. These functions are documented on page 25.)
Same as above but execute first argument after gobbling. Very useful when you need to skip the rest of a mapping sequence but want an easy way to control what should be expanded next.

\cs_set:Npn \use_i_delimit_by_q_nil:nw #1 \q_nil {#1}
\cs_set:Npn \use_i_delimit_by_q_stop:nw #1 \q_stop {#1}
\cs_set:Npn \use_i_delimit_by_q_recursion_stop:nw #1 \q_recursion_stop {#1}

(End definition for \use_i_delimit_by_q_nil:nw, \use_i_delimit_by_q_stop:nw, and \use_i_delimit_by_q_recursion_stop:nw. These functions are documented on page 25.)

### 41.5 Gobbling tokens from input

To gobble tokens from the input we use a standard naming convention: the number of tokens gobbled is given by the number of n's following the : in the name. Although we could define functions to remove ten arguments or more using separate calls of \use_:nnnnnnn, this is very non-intuitive to the programmer who will assume that expanding such a function once takes care of gobbling all the tokens in one go.

\cs_set:Npn \use_none:n #1 { }
\cs_set:Npn \use_none:nn #1#2 { }
\cs_set:Npn \use_none:nnn #1#2#3 { }
\cs_set:Npn \use_none:nnnn #1#2#3#4 { }
\cs_set:Npn \use_none:nnnnn #1#2#3#4#5 { }
\cs_set:Npn \use_none:nnnnnn #1#2#3#4#5#6 { }
\cs_set:Npn \use_none:nnnnnnn #1#2#3#4#5#6#7 { }
\cs_set:Npn \use_none:nnnnnnnn #1#2#3#4#5#6#7#8 { }
\cs_set:Npn \use_none:nnnnnnnn #1#2#3#4#5#6#7#8#9 { }

(End definition for \use_none:n and others. These functions are documented on page 25.)

### 41.6 Debugging and patching later definitions

\__kernel_if_debug:TF

A more meaningful test of whether debugging is enabled than messeng up with guards. We can also more easily change the logic in one place then. This is needed primarily for deprecations.

\cs_set_protected:Npn \__kernel_if_debug:TF #1#2 {#2}

(End definition for \__kernel_if_debug:TF.)

\debug_on:n
\debug_off:n

Stubs.

\cs_set_protected:Npn \debug_on:n #1
\msg_error:nxn { debug } { enable-debug }
\tl_to_str:n { \debug_on:n (#1) }
\cs_set_protected:Npn \debug_off:n #1
\msg_error:nxn { debug } { enable-debug }
\tl_to_str:n { \debug_off:n (#1) }

357
41.7 Conditional processing and definitions

Underneath any predicate function (\_p) or other conditional forms (TF, etc.) is a built-in logic saying that it after all of the testing and processing must return the ⟨state⟩ this leaves \TeX{} in. Therefore, a simple user interface could be something like

\begin{verbatim}
\if_meaning:w #1 #2
  \prg_return_true:
\else:
  \if_meaning:w #1 #3
    \prg_return_true:
  \else:
    \prg_return_false:
  \fi:
\fi:
\end{verbatim}

Usually, a \TeX{} programmer would have to insert a number of \texttt{\exp_after:wN}s to ensure the state value is returned at exactly the point where the last conditional is finished. However, that obscures the code and forces the \TeX{} programmer to prove that he/she knows the 2^n − 1 table. We therefore provide the simpler interface.

\begin{verbatim}
\cs_set:Npn \prg_return_true: { \exp_after:wN \use_i:nn \exp:w }
\cs_set:Npn \prg_return_false: { \exp_after:wN \use_ii:nn \exp:w }
\end{verbatim}

The idea here is that \texttt{\exp:w} expands fully any \texttt{\else:} and \texttt{\fi:} that are waiting to be discarded, before reaching the \texttt{\exp_end:} which leaves an empty expansion. The code can then leave either the first or second argument in the input stream. This means that all of the branching code has to contain at least two tokens: see how the logical tests are actually implemented to see this.

\begin{verbatim}
\cs_set:Npn \prg_return_true: { \exp_after:wN \use_i:nn \exp:w }
\cs_set:Npn \prg_return_false: { \exp_after:wN \use_ii:nn \exp:w }
\end{verbatim}
An extended state space could be implemented by including a more elaborate function in place of `\use_i:nn/\use_i:nn`. Provided two arguments are absorbed then the code would work.

(End definition for `\prg_return_true` and `\prg_return_false`. These functions are documented on page 64.)

Private version of `\use_none_delimit_by_q_recursion_stop:w`.

```latex
\cs_set:Npn \__prg_use_none_delimit_by_q_recursion_stop:w ...
```

(End definition for `\__prg_use_none_delimit_by_q_recursion_stop:w`.)

The user functions for the types using parameter text from the programmer. The various functions only differ by which function is used for the assignment. For those `Npnn` type functions, we must grab the parameter text, reading everything up to a left brace before continuing. Then split the base function into name and signature, and feed `{⟨name⟩}{⟨signature⟩}{⟨boolean⟩}{⟨set or new⟩}{⟨maybe protected⟩}{⟨parameters⟩}{⟨TF,...⟩}{⟨code⟩}` to the auxiliary function responsible for defining all conditionals. Note that `e` stands for expandable and `p` for protected.

```latex
\cs_set_protected:Npn \prg_set_conditional:Npnn ...
```

```latex
\cs_set_protected:Npn \prg_new_conditional:Npnn ...
```

```latex
\cs_set_protected:Npn \prg_set_protected_conditional:Npnn ...
```

```latex
\cs_set_protected:Npn \prg_new_protected_conditional:Npnn ...
```

The user functions for the types automatically inserting the correct parameter text based on the signature. The various functions only differ by which function is used for the assignment. Split the base function into name and signature. The second auxiliary generates the parameter text from the number of letters in the signature. Then feed `{⟨name⟩}{⟨signature⟩}{⟨boolean⟩}{⟨set or new⟩}{⟨maybe protected⟩}{⟨parameters⟩}{⟨TF,...⟩}{⟨code⟩}` to the auxiliary function responsible for defining all conditionals. If the signature has more than 9 letters, the definition is aborted since TEX macros have at most 9 arguments. The erroneous case where the function name contains no colon is captured later.

```latex
\cs_set_protected:Npn \prg_set_conditional:Npnn ...
```

```latex
\cs_set_protected:Npn \prg_new_conditional:Npnn ...
```

```latex
\cs_set_protected:Npn \prg_set_protected_conditional:Npnn ...
```

```latex
\cs_set_protected:Npn \prg_new_protected_conditional:Npnn ...
```
The workhorse here is going through a list of desired forms, \emph{i.e.}, p, TF, T and F. The first three arguments come from splitting up the base form of the conditional, which gives the name, signature and a boolean to signal whether or not there was a colon in the name. In the absence of a colon, we throw an error and don’t define any conditional. The fourth and fifth arguments build up the defining function. The sixth is the parameters to use (possibly empty), the seventh is the list of forms to define, the eighth is the replacement text which we will augment when defining the forms. The use of \verb+	l_to_str:n+ makes the later loop more robust.

A large number of our low-level conditionals look like \langle code \rangle \texttt{\textbackslash prg\_return\_true:} \texttt{\textbackslash else: \texttt{\textbackslash prg\_return\_false:} \texttt{\textbackslash fi:}} so we optimize this special case by calling \verb+\__prg\_generate\_conditional\_fast:nw+ \{(code)\}. This passes \verb+use_i:nn+ instead of \verb+use_i_i:nnn+ to functions such as \verb+\__prg\_generate\_p\_form:wNNnnnnN+. 

(End definition for \texttt{\textbackslash prg\_set\_conditional:Nnn} and others. These functions are documented on page 62.)
Loopying through the list of desired forms. First are six arguments and seventh is the form. Use the form to call the correct type. If the form does not exist, the \use:c construction results in \relax, and the error message is displayed (unless the form is empty, to allow for \{T, , F\}), then \use:none:nnnnnnnn cleans up. Otherwise, the error message is removed by the variant form.

\cs_set_protected:Npn \__prg_generate_conditional:NNnnnnNw #1#2#3#4#5#6#7#8 ,
{\iff_meaning:w \q__prg_recursion_tail #8 \exp_after:wN \__prg_use_none_delimit_by_q_recursion_stop:w \fi: \use:c { __prg_generate_ #8 _form:wNNnnnnN }
 \tl_if_empty:nF {#8} {
 \msg_error:nnxx { kernel } { conditional-form-unknown }
 {#8} { \token_to_str:c { #3 : #4 } }
 }
 \use:none:nnnnnnn
 \s__prg_stop
 \#1 \#2 \#3 \#4 \#5 \#6 \#7
 \__prg_generate_conditional:NNnnnnNw #1 \#2 \#3 \#4 \#5 \#6 \#7
}

(End definition for \__prg_generate_conditional:nnNNnnn and others.)

How to generate the various forms. Those functions take the following arguments: 1: junk, 2: \cs_set:Npn or similar, 3: p (for protected conditionals) or e, 4: function name, 5: signature, 6: parameter text, 7: replacement (possibly trimmed by \__prg_generate_conditional_fast:nw), 8: \use_i_ii:nnn or \use_i:nn (for “fast” conditionals). Remember that the logic-returning functions expect two arguments to be present after \exp_end:: notice the construction of the different variants relies on this, and that the TF and F variants will be slightly faster than the T version. The p form is only valid for expandable tests, we check for that by making sure that the second argument is empty. For “fast” conditionals, \#7 has an extra \if_. To optimize a bit further we don’t use \exp_after:wN \use_i_ii:nnn and similar but instead use \__prg_TF_true:w and similar to swap out the macro after \fi:. It would be a tiny bit faster if we directly grabbed the T and F arguments there, but if those are actually missing, the recovery from the runaway argument would not insert \fi: back, messing up nesting of conditionals.
The setting-equal functions. Split both functions and feed \( \langle \text{name} \rangle \) \{ \langle \text{signature} \rangle \} \langle \text{boolean} \rangle \{ \langle \text{name} \rangle \} \{ \langle \text{signature} \rangle \} \langle \text{boolean} \rangle \langle \text{copying function} \rangle \langle \text{conditions} \rangle \), \text{q\_prg\_recursion\_tail} \ , \text{q\_prg\_recursion\_stop} to a first auxiliary.
\cs_set_protected:Npn \prg_set_eq_conditional:NNn { \__prg_set_eq_conditional:NNNn \cs_set_eq:cc }
\cs_set_protected:Npn \prg_new_eq_conditional:NNn { \__prg_set_eq_conditional:NNNn \cs_new_eq:cc }
\cs_set_protected:Npn \__prg_set_eq_conditional:NNNn #1#2#3#4
{
\use:x
{
\exp_not:N \__prg_set_eq_conditional:nnNnnNNw
\cs_split_function:N #2
\cs_split_function:N #3
\exp_not:N #1
\tl_to_str:n {#4}
\exp_not:n {, \q__prg_recursion_tail , \q__prg_recursion_stop }
}
}
(End definition for \prg_set_eq_conditional:NNn, \prg_new_eq_conditional:NNn, and \__prg_set_eq_conditional:NNNn. These functions are documented on page 64.)

\cs_set_protected:Npn \__prg_set_eq_conditional:nnNnnNNw #1#2#3#4#5#6
{
\if_meaning:w \c_false_bool #3
\msg_error:nnx { kernel } { missing-colon }
{ \token_to_str:c {#1} }
\fi:
\if_meaning:w \c_false_bool #6
\msg_error:nnx { kernel } { missing-colon }
{ \token_to_str:c {#4} }
\fi:
\__prg_set_eq_conditional_loop:nnnnNw {#1} {#2} {#4} {#5}
\}
\cs_set_protected:Npn \__prg_set_eq_conditional_loop:nnnnNw #1#2#3#4#5#6 ,
{
\if_meaning:w \q__prg_recursion_tail #6
\exp_after:wN \__prg_use_none_delimit_by_q_recursion_stop:w
\fi:
\use:c { \__prg_set_eq_conditional_ #6_form:wNnnn }
\tl_if_empty:nF {#6}
{ \msg_error:nnx
{ kernel } { conditional-form-unknown }
{#6} { \token_to_str:c { #1 : #2 } }
}
\use_none:nnnnnn \s__prg_stop

Split the function to be defined, and setup a manual clist loop over argument #6 of the first auxiliary. The second auxiliary receives twice three arguments coming from splitting the function to be defined and the function to copy. Make sure that both functions contained a colon, otherwise we don’t know how to build conditionals, hence abort. Call the looping macro, with arguments {⟨name1⟩} {⟨signature1⟩} {⟨name2⟩} {⟨signature2⟩} {copying function} and followed by the comma list. At each step in the loop, make sure that the conditional form we copy is defined, and copy it, otherwise abort.
All that is left is to define the canonical boolean true and false. I think Michael originated the idea of expandable boolean tests. At first these were supposed to expand into either \textit{TT} or \textit{TF} to be tested using \texttt{if:w} but this was later changed to \texttt{00} and \texttt{01}, so they could be used in logical operations. Later again they were changed to being numerical constants with values of 1 for true and 0 for false. We need this from the get-go.

\texttt{\c_true_bool} \hspace{1cm} \texttt{\c_false_bool}

Here are the canonical boolean values.

\begin{align*}
\backslash \text{tex_chardef:D} \backslash \text{\c_true_bool} &= 1 - \\
\backslash \text{tex_chardef:D} \backslash \text{\c_false_bool} &= 0 -
\end{align*}

\textit{(End definition for \texttt{\c_true_bool} and \texttt{\c_false_bool}. These variables are documented on page 26.)}

\section{Dissecting a control sequence}

\texttt{\langle@@=cs\rangle}

\texttt{\__cs_count_signature:N} \hspace{1cm} \texttt{\__cs_count_signature:N \langle function \rangle}

Splits the \textit{(function)} into the \textit{(name)} (\textit{i.e.} the part before the colon) and the \textit{(signature)} (\textit{i.e.} after the colon). The \textit{(number)} of tokens in the \textit{(signature)} is then left in the input stream. If there was no \textit{(signature)} then the result is the marker value $-1$.

\texttt{\__cs_get_function_name:N \star} \hspace{1cm} \texttt{\__cs_get_function_name:N \langle function \rangle}

Splits the \textit{(function)} into the \textit{(name)} (\textit{i.e.} the part before the colon) and the \textit{(signature)} (\textit{i.e.} after the colon). The \textit{(name)} is then left in the input stream without the escape character present made up of tokens with category code 12 (other).

\texttt{\__cs_get_function_signature:N \star} \hspace{1cm} \texttt{\__cs_get_function_signature:N \langle function \rangle}

Splits the \textit{(function)} into the \textit{(name)} (\textit{i.e.} the part before the colon) and the \textit{(signature)} (\textit{i.e.} after the colon). The \textit{(signature)} is then left in the input stream made up of tokens with category code 12 (other).

\texttt{\__cs_tmp:w}

Function used for various short-term usages, for instance defining functions whose definition involves tokens which are hard to insert normally (spaces, characters with category other).
\cs_to_str:N  This converts a control sequence into the character string of its name, removing the leading escape character. This turns out to be a non-trivial matter as there are different cases:

- The usual case of a printable escape character;
- the case of a non-printable escape characters, e.g., when the value of the \escapechar is negative;
- when the escape character is a space.

One approach to solve this is to test how many tokens result from \token_to_str:N \a. If there are two tokens, then the escape character is printable, while if it is non-printable then only one is present.

However, there is an additional complication: the control sequence itself may start with a space. Clearly that should not be lost in the process of converting to a string. So the approach adopted is a little more intricate still. When the escape character is printable, \token_to_str:N \a yields the escape character itself and a space. The character codes are different, thus the \if:w test is false, and \TeX reads \__cs_to_str:N after turning the following control sequence into a string; this auxiliary removes the escape character, and stops the expansion of the initial \tex_romannumeral:D. The second case is that the escape character is not printable. Then the \if:w test is unfinished after reading a the space from \token_to_str:N, and the auxiliary \__cs_to_str:w is expanded, feeding - as a second character for the test; the test is false, and \TeX skips to \fi:, then performs \token_to_str:N, and stops the \tex_romannumeral:D with \c_zero_int. The last case is that the escape character is itself a space. In this case, the \if:w test is true, and the auxiliary \__cs_to_str:w comes into play, inserting -\int_value:w, which expands \c_zero_int to the character 0. The initial \tex_romannumeral:D then sees 0, which is not a terminated number, followed by the escape character, a space, which is removed, terminating the expansion of \tex_romannumeral:D. In all three cases, \cs_to_str:N takes two expansion steps to be fully expanded.

\cs_set:Npn \cs_to_str:N
{ }

We implement the expansion scheme using \tex_romannumeral:D terminating it with \c_zero_int rather than using \exp:w and \exp_end: as we normally do. The reason is that the code heavily depends on terminating the expansion with \c_zero_int so we make this dependency explicit.

\cs_set:Npn \__cs_to_str:N #1 { \c_zero_int }
\cs_set:Npn \__cs_to_str:w #1 \__cs_to_str:N
{ - \int_value:w \fi: \exp_after:wN \c_zero_int }

If speed is a concern we could use \csstring in \LuaTeX. For the empty csname that primitive gives an empty result while the current \cs_to_str:N gives incorrect results in all engines (this is impossible to fix without huge performance hit).

(End definition for \cs_to_str:N, \__cs_to_str:N, and \__cs_to_str:w. This function is documented on page 21.)
This function takes a function name and splits it into name with the escape char removed and argument specification. In addition to this, a third argument, a boolean \texttt{(true)} or \texttt{(false)} is returned with \texttt{(true)} for when there is a colon in the function and \texttt{(false)} if there is not.

We cannot use \texttt{:} directly as it has the wrong category code so an \texttt{x}-type expansion is used to force the conversion.

First ensure that we actually get a properly evaluated string by expanding \texttt{cs\_to\_str:N} twice. If the function contained a colon, the auxiliary takes as \#1 the function name, delimited by the first colon, then the signature \#2, delimited by \texttt{cs\_mark}, then \texttt{true\_bool} as \#3, and \#4 cleans up until \texttt{cs\_stop}. Otherwise, the \#1 contains the function name and \texttt{cs\_mark true\_bool} \#2 is empty, \#3 is \texttt{false\_bool}, and \#4 cleans up. The second auxiliary trims the trailing \texttt{cs\_mark} from the function name if present (that is, if the original function had no colon).

\begin{verbatim}
\cs_set_protected:Npn \__cs_tmp:w #1
  \cs_set:Npn \cs_split_function:N ##1
    \exp_after:wN \exp_after:wN \exp_after:wN
    \__cs_split_function_auxi:w
    \cs_to_str:N ##1 \cs_mark \c_true_bool
    #1 \cs_mark \c_false_bool \cs_stop
\end{verbatim}

(End definition for \texttt{cs\_split\_function:N}, \texttt{cs\_split\_function\_auxi:w}, and \texttt{cs\_split\_function\_auxii:w}. This function is documented on page \pageref{cs_split_function:N}.)

## 41.9 Exist or free

A control sequence is said to \textit{exist} (to be used) if has an entry in the hash table and its meaning is different from the primitive \texttt{relax} token. A control sequence is said to be \textit{free} (to be defined) if it does not already exist.

Two versions for checking existence. For the \texttt{N} form we firstly check for \texttt{scan\_stop:} and then if it is in the hash table. There is no problem when inputting something like \texttt{else:} or \texttt{fi:} as \texttt{TeX} will only ever skip input in case the token tested against is \texttt{scan\_stop:}.

\begin{verbatim}
\prg_set_conditional:Npnn \cs_if_exist:N #1 { p , T , F , TF }
  \{ \if_meaning:w #1 \scan_stop:
    \prg_return_false:
  \else:
    \if_cs_exist:N #1
      \prg_return_true:
    \else:
      \prg_return_false:
  \fi:
\end{verbatim}

366
For the \texttt{c} form we firstly check if it is in the hash table and then for \texttt{scan_stop}: so that we do not add it to the hash table unless it was already there. Here we have to be careful as the text to be skipped if the first test is false may contain tokens that disturb the scanner. Therefore, we ensure that the second test is performed after the first one has concluded completely.

\begin{verbatim}
\prg_set_conditional:Npnn \cs_if_exist:c #1 { p , T , F , TF } 
{ 
\if_cs_exist:w #1 \cs_end: 
\exp_after:wN \use_i:nn 
\else: 
\exp_after:wN \use_ii:nn 
\fi:
{ 
\exp_after:wN \if_meaning:w \cs:w #1 \cs_end: \scan_stop: 
\prg_return_false: 
\else: 
\prg_return_true: 
\fi:
}
\prg_return_false: 
}
\end{verbatim}

(End definition for \texttt{cs_if_exist:NTF}. This function is documented on page \pageref{cs_if_exist:NTF}.)

\begin{verbatim}
\cs_if_free_p:N \cs_if_free_p:c \cs_if_free:NTF \cs_if_free:cTF
\prg_set_conditional:Npnn \cs_if_free:N #1 { p , T , F , TF } 
{ 
\if_meaning:w #1 \cs_end: \scan_stop: 
\prg_return_true: 
\else: 
\if_cs_exist:N #1 
\prg_return_false: 
\else: 
\prg_return_true: 
\fi:
\fi:
}
\prg_set_conditional:Npnn \cs_if_free:c #1 { p , T , F , TF } 
{ 
\if_cs_exist:w #1 \cs_end: 
\exp_after:wN \use_i:nn 
\else: 
\exp_after:wN \use_ii:nn 
\fi:
{ 
\exp_after:wN \if_meaning:w \cs:w #1 \cs_end: \scan_stop: 
\prg_return_true: 
\else: 
\prg_return_false: 
\fi:
}
\end{verbatim}
The \cs_if_exist_use:... functions cannot be implemented as conditionals because the true branch must leave both the control sequence itself and the true code in the input stream. For the \texttt{c} variants, we are careful not to put the control sequence in the hash table if it does not exist. In \LaTeX{} we could use the \verb|\lastnamedcs| primitive.

\begin{verbatim}
\cs_set:Npn \cs_if_exist_use:NTF #1#2 { \cs_if_exist:NTF #1 { #1 #2 } }
\cs_set:Npn \cs_if_exist_use:N { \cs_if_exist:NTF #1 { #1 } }
\cs_set:Npn \cs_if_exist_use:NTF #1 #2 { \cs_if_exist:NTF #1 { #1 #2 } { } }
\cs_set:Npn \cs_if_exist_use:cTF #1#2 { \cs_if_exist:cTF {#1} { \use:c {#1} #2 } }
\cs_set:Npn \cs_if_exist_use:cF #1 { \cs_if_exist:cTF {#1} { \use:c {#1} } }
\cs_set:Npn \cs_if_exist_use:cT #1#2 { \cs_if_exist:cTF {#1} { \use:c {#1} #2 } { } }
\cs_set:Npn \cs_if_exist_use:c #1 { \cs_if_exist:cTF {#1} { \use:c {#1} } { } }
\end{verbatim}

(End definition for \cs_if_exist_use:NTF. This function is documented on page 21.)

\section{41.10 Preliminaries for new functions}

We provide two kinds of functions that can be used to define control sequences. On the one hand we have functions that check if their argument doesn’t already exist, they are called \texttt{..._new}. The second type of defining functions doesn’t check if the argument is already defined.

Before we can define them, we need some auxiliary macros that allow us to generate error messages. The next few definitions here are only temporary, they will be redefined later on.

\begin{verbatim}
\msg_error:nnxx
\msg_error:nxx
\msg_error:nn
\end{verbatim}

If an internal error occurs before \LaTeX{} has loaded \texttt{l3msg} then the code should issue a usable if terse error message and halt. This can only happen if a coding error is made by the team, so this is a reasonable response. Setting the \texttt{\newlinechar} is needed, to turn \verb|^J| into a proper line break in plain \TeX{}.

\begin{verbatim}
\cs_set_protected:Npn \msg_error:nnxx #1#2#3#4
\tex_newlinechar:D = '^\^^J \scan_stop:
\tex_errmessage:D
{
  !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! ^\^^J 
  Argh,-internal-LaTeX3-error! ^\^^J ^\^^J 
  Module ~ #1 , ~ message-name="#2": ^\^^J 
  Arguments~"#3'~and~"#4' ~^\^^J 
  This-is-one-for-The-LaTeX3-Project:-bailing-out
\end{verbatim}

368
\msg_line_context:  Another one from \l3msg which will be altered later.

\iow_log:x \iow_term:x  We define a routine to write only to the log file. And a similar one for writing to both the log file and the terminal. These will be redefined later by \l3io.

\__kernel_chk_if_free_cs:N \__kernel_chk_if_free_cs:c  This command is called by \cs_new_nopar:Npn and \cs_new_eq:NN etc. to make sure that the argument sequence is not already in use. If it is, an error is signalled. It checks if \csname is undefined or \scan_stop:. Otherwise an error message is issued. We have to make sure we don’t put the argument into the conditional processing since it may be an \if... type function!

41.11 Defining new functions

\cs_new_nopar:Npn \cs_new_nopar:Npx \cs_new:Npn \cs_new:Npx \cs_new_protected_nopar:Npn \cs_new_protected:Npn \cs_new_protected:Npx \cs_new_protected:Npx  Function which check that the control sequence is free before defining it.
\cs_set_nopar:cpn \cs_set_nopar:cpx \cs_gset_nopar:cpn \cs_gset_nopar:cpx \cs_new_nopar:cpn \cs_new_nopar:cpx

Variants of the \cs_set_nopar:Npn versions which make a csname out of the first arguments. We may also do this globally.


Variants of the \cs_set_protected_nopar:Npn versions which make a csname out of the first arguments. We may also do this globally.

(End definition for \cs_set_nopar:Npn. This function is documented on page 16.)

(End definition for \cs_set_protected_nopar:Npn. This function is documented on page 16.)

(End definition for \cs_set_nopar:cpn and others. These functions are documented on page 15.)
Variants of the \cs_set_protected:Npn versions which make a csname out of the first arguments. We may also do this globally.

1953 \_cs_tmp:w \cs_set_protected:cpn \cs_set_protected:Npn
1954 \_cs_tmp:w \cs_set_protected:cpx \cs_set_protected:Npx
1955 \_cs_tmp:w \cs_gset_protected:cpn \cs_gset_protected:Npn
1956 \_cs_tmp:w \cs_gset_protected:cpx \cs_gset_protected:Npx
1957 \_cs_tmp:w \cs_new_protected:cpn \cs_new_protected:Npn
1958 \_cs_tmp:w \cs_new_protected:cpx \cs_new_protected:Npx

(End definition for \cs_set_protected:Npn. This function is documented on page 16.)

41.12 Copying definitions

These macros allow us to copy the definition of a control sequence to another control sequence.

The = sign allows us to define funny char tokens like = itself or \_ with this function. For the definition of \c_space_char{-} to work we need the - after the =.

\cs_set_eq:NN \cs_set_eq:cN \cs_set_eq:Nc \cs_set_eq:cc
\cs_gset_eq:NN \cs_gset_eq:cN \cs_gset_eq:Nc \cs_gset_eq:cc
\cs_new_eq:NN \cs_new_eq:cN \cs_new_eq:Nc \cs_new_eq:cc

(End definition for \cs_set_eq:NN, \cs_gset_eq:NN, and \cs_new_eq:NN. These functions are documented on page 19.)

41.13 Undefining functions

The following function is used to free the main memory from the definition of some function that isn’t in use any longer. The c variant is careful not to add the control sequence to the hash table if it isn’t there yet, and it also avoids nesting \TeX conditionals in case #1 is unbalanced in this matter.

\cs_new_protected:Npn \cs_undefine:N \cs_undefine:c

{ \if_cs_exist:w #1 \cs_end: }

(371)
41.14 Generating parameter text from argument count

\[ \text{LTex3 provides shorthands to define control sequences and conditionals with a simple parameter text, derived directly from the signature, or more generally from knowing the number of arguments, between 0 and 9. This function expands to its first argument, untouched, followed by a brace group containing the parameter text}{\{#1...#n}\} , where} \ n \ \text{is the result of evaluating the second argument (as described in}{\text{int_eval:n}}. \ \text{If the second argument gives a result outside the range [0, 9], the third argument is returned instead, normally an error message. Some of the functions use here are not defined yet, but will be defined before this function is called.}

\[ \text{End definition for}{\cs_undefine:N}. \ \text{This function is documented on page 20.}\]
41.15 Defining functions from a given number of arguments

Counting the number of tokens in the signature, i.e., the number of arguments the function should take. Since this is not used in any time-critical function, we simply use \_\_tl_count:n if there is a signature, otherwise −1 arguments to signal an error. We need a variant form right away.

\cs_new:Npn \__cs_count_signature:N #1
{ \exp_args:Nf \__cs_count_signature:n { \cs_split_function:N #1 } }
\cs_new:Npn \__cs_count_signature:n #1
{ \int_eval:n { \__cs_count_signature:nnN #1 } }
\cs_new:Npn \__cs_count_signature:nnN #1#2#3
{\if_meaning:w \c_true_bool #3 \tl_count:n {#2} \else: -1 \fi: }
\cs_new:Npn \__cs_count_signature:c
{ \exp_args:Nc \__cs_count_signature:N }

We provide a constructor function for defining functions with a given number of arguments. For this we need to choose the correct parameter text and then use that when defining. Since \TeX supports from zero to nine arguments, we use a simple switch to choose the correct parameter text, ensuring the result is returned after finishing the conditional. If it is not between zero and nine, we throw an error.

1: function to define, 2: with what to define it, 3: the number of args it requires and 4: the replacement text

\cs_new_protected:Npn \cs_generate_from_arg_count:NNnn #1#2#3#4
{ \__kernel_cs_parm_from_arg_count:nnF { \use:nnn #2 #1 } {#3} \msg_error:nnxx { kernel } { bad-number-of-arguments } { \token_to_str:N #1 } { \int_eval:n {#3} } \use_none:n }\{#4\}

A variant form we need right away, plus one which is used elsewhere but which is most logically created here.

\cs_new_protected:Npn \cs_generate_from_arg_count:cNnn
{ \exp_args:Nc \cs_generate_from_arg_count:NNnn }
\cs_new_protected:Npn \cs_generate_from_arg_count:Ncnn
{ \exp_args:NNc \cs_generate_from_arg_count:NNnn }

(End definition for \_\_\_cs_count_signature:N, \_\_\_cs_count_signature:n, and \_\_\_cs_count_signature:nnN.)

(End definition for \_\_\_cs_count_signature:N, \_\_\_cs_count_signature:n, and \_\_\_cs_count_signature:nnN.)

This function is documented on page 19.)
41.16 Using the signature to define functions

We can now combine some of the tools we have to provide a simple interface for defining functions, where the number of arguments is read from the signature. For instance, \texttt{\cs_set:Nn \foo_bar:nn \{#1,#2\}}.

We want to define \texttt{\cs_set:Nn} as

\begin{verbatim}
\end{verbatim}

We want to define \texttt{\cs_set:Nn} as

\begin{verbatim}
\cs_set_protected:Npn \cs_set:Nn \{#1\} {#2}
\end{verbatim}

In short, to define \texttt{\cs_set:Nn} we need just use \texttt{\cs_set:Npn}, everything else is the same for each variant. Therefore, we can make it simpler by temporarily defining a function to do this for us.

\begin{verbatim}
\cs_set:Npn \__cs_tmp:w \{#1\} {#2}
\exp_not:N \__cs_generate_from_signature:NNn \exp_after:wN \exp_not:N \cs:w \cs_ \end:}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__cs_generate_from_signature:NNn \{ \exp_after:wN \exp_not:N \cs:w \cs_ \end:}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \__cs_generate_from_signature:NNn \{ \exp_after:wN \exp_not:N \cs:w \cs_ \end:}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \__cs_generate_from_signature:NNn \{ \exp_after:wN \exp_not:N \cs:w \cs_ \end:}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \__cs_generate_from_signature:NNn \{ \exp_after:wN \exp_not:N \cs:w \cs_ \end:}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \__cs_generate_from_signature:NNn \{ \exp_after:wN \exp_not:N \cs:w \cs_ \end:}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \__cs_generate_from_signature:NNn \{ \exp_after:wN \exp_not:N \cs:w \cs_ \end:}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \__cs_generate_from_signature:NNn \{ \exp_after:wN \exp_not:N \cs:w \cs_ \end:}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \__cs_generate_from_signature:NNn \{ \exp_after:wN \exp_not:N \cs:w \cs_ \end:}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \__cs_generate_from_signature:NNn \{ \exp_after:wN \exp_not:N \cs:w \cs_ \end:}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \__cs_generate_from_signature:NNn \{ \exp_after:wN \exp_not:N \cs:w \cs_ \end:}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \__cs_generate_from_signature:NNn \{ \exp_after:wN \exp_not:N \cs:w \cs_ \end:}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \__cs_generate_from_signature:NNn \{ \exp_after:wN \exp_not:N \cs:w \cs_ \end:}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \__cs_generate_from_signature:NNn \{ \exp_after:wN \exp_not:N \cs:w \cs_ \end:}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \__cs_generate_from_signature:NNn \{ \exp_after:wN \exp_not:N \cs:w \cs_ \end:}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \__cs_generate_from_signature:NNn \{ \exp_after:wN \exp_not:N \cs:w \cs_ \end:}
\end{verbatim}

\begin{verbatim}
\bool_if:NTF \#3
\end{verbatim}

\begin{verbatim}
\cs_set:nopar:Npx \__cs_tmp:w \{ \tl_map_function:nN \{#2\} \__cs_generate_from_signature:n \}
\tl_if_empty:oF \__cs_tmp:w \\
\msg_error:nxxx \{ kernel \} \{ non-base-function \}
\end{verbatim}

\begin{verbatim}
\cs_generate_from_arg_count:NNnn \#5 \{ \tl_count:n {#2} \} \{#6\}
\end{verbatim}

\begin{verbatim}
\msg_error:nxx \{ kernel \} \{ missing-colon \}
\end{verbatim}

\begin{verbatim}
\token_to_str:N \#5
\end{verbatim}

\begin{verbatim}
\end{verbatim}

374
Then we define the 24 variants beginning with \(N\).

\[
\text{\_cs_tmp:w \{ set \} \{ Nn \} \{ Npn \}}
\]
\[
\text{\_cs_tmp:w \{ set \} \{ Nx \} \{ Npx \}}
\]
\[
\text{\_cs_tmp:w \{ set_nopar \} \{ Nn \} \{ Npn \}}
\]
\[
\text{\_cs_tmp:w \{ set_nopar \} \{ Nx \} \{ Npx \}}
\]
\[
\text{\_cs_tmp:w \{ set_protected \} \{ Nn \} \{ Npn \}}
\]
\[
\text{\_cs_tmp:w \{ set_protected \} \{ Nx \} \{ Npx \}}
\]
\[
\text{\_cs_tmp:w \{ set_protected_nopar \} \{ Nn \} \{ Npn \}}
\]
\[
\text{\_cs_tmp:w \{ set_protected_nopar \} \{ Nx \} \{ Npx \}}
\]
\[
\text{\_cs_tmp:w \{ gset \} \{ Nn \} \{ Npn \}}
\]
\[
\text{\_cs_tmp:w \{ gset \} \{ Nx \} \{ Npx \}}
\]
\[
\text{\_cs_tmp:w \{ gset_nopar \} \{ Nn \} \{ Npn \}}
\]
\[
\text{\_cs_tmp:w \{ gset_nopar \} \{ Nx \} \{ Npx \}}
\]
\[
\text{\_cs_tmp:w \{ gset_protected \} \{ Nn \} \{ Npn \}}
\]
\[
\text{\_cs_tmp:w \{ gset_protected \} \{ Nx \} \{ Npx \}}
\]
\[
\text{\_cs_tmp:w \{ gset_protected_nopar \} \{ Nn \} \{ Npn \}}
\]
\[
\text{\_cs_tmp:w \{ gset_protected_nopar \} \{ Nx \} \{ Npx \}}
\]
\[
\text{\_cs_tmp:w \{ new \} \{ Nn \} \{ Npn \}}
\]
\[
\text{\_cs_tmp:w \{ new \} \{ Nx \} \{ Npx \}}
\]
\[
\text{\_cs_tmp:w \{ new_nopar \} \{ Nn \} \{ Npn \}}
\]
\[
\text{\_cs_tmp:w \{ new_nopar \} \{ Nx \} \{ Npx \}}
\]
\[
\text{\_cs_tmp:w \{ new_protected \} \{ Nn \} \{ Npn \}}
\]
\[
\text{\_cs_tmp:w \{ new_protected \} \{ Nx \} \{ Npx \}}
\]
\[
\text{\_cs_tmp:w \{ new_protected_nopar \} \{ Nn \} \{ Npn \}}
\]
\[
\text{\_cs_tmp:w \{ new_protected_nopar \} \{ Nx \} \{ Npx \}}
\]

(End definition for \cs_set:Nn and others. These functions are documented on page 17.)

The 24 \(c\) variants simply use \exp_args:Nc.

\[
\text{\_cs_set:cn}
\]
\[
\text{\_cs_set:cx}
\]
\[
\text{\_cs_set_nopar:cn}
\]
\[
\text{\_cs_set_nopar:cx}
\]
\[
\text{\_cs_set_protected:cn}
\]
\[
\text{\_cs_set_protected:cx}
\]
\[
\text{\_cs_gset:cn}
\]
\[
\text{\_cs_gset:cx}
\]
\[
\text{\_cs_gset_nopar:cn}
\]
\[
\text{\_cs_gset_nopar:cx}
\]
\[
\text{\_cs_gset_protected:cn}
\]
\[
\text{\_cs_gset_protected:cx}
\]
\[
\text{\_cs_new:cn}
\]
\[
\text{\_cs_new:cx}
\]
\[
\text{\_cs_new_nopar:cn}
\]
\[
\text{\_cs_new_nopar:cx}
\]
\[
\text{\_cs_new_protected:cn}
\]
\[
\text{\_cs_new_protected:cx}
\]
\[
\text{\_cs_new_protected_nopar:cn}
\]
\[
\text{\_cs_new_protected_nopar:cx}
\]
41.17 Checking control sequence equality

Check if two control sequences are identical.

\begin{verbatim}
\cs_if_eq_p:NN \cs_if_eq_p:cN \cs_if_eq_p:cc \cs_if_eq:NN
\cs_if_eq_p:cNTF \cs_if_eq_p:cNT \cs_if_eq_p:cNF \cs_if_exist:NTF #1\#2
\cs_new:NI \cs_new:Nn \cs_new:Np \cs_new:Npc \cs_new:Npc \cs_new:Npc
\cs_new:Npn \cs_new:Npn \cs_new:Npn \cs_new:Npn \cs_new:Npn \cs_new:Npn
\cs_new:Npn \cs_new:Npn \cs_new:Npn \cs_new:Npn \cs_new:Npn \cs_new:Npn
\msg_error:nnx { kernel } { variable-not-defined } { \token_to_str:N #1 }
\end{verbatim}

(End definition for \cs_set:Nn. This function is documented on page 17.)

41.18 Diagnostic functions

\begin{verbatim}
\_kernel_chk_defined:NT
\cs_new_protected:Np \_kernel_chk_defined:NT #1\#2
\cs_if_exist:NTF #1 \#{2}
\msg_error:nnx { kernel } { variable-not-defined }
{ \token_to_str:N #1 }
\end{verbatim}

(End definition for \cs_if_eq:NNTF. This function is documented on page 27.)
Simply using the `\show` primitive does not allow for line-wrapping, so instead use `\tl_show:n` and `\tl_log:n` (defined in `l3tl` and that performs line-wrapping). This displays `$\langle \text{variable} \rangle \equiv \langle \text{value} \rangle$`. We expand the value beforehand as otherwise some integers (such as `\currentgrouplevel` or `\currentgrouptype`) altered by the line-wrapping code would show wrong values.

2171 \cs_new_protected:Npn \__kernel_register_show:N
2172 { \__kernel_register_show_aux:NN \tl_show:n }
2173 \cs_new_protected:Npn \__kernel_register_show:c
2174 { \exp_args:Nc \__kernel_register_show:N }
2175 \cs_new_protected:Npn \__kernel_register_log:N
2176 { \__kernel_register_show_aux:NN \tl_log:n }
2177 \cs_new_protected:Npn \__kernel_register_log:c
2178 { \exp_args:Nc \__kernel_register_log:N }
2179 \cs_new_protected:Npn \__kernel_register_show_aux:NN #1#2
2180 { \__kernel_chk_defined:NT #2
2181 { \exp_args:No \__kernel_register_show_aux:nNN
2182 { \tex_the:D #2 } #2 #1
2183 }
2184 }
2185 \cs_new_protected:Npn \__kernel_register_show_aux:nNN #1#2#3
2186 { \exp_args:No #3 { \token_to_str:N #2 = \cs_meaning:N #2 } }

(End definition for \__kernel_register_show:N and others.)

Some control sequences have a very long name or meaning. Thus, simply using \TeX’s primitive \show could lead to overlong lines. The output of this primitive is mimicked to some extent, then the re-built string is given to \tl_show:n or \tl_log:n for line-wrapping. We must expand the meaning before passing it to the wrapping code as otherwise we would wrongly see the definitions that are in place there. To get correct escape characters, set the `\escapechar` in a group; this also localizes the assignment performed by \x-expansion. The `\cs_show:c` and `\cs_log:c` commands convert their argument to a control sequence within a group to avoid showing \relax for undefined control sequences.

2189 \cs_new_protected:Npn \cs_show:N { \__kernel_show:NN \tl_show:n }
2190 \cs_new_protected:Npn \cs_show:c
2191 { \group_begin: \exp_args:NNc \group_end: \cs_show:N }
2192 \cs_new_protected:Npn \cs_log:N { \__kernel_show:NN \tl_log:n }
2193 \cs_new_protected:Npn \cs_log:c
2194 { \group_begin: \exp_args:NNc \group_end: \cs_log:N }
2195 \cs_new_protected:Npn \__kernel_show:NN #1#2
2196 { \\group_begin:
2197 \\int_set:Nn \\tex_escapechar:D { \textbackslash \textbackslash }
2198 \exp_args:NNx
2199 \group_end:
2200 \#1 { \token_to_str:N #2 = \cs_meaning:N #2 } }

(End definition for \cs_show:N, \cs_log:N, and \__kernel_show:N. These functions are documented on page 20.)
Wrapper around `\showgroups`. Getting \TeX{} to write to the log without interruption the run is done by altering the interaction mode.

\begin{verbatim}
\cs_new_protected:Npn \group_show_list: { \__kernel_group_show:NN \use_none:n 1 }
\cs_new_protected:Npn \group_log_list: { \__kernel_group_show:NN \int_zero:N 0 }
\cs_new_protected:Npn \__kernel_group_show:NN #1#2 {
  \use:x {
    \__kernel_prefix_arg_replacement:wN \exp_not:n { \cs_new:Npn \__kernel_prefix_arg_replacement:wN #1 }
    \tl_to_str:n { macro : } \exp_not:n { #2 -> #3 \s__kernel_stop #4 }
  }
}
\end{verbatim}

\textit{(End definition for \texttt{\group_show_list:}, \texttt{\group_log_list:}, and \texttt{\__kernel_group_show:NN}. These functions are documented on page 14.)}

### 41.19 Decomposing a macro definition

We sometimes want to test if a control sequence can be expanded to reveal a hidden value. However, we cannot just expand the macro blindly as it may have arguments and none might be present. Therefore we define these functions to pick either the prefix(es), the argument specification, or the replacement text from a macro. All of this information is returned as characters with catcode \texttt{12}. If the token in question isn’t a macro, the token `\scan_stop:` is returned instead.

\begin{verbatim}
\cs_prefix_spec:N \cs_argument_spec:N \cs_replacement_spec:N \__kernel_prefix_arg_replacement:wN
\use:x {
  \token_if_macro:NTF #1 {
    \exp_not:n { \cs_new:Npn \__kernel_prefix_arg_replacement:wN #1 }
    \tl_to_str:n { macro : } \exp_not:n { #2 -> #3 \s__kernel_stop #4 }
  }
  \exp_not:n { \token_if_macro:NTF #1 }
}
\end{verbatim}
41.20 Doing nothing functions

\texttt{\textbackslash prg\_do\_nothing:} This does not fit anywhere else!

\texttt{\textbackslash cs\_new:Np\textbackslash prg\_do\_nothing: \{ \}}

(End definition for \texttt{\cs\_prefix\_spec:N} and others. These functions are documented on page 22.)

41.21 Breaking out of mapping functions

\texttt{\textbackslash prg\_break\_point:Nn \textbackslash prg\_map\_break:Nn} In inline mappings, the nesting level must be reset at the end of the mapping, even when the user decides to break out. This is done by putting the code that must be performed as an argument of \texttt{\_\_prg\_break\_point:Nn}. The breaking functions are then defined to jump to that point and perform the argument of \texttt{\_\_prg\_break\_point:Nn}, before the user's code (if any). There is a check that we close the correct loop, otherwise we continue breaking.

\texttt{\cs\_new_eq:NN \prg\_break\_point:Nn \use\_ii:nn}

\texttt{\cs\_new:Npn \prg\_map\_break:Nn \texttt{\#1\texttt{\#2\texttt{\#3 \prg\_break\_point:Nn \texttt{\#4\texttt{\#5}}}}}

\texttt{\#5}

\texttt{\if\_meaning:w \#1 \#4}

\texttt{\exp\_after:wN \use\_iii:nn}

\texttt{\texttt{\f1}}

\texttt{\texttt{\prg\_map\_break:Nn \texttt{\#1 \#2}}}

\texttt{\#2}

\texttt{\#7}

(End definition for \texttt{\prg\_break\_point:Nn} and \texttt{\prg\_map\_break:Nn}. These functions are documented on page 70.)

\texttt{\prg\_break\_point: \texttt{\prg\_break: \texttt{\prg\_break:n}} Very simple analogues of \texttt{\prg\_break\_point:Nn} and \texttt{\prg\_map\_break:Nn}, for use in fast short-term recursions which are not mappings, do not need to support nesting, and in which nothing has to be done at the end of the loop.

\texttt{\cs\_new_eq:NN \prg\_break\_point: \prg\_do\_nothing:}

\texttt{\cs\_new:Npn \prg\_break: \texttt{\#1 \prg\_break\_point: \{}\}}

\texttt{\cs\_new:Npn \prg\_break:n \texttt{\#1\#2 \prg\_break\_point: \{}\#1\}}
41.22 Starting a paragraph

\mode_leave_vertical: The approach here is different to that used by \LaTeX2ε or plain \TeX, which unbox a void box to force horizontal mode. That inserts the %everypar tokens before the re-inserted unboxing tokens. The approach here uses either the \quitvmode primitive or the equivalent protected macro. In vertical mode, the %indent primitive is inserted: this will switch to horizontal mode and insert %everypar tokens and nothing else. Unlike the \LaTeX2ε version, the availability of \verb|ε-TeX| means using a mode test can be done at for example the start of an %halign.

\begin{verbatim}
\cs_new_protected:Npn \mode_leave_vertical: \
\{ 
\if_mode_vertical: 
\exp_after:wN \tex_indent:D 
\fi: 
\}
\end{verbatim}

(End definition for %\mode_leave_vertical:. This function is documented on page 28.)
Chapter 42

\l3expan implementation

\l__exp_internal_tl

\exp_module has its private variable to temporarily store the result of x-type argument expansion. This is done to avoid interference with other functions using temporary variables.

(End definition for \l__exp_internal_tl.)

\exp_after:wN \exp_not:N \exp_not:n

These are defined in \l3basics, as they are needed “early”. This is just a reminder of that fact!

(End definition for \exp_after:wN, \exp_not:N, and \exp_not:n. These functions are documented on page 38.)

42.1 General expansion

In this section a general mechanism for defining functions that handle arguments is defined. These general expansion functions are expandable unless x is used. (Any version of x is going to have to use one of the \Eproc\TeX\ names for \cs_set:Npx at some point, and so is never going to be expandable.)

The definition of expansion functions with this technique happens in section 42.8.

In section 42.2 some common cases are coded by a more direct method for efficiency, typically using calls to \exp_after:wN.

\l__exp_internal_tl

This scratch token list variable is defined in \l3basics.

(End definition for \l__exp_internal_tl.)

This code uses internal functions with names that start with \:: to perform the expansions. All macros are long since the tokens undergoing expansion may be arbitrary user input.

An argument manipulator \::(Z) always has signature #1::+#2#3 where #1 holds the remaining argument manipulations to be performed, \:: serves as an end marker for the list of manipulations, #2 is the carried over result of the previous expansion steps and #3 is the argument about to be processed. One exception to this rule is \::p, which has to grab an argument delimited by a left brace.
#1 is the result of an expansion step, #2 is the remaining argument manipulations and #3 is the current result of the expansion chain. This auxiliary function moves #1 back after #3 in the input stream and checks if any expansion is left to be done by calling #2. In by far the most cases we need to add a set of braces to the result of an argument manipulation so it is more effective to do it directly here. Actually, so far only the \texttt{c} of the final argument manipulation variants does not require a set of braces.

\begin{verbatim}
\cs_new:Npn \__exp_arg_next:nnn #1#2#3 { #2 \::: { #3 {#1} } }
\cs_new:Npn \__exp_arg_next:Nnn #1#2#3 { #2 \::: { #3 #1 } }
\end{verbatim}

\texttt{\:::} The end marker is just another name for the identity function.

\begin{verbatim}
\cs_new:Npn \::: #1 {#1}
\end{verbatim}

\texttt{\::n} This function is used to skip an argument that doesn’t need to be expanded.

\begin{verbatim}
\cs_new:Npn \::n #1 \::: #2#3 { #1 \::: { #2 {#3} } }
\end{verbatim}

\texttt{\::N} This function is used to skip an argument that consists of a single token and doesn’t need to be expanded.

\begin{verbatim}
\cs_new:Npn \::N #1 \::: #2#3 { #1 \::: {#2#3} }
\end{verbatim}

\texttt{\::p} This function is used to skip an argument that is delimited by a left brace and doesn’t need to be expanded. It is not wrapped in braces in the result.

\begin{verbatim}
\cs_new:Npn \::p #1 \::: #2#3# { #1 \::: {#2#3} }
\end{verbatim}

\texttt{\::c} This function is used to skip an argument that is turned into a control sequence without expansion.

\begin{verbatim}
\cs_new:Npn \::c #1 \::: #2#3
\end{verbatim}

\texttt{\::o} This function is used to expand an argument once.

\begin{verbatim}
\cs_if_exist:NTF \tex_expanded:D
\end{verbatim}

\texttt{\::e} With the \texttt{\expanded} primitive available, just expand. Otherwise defer to \texttt{\exp_args:Ne} implemented later.
This function is used to expand a token list until the first unexpandable token is found. This is achieved through \exp:w \exp_end_continue_f:w that expands everything in its way following it. This scanning procedure is terminated once the expansion hits something non-expandable (if that is a space it is removed). We introduce \exp_stop_f: to mark such an end-of-expansion marker. For example, f-expanding \cs_set_eq:Nc \aaa { b \l_tmpa_tl b } where \l_tmpa_tl contains the characters lur gives \texttt{tex\_let:D} \aaa = \blurb which then turns out to start with the non-expandable token \texttt{tex\_let:D}. Since the expansion of \exp:w \exp_end_continue_f:w is empty, we wind up with a fully expanded list, only \TeX{} has not tried to execute any of the non-expandable tokens. This is what differentiates this function from the \texttt{x} argument type.

\begin{verbatim}
\cs_new:Npn \::f #1 \::: #2#3
{ \exp_after:wN \__exp_arg_next:nnn
\exp_after:wN { \exp:w \exp_end_continue_f:w #3 }
{#1} {#2}
}
\use:nn { \cs_new_eq:NN \exp_stop_f: } { ~ }
\end{verbatim}

(End definition for \::f and \exp_stop_f:. These functions are documented on page 42.)

\texttt{\::x} This function is used to expand an argument fully. We build in the expansion of \texttt{\__:exp\_arg\_next:nnn}.

\begin{verbatim}
\cs_new_protected:Npn \::x #1 \::: #2#3
{ \cs_set_nopar:Npx \l__exp_internal_tl
\exp_not:n { #1 \::: } { \exp_not:n {#2} {#3} } }
\l__exp_internal_tl
\end{verbatim}

(End definition for \::x. This function is documented on page 42.)

\texttt{\::v} These functions return the value of a register, i.e., one of \texttt{tl}, \texttt{clist}, \texttt{int}, \texttt{skip}, \texttt{dim}, \texttt{muskip}, or built-in \TeX{} register. The \texttt{V} version expects a single token whereas \texttt{v} like \texttt{c} creates a csname from its argument given in braces and then evaluates it as if it was a \texttt{V}. The \texttt{\exp:w} sets off an expansion similar to an \texttt{f}-type expansion, which we terminate using \texttt{\exp_end:}. The argument is returned in braces.

\begin{verbatim}
\cs_new:Npn \::v #1 \::: #2#3
{ \exp_after:wN \__exp_arg_next:nnn
\exp_after:wN { \exp:w \__exp_eval_register:N #3 }
{#1} {#2}
}
\cs_new:Npn \::V #1 \::: #2#3
{ \exp_after:wN \__exp_arg_next:nnn
\exp_after:wN { \exp:w \__exp_eval_register:c {#3} }
{#1} {#2}
}
\end{verbatim}

(End definition for \::v and \::V. These functions are documented on page 42.)
This function evaluates a register. Now a register might exist as one of two things: A parameter-less macro or a built-in \TeX register such as \count. For the \TeX registers we have to utilize a \the whereas for the macros we merely have to expand them once. The trick is to compare the meaning of the token in question when it has been prefixed with \exp_not:N and the token itself. If it is a macro, the prefixed \exp_not:N temporarily turns it into the primitive \scan_stop:

\begin{verbatim}
cs_new:Npn \__exp_eval_register:N #1
{ \exp_after:wN \if_meaning:w \exp_not:N #1 #1

\if_meaning:w \scan_stop: #1
__exp_eval_error_msg:w
\fi:
\fi:
\msg_expandable_error:nnn { kernel } { bad-variable } {#2}
\exp_end:
}
cs_new:Npn \__exp_eval_register:c #1
{ \exp_after:wN \__exp_eval_register:N \cs:w #1 \cs_end: }
\end{verbatim}

If the token was not a macro it may be a malformed variable from a c expansion in which case it is equal to the primitive \scan_stop:. In that case we throw an error. We could let \TeX do it for us but that would result in the rather obscure

\begin{verbatim}
! You can’t use ‘relax’ after \the.
\end{verbatim}

which while quite true doesn’t give many hints as to what actually went wrong. We provide something more sensible.

\begin{verbatim}
\if_meaning:w \scan_stop: #1
__exp_eval_error_msg:w
\fi:
\fi:
\msg_expandable_error:nnn { kernel } { bad-variable } {#2}
\exp_end:
}
\end{verbatim}

The next bit requires some explanation. The function must be initiated by \exp:w and we want to terminate this expansion chain by inserting the \exp_end: token. However, we have to expand the register \#1 before we do that. If it is a \TeX register, we need to execute the sequence \exp_after:wN \exp_end: \tex_the:D #1 and if it is a macro we need to execute \exp_after:wN \exp_end: #1. We therefore issue the longer of the two sequences and if the register is a macro, we remove the \tex_the:D.

\begin{verbatim}
\else:
\exp_after:wN \use_i_ii:nnn
\fi:
\exp_after:wN \exp_end: \tex_the:D #1
}
cs_new:Npn \__exp_eval_register:c #1
{ \exp_after:wN \__exp_eval_register:N \cs:w #1 \cs_end: }
\end{verbatim}

Clean up nicely, then call the undefined control sequence. The result is an error message looking like this:

\begin{verbatim}
! Undefined control sequence.
<argument> \LaTeX3 error: Erroneous variable used!
1.55 \tl_set:Nv \l_tmpa_tl {undefined_tl}
\end{verbatim}

\begin{verbatim}
\msg_expandable_error:nnn { kernel } { bad-variable } {#2}
\exp_end:
}
\end{verbatim}

(End definition for \__exp_eval_register:N and \__exp_eval_error_msg:w.)
42.2 Hand-tuned definitions

One of the most important features of these functions is that they are fully expandable.

\begin{verbatim}
\exp_args:Nc  \exp_args:cc  (End definition for \exp_args:Nc and \exp_args:cc. These functions are documented on page 34.)
\exp_args:NNc \exp_args:Ncc \exp_args:Nccc  Here are the functions that turn their argument into csnames but are expandable.
\begin{verbatim}
\cs_new:Npn \exp_args:NNc #1#2#3 { \exp_after:wN #1 \exp_after:wN \cs:w # 3\cs_end: }
\cs_new:Npn \exp_args:Ncc #1#2#3 { \exp_after:wN #1 \cs:w #2 \exp_after:wN \cs_end: \cs:w #3 \cs_end: }
\cs_new:Npn \exp_args:Nccc #1#2#3#4 { \exp_after:wN #1 \exp_after:wN \cs:w #2 \exp_after:wN \cs_end: \exp_after:wN \cs:w #3 \exp_after:wN \cs_end: \exp_after:wN \cs:w #4\cs_end: }
(End definition for \exp_args:NNc, \exp_args:Ncc, and \exp_args:Nccc. These functions are documented on page 36.)
\exp_args:No  \exp_args:NNo \exp_args:NNNo  Those lovely runs of expansion!
\begin{verbatim}
\cs_new:Npn \exp_args:No #1#2 { \exp_after:wN #1 \exp_after:wN {#2} }
\cs_new:Npn \exp_args:NNo #1#2#3 { \exp_after:wN #1 \exp_after:wN #2 \exp_after:wN {#3} }
\cs_new:Npn \exp_args:NNNo #1#2#3#4 { \exp_after:wN #1 \exp_after:wN #2 \exp_after:wN #3 \exp_after:wN {#4} }
(End definition for \exp_args:No, \exp_args:NNo, and \exp_args:NNNo. These functions are documented on page 35.)
\exp_args:Ne  When the \texttt{expanded} primitive is available, use it. Otherwise use \texttt{\_\_exp_e:nn}, defined later, to fully expand tokens.
\begin{verbatim}
\cs_if_exist:NTF \tex_expanded:D { \cs_new:Npn \exp_args:Ne #1#2 { \exp_after:wN #1 \tex_expanded:D \exp_after:wN \cs:w #2\cs_end: } }
\cs_new:Npn \exp_args:Ne #1#2 { \exp_after:wN #1 \exp_after:wN \exp:w \__exp_e:nn {#2} { } }
(End definition for \exp_args:Ne. This function is documented on page 35.)
\end{verbatim}
\end{verbatim}

385
\exp_args:Nf
\exp_args:NV
\exp_args:Nv
\exp_after:wN \exp_after:wN \exp_after:wN { \exp:w \exp_end_continue_f:w #2 }
\exp_after:wN \exp_after:wN { \exp:w \__exp_eval_register:c {#2} }
\exp_after:wN \exp_after:wN { \exp:w \__exp_eval_register:N #2 }

(End definition for \exp_args:Nf, \exp_args:NV, and \exp_args:Nv. These functions are documented on page 35.)

\exp_args:NNV
\exp_args:NNv
\exp_args:NNe
\exp_args:NNf
\exp_args:Nco
\exp_args:NcV
\exp_args:Ncv
\exp_args:Ncf
\exp_args:NVV

Some more hand-tuned function with three arguments. If we forced that an \o argument always has braces, we could implement \exp_args:Nco with less tokens and only two arguments.

\exp_after:wN \exp_after:wN \exp_after:wN { \exp:w \__exp_eval_register:N #3 }
\exp_after:wN \exp_after:wN { \exp:w \__exp_eval_register: N #3 }
\cs_new:Npn \exp_args:NcV #1#2#3
{\exp_after:wN #1 \cs:w #2 \exp_after:wN \cs_end: \exp_after:wN { \exp:w \__exp_eval_register:N #3 } }
\cs_new:Npn \exp_args:Ncv #1#2#3
{\exp_after:wN #1 \cs:w #2 \exp_after:wN \cs_end: \exp_after:wN { \exp:w \__exp_eval_register:c {#3} } }
\cs_new:Npn \exp_args:Ncf #1#2#3
{\exp_after:wN #1 \cs:w #2 \exp_after:wN \cs_end: \exp_after:wN { \exp:w \exp_end_continue_f:w #3 } }
\cs_new:Npn \exp_args:NVV #1#2#3 #4
{\exp_after:wN #1 \exp_after:wN { \exp:w \exp_after:wN \__exp_eval_register:N \exp_after:wN #2 \exp_after:wN } \exp_after:wN { \exp:w \__exp_eval_register:N #3 } }
\exp_args:NNNV
\exp_args:NNNv
\exp_args:NcNc
\exp_args:NcNo
\exp_args:Ncco

A few more that we can hand-tune.

\exp_args:NNNV
\exp_args:NNNv
\exp_args:NcNc
\exp_args:NcNo
\exp_args:Ncco

(End definition for \exp_args:NNV and others. These functions are documented on page 36.)

387
42.3 Last-unbraced versions

There are a few places where the last argument needs to be available unbraced. First
some helper macros.

\exp_after:wN \cs_new:Npn \exp_after:wN \cs_new:Npn \exp_after:wN \cs_new:Npn \exp_after:wN \cs_new:Npn \exp_after:wN \cs_new:Npn \exp_after:wN \cs_new:Npn \exp_after:wN \cs_new:Npn \exp_after:wN \cs_new:Npn \exp_after:wN \cs_new:Npn \exp_after:wN \cs_new:Npn \exp_after:wN \cs_new:Npn \exp_after:wN

\cs_new:Npn \__exp_arg_last_unbraced:nn #1#2 { #2#1 }
\cs_new:Npn \::o_unbraced \::: #1#2 { \exp_after:wN \__exp_arg_last_unbraced:nn \exp_after:wN {#2} {#1} }
\cs_new:Npn \::V_unbraced \::: #1#2 { \exp_after:wN \__exp_arg_last_unbraced:nn \exp_after:wN { \exp:w \__exp_eval_register:N #2 } {#1} }
\cs_new:Npn \::v_unbraced \::: #1#2 { \exp_after:wN \__exp_arg_last_unbraced:nn \exp_after:wN { \exp:w \__exp_eval_register:c {#2} } {#1} }
\cs_if_exist:NTF \tex_expanded:D { \exp_not:n {#1} #2 } }
\cs_new:Npn \::e_unbraced \::: #1#2 { \tex_expanded:D { \exp_not:n {#1} #2 } }
\cs_new:Npn \::f_unbraced \::: #1#2 { \exp:w \__exp_e:nn {#2} {#1} }
\cs_new_protected:Npn \::x_unbraced \::: #1#2 { \cs_set_nopar:Npx \l__exp_internal_tl { \exp_not:n {#1} #2 } }

(End definition for \exp_args:NV and others. These functions are documented on page 36.)

\exp_args:Nx

\cs_new_protected:Npn \exp_args:Nx #1#2 { \use:x { \exp_not:N #1 {#2} } }
(End definition for \exp_args:Nx. This function is documented on page 35.)
Now the business end: most of these are hand-tuned for speed, but the general system is in place.

\cs_new:Npn \exp_last_unbraced:No #1#2 { \exp_after:wN #1 #2 }
\cs_new:Npn \exp_last_unbraced:NV #1#2 { \exp_after:wN #1 \exp:w \__exp_eval_register:N #2 }
\cs_new:Npn \exp_last_unbraced:Nv #1#2 { \exp_after:wN #1 \exp:w \__exp_eval_register:c {#2} }
\cs_if_exist:NTF \tex_expanded:D
{ \cs_new:Npn \exp_last_unbraced:Ne #1#2 { \exp_after:wN #1 \tex_expanded:D {#2} }
  \cs_new:Npn \exp_last_unbraced:Ne { \::e_unbraced \::: } }
\cs_new:Npn \exp_last_unbraced:Nf #1#2#3 { \exp_after:wN #1 \exp:w \exp_end_continue_f:w #3 }
\cs_new:Npn \exp_last_unbraced:NNo #1#2#3#4 { \exp_after:wN #1 \exp_after:wN #2 \exp_after:wN #3 #4 }
\cs_new:Npn \exp_last_unbraced:NNV #1#2#3 { \exp_after:wN #1 \exp_after:wN #2 \exp:w \__exp_eval_register:N #3 }
\cs_new:Npn \exp_last_unbraced:NNf #1#2#3#4 { \exp_after:wN #1 \exp_after:wN #2 \exp_after:wN #3 \exp:w \__exp_eval_register:N #4 }
\cs_new:Npn \exp_last_unbraced:Nx #1#2#3#4 { \exp_after:wN #1 \exp_after:wN #2 \exp_after:wN #3 \exp_after:wN #4 }
\cs_new:Npn \exp_last_unbraced:NNNo #1#2#3#4 { \exp_after:wN #1 \exp_after:wN #2 \exp_after:wN #3 \exp_after:wN #4 }
\cs_new:Npn \exp_last_unbraced:NNNV #1#2#3#4 { \exp_after:wN #1 \exp_after:wN #2 \exp_after:wN #3 \exp:w \__exp_eval_register:N #4 }
\cs_new:Npn \exp_last_unbraced:NNo #1#2#3 { \exp_after:wN #1 \exp_after:wN #2 \exp:w \__exp_eval_register:N #3 }
\cs_new:Npn \exp_last_unbraced:Nco #1#2#3#4 { \exp_after:wN #1 \cs:w #2 \exp_after:wN \cs_end: #3 \exp:w \__exp_eval_register:N #4 }
\cs_new:Npn \exp_last_unbraced:NcV #1#2#3 { \exp_after:wN #1 \cs:w #2 \exp_after:wN \cs_end: \exp:w \__exp_eval_register:N #3 }
\cs_new:Npn \exp_last_unbraced:NNf #1#2#3#4 { \exp_after:wN #1 \exp_after:wN #2 \exp_after:wN #3 \exp:w \exp_end_continue_f:w #4 }
\cs_new:Npn \exp_last_unbraced:Nx #1#2#3#4 { \exp_after:wN #1 \exp_after:wN #2 \exp_after:wN #3 \exp_after:wN #4 }
\cs_new:Npn \exp_last_unbraced:Nx #1#2#3#4 { \exp_after:wN #1 \exp_after:wN #2 \exp_after:wN #3 \exp_after:wN #4 }

(End definition for \_\_exp_arg_last_unbraced and others. These functions are documented on page 42.)
If \#2 is a single token then this can be implemented as
\begin{verbatim}
\cs_new:Npn \exp_last_two_unbraced:Noo \#1 \#2 \#3
{ \exp_after:wN \exp_after:wN \exp_after:wN \#1 \exp_after:wN \#2 \#3 }
\end{verbatim}
However, for robustness this is not suitable. Instead, a bit of a shuffle is used to ensure that \#2 can be multiple tokens.
\begin{verbatim}
\cs_new:Npn \exp_last_two_unbraced:Noo \#1\#2\#3
{ \exp_after:wN \__exp_last_two_unbraced:noN \exp_after:wN {\#3} {\#2} \#1 }
\end{verbatim}
(End definition for \texttt{exp_last_unbraced:No} and \texttt{\_exp_last_two_unbraced:nO}. This function is documented on page 38.)

## 42.4 Preventing expansion

At the kernel level, we need the primitive behaviour to allow expansion before the brace group.
\begin{verbatim}
\cs_new_eq:NN \_kernel_exp_not:w \tex_unexpanded:D
\end{verbatim}
(End definition for \texttt{\_kernel_exp_not:w}.)

All these except \texttt{\_exp_not:c} call the kernel-internal \texttt{\_kernel_exp_not:w} namely
\begin{verbatim}
\exp_not:c \#1 { \exp_after:wN \exp_not:N \cs:w \#1 \cs_end: }
\end{verbatim}
\begin{verbatim}
\exp_not:o \tex_unexpanded:D.
\end{verbatim}
\begin{verbatim}
\exp_not:e \#1 { \exp_after:wN \exp_not:N \cs:w \#1 \cs_end: }
\end{verbatim}
\begin{verbatim}
\exp_not:f \#1 { \_kernel_exp_not:w \exp_after:wN {\#1} }
\end{verbatim}
\begin{verbatim}
\exp_not:v \#1 \tex_expanded:D
\end{verbatim}
\begin{verbatim}
\exp_not:v \#1 \tex_expanded:D { {\#1} }
\end{verbatim}
To trigger a sequence of “arbitrarily” many expansions we need a method to invoke \TeX’s expansion mechanism in such a way that (a) we are able to stop it in a controlled manner and (b) the result of what triggered the expansion in the first place is null, i.e., that we do not get any unwanted side effects. There aren’t that many possibilities in \TeX; in fact the one explained below might well be the only one (as normally the result of expansion is not null).

The trick here is to make use of the fact that \texttt{\@romannumeral:D} expands the tokens following it when looking for a number and that its expansion is null if that number turns out to be zero or negative. So we use that to start the expansion sequence: \exp:w is set equal to \texttt{\@romannumeral:D} in \texttt{l3basics}. To stop the expansion sequence in a controlled way all we need to provide is a constant integer zero as part of expanded tokens. As this is an integer constant it immediately stops \texttt{\@romannumeral:D}’s search for a number. Again, the definition of \exp_end: as the integer constant zero is in \texttt{l3basics}. (Note that according to our specification all tokens we expand initiated by \exp:w are supposed to be expandable (as well as their replacement text in the expansion) so we will not encounter a “number” that actually result in a roman numeral being generated. Or if we do then the programmer made a mistake.)

If on the other hand we want to stop the initial expansion sequence but continue with an f-type expansion we provide the alphabetic constant \texttt{'^^@} that also represents 0 but this time \TeX’s syntax for a \langle number \rangle continues searching for an optional space (and it continues expansion doing that) — see \TeXbook page 269 for details.

If the above definition ever appears outside its proper context the active character \texttt{'^^@} will be executed so we turn this into an error. The test for existence covers the (unlikely) case that some other code has already defined \texttt{'^^@}; this is true for example for \texttt{xmlyx.tex}.

\cs_new_protected:Npn \exp_end_continue_f:w \{ \exp:w \__exp_eval_register:c {#1} \}
\cs_new_protected:Npn \exp_end_continue_f:nw \{ \exp:w \__exp_eval_register:N #1 \}
The same but grabbing an argument to remove spaces and braces.

\cs_new:Npn \exp_end_continue_f:nw #1 { '^^@ #1 }

\group_end:

(End definition for \exp:w and others. These functions are documented on page 40.)

42.6 Emulating e-type expansion

When the \expanded primitive is available it is used to implement e-type expansion; otherwise we emulate it.

\cs_if_exist:NF \tex_expanded:D

\__exp_e:nn
\__exp_e_end:nn

Repeatedly expand tokens, keeping track of fully-expanded tokens in the second argument to \__exp_e:nn; this function eventually calls \__exp_e_end:nn to leave \exp_end: in the input stream, followed by the result of the expansion. There are many special cases: spaces, brace groups, \noexpand, \unexpanded, \the, \primitive. While we use brace tricks \if_false: \fi:, the expansion of this function is always triggered by \exp:w so brace balance is eventually restored after that is hit with a single step of expansion. Otherwise we could not nest e-type expansions within each other.

\cs_new:Npn \__exp_e:nn #1

{ \tl_if_head_is_N_type:nTF {#1} { \__exp_e:N } { \tl_if_head_is_group:nTF {#1} { \__exp_e_group:n } { \tl_if_empty:nTF {#1} { \exp_after:wN \__exp_e_end:nn } { \exp_after:wN \__exp_e_space:nn } } \exp_after:wN { \if_false: } \fi: \fi: #1 } }

\cs_new:Npn \__exp_e_end:nn #1#2 { \exp_end: #2 }

(End definition for \__exp_e:nn and \__exp_e_end:nn.)

\__exp_e_space:nn

For an explicit space character, remove it by f-expansion and put it in the (future) output.

\cs_new:Npn \__exp_e_space:nn #1#2 { \exp_args:Nf \__exp_e:nn {#1} { #2 - } }

(End definition for \__exp_e_space:nn.)
For a group, expand its contents, wrap it in two pairs of braces, and call \_exp_e_group:n. This function places the first item (the double-brace wrapped result) into the output. Importantly, \_tl_head:n works even if the input contains quarks.

```
cs_new:Npn \_exp_e_group:n #1
{
  \exp_after:wN \_exp_e_put:nn
  \exp_after:wN { \exp_after:wN { \exp_after:wN { }
    \exp:w \if_false: } \fi:
      \_exp_e:nn (#1) { } }
}
```

2639 \_exp_e_group:n \_exp_e_put:nn \_exp_e_put:nn For an \emph{N}-type token, call \_exp_e:Nnn with arguments the \emph{first token}, the remaining tokens to expand and what's already been expanded. If the \emph{first token} is non-expandable, including \protected (\long or not) macros, it is put in the result by \_exp_e_protected:Nnn. The four special primitives \unexpanded, \noexpand, \the, \\primitive are detected; otherwise the token is expanded by \_exp_e_expandable:Nnn.

```
cs_new:Npn \_exp_e:N #1
{
  \exp_after:wN \_exp_e:Nnn
  \exp_after:wN #1
  \exp_after:wN { \if_false: } \fi:
}
```

(End definition for \_exp_e_group:n, \_exp_e_put:nn, and \_exp_e_put:nnn.)
We don’t try hard to make sensible error recovery since the error recovery of \text\_primitive:D when followed by something else than a primitive depends on the engine. The only valid case is when what follows is \text N-type. Then distinguish special primitives \unexpanded, \noexpand, \the, \primitive from other primitives. In the “other” case, the only reasonable way to check if the primitive that follows \text primitive:D is expandable is to expand and compare the before-expansion and after-expansion results. If they coincide then probably the primitive is non-expandable and should be put in the output together with \text primitive:D (one can cook up contrived counter-examples where the true \expanded would have an infinite loop), and otherwise one should continue expanding.
The \noexpand primitive has no effect when followed by a token that is not N-type; otherwise \__exp_e_put:nn can grab the next token and put it in the result unchanged.

The \unexpanded primitive expands and ignores any space, \scan_stop:, or token affected by \exp_not:N, then expects a brace group. Since we only support brace-balanced token lists it is impossible to support the case where the argument of \unexpanded starts with an implicit brace. Even though we want to expand and ignore spaces we cannot blindly f-expand because tokens affected by \exp_not:N should discarded without being expanded further.

As usual distinguish four cases: brace group (the normal case, where we just put the item in the result), space (just f-expand to remove the space), empty (an error), or N-type ⟨token⟩. In the last case call \__exp_e_unexpanded:nN triggered by an f-expansion. Having a non-expandable ⟨token⟩ after \unexpanded is an error (we recover by passing {} to \unexpanded; this is different from \TeX{} because the error recovery of \unexpanded changes the balance of braces), unless that ⟨token⟩ is \scan_stop: or a space (recall that we don’t implement the case of an implicit begin-group token). An expandable ⟨token⟩ is instead expanded, unless it is \noexpand. The latter primitive can be followed by an expandable N-type token (removed), by a non-expandable one (kept and later causing an error), by a space (removed by f-expansion), or by a brace group or nothing (later causing an error).
Finally implement \texttt{\_\_exp_e_the:NN}. Followed by anything other than an \texttt{N}-type \langle\textit{token}\rangle this causes an error (we just let \TeX make one), otherwise we test the \langle\textit{token}\rangle. If the \langle\textit{token}\rangle is expandable, expand it. Otherwise it could be any kind of register, or things like \texttt{\numexpr}, so there is no way to deal with all cases. Thankfully, only \texttt{\toks} data needs to be protected from expansion since everything else gives a string of characters. If the \langle\textit{token}\rangle is \texttt{\toks} we find a number and unpack using the \texttt{\textit{the_toks}} functions. If it is a token register we unpack it in a brace group and call \texttt{\_\_exp_e_put:nn} to move it to
the result. Otherwise we unpack and continue expanding (useless but safe) since it is basically impossible to have a handle on where the result of \texttt{\the} ends.

\begin{verbatim}
\cs_new:Npn \__exp_e_the:Nnn \texttt{#1} #2
\begin{verbatim}
\tl_if_head_is_N_type:nTF \texttt{#2} \n\{ \iffalse: \fi: \__exp_e_the:N \texttt{#2} \}
\end{verbatim}
\end{verbatim}
\cs_new:Npn \__exp_e_the:N #1 \n\{ \exp_after:wN \if_meaning:w \exp_not:N \texttt{#1} \texttt{#1} \\
\exp_after:wN \use_i:nn \else: \exp_after:wN \use_ii:nn \fi: \n\{ \if_meaning:w \tex_toks:D \texttt{#1} \exp_after:wN \__exp_e_the_toks:wnn \int_value:w \exp_after:wN \__exp_e_the_toks:n \exp_after:wN { \int_value:w \iffalse: } \fi: \n\else: \__exp_e_if_toks_register:NTF \texttt{#1} \n\{ \exp_after:wN \__exp_e_the_toks_reg:N \}
\end{verbatim}
\end{verbatim}
\cs_new:Npn \__exp_e_the_toks_reg:N #1 \n\{ \exp_after:wN \__exp_e_put:nn \exp_after:wN { \exp_after:wN { \tex_the:D \iffalse: } \fi: \texttt{#1} } \}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verb}
\cs_new:Npn \__exp_e_the_toks:wnn #1; #2
{
\exp_args:No \__exp_e_put:nnn
{ \tex_the:D \tex_toks:D #1 } { ? #2 }
}
\cs_new:Npn \__exp_e_the_toks:n #1
{
\tl_if_head_is_N_type:nTF {#1}
{ \exp_after:wN \__exp_e_the_toks:N \if_false: { \fi: #1 } }
{ ; {#1} }
}
\cs_new:Npn \__exp_e_the_toks:N #1
{
\if_int_compare:w 10 < 9 \token_to_str:N #1 \exp_stop_f:
\exp_after:wN \use_i:nn
\else:
\exp_after:wN \use_ii:nn
\fi:
#1
\exp_after:wN \__exp_e_the_toks:n
\exp_after:wN { \if_false: } \fi:
}
\prg_new_conditional:Npnn \__exp_e_if_toks_register:N #1 { TF }
{
\token_if_toks_register:NTF #1 { \prg_return_true: }
{ \cs_if_exist:cTF
\{ \exp_after:wN \cs_to_meaning:N #1:
\} { \prg_return_true: } { \prg_return_false: }
}
\cs_new_eq:NN \__exp_e_the_XeTeXinterchartoks: ?
\cs_new_eq:NN \__exp_e_the_errhelp: ?
\cs_new_eq:NN \__exp_e_the_everycr: ?
\cs_new_eq:NN \__exp_e_the_everydisplay: ?
\cs_new_eq:NN \__exp_e_the_everyeof: ?
\cs_new_eq:NN \__exp_e_the_everyhbox: ?
\cs_new_eq:NN \__exp_e_the_everyjob: ?
\cs_new_eq:NN \__exp_e_the_everymath: ?
\cs_new_eq:NN \__exp_e_the_everypar: ?
\cs_new_eq:NN \__exp_e_the_pdfpageattr: ?
\cs_new_eq:NN \__exp_e_the_pdfpagesattr: ?
\cs_new_eq:NN \__exp_e_the_pdfpkmode: ?

(End definition for \__exp_e_the_toks:wnn, \__exp_e_the_toks:n, and \__exp_e_the_toks:N.)

We need to detect both \toks registers like \toks@ in \LaTeX{} and parameters such as \everypar, as the result of unpacking the register should not expand further. Registers are found by \token_if_toks_register:NTF by inspecting the meaning. The list of parameters is finite so we just use a \cs_if_exist:cTF test to look up in a table. We abuse \cs_to_str:N's ability to remove a leading escape character whatever it is.

\prg_new_conditional:Npnn \__exp_e_if_toks_register:NTF #1 { TF }
{
\token_if_toks_register:NTF #1 { \prg_return_true: }
{ \cs_if_exist:cTF
\{ \exp_after:wN \cs_to_str:N
\token_to_meaning:N #1:
\} { \prg_return_true: } { \prg_return_false: }
}
\endinput
We are done emulating e-type argument expansion when \texttt{\expanded} is unavailable.

\subsection{Defining function variants}

\texttt{\cs_new_eq:NN \__exp_e_the_everyhbox: ?} \texttt{\cs_new_eq:NN \__exp_e_the_everyjob: ?} \texttt{\cs_new_eq:NN \__exp_e_the_everymath: ?} \texttt{\cs_new_eq:NN \__exp_e_the_everypar: ?} \texttt{\cs_new_eq:NN \__exp_e_the_everyvbox: ?} \texttt{\cs_new_eq:NN \__exp_e_the_output: ?} \texttt{\cs_new_eq:NN \__exp_e_the_pdfpageattr: ?} \texttt{\cs_new_eq:NN \__exp_e_the_pdfpageresources: ?} \texttt{\cs_new_eq:NN \__exp_e_the_pdfpagesattr: ?} \texttt{\cs_new_eq:NN \__exp_e_the_pdfpkmode: ?} \texttt{(\texttt{End definition for \__exp_e_if_toks_register:NTF} and others.)}

We are done emulating e-type argument expansion when \texttt{\expanded} is unavailable.

\begin{verbatim}
\end{verbatim}

\texttt{\cs_new_eq:NN \s__cs_mark \scan_stop:} \texttt{\cs_new_eq:NN \s__cs_stop \scan_stop:} \texttt{(\texttt{End definition for \s__cs_mark and \s__cs_stop}.)}

\texttt{\cs_new:Npn \q__cs_recursion_stop { \q__cs_recursion_stop }} \texttt{(\texttt{End definition for \q__cs_recursion_stop}.)}

\texttt{\cs_new:Npn \__cs_use_none_delimit_by_s_stop:w #1 \s__cs_stop { }} \texttt{\cs_new:Npn \__cs_use_i_delimit_by_s_stop:nw #1 #2 \s__cs_stop {#1}} \texttt{\cs_new:Npn \__cs_use_none_delimit_by_q_recursion_stop:w \__cs_use_none_delimit_by_s_stop:w \__cs_use_i_delimit_by_s_stop:nw \__cs_use_none_delimit_by_q_recursion_stop:w \__cs_use_none_delimit_by_q_recursion_stop:w} \texttt{\__cs_use_none_delimit_by_s_stop:w \__cs_use_i_delimit_by_s_stop:nw \__cs_use_none_delimit_by_q_recursion_stop:w} \texttt{(\texttt{End definition for \__cs_use_none_delimit_by_s_stop:w, \__cs_use_i_delimit_by_s_stop:nw, and \__cs_use_none_delimit_by_q_recursion_stop:w.})}

\texttt{\cs_new_protected:Npn \cs_generate_variant:Nn #1#2} \texttt{\cs_new_protected:Npn \cs_generate_variant:cn #1: Base form of a function; e.g., \texttt{\tl_set:Nn}} \texttt{\cs_new_protected:Npx or \texttt{\cs_new_protected:Np}, which is then used to define all the variants (except those involving x-expansion, always protected). Split up the original base function only once, to grab its name and signature. Then we wish to iterate through the comma list of variant argument specifiers, which we first convert to a string: the reason is explained later.

\begin{verbatim}
\end{verbatim}

\begin{verbatim}
\end{verbatim}
The goal here is to pick up protected parent functions. There are four cases: the parent function can be a primitive or a macro, and can be expandable or not. For non-expandable primitives, all variants should be protected; skipping the \else: branch is safe because non-expandable primitives cannot be \TeX{} conditionals.

The other case where variants should be protected is when the parent function is a protected macro: then \texttt{protected} appears in the meaning before the first occurrence of \texttt{macro}. The \texttt{ww} auxiliary removes everything in the meaning string after the first \texttt{ma}. We use \texttt{ma} rather than the full \texttt{macro} because the meaning of the \texttt{\firstmark} primitive (and four others) can contain an arbitrary string after a leading \texttt{firstmark}:. Then, look for \texttt{pr} in the part we extracted: no need to look for anything longer: the only strings we can have are an empty string, \texttt{\long}, \texttt{\protected}, \texttt{\protected\long}, \texttt{\first}, \texttt{\top}, \texttt{\bot}, \texttt{\splittop}, or \texttt{\splitbot}, with \texttt{\} replaced by the appropriate escape character. If \texttt{pr} appears in the part before \texttt{ma}, the first \texttt{s\_cs\_mark} is taken as an argument of the \texttt{wwNw} auxiliary, and \texttt{#3} is \texttt{\cs\_new\_protected\_Np}x, otherwise it is \texttt{\cs\_new\_protected\_Np}x.
If the boolean is \texttt{false}, the base function has no colon and we abort with an error; otherwise, set off a loop through the desired variant forms. The original function is retained as \#4 for efficiency.

```latex
\texttt{\cs_new_protected:Npn \__cs_generate_variant:nnNN \#1\#2\#3\#4}
{
  \if_meaning:w \c\texttt{false}\texttt{bool} \#3
  \msg_error:nnx { kernel } { missing-colon }
  \token_to_str:c {\#1} \}
  \exp_after:wN \__cs_use_none_delimit_by_q_recursion_stop:w
  \fi:
  \__cs_generate_variant:Nnnw \#4 {\#1}{\#2}
}
```

(End definition for \texttt{\__cs_generate_variant:nnNN}.)

\texttt{\__cs_generate_variant:NNnN \#1 \#2 \#3 \#4}

\#1: Base function.
\#2: Base name.
\#3: Base signature.
\#4: Beginning of variant signature.

First check whether to terminate the loop over variant forms. Then, for each variant form, construct a new function name using the original base name, the variant signature consisting of \( l \) letters and the last \( k - l \) letters of the base signature (of length \( k \)). For example, for a base function \texttt{prop\_put:Nnn} which needs a \texttt{cV} variant form, we want the new signature to be \texttt{cVn}.

There are further subtleties:

- In \texttt{\cs_generate_variant:Nn\foo:nnTF \{xxTF\}}, we must define \texttt{\foo:xxTF} using \texttt{\exp_args:Nxx}, rather than a hypothetical \texttt{\exp_args:NxxTF}. Thus, we wish to trim a common trailing part from the base signature and the variant signature.

- In \texttt{\cs_generate_variant:Nn\foo:on \{ox\}}, the function \texttt{\foo:ox} must be defined using \texttt{\exp_args:Nnx}, not \texttt{\exp_args:Nox}, to avoid double \texttt{o} expansion.

- Lastly, \texttt{\cs_generate_variant:Nn \foo:on \{xn\}} must trigger an error, because we do not have a means to replace \texttt{o}-expansion by \texttt{x}-expansion. More generally, we can only convert \texttt{N} to \texttt{c}, or convert \texttt{n} to \texttt{V}, \texttt{v}, \texttt{o}, \texttt{f}, \texttt{x}.

All this boils down to a few rules. Only \texttt{n} and \texttt{N}-type arguments can be replaced by \texttt{\cs_generate_variant:Nn}. Other argument types are allowed to be passed unchanged from the base form to the variant: in the process they are changed to \texttt{n} except for \texttt{N} and \texttt{p}-type arguments. A common trailing part is ignored.

We compare the base and variant signatures one character at a time within \texttt{x}-expansion. The result is given to \texttt{\__cs_generate_variant:wwNN} (defined later) in the form (\texttt{processed variant signature}) \texttt{\_\_cs\_mark \{errors\} \_\_cs\_stop \{base function\} \{new function\}}. If all went well, \texttt{\{errors\}} is empty; otherwise, it is a kernel error message and some clean-up code.

Note the space after \#3 and after the following brace group. Those are ignored by \texttt{TEX} when fetching the last argument for \texttt{\__cs_generate_variant_loop:nNnN}, but can be used as a delimiter for \texttt{\_\_cs\_generate\_variant\_loop\_end:nwNNnnN}.

```latex
\texttt{\cs_new_protected:Npn \_\_cs_generate_variant:Nnnn \#1\#2\#3\#4}
```

401
\begin{verbatim}
{| 
  \if_meaning:w \scan_stop: #4 
  \exp_after:wN \__cs_use_none_delimit_by_q_recursion_stop:w 
  \fi: 
  \use:x 
  { 
    \exp_not:N \__cs_generate_variant:wwNN 
    \__cs_generate_variant_loop:nNwN 
    \s__cs_mark 
    #3 - 
    { - { } \fi: \__cs_generate_variant_loop_long:wNNnn } - 
    { } 
    \s__cs_stop 
    \exp_not:N #1 {#2} {#4} 
  } 
  \__cs_generate_variant:Nnnw #1 {#2} {#3} 
|}

(End definition for \__cs_generate_variant:Nnnw.)

\__cs_generate_variant_loop:nNwN 
\__cs_generate_variant_loop_base:N 
\__cs_generate_variant_loop_same:w 
\__cs_generate_variant_loop_end:nwwwNNnn 
\__cs_generate_variant_loop_invalid:NNwNNnn 
\__cs_generate_variant_loop_special:NNwNNnn

#1:  Last few consecutive letters common between the base and variant (more precisely, \__cs_generate_variant_same:N \{letter\} for each letter).
#2:  Next variant letter.
#3:  Remainder of variant form.
#4:  Next base letter.

The first argument is populated by \__cs_generate_variant_loop_same:w when a variant letter and a base letter match. It is flushed into the input stream whenever the two letters are different: if the loop ends before, the argument is dropped, which means that trailing common letters are ignored.

The case where the two letters are different is only allowed if the base is \texttt{N} and the variant is \texttt{c}, or when the base is \texttt{n} and the variant is \texttt{o}, \texttt{V}, \texttt{v}, \texttt{f} or \texttt{x}. Otherwise, call \__cs_generate_variant_loop_invalid:NNwNNnn to remove the end of the loop, get arguments at the end of the loop, and place an appropriate error message as a second argument of \__cs_generate_variant:wwNN.

If the letters are distinct and the base letter is indeed \texttt{n} or \texttt{N}, leave in the input stream whatever argument \#1 was collected, and the next variant letter \#2, then loop by calling \__cs_generate_variant_loop:nNwN.

The loop can stop in three ways.

- If the end of the variant form is encountered first, \#2 is \__cs_generate_variant_loop_end:nwwwNNnn (expanded by the conditional \if:w), which inserts some tokens to end the conditional; grabs the \{base name\} as \#7, the \{variant signature\} \#8, the \{next base letter\} \#1 and the part \#3 of the base signature that wasn’t read yet; and combines those into the \{new function\} to be defined.

- If the end of the base form is encountered first, \#4 is \{-\}\fi: which ends the conditional (with an empty expansion), followed by \__cs_generate_variant_loop_long:wNNnn, which places an error as the second argument of \__cs_generate_variant:wwNN.

- The loop can be interrupted early if the requested expansion is unavailable, namely when the variant and base letters differ and the base is not the right one (\texttt{n} or \texttt{N} to
\end{verbatim}
support the variant). In that case too an error is placed as the second argument of \_\_cs_generate_variant:wwNN.

Note that if the variant form has the same length as the base form, \#2 is as described in the first point, and \#4 as described in the second point above. The \_\_cs_generate_variant_loop_end:nwwNnn breaking function takes the empty brace group in \#4 as its first argument: this empty brace group produces the correct signature for the full variant.

\begin{verbatim}
\cs_new:Npn \_\_cs_generate_variant_loop:nNwN #1#2#3 \s__cs_mark #4 
{  
  \if:w #2 #4 
  \exp_after:wN \_\_cs_generate_variant_loop_same:w 
  \else:  
  \if:w #4 \_\_cs_generate_variant_loop_base:N #2 \else: 
    \if:w 0 
    \if:w N \#4 \else: \if:w n \#4 \else: 1 \fi: \fi: 
    \if:w \scan_stop: \_\_cs_generate_variant_loop_base:N #2 1 \fi: 
    0 
  \_\_cs_generate_variant_loop_special:NwwNnn #4\#2 
  \else: 
  \_\_cs_generate_variant_loop_invalid:NwwNnn #4\#2 
  \fi: 
  \fi: 
  \fi: 
  \prg_do_nothing: 
  \fi: 
  \_\_cs_generate_variant_loop_same:w #1 \prg_do_nothing: #2#3#4 
  \} \s__cs_mark } 
\cs_new:Npn \_\_cs_generate_variant_loop_base:N #1 
{  
  \if:w c #1 N \else: 
    \if:w o #1 n \else: 
      \if:w V #1 n \else: 
        \if:w v #1 n \else: 
          \if:w f #1 n \else: 
            \if:w x #1 n \else: 
              \if:w n \#1 n \else: 
                \if:w N \#1 N \else: 
                  \scan_stop: 
                  \fi: 
                \fi: 
              \fi: 
            \fi: 
          \fi: 
        \fi: 
      \fi: 
    \fi: 
  \fi: 
  } 
\cs_new:Npn \_\_cs_generate_variant_loop_same:w #1 \prg_do_nothing: #2#3#4 
{ \s__cs_mark } 
\end{verbatim}

403
When the base and variant letters are identical, don’t do any expansion. For most argument types, we can use the \texttt{n}-type no-expansion, but the \texttt{N} and \texttt{p} types require a slightly different behaviour with respect to braces. For \texttt{V}-type this function could output \texttt{N} to avoid adding useless braces but that is not a problem.
\begin{verbatim}
3060 \cs_new:Npn \__cs_generate_variant_same:N #1 #1
3061 { \if:w N #1 #1 \else: \if:w p #1 #1 \else: \token_to_str:N n \if:w n #1 \else: \__cs_generate_variant_loop_special:NNnNN #1#1 \fi: \fi: \fi: \fi:
3070 }

(End definition for \__cs_generate_variant_same:N.)
\end{verbatim}

If the variant form has already been defined, log its existence (provided \texttt{log-functions} is active). Otherwise, make sure that the \texttt{\exp_args:N \#3} form is defined, and if it contains \texttt{x}, change \texttt{\cs_new_protected:Npx} locally to \texttt{\cs_new_protected:cpn}. Then define the variant by combining the \texttt{\exp_args:N \#3} variant and the base function.

\begin{verbatim}
3071 \cs_new_protected:Npn \__cs_generate_variant:wwNN #1 \s__cs_mark #2 \s__cs_stop #3#4
3072 { #2 \cs_if_free:NT #4 {
3073 \group_begin:
3074 \__cs_generate_internal_variant:n {#1}
3075 \__cs_tmp:w #4 { \exp_not:c { \exp_args:N #1 } \exp_not:N #3 }
3076 \group_end:
3077 }
3078 }

(End definition for \__cs_generate_variant:wwNN.)
\end{verbatim}

First test for the presence of \texttt{x} (this is where working with strings makes our lives easier), as the result should be protected, and the next variant to be defined using that internal variant should be protected (done by setting \texttt{\cs_new_protected:Npx}). Then call \texttt{\__cs_generate_internal_variant:nn} with arguments \texttt{\cs_new_protected:cpn} \texttt{\use:x} (for protected) or \texttt{\cs_new:cpn \tex_expanded:D} (expandable) and the signature. If \texttt{p} appears in the signature, or if the function to be defined is expandable and the primitive \texttt{\expanded} is not available, or if there are more than 8 arguments, call some fall-back code that just puts the appropriate \texttt{\cs} commands. Otherwise, call \texttt{\__cs_generate_internal_one_go:NNN} to construct the \texttt{\exp_args:N\ldots} function as a macro taking up to 9 arguments and expanding them using \texttt{\use:x} or \texttt{\tex_expanded:D}.

\begin{verbatim}
3083 \cs_new_protected:Npn \__cs_generate_internal_variant:n #1
3084 { \exp_not:N \__cs_generate_internal_variant:wwnNwn #1 \s__cs_mark
3086 \__cs_set_eq:NN \exp_not:N \__cs_tmp:w \cs_new_protected:Npx }
3088 \cs_new_protected:cpn \use:x \token_to_str:N x \s__cs_mark
3091 { }
3092 \cs_new:cpn
\end{verbatim}

405
This command grabs char by char outputting \::#1 (not expanded further). We avoid tests by putting a trailing : \use_i:nn, which leaves \cs_end: and removes the looping
macro. The colon is fact also turned into \\colon so that the required structure for \exp_args:N... commands is correctly terminated.

\cs_new:Npn \__cs_generate_internal_variant_loop:n #1
\begin{verbatim}
\exp_after:wN \exp_not:N \cs:w :: #1 \cs_end:
\__cs_generate_internal_variant_loop:n
\end{verbatim}

(End definition for \__cs_generate_internal_variant:n and \__cs_generate_internal_variant_loop:n.)
\cs_new_protected:Npn \__cs_generate_variant_F_form:nnn #1#2 { \cs_generate_variant:cn { #1 : #2 F } }
\cs_new_protected:Npn \__cs_generate_variant_TF_form:nnn #1#2 { \cs_generate_variant:cn { #1 : #2 TF } }

(End definition for \prg_generate_conditional_variant:Nnn and others. This function is documented on page 64.)

This function is not used in the kernel hence we can use functions that are defined in later modules. It also does not need to be fast so use inline mappings. For each requested variant we check that there are no characters besides \textit{NnpcofVvx}, in particular that there are no spaces. Then we just call the internal function.

\cs_new_protected:Npn \exp_args_generate:n #1 { \exp_args:No \clist_map_inline:nn { \tl_to_str:n {#1} } { \str_map_inline:nn {##1} { \str_if_in:nnF { NnpcofeVvx } {####1} { \msg_error:nnnn { kernel } { invalid-exp-args } {####1} {##1} \str_map_break:n { \use_none:nn } } } \__cs_generate_internal_variant:n {##1} } }

(End definition for \exp_args_generate:n. This function is documented on page 301.)

42.8 Definitions with the automated technique

Some of these could be done more efficiently, but the complexity of coding then becomes an issue. Notice that the auto-generated functions actually take no arguments themselves.

\cs_set_protected:Npn \__cs_tmp:w #1 { \group_begin: \exp_args:No \__cs_tmp:w #1 \group_end: \__cs_tmp:w { nc } \__cs_tmp:w { no } \__cs_tmp:w { nV } \__cs_tmp:w { nv } \__cs_tmp:w { ne } \__cs_tmp:w { Nf } \__cs_tmp:w { oc } \__cs_tmp:w { oo } }

Here are the actual function definitions, using the helper functions above. The group is used because \__cs_generate_internal_variant:n redefines \__cs_tmp:w locally.
(End definition for \exp_args:Nnc and others. These functions are documented on page 36.)

\exp_args:NNcf
\exp_args:NNno
\exp_args:NNnV
\exp_args:NNoo
\exp_args:NNVV
\exp_args:Ncno
\exp_args:NcnV
\exp_args:Ncoo
\exp_args:NcVV
\exp_args:Nncn
\exp_args:Nnno
\exp_args:Nnnf
\exp_args:Nnff
\exp_args:Nooo
\exp_args:Noof
\exp_args:Neee
\exp_args:Nffo
\exp_args:NNNx
\exp_args:NNNx
\exp_args:NNox
\exp_args:Nccx
\exp_args:Ncnx
\exp_args:Nnxx
\exp_args:Nnx
\exp_args:Nnox
\exp_args:Noox

(End definition for \exp_args:Ncf and others. These functions are documented on page 37.)

//package
Chapter 43

13sort implementation

### 43.1 Variables

<table>
<thead>
<tr>
<th>\texttt{_sort_internal_seq}</th>
<th>\texttt{_sort_internal_tl}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sorting happens in a group; the result is stored in those global variables before being copied outside the group to the proper places. For seq and tl this is more efficient than using \texttt{\use:x} (or some \texttt{\exp_args:NNN}) to smuggle the definition outside the group since \TeX{} does not need to re-read tokens. For clist we don’t gain anything since the result is converted from seq to clist anyways.</td>
<td></td>
</tr>
</tbody>
</table>

\begin{Verbatim}
\seq_new:N \g__sort\_internal\_seq
\tl_new:N \g__sort\_internal\_tl
\end{Verbatim}

(End definition for \texttt{\_sort\_internal\_seq} and \texttt{\_sort\_internal\_tl}.)

<table>
<thead>
<tr>
<th>\texttt{_sort_length_int}</th>
<th>\texttt{_sort_min_int}</th>
</tr>
</thead>
</table>
| The sequence has \texttt{\_sort\_length\_int} items and is stored from \texttt{\_sort\_min\_int} to \texttt{\_sort\_top\_int} − 1. While reading the sequence in memory, we check that \texttt{\_sort\_top\_int} remains at most \texttt{\_sort\_max\_int}, precomputed by \texttt{\_sort\_-\_compute\_range}.

\begin{Verbatim}
\int_new:N \l__sort\_length\_int
\int_new:N \l__sort\_min\_int
\int_new:N \l__sort\_top\_int
\int_new:N \l__sort\_max\_int
\int_new:N \l__sort\_true\_max\_int
\end{Verbatim}

(End definition for \texttt{\_sort\_length\_int} and others.)

<table>
<thead>
<tr>
<th>\texttt{_sort_block_int}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merge sort is done in several passes. In each pass, blocks of size \texttt{_sort_block_int} are merged in pairs. The block size starts at 1, and, for a length in the range ([2^k + 1, 2^{k+1}]), reaches (2^k) in the last pass.</td>
</tr>
</tbody>
</table>

\begin{Verbatim}
\int_new:N \l__sort\_block\_int
\end{Verbatim}

(End definition for \texttt{\_sort\_block\_int}.)
When merging two blocks, \l__sort_begin_int marks the lowest index in the two blocks, and \l__sort_end_int marks the highest index, plus 1.

\int_new:N \l__sort_begin_int
\int_new:N \l__sort_end_int

(End definition for \l__sort_begin_int and \l__sort_end_int.)

When merging two blocks (whose end-points are beg and end), A starts from the high end of the low block, and decreases until reaching beg. The index B starts from the top of the range and marks the register in which a sorted item should be put. Finally, C points to the copy of the high block in the interval of registers starting at \l__sort_length_int, upwards. C starts from the upper limit of that range.

\int_new:N \l__sort_A_int
\int_new:N \l__sort_B_int
\int_new:N \l__sort_C_int

(End definition for \l__sort_A_int, \l__sort_B_int, and \l__sort_C_int.)

\s__sort_mark Internal scan marks.
\s__sort_stop

(End definition for \s__sort_mark and \s__sort_stop.)

43.2 Finding available \toks registers

\_sort_shrink_range: \_sort_shrink_range_loop: After \_sort_compute_range: (defined below) determines that \toks registers between \l__sort_min_int (included) and \l__sort_true_max_int (excluded) have not yet been assigned, \_sort_shrink_range: computes \l__sort_max_int to reflect the need for a buffer when merging blocks in the merge sort. Given $2^n \leq A \leq 2^n + 2^{n-1}$ registers we can sort $\lfloor A/2 \rfloor + 2^{n-2}$ items while if we have $2^n + 2^{n-1} \leq A \leq 2^{n+1}$ registers we can sort $A - 2^{n-1}$ items. We first find out a power $2^n$ such that $2^n \leq A \leq 2^{n+1}$ by repeatedly halving \l__sort_block_int, starting at $2^{15}$ or $2^{14}$ namely half the total number of registers, then we use the formulas and set \l__sort_max_int.

\cs_new_protected:Npn \_sort_shrink_range:
\{ \int_set:Nn \l__sort_A_int { \l__sort_true_max_int - \l__sort_min_int + 1 } \int_set:Nn \l__sort_block_int { \c_max_register_int / 2 } \_sort_shrink_range_loop:
\int_set:Nn \l__sort_max_int
\{ \int_compare:nNnTF { \l__sort_block_int * 3 / 2 } > \l__sort_A_int \{ \_sort_min_int
+ ( \l__sort_block_int / 4
- 1
\} \l__sort_true_max_int - \l__sort_block_int / 2
\}
\}

412
\cs_new_protected:Npn \__sort_shrink_range_loop: 
\{
  \if_int_compare:w \l__sort_A_int < \l__sort_block_int
  \tex_divide:D \l__sort_block_int 2 \exp_stop_f:
  \exp_after:wN \__sort_shrink_range_loop:
  \fi:
  \fi:
\}

(End definition for \_sort_shrink_range: and \_sort_shrink_range_loop:.)

\_sort_compute_range: \_sort_redefine_compute_range: \_sort_max_length_int

First find out what \toks have not yet been assigned. There are many cases. In \LaTeX2ε
with no package, available \toks range from \count15 + 1 to \c_max_register_int included
(this was not altered despite the 2015 changes). When \loctoks is defined, namely
in plain (e)\TeX, or when the package etex is loaded in \LaTeX2ε, redefine \_sort_compute_range:
\count265 to \count275 – 1. The elocalloc package also defines \loctoks but uses yet another number for the upper bound, namely \e@alloc@top (minus one). We must check for \loctoks every time a sorting function
is called, as etex or elocalloc could be loaded.

In \Con\TeX MkIV the range is from \c_syst_last_allocated_toks + 1 to \c_max_register_int, and in MkII it is from \lastallocatedtoks + 1 to \c_max_register_int. In all these cases, call \_sort_shrink_range:

\cs_new_protected:Npn \_sort_compute_range: 
\{
  \int_set:Nn \l__sort_min_int { \tex_count:D 15 + 1 }
  \int_set:Nn \l__sort_true_max_int { \c_max_register_int + 1 }
  \_sort_shrink_range:
  \if_meaning:w \loctoks \tex_undefined:D \else:
    \_sort_redefine_compute_range:
    \_sort_compute_range:
  \fi:
\}

\cs_new_protected:Npn \_sort_redefine_compute_range: 
\{
  \cs_if_exist:cTF { ver@elocalloc.sty }
  \{
    \cs_gset_protected:Npn \_sort_compute_range: 
    \{
      \int_set:Nn \l__sort_min_int { \tex_count:D 265 }
      \int_set_eq:NN \l__sort_true_max_int \e@alloc@top
      \_sort_shrink_range:
    \}
  \}
  \{
    \cs_gset_protected:Npn \_sort_compute_range: 
    \{
      \int_set:Nn \l__sort_min_int { \tex_count:D 275 }
      \int_set:Nn \l__sort_true_max_int { \tex_count:D 275 }
      \_sort_shrink_range:
    \}
  \}
\}
3392 \cs_if_exist:NT \loctoks \__sort_redefine_compute_range: \}
3393 \tl_map_inline:nn \{ \lastallocatedtoks \c_syst_last_allocated_toks \}
3394 \{
3395 \cs_if_exist:NT #1
3396 \{
3397 \cs_gset_protected:Npn \__sort_compute_range: \}
3398 \{
3399 \int_set:Nn \l__sort_min_int { #1 + 1 }
3400 \int_set:Nn \l__sort_true_max_int { \c_max_register_int + 1 }
3401 \__sort_shrink_range:
3402 \}
3403 \}
3404 \}
3405 )
(End definition for \__sort_compute_range:, \__sort_redefine_compute_range:, and \c__sort_max_length_int.)

43.3 Protected user commands

\__sort_main:NNNn

Sorting happens in three steps. First store items in \toks registers ranging from \l__sort_min_int to \l__sort_top_int − 1, while checking that the list is not too long. If we reach the maximum length, that’s an error; exit the group. Secondly, sort the array of \toks registers, using the user-defined sorting function: \__sort_level: calls \__sort_compare:nn as needed. Finally, unpack the \toks registers (now sorted) into the target \tl, or into \g__sort_internal_seq for seq and clist. This is done by \__sort_seq:NNNNn and \__sort_tl:NNn.

\cs_new_protected:Npn \__sort_main:NNNn \#1\#2\#3\#4
\{
\__sort_disable_toksdef:
\__sort_compute_range:
\int_set_eq:NN \l__sort_top_int \l__sort_min_int \#1 \#3
\{
\if_int_compare:w \l__sort_top_int = \l__sort_max_int
\__sort_too_long_error:NNw \#2 \#3
\fi:
\tex_toks:D \l__sort_top_int {##1}
\int_incr:N \l__sort_top_int
\}
\int_set:Nn \l__sort_length_int
\{
\l__sort_top_int - \l__sort_min_int \}
\cs_set:Npn \__sort_compare:nn \#1 \#2 \#4
\int_set:Nn \l__sort_block_int { 1 }
\__sort_level:
\}
(End definition for \__sort_main:NNNn.)

\tl_sort:Nn \tl_sort:cn \tl_gsort:Nn \tl_gsort:cn
\__sort_tl:NNn \__sort_tl_toks:w

Call the main sorting function then unpack \toks registers outside the group into the target token list. The unpacking is done by \__sort_tl_toks:w; registers are numbered from \l__sort_min_int to \l__sort_top_int − 1. For expansion behaviour we need
a couple of primitives. The \texttt{\_gclear:N} reduces memory usage. The \texttt{\_gclear:N} reduces memory usage. The \texttt{\prg_break_point:} is used by \texttt{\_sort_main:NNNn} when the list is too long.

\begin{verbatim}
\cs_new_protected:Npn \seq_sort:Nn { \__sort_seq:NNNNn \seq_map_inline:Nn \seq_map_break:n \seq_set_eq:NN }
\cs_generate_variant:Nn \seq_sort:Nn { c }
\cs_new_protected:Npn \seq_gsort:Nn { \__sort_seq:NNNNn \seq_map_inline:Nn \seq_map_break:n \seq_gset_eq:NN }
\cs_generate_variant:Nn \seq_gsort:Nn { c }
\cs_new_protected:Npn \clist_sort:Nn { \__sort_seq:NNNNn \clist_map_inline:Nn \clist_map_break:n \clist_set_from_seq:NN }
\cs_generate_variant:Nn \clist_sort:Nn { c }
\cs_new_protected:Npn \clist_gsort:Nn { \__sort_seq:NNNNn \clist_map_inline:Nn \clist_map_break:n \clist_gset_from_seq:NN }
\cs_generate_variant:Nn \clist_gsort:Nn { c }
\cs_new_protected:Npn \__sort_seq:NNNNn #1#2#3#4#5
\group_begin:
\__sort_main:NNNn #1 #2 #4 {#5}
\__kernel_tl_gset:Nx \g__sort_internal_tl
\__sort_tl_toks:w \__sort_min_int ;
\group_end:
\__if_int_compare:w #1 < \__sort_top_int
{ \tex_the:D \tex_toks:D #1 }
\exp_after:wN \__sort_tl_toks:w
\int_value:w \int_eval:n { #1 + 1 } \exp_after:wN ;
\fi:
}
\cs_new:Npn \__sort_tl_toks:w #1 ;
{ \if_int_compare:w #1 < \__sort_top_int
{ \tex_the:D \tex_toks:D #1 }
\exp_after:wN \__sort_tl_toks:w
\int_value:w \int_eval:n { #1 + 1 } \exp_after:wN ;
\fi:
}
\end{verbatim}

\texttt{\seq_sort:Nn} and others. These functions are documented on page 116. Use the same general framework for seq and clist. Apply the general sorting code, then unpack \texttt{\_sort_tl_toks} into \texttt{\_sort_internal_seq}. Outside the group copy or convert (for clist) the data to the target variable. The \texttt{\_gclear:N} reduces memory usage. The \texttt{\prg_break_point:} is used by \texttt{\_sort_main:NNNn} when the list is too long.
43.4 Merge sort

\_\_sort_level: This function is called once blocks of size \l\_\_sort_block_int (initially 1) are each sorted. If the whole list fits in one block, then we are done (this also takes care of the case of an empty list or a list with one item). Otherwise, go through pairs of blocks starting from 0, then double the block size, and repeat.

\cs_new_protected:Npn \_\_sort_level:n {\if_int_compare:w \l\_\_sort_block_int < \l\_\_sort_length_int \l\_\_sort_end_int \l\_\_sort_min_int \_\_sort_merge_blocks: \tex_advance:D \l\_\_sort_end_int \l\_\_sort_block_int \exp_after:wN \_\_sort_level:n \fi: }

(End definition for \_\_sort_level:n.)

\_\_sort_merge_blocks: This function is called to merge a pair of blocks, starting at the last value of \l\_\_sort_end_int (end-point of the previous pair of blocks). If shifting by one block to the right we reach the end of the list, then this pass has ended: the end of the list is sorted already. Otherwise, store the result of that shift in A, which indexes the first block starting from the top end. Then locate the end-point (maximum) of the second block: shift end upwards by one more block, but keeping it \leq top. Copy this upper block of \toks registers in registers above length, indexed by C: this is covered by \_\_sort_copy_block:. Once this is done we are ready to do the actual merger using \_\_sort_merge_blocks_aux:, after shifting A, B and C so that they point to the largest index in their respective ranges rather than pointing just beyond those ranges. Of course, once that pair of blocks is merged, move on to the next pair.

\cs_new_protected:Npn \_\_sort_merge_blocks:n { \l\_\_sort_begin_int \l\_\_sort_end_int \tex_advance:D \l\_\_sort_end_int \l\_\_sort_block_int \if_int_compare:w \l\_\_sort_end_int < \l\_\_sort_top_int \l\_\_sort_A_int \l\_\_sort_end_int \tex_advance:D \l\_\_sort_end_int \l\_\_sort_block_int \if_int_compare:w \l\_\_sort_end_int > \l\_\_sort_top_int \l\_\_sort_end_int \l\_\_sort_top_int \end{verbatim}
\_sort\_copy\_block: We wish to store a copy of the “upper” block of \toks registers, ranging between the initial value of \_sort\_B\_int (included) and \_sort\_end\_int (excluded) into a new range starting at the initial value of \_sort\_C\_int, namely \_sort\_top\_int.

\cs_new_protected:Npn \_sort\_copy\_block: {
\tex_toks:D \_sort\_C\_int \tex_toks:D \_sort\_B\_int
\int_incr:N \_sort\_C\_int
\int_incr:N \_sort\_B\_int
\if_int_compare:w \_sort\_B\_int = \_sort\_end\_int
\use_i:nn
\fi:
\_sort\_copy\_block:
}

\_sort\_merge\_blocks\_aux: At this stage, the first block starts at \_sort\_begin\_int, and ends at \_sort\_A\_int, and the second block starts at \_sort\_top\_int and ends at \_sort\_C\_int. The result of the merger is stored at positions indexed by \_sort\_B\_int, which starts at \_sort\_end\_int − 1 and decreases down to \_sort\_begin\_int, covering the full range of the two blocks. In other words, we are building the merger starting with the largest values. The comparison function is defined to return either swapped or same. Of course, this means the arguments need to be given in the order they appear originally in the list.

\cs_new_protected:Npn \_sort\_merge\_blocks\_aux: {
\exp_after:wN \_sort\_compare:nn \exp_after:wN \{ \tex_the:D \tex_toks:D \exp_after:wN \_sort\_A\_int \exp_after:wN \}
\exp_after:wN \{ \tex_the:D \tex_toks:D \_sort\_C\_int \}
\prg_do_nothing:
\_sort\_return\_mark:w
\_sort\_return\_mark:w
\_sort\_mark
\_sort\_return\_none\_error:
}

(End definition for \_sort\_merge\_blocks\_aux:)
Each comparison should call \texttt{\sort_return_same:} or \texttt{\sort_return_swapped:} exactly once. If neither is called, \texttt{\__sort_return_none_error:} is called, since the \texttt{return_mark} removes tokens until \texttt{\s__sort_mark}. If one is called, the \texttt{return_mark} auxiliary removes everything except \texttt{\__sort_return_same:w} (or its \texttt{swapped} analogue) followed by \texttt{\__sort_return_none_error:}. Finally if two or more are called, \texttt{\__sort_return_two_error:} ends up before any \texttt{\__sort_return_mark:w}, so that it produces an error.

\begin{verbatim}
\cs_new_protected:Npn \sort_return_same: #1 \__sort_return_mark:w #2 \s__sort_mark {
    #1 \__sort_return_two_error:
    \__sort_return_mark:w \s__sort_mark \__sort_return_same:w
}
\cs_new_protected:Npn \sort_return_swapped: #1 \__sort_return_mark:w #2 \s__sort_mark {
    #1 \__sort_return_two_error:
    \__sort_return_mark:w \s__sort_mark \__sort_return_swapped:w
}
\cs_new_protected:Npn \__sort_return_mark:w #1 \s__sort_mark { }
\cs_new_protected:Npn \__sort_return_none_error: {
    \msg_error:nnxx { sort } { return-none }
    \tex_toks:D \l__sort_B_int \tex_toks:D \l__sort_C_int
    \int_decr:N \l__sort_B_int
}
\cs_new_protected:Npn \__sort_return_two_error: {
    \msg_error:nnxx { sort } { return-two }
    \tex_toks:D \l__sort_B_int \tex_toks:D \l__sort_C_int
    \tex_toks:D \l__sort_B_int
}
\end{verbatim}

\texttt{\__sort_return_same:w} If the comparison function returns \texttt{same}, then the second argument fed to \texttt{\__sort_compare:nn} should remain to the right of the other one. Since we build the merger starting from the right, we copy that \texttt{\toks} register into the allotted range, then shift the pointers \texttt{B} and \texttt{C}, and go on to do one more step in the merger, unless the second block has been exhausted: then the remainder of the first block is already in the correct registers and we are done with merging those two blocks.

\begin{verbatim}
\cs_new_protected:Npn \__sort_return_same:w #1 \__sort_return_none_error: {
    \text_toks:D \l__sort_B_int \text_toks:D \l__sort_C_int
    \int_decr:N \l__sort_B_int
}
\end{verbatim}
If the comparison function returns \texttt{swapped}, then the next item to add to the merger is the first argument, contents of the \verb|\toks| register \texttt{A}. Then shift the pointers \texttt{A} and \texttt{B} to the left, and go for one more step for the merger, unless the left block was exhausted (\texttt{A} goes below the threshold). In that case, all remaining \verb|\toks| registers in the second block, indexed by \texttt{C}, are copied to the merger by \verb|\__sort_merge_blocks_end:|.

\begin{verbatim}
\cs_new_protected:Npn \__sort_return_swapped:w #1 \__sort_return_none_error:
{ \tex_toks:D \l__sort_B_int \tex_toks:D \l__sort_A_int \int_decr:N \l__sort_B_int \int_decr:N \l__sort_A_int \if_int_compare:w \l__sort_A_int < \l__sort_begin_int \__sort_merge_blocks_end: \use_i:nn \fi: \__sort_merge_blocks_aux: }
\end{verbatim}

\texttt{\__sort_merge_blocks_end:} This function's task is to copy the \verb|\toks| registers in the block indexed by \texttt{C} to the merger indexed by \texttt{B}. The end can equally be detected by checking when \texttt{B} reaches the threshold \texttt{begin}, or when \texttt{C} reaches \texttt{top}.

\begin{verbatim}
\cs_new_protected:Npm \__sort_merge_blocks_end: 
{ \tex_toks:D \l__sort_B_int \tex_toks:D \l__sort_C_int \int_decr:N \l__sort_B_int \int_decr:N \l__sort_C_int \if_int_compare:w \l__sort_B_int < \l__sort_begin_int \__sort_merge_blocks_end: \use_i:nn \fi: \__sort_merge_blocks_end: }
\end{verbatim}

\subsection{Expandable sorting}

Sorting expandably is very different from sorting and assigning to a variable. Since tokens cannot be stored, they must remain in the input stream, and be read through at every step. It is thus necessarily much slower (at best $O(n^2 \ln n)$) than non-expandable sorting functions ($O(n \ln n)$).

A prototypical version of expandable quicksort is as follows. If the argument has no item, return nothing, otherwise partition, using the first item as a pivot (argument \texttt{#4} of \verb|\__sort:nnNnn|). The arguments of \verb|\__sort:nnNnn| are 1. items less than \texttt{#4}, 2. items greater or equal to \texttt{#4}, 3. comparison, 4. pivot, 5. next item to test. If \texttt{#5} is the tail of
the list, call \tl_sort:nN on \#1 and on \#2, placing \#4 in between; \use:ff expands the
parts to make \tl_sort:nN f-expandable. Otherwise, compare \#4 and \#5 using \#3. If
they are ordered, place \#5 amongst the “greater” items, otherwise amongst the “lesser”
items, and continue partitioning.

\cs_new:Npn \tl_sort:nN \#1\#2
{\tl_if_blank:nF {\#1}
 {\__sort:nnNnn { } { } \#2
 \#1 \q__sort_recursion_tail \q__sort_recursion_stop}
}

\cs_new:Npn \__sort:nnNnn #1#2#3#4#5
{\quark_if_recursion_tail_stop_do:nn {#5}
 { \use:ff { \tl_sort:nN {#1} \#3 \#4 } { \tl_sort:nN {#2} \#3 } }
\#3 \#4 \#5
{ \__sort:nnNnn {#1} \#2 \#5 \#3 \#4 }
{ \__sort:nnNnn {#1 \#5} \#2 \#3 \#4 }
}
\cs_generate_variant:Nn \use:nn { ff }

There are quite a few optimizations available here: the code below is less legible, but
more than twice as fast.
In the simple version of the code, \__sort:nnNnn is called \(O(n \ln n)\) times on average
(the number of comparisons required by the quicksort algorithm). Hence most of our focus
is on optimizing that function.

The first speed up is to avoid testing for the end of the list at every call to
\__sort:nnNnn. For this, the list is prepared by changing each \langle item \rangle of the original
token list into \langle command \rangle \{\langle item \rangle \}, just like sequences are stored. We arrange things
such that the \langle command \rangle is the \langle conditional \rangle provided by the user: the loop over the
\langle prepared tokens \rangle then looks like

\cs_new:Npn \__sort_loop:wNn ... #6#7
{ #6 \langle pivot \rangle \{#7 \langle loop big \rangle \langle loop small \rangle \langle extra arguments \rangle \}
 \__sort_loop:wNn ... \langle prepared tokens \rangle
 \end-loop } \s__sort_stop

In this example, which matches the structure of \__sort_quick_split_i:NnnmmNn and a
few other functions below, the \__sort_loop:wNn auxiliary normally receives the user’s
\langle conditional \rangle as \#6 and an \langle item \rangle as \#7. This is compared to the \langle pivot \rangle (the argument
\#5, not shown here), and the \langle conditional \rangle leaves the \langle loop big \rangle or \langle loop small \rangle
auxiliary, which both have the same form as \__sort_loop:wNn, receiving the next pair
\langle conditional \rangle \langle \langle item \rangle \rangle as \#6 and \#7. At the end, \#6 is the \langle end-loop \rangle function, which
terminates the loop.

The second speed up is to minimize the duplicated tokens between the \texttt{true} and
\texttt{false} branches of the conditional. For this, we introduce two versions of \__sort:nnNnn,
which receive the new item as \#1 and place it either into the list \#2 of items less than
the pivot \#4 or into the list \#3 of items greater or equal to the pivot.

\cs_new:Npn \_sort_i:nnnnNn #1#2#3#4#5#6
\{
    \#5 \{\#4\} \_sort_ii:nnnnNn \_sort_i:nnnnNn
    \{\#6\} \{\#2 \{\#1\} \} \{\#3\} \{\#4\}
\}
\cs_new:Npn \_sort_ii:nnnnNn #1#2#3#4#5#6
\{
    \#5 \{\#4\} \_sort_ii:nnnnNn \_sort_i:nnnnNn
    \{\#6\} \{\#2\} \{\#3 \{\#1\} \} \{\#4\}
\}

Note that the two functions have the form of \_sort_loop:wNn above, receiving as \#5
the conditional or a function to end the loop. In fact, the lists \#2 and \#3 must be made
of pairs \langle conditional \rangle \{\langle item \rangle \}, so we have to replace \{\#6\} above by \{ \#5 \{\#6\} \}, and
\{\#1\} by \#1. The actual functions have one more argument, so all argument numbers are
shifted compared to this code.

The third speed up is to avoid \use:ff using a continuation-passing style:
\_sort_quick_split:NnNn expects a list followed by \s__sort_mark \{\langle code \rangle\}, and
expands to \langle code \rangle \{sorted list\}. Sorting the two parts of the list around the pivot is done
with

\_sort_quick_split:NnNn \#2 \ldots \_sort_mark
\{
    \_sort_quick_split:NnNn \#1 \ldots \_sort_mark \{\langle code \rangle\}
    \{\langle pivot \rangle\}
\}

Items which are larger than the \langle pivot \rangle are sorted, then placed after code that sorts the
smaller items, and after the (braced) \langle pivot \rangle.

The fourth speed up is avoid the recursive call to \tl_sort:nN with an empty first
argument. For this, we introduce functions similar to the \_sort_i:nnnnNn of the
last example, but aware of whether the list of \langle conditional \rangle \{\langle item \rangle \} read so far that
are less than the pivot, and the list of those greater or equal, are empty or not: see
\_sort_quick_split:NnNn and functions defined below. Knowing whether the lists
are empty or not is useless if we do not use distinct ending codes as appropriate. The
splitting auxiliaries communicate to the \langle end-loop \rangle function (that is initially placed after
the “prepared” list) by placing a specific ending function, ignored when looping, but
useful at the end. In fact, the \langle end-loop \rangle function does nothing but place the appropriate
ending function in front of all its arguments. The ending functions take care of sorting
non-empty sublists, placing the pivot in between, and the continuation before.

The final change in fact slows down the code a little, but is required to avoid memory
issues: schematically, when \TeX encounters

\use:n \{ \use:n \{ \use:n \{ \ldots \} \ldots \} \ldots \}

the argument of the first \use:n is not completely read by the second \use:n, hence
must remain in memory; then the argument of the second \use:n is not completely read
when grabbing the argument of the third \use:n, hence must remain in memory, and so
on. The memory consumption grows quadratically with the number of nested \use:n. In
practice, this means that we must read everything until a trailing `\_sort_stop` once in a while, otherwise sorting lists of more than a few thousand items would exhaust a typical \TeX{}’s memory.

The code within the `\exp_not:f` sorts the list, leaving in most cases a leading `\exp_not:n`, which stops the expansion, letting the result be return within `\exp_not:n`. We filter out the case of a list with no item, which would otherwise cause problems. Then prepare the token list #1 by inserting the conditional #2 before each item. The `\prepare` auxiliary receives the conditional as #1, the prepared token list so far as #2, and the item after that as #4. The loop ends when #4 contains `\prg_break_point:`, then the `\prepare_end` auxiliary finds the prepared token list as #4.

The scene is then set up for `\_sort_quick_split:NnNn`, which sorts the prepared list and perform the post action placed after `\_sort_mark`, namely removing the trailing `\_sort_stop` and leaving `\exp_stop_f:` to stop \-expansion.

The only_i, only_ii, split_i and split_ii auxiliaries receive a useless first argument, the new item #2 (that they append to either one of the next two arguments), the list #3 of items less than the pivot, bigger items #4, the pivot #5, a ⟨function⟩ #6, and an item #7. The ⟨function⟩ is the user’s ⟨conditional⟩ except at the end of the list where it is `\_sort_quick_end:nNTFNN` . The comparison is applied to the ⟨pivot⟩ and the ⟨item⟩, and calls the only_i or split_i auxiliaries if the ⟨item⟩ is smaller, and the only_ii or split_ii auxiliaries otherwise. In both cases, the next auxiliary goes to work right away, with no intermediate expansion that would slow down operations. Note that the argument #2 left for the next call has the form ⟨conditional⟩ {⟨item⟩}, so that the lists #3 and #4 keep the right form to be fed to the next sorting function. The split auxiliary
differs from these in that it is missing three of the arguments, which would be empty, and its first argument is always the user’s (conditional) rather than an ending function.

\begin{verbatim}
\cs_new:Npn \__sort_quick_split:NnNn #1#2#3#4
\{ #3 {#2} {#4} \__sort_quick_only_ii:NnnnnNn \__sort_quick_only_i:NnnnnNn
\__sort_quick_single_end:nw
\{ #3 {#4} } { } { } {#2}
\}
\cs_new:Npn \__sort_quick_only_i:NnnnnNn #1#2#3#4#5#6#7
\{ #6 {#5} {#7} \__sort_quick_split_ii:NnnnnNn \__sort_quick_split_i:NnnnnNn
\__sort_quick_split_only_i_end:nw
\{ #6 {#7} } { } {#4 #2} {#5}
\}
\cs_new:Npn \__sort_quick_only_ii:NnnnnNn #1#2#3#4#5#6#7
\{ #6 {#5} {#7} \__sort_quick_split_ii:NnnnnNn \__sort_quick_split_i:NnnnnNn
\__sort_quick_split_only_ii_end:nw
\{ #6 {#7} } { #3 #2 } { } {#5}
\}
\cs_new:Npn \__sort_quick_split_ii:NnnnnNn #1#2#3#4#5#6#7
\{ #6 {#5} {#7} \__sort_quick_split_ii:NnnnnNn \__sort_quick_split_i:NnnnnNn
\__sort_quick_split_split_end:nw
\{ #6 {#7} } { } { #4 #2 } {#5}
\}
\cs_new:Npn \__sort_quick_split_i:NnnnnNn #1#2#3#4#5#6#7
\{ #6 {#5} {#7} \__sort_quick_split_i:NnnnnNn \__sort_quick_split_i:NnnnnNn
\__sort_quick_single_end:nw
\{ #6 {#7} } {#3} {#4 #2} {#5}
\}
\end{verbatim}

(End definition for \__sort_quick_split:NnNn and others.)

\begin{verbatim}
\__sort_end:nwTFNn \__sort_single_end:nw \__sort_only_i_end:nw \__sort_only_ii_end:nw
\__sort_split_end:nw
\end{verbatim}

The \__sort_end:nw appears instead of the user’s conditional, and receives as its arguments the pivot {#1}, a fake item {#2}, a true and a false branches {#3} and {#4}, followed by an ending function {#5} (one of the four auxiliaries here) and another copy {#6} of the fake item. All those are discarded except the function {#5}. This function receives lists {#1} and {#2} of items less than or greater than the pivot {#3}, then a continuation code {#5} just after \s__sort_mark. To avoid a memory problem described earlier, all of the ending functions read {#6} until \s__sort_stop and place {#6} back into the input stream. When the lists {#1} and {#2} are empty, the single auxiliary simply places the continuation {#5} before the pivot {#3}. When {#2} is empty, {#1} is sorted and placed before the pivot {#3}, taking care to feed the continuation {#5} as a continuation for the function sorting {#1}. When {#1} is empty, {#2} is sorted, and the continuation argument is used to place the continuation {#5} and the pivot {#3} before the sorted result. Finally, when both
lists are non-empty, items larger than the pivot are sorted, then items less than the pivot,
and the continuations are done in such a way to place the pivot in between.

\cs_new:Npn \__sort_quick_end:nnTFNn #1#2#3#4#5#6 \s__sort_stop
\{ #5 \{#3 \} #6 \s__sort_stop \}
\cs_new:Npn \__sort_quick_split:NnNn #1 #2 \s__sort_stop
\{ \s__sort_split:NnNn #1 \s__sort_end:nnTFNn \s__sort_stop \}
\cs_new:Npn \__sort_quick_split_end:nnnwnw #1#2#3#4 \s__sort_mark #5#6 \s__sort_stop
\{ \s__sort_split:NnNn #2 \s__sort_split:NnNn #1 \s__sort_end:nnTFNn \s__sort_mark \s__sort_stop \}

(End definition for \__sort_quick_end:nnTFNn \__sort_split:NnNn, and others.)

43.6 Messages

\__sort_error: Bailing out of the sorting code is a bit tricky. It may not be safe to use a delimited
argument, so instead we redefine many l3sort commands to be trivial, with \__sort_level:
jumping to the break point. This error recovery won’t work in a group.

\cs_new_protected:Npn \__sort_error:
\{ \cs_set_eq:NN \__sort_merge_blocks_aux: \prg_do_nothing:
\cs_set_eq:NN \__sort_merge_blocks: \prg_do_nothing:
\cs_set_protected:Npn \__sort_level: \{ \group_end: \prg_break: \}
\}

(End definition for \__sort_error:)
\msg_error:nnx { \text{ sort } } { \text{ toksdef } }
\token_to_str: \#1
\__sort_error:
\\text{toksdef:D } \#1
\msg_new:nnnn { \text{ sort } } { \text{ toksdef } }
\{ \text{ Allocation-of-}\text{iow_char:N}\text{\toks-registers-impossible-while-sorting.} \}
\\text{The-comparison-code-used-for-sorting-a-list-has-attempted-to-}
\\text{define-}\#1\text{-as-a-new}\text{-}\text{iow_char:N}\text{\toks-register-using-}
\\text{iow_char:N}\text{\newtoks-}
\\text{or-a-similar-command.-The-list-will-not-be-sorted.} \}
(End definition for \text{\_sort_disable_toksdef:} and \text{\_sort_disabled_toksdef:n}.)
\__sort_too_long_error:NNw
When there are too many items in a sequence, this is an error, and we clean up properly
the mapping over items in the list: break using the type-specific breaking function \#1.
\cs_new_protected:Npn \__sort_too_long_error:NNw #1#2 \fi:
\fi:
\msg_error:nnxxx { \text{ sort } } { \text{ too-large } }
\{ \text{ \token_to_str:N } \#2 \}
\{ \text{ \int_eval:n { \l__sort_true_max_int - \l__sort_min_int } } \}
\{ \text{ \int_eval:n { \l__sort_top_int - \l__sort_min_int } } \}
\#1 \\text{\__sort_error:}
\msg_new:nnnn { \text{ sort } } { \text{ too-large } }
\{ \text{ The-list-}\#1\text{-is-too-long-to-be-sorted-by-\text{\TeX{}}.} \}
\{ \text{ \TeX{}-has-}\#2\text{-toks-registers-still-available:-}
\text{this-only-allows-to-sort-with-up-to-}\#3\text{-items.-The-list-will-not-be-sorted.} \}
(End definition for \text{\_sort_too_long_error:NNw}.)
\msg_new:nnnn { \text{ sort } } { \text{ return-none } }
\{ \text{ The-comparison-code-did-not-return. } \}
\{ \text{ When-sorting-a-list,-the-code-to-compare-items-}\#1\text{-and-}\#2\text{-did-not-call-}
\text{iow_char:N}\text{\sort_return_same: -nor-}
\text{iow_char:N}\text{\sort_return_swapped: -}.
\text{Exactly-one-of-these-should-be-called.} \}
\msg_new:nnnn { \text{ sort } } { \text{ return-two } }
\{ \text{ The-comparison-code-returned-multiple-times. } \}
\{ \text{ When-sorting-a-list,-the-code-to-compare-items-}\#1\text{-and-}\#2\text{-called-}
\text{iow_char:N}\text{\sort_return_same: -or-}
\text{iow_char:N}\text{\sort_return_swapped: -multiple-times.-}
\text{Exactly-one-of-these-should-be-called.} \}
\prop_gput:Nnn \g_msg_module_name_prop { \text{ sort } } { \text{ LaTeX3 } }
\prop_gput:Nnn \g_msg_module_type_prop { \text{ sort } } { }
3739 ⟨/package⟩
Chapter 44

\texttt{l3tl-analysis implementation}

\section*{44.1 Internal functions}
\\texttt{\textbackslash s\_\_tl}

The format used to store token lists internally uses the scan mark \texttt{\textbackslash s\_\_tl} as a delimiter.
\hfill (End definition for $\texttt{\textbackslash s\_\_tl}$.)

\section*{44.2 Internal format}

The task of the \texttt{l3tl-analysis} module is to convert token lists to an internal format which allows us to extract all the relevant information about individual tokens (category code, character code), as well as reconstruct the token list quickly. This internal format is used in \texttt{l3regex} where we need to support arbitrary tokens, and it is used in conversion functions in \texttt{l3str-convert}, where we wish to support clusters of characters instead of single tokens.

We thus need a way to encode any \texttt{token} (even begin-group and end-group character tokens) in a way amenable to manipulating tokens individually. The best we can do is to find \texttt{tokens} which both $\texttt{o}$-expand and $\texttt{x}$-expand to the given \texttt{token}. Collecting more information about the category code and character code is also useful for regular expressions, since most regexes are catcode-agnostic. The internal format thus takes the form of a succession of items of the form

\begin{align*}
\texttt{tokens} \ \texttt{\_\_tl} \ \texttt{catcode} \ \texttt{char code} \ \texttt{\_\_tl}
\end{align*}

The \texttt{tokens} $\texttt{o}$- and $\texttt{x}$-expand to the original token in the token list or to the cluster of tokens corresponding to one Unicode character in the given encoding (for \texttt{l3str-convert}). The \texttt{catcode} is given as a single hexadecimal digit, 0 for control sequences. The \texttt{char code} is given as a decimal number, −1 for control sequences.

Using delimited arguments lets us build the \texttt{tokens} progressively when doing an encoding conversion in \texttt{l3str-convert}. On the other hand, the delimiter \texttt{\_\_tl} may not appear unbraced in \texttt{tokens}. This is not a problem because we are careful to wrap control sequences in braces (as an argument to \texttt{\textbackslash exp not:n}) when converting from a general token list to the internal format.

The current rule for converting a \texttt{token} to a balanced set of \texttt{tokens} which both $\texttt{o}$-expands and $\texttt{x}$-expands to it is the following.
• A control sequence \cs becomes \exp_not:n \{ \cs \} \s__tl 0 − 1 \s__tl.

• A begin-group character { becomes \exp_after:wN \{ \iffalse: \fi: \s__tl 1 \char code \} \s__tl.

• An end-group character } becomes \iffalse: \{ \fi: \s__tl 2 \char code \} \s__tl.

• A character with any other category code becomes \exp_not:n \{ \char \} \s__tl.

### 44.3 Variables and helper functions

\s__tl

The scan mark \s__tl is used as a delimiter in the internal format. This is more practical than using a quark, because we would then need to control expansion much more carefully: compare \int_value:w ‘#1 \s__tl with \int_value:w ‘#1 \exp_stop_f: \exp_not:N \q_mark to extract a character code followed by the delimiter in an x-expansion.

\s__tl

(End definition for \s__tl.)

\l__tl_analysis_token\l__tl_analysis_char_token\l__tl_analysis_next_token

The tokens in the token list are probed with the \TeX primitive \futurelet. We use \l__tl_analysis_token in that construction. In some cases, we convert the following token to a string before probing it: then the token variable used is \l__tl_analysis_char_token. When getting tokens from the input stream we may need to look two tokens ahead, for which we use \l__tl_analysis_next_token.

\l__tl_analysis_token

(End definition for \l__tl_analysis_token, \l__tl_analysis_char_token, and \l__tl_analysis_next_token.)

\l__tl_peek_code_tl

Holds some code to be run once the next token has been fully analysed in \peek_analysis_map_inline:n.

\l__tl_peek_code_tl

(End definition for \l__tl_peek_code_tl.)

\c__tl_peek_catcodes_tl

A token list containing the character number 32 (space) with all possible category codes except 1 and 2 (begin-group and end-group). Why 32? Because some \LuaTeX versions only allow creation of \texttt{10} tokens with this character code, and because even in other engines it is much easier to produce since \texttt{char generate:nn} refuses to produce spaces.

\c__tl_peek_catcodes_tl

(group_begin:

\char_set_active_eq:NN \scan_stop:
\tl_const:Nx \c__tl_peek_catcodes_tl
{
\char_generate:nn { 32 } { 3 } 3
\char_generate:nn { 32 } { 4 } 4
# \char_generate:nn { 32 } { 6 } 6

428
\l__tl_analysis_normal_int \textit{The number of normal ($N$-type argument) tokens since the last special token.}
\int_new:N \l__tl_analysis_normal_int
\textit{(End definition for \l__tl_analysis_normal_int.)}

\l__tl_analysis_index_int \textit{During the first pass, this is the index in the array being built. During the second pass, it is equal to the maximum index in the array from the first pass.}
\int_new:N \l__tl_analysis_index_int
\textit{(End definition for \l__tl_analysis_index_int.)}

\l__tl_analysis_nesting_int \textit{Nesting depth of explicit begin-group and end-group characters during the first pass. This lets us detect the end of the token list without a reserved end-marker.}
\int_new:N \l__tl_analysis_nesting_int
\textit{(End definition for \l__tl_analysis_nesting_int.)}

\l__tl_analysis_type_int \textit{When encountering special characters, we record their “type” in this integer.}
\int_new:N \l__tl_analysis_type_int
\textit{(End definition for \l__tl_analysis_type_int.)}

\g__tl_analysis_result_tl \textit{The result of the conversion is stored in this token list, with a succession of items of the form (tokens) $\langle$ catcode $\rangle$ $\langle$ char code $\rangle$ $\langle$ catcode $\rangle$.}
\tl_new:N \g__tl_analysis_result_tl
\textit{(End definition for \g__tl_analysis_result_tl.)}

\__tl_analysis_extract_charcode: \textit{Extracting the character code from the meaning of \l__tl_analysis_token. This has no error checking, and should only be assumed to work for begin-group and end-group character tokens. It produces a number in the form ‘$(char)$.}
\cs_new:Npn \__tl_analysis_extract_charcode: { \exp_after:wN \__tl_analysis_extract_charcode_aux:w \token_to_meaning:N \l__tl_analysis_token }
\cs_new:Npn \__tl_analysis_extract_charcode_aux:w #1 - #2 - { ‘ }
\textit{(End definition for \__tl_analysis_extract_charcode: and \__tl_analysis_extract_charcode_aux:w.)}
Counts the number of spaces in the string representation of its second argument, as well as the number of characters following the last space in that representation, and feeds the two numbers as semicolon-delimited arguments to the first argument. When this function is used, the escape character is printable and non-space.

```latex
\cs_new:Npn \_\_tl_analysis_cs_space_count:NN #1 #2
{\exp_after:wN #1 \int_value:w \int_eval:w 0 \exp_after:wN \_\_tl_analysis_cs_space_count:w \token_to_str:N #2 \fi: \_\_tl_analysis_cs_space_count_end:w ; - !}
```

(End definition for \_\_tl_analysis_cs_space_count:NN, \_\_tl_analysis_cs_space_count:w, and \_\_tl_analysis_cs_space_count_end:w)

### 44.4 Plan of attack

Our goal is to produce a token list of the form roughly

\[
\langle \text{token 1} \rangle \text{\_\_s\_tl} \langle \text{catcode 1} \rangle \text{\_\_s\_tl} \langle \text{char code 1} \rangle \text{\_\_s\_tl} \\
\langle \text{token 2} \rangle \text{\_\_s\_tl} \langle \text{catcode 2} \rangle \text{\_\_s\_tl} \langle \text{char code 2} \rangle \text{\_\_s\_tl} \\
\vdots \langle \text{token N} \rangle \text{\_\_s\_tl} \langle \text{catcode N} \rangle \text{\_\_s\_tl} \langle \text{char code N} \rangle \text{\_\_s\_tl}
\]

Most but not all tokens can be grabbed as an undelimited (N-type) argument by \TeX. The plan is to have a two pass system. In the first pass, locate special tokens, and store them in various \toks registers. In the second pass, which is done within an x-expanding assignment, normal tokens are taken in as N-type arguments, and special tokens are retrieved from the \toks registers, and removed from the input stream by some means. The whole process takes linear time, because we avoid building the result one item at a time.

We make the escape character printable (backslash, but this later oscillates between slash and backslash): this allows us to distinguish characters from control sequences.

A token has two characteristics: its \text{\textbackslash meaning}, and what it looks like for \TeX when it is in scanning mode (\textit{e.g.}, when capturing parameters for a macro). For our purposes, we distinguish the following meanings:

- begin-group token (category code 1), either space (character code 32), or non-space;
- end-group token (category code 2), either space (character code 32), or non-space;
- space token (category code 10, character code 32);
- anything else (then the token is always an N-type argument).

The token itself can “look like” one of the following
• a non-active character, in which case its meaning is automatically that associated to its character code and category code, we call it “true” character;
• an active character;
• a control sequence.

The only tokens which are not valid \texttt{N}-type arguments are true begin-group characters, true end-group characters, and true spaces. We detect those characters by scanning ahead with \texttt{futurelet}, then distinguishing true characters from control sequences set equal to them using the \texttt{string} representation.

The second pass is a simple exercise in expandable loops.

\begin{verbatim}
\_tl_analysis:n Everything is done within a group, and all definitions are local. We use \texttt{group_align_safe_begin/end} to avoid problems in case \_\_tl_analysis:n is used within an alignment and its argument contains alignment tab tokens.
3789 \cs_new_protected:Npn \_\_tl_analysis:n #1
3790 \{
3791 \group_begin:
3792 \group_align_safe_begin:
3793 \_\_tl_analysis_a:n {#1}
3794 \_\_tl_analysis_b:n {#1}
3795 \group_align_safe_end:
3796 \group_end:
3797 \}
(End definition for \_\_tl_analysis:n.)
\end{verbatim}

\subsection{Disabling active characters}

\begin{verbatim}
\_tl_analysis_disable:n Active characters can cause problems later on in the processing, so we provide a way to disable them, by setting them to \texttt{undefined}. Since Unicode contains too many characters to loop over all of them, we instead do this whenever we encounter a character. For \TeX{} and up\TeX{} we skip characters beyond \(0, 255\) because \texttt{lccode} only allows those values.
3798 \group_begin:
3799 \char_set_catcode_active:N \^^@ \\
3800 \cs_new_protected:Npn \_\_tl_analysis_disable:n #1
3801 \{
3802 \tex_lccode:D 0 = #1 \exp_stop_f:
3803 \tex_lowercase:D { \tex_let:D ^^@ } \tex_undefined:D
3804 \}
3805 \bool_lazy_or:nnT
3806 { \sys_if_engine_ptex_p: }
3807 { \sys_if_engine_uptex_p: }
3808 \{
3809 \cs_gset_protected:Npn \_\_tl_analysis_disable:n #1
3810 \{
3811 \if_int_compare:w 256 > #1 \exp_stop_f:
3812 \tex_lccode:D 0 = #1 \exp_stop_f:
3813 \tex_lowercase:D { \tex_let:D ^^@ } \tex_undefined:D
3814 \fi:
3815 \}
3816 \}
3817 \group_end:
\end{verbatim}

431
44.6 First pass

The goal of this pass is to detect special (non-N-type) tokens, and count how many N-type tokens lie between special tokens. Also, we wish to store some representation of each special token in a \toks register.

We have 11 types of tokens:

1. a true non-space begin-group character;
2. a true space begin-group character;
3. a true non-space end-group character;
4. a true space end-group character;
5. a true space blank space character;
6. an active character;
7. any other true character;
8. a control sequence equal to a begin-group token (category code 1);
9. a control sequence equal to an end-group token (category code 2);
10. a control sequence equal to a space token (character code 32, category code 10);
11. any other control sequence.

Our first tool is \futurelet. This cannot distinguish case 8 from 1 or 2, nor case 9 from 3 or 4, nor case 10 from case 5. Those cases are later distinguished by applying the \string primitive to the following token, after possibly changing the escape character to ensure that a control sequence’s string representation cannot be mistaken for the true character.

In cases 6, 7, and 11, the following token is a valid N-type argument, so we grab it and distinguish the case of a character from a control sequence: in the latter case, \str_tail:n \{ \langle token \rangle \} is non-empty, because the escape character is printable.

\_\_tl_analysis_a:n We read tokens one by one using \futurelet. While performing the loop, we keep track of the number of true begin-group characters minus the number of true end-group characters in \_\_tl_analysis_nesting_int. This reaches −1 when we read the closing brace.
\__tl_analysis_a_loop:w  Read one character and check its type.
\cs_new_protected:Npn \__tl_analysis_a_loop:w
\tex_futurelet:D \l__tl_analysis_token \__tl_analysis_a_type:w
\__tl_analysis_a_type:w  At this point, \l__tl_analysis_token holds the meaning of the following token. We store in \l__tl_analysis_type_int information about the meaning of the token ahead:
• 0 space token;
• 1 begin-group token;
• -1 end-group token;
• 2 other.

The values 0, 1, −1 correspond to how much a true such character changes the nesting level (2 is used only here, and is irrelevant later). Then call the auxiliary for each case. Note that nesting conditionals here is safe because we only skip over \l__tl_analysis_token if it matches with one of the character tokens (hence is not a primitive conditional).
\cs_new_protected:Npn \__tl_analysis_a_type:w
\l__tl_analysis_type_int =
\if_meaning:w \l__tl_analysis_token \c_space_token
0
\else:
\if_catcode:w \exp_not:N \l__tl_analysis_token \c_group_begin_token
1
\else:
\if_catcode:w \exp_not:N \l__tl_analysis_token \c_group_end_token
-1
\else:
2
\fi:
\fi:
\exp_stop_f:
\if_case:w \l__tl_analysis_type_int
\exp_after:wN \__tl_analysis_a_space:w
\or: \exp_after:wN \__tl_analysis_a_bgroup:w
\or: \exp_after:wN \__tl_analysis_a_safe:N
\else: \exp_after:wN \__tl_analysis_a_egroup:w
\fi:
}

\__tl_analysis_a_space:w  In this branch, the following token’s meaning is a blank space. Apply \string to that token: a true blank space gives a space, a control sequence gives a result starting with the escape character, an active character gives something else than a space since we disabled the space. We grab as \l__tl_analysis_char_token the first character of the string representation then test it in \__tl_analysis_a_space_test:w. Also, since
\_tl_analysis\_a\_store: expects the special token to be stored in the relevant \toks register, we do that. The extra \exp\_not\_n is unnecessary of course, but it makes the treatment of all tokens more homogeneous. If we discover that the next token was actually a control sequence or an active character instead of a true space, then we step the counter of normal tokens. We now have in front of us the whole string representation of the control sequence, including potential spaces; those will appear to be true spaces later in this pass. Hence, all other branches of the code in this first pass need to consider the string representation, so that the second pass does not need to test the meaning of tokens, only strings.

\cs_new_protected:Npn \__tl_analysis\_a\_space:w
\{\tex\_afterassignment:D \__tl_analysis\_a\_space\_test:w
\exp\_after:WN \cs\_set\_eq:NN
\exp\_after:WN \l__tl_analysis\_char\_token
\token\_to\_str:N
\}
\cs_new_protected:Npn \__tl_analysis\_a\_space\_test:w
\{\if\meaning:w \l__tl_analysis\_char\_token \c\_space\_token
\tex\_toks:D \l__tl_analysis\_index\_int \{ \exp\_not\_n \{ \} \}
\__tl_analysis\_a\_store:\n\else:
\int\_incr:N \l__tl_analysis\_normal\_int
\fi:
\__tl_analysis\_a\_loop:w
\}

(End definition for \__tl_analysis\_a\_space:w and \__tl_analysis\_a\_space\_test:w.)

The token is most likely a true character token with catcode 1 or 2, but it might be a control sequence, or an active character. Optimizing for the first case, we store in a toks register some code that expands to that token. Since we will turn what follows into a string, we make sure the escape character is different from the current character code (by switching between solidus and backslash). To detect the special case of an active character let to the catcode 1 or 2 character with the same character code, we disable the active character with that character code and re-test: if the following token has become undefined we can in fact safely grab it. We are finally ready to turn what follows to a string and test it. This is one place where we need \l__tl_analysis\_char\_token to be a separate control sequence from \l__\_tl_analysis\_char\_token, to compare them.

\group\_begin:
\char\_set\_catcode\_group\_begin:N \"\^\@ \%
{\cs\_new\_protected:Npn \\_\_tl_analysis\_a\_bgroup:w
\{\\_\_tl_analysis\_a\_group:nw \{ \exp\_after:WN \^^@ \if\_false: \} \fi: \}}
\char\_set\_catcode\_group\_end:N \"\^\@\n\cs\_new\_protected:Npn \\_\_tl_analysis\_a\_egroup:w
{\\_\_tl_analysis\_a\_group:nw \{ \if\_false: \{ \fi: \^^@ \} \} \%
\group\_end:
\cs\_new\_protected:Npn \\_\_tl_analysis\_a\_group\_nw \#1
{\tex\_lccode:D 0 = \\_\_tl_analysis\_extract\_char\_code:\ \scan\_stop:
\tex\_lowercase:D \{ \tex\_toks:D \l__tl_analysis\_index\_int \{\#1 \} \}
\if\_int\_compare:w \tex\_lccode:D 0 = \tex\_escape\_char:D
\int\_set:Nn \tex\_escape\_char:D \{ 139 - \tex\_escape\_char:D \}
\_tl_analysis_disable:n \{ \tex_lccode:D 0 \}
\tex_futurelet:D \_\_tl_analysis_token \_\_tl_analysis_a_group_aux:w
}
\cs_new_protected:Npn \_\_tl_analysis_a_group_aux:w
{
  \if_meaning:w \l__tl_analysis_token \tex_undefined:D
    \exp_after:wN \_\_tl_analysis_a_safe:N
  \else:
    \exp_after:wN \_\_tl_analysis_a_group_auxii:w
  \fi:
}
\cs_new_protected:Npn \_\_tl_analysis_a_group_auxii:w
{
  \tex_afterassignment:D \_\_tl_analysis_a_group_test:w
  \exp_after:wN \cs_set_eq:NN \exp_after:wN \l__tl_analysis_char_token
  \token_to_str:N
}
\cs_new_protected:Npn \_\_tl_analysis_a_group_test:w
{
  \if_charcode:w \l__tl_analysis_token \l__tl_analysis_char_token
    \_\_tl_analysis_a_store:
  \else:
    \int_incr:N \l\_tl_analysis_normal_int
    \_\_tl_analysis_a_loop:w
  \fi:
}

(End definition for \_\_tl_analysis_a_bgroup:w and others.)

\_\_tl_analysis_a_store: This function is called each time we meet a special token; at this point, the \toks register \_\_\_tl_analysis_index_int holds a token list which expands to the given special token. Also, the value of \_\_\_tl_analysis_type_int indicates which case we are in:

- -1 end-group character;
- 0 space character;
- 1 begin-group character.

We need to distinguish further the case of a space character (code 32) from other character codes, because those behave differently in the second pass. Namely, after testing the \lccode of 0 (which holds the present character code) we change the cases above to

- -2 space end-group character;
- -1 non-space end-group character;
- 0 space blank space character;
- 1 non-space begin-group character;
- 2 space begin-group character.
This has the property that non-space characters correspond to odd values of \l__tl_analysis_type_int. The number of normal tokens until here and the type of special token are packed into a \skip register. Finally, we check whether we reached the last closing brace, in which case we stop by disabling the looping function (locally).

```
cs_new_protected:Npn \__tl_analysis_a_store:  
  {  
    \tex_advance:D \l__tl_analysis_nesting_int \l__tl_analysis_type_int  
    \if_int_compare:w \tex_lccode:D 0 = '\exp_stop_f:  
      \tex_advance:D \l__tl_analysis_type_int \l__tl_analysis_type_int  
    \fi:  
    \tex_skip:D \l__tl_analysis_index_int  
    = \l__tl_analysis_normal_int sp  
    plus \l__tl_analysis_type_int sp \scan_stop:  
    \int_incr:N \l__tl_analysis_index_int  
    \int_zero:N \l__tl_analysis_normal_int  
    \if_int_compare:w \l__tl_analysis_nesting_int = - \c_one_int  
      \cs_set_eq:NN \__tl_analysis_a_loop:w \scan_stop:  
    \fi:  
  }
```

(End definition for \__tl_analysis_a_store:)

This should be the simplest case: since the upcoming token is safe, we can simply grab it in a second pass. If the token is a single character (including space), the \if_charcode:w test yields true; we disable a potentially active character (that could otherwise masquerade as the true character in the next pass) and we count one “normal” token. On the other hand, if the token is a control sequence, we should replace it by its string representation for compatibility with other code branches. Instead of slowly looping through the characters with the main code, we use the knowledge of how the second pass works: if the control sequence name contains no space, count that token as a number of normal tokens equal to its string length. If the control sequence contains spaces, they should be registered as special characters by increasing \l__tl_analysis_index_int (no need to carefully count character between each space), and all characters after the last space should be counted in the following sequence of “normal” tokens.

```
cs_new_protected:Npn \__tl_analysis_a_safe:N \__tl_analysis_a_cs:ww  
  {  
    \if_charcode:w \scan_stop:  
      \exp_after:wN \use_none:n \token_to_str:N #1 \prg_do_nothing:  
      \scan_stop:  
      \exp_after:wN \use_i:nn  
    \else:  
      \exp_after:wN \use_ii:nn  
    \fi:  
    {  
      \__tl_analysis_disable:n { '#1 }  
      \int_incr:N \l__tl_analysis_normal_int  
    }  
    \__tl_analysis_cs_space_count:NN \__tl_analysis_a_cs:ww #1  
  }  
```

\__tl_analysis_a_loop:w

436
44.7 Second pass

The second pass is an exercise in expandable loops. All the necessary information is stored in \skip and \toks registers.

\__tl_analysis_b:n
\__tl_analysis_b_loop:w

Start the loop with the index 0. No need for an end-marker: the loop stops by itself when the last index is read. We repeatedly oscillate between reading long stretches of normal tokens, and reading special tokens.

\__tl_analysis_b_normals:ww
\__tl_analysis_b_normal:wwN

The first argument is the number of normal tokens which remain to be read, and the second argument is the index in the array produced in the first step. A character's string representation is always one character long, while a control sequence is always longer (we have set the escape character to a printable value). In both cases, we leave \exp_not:n \{⟨token⟩\} \s__tl in the input stream (after x-expansion). Here, \exp_not:n is used rather than \exp_not:N because #3 could be a macro parameter character or could be \s__tl (which must be hidden behind braces in the result).
\if_charcode:w \
\scan_stop: \
\exp_after:wN \use_none:n \token_to_str:N #3 \prg_do_nothing: \
\scan_stop: \
\exp_after:wN \_t_analysis_b_char:Nww \
\else: \
\exp_after:wN \_t_analysis_b_char:Nww \
\fi: 
#3 #1; #2; 
}

(End definition for \_t_analysis_b_normals:ww and \_t_analysis_b_normal:ww.)

\_t_analysis_b_char:Nww

If the normal token we grab is a character, leave \{ \langle \text{catcode} \rangle \langle \text{charcode} \rangle \} followed by \s__tl in the input stream, and call \_t_analysis_b_normals:ww with its first argument decremented.

\cs_new:Npx \_t_analysis_b_char:Nww #1
{ 
\exp_not:N \if_meaning:w #1 \exp_not:N \tex_undefined:D 
\token_to_str:N D \exp_not:N \else: 
\exp_not:N \if_catcode:w #1 \c_catcode_other_token 
\token_to_str:N C \exp_not:N \else: 
\exp_not:N \if_catcode:w #1 \c_catcode_letter_token 
\token_to_str:N B \exp_not:N \else: 
\exp_not:N \if_catcode:w #1 \c_math_toggle_token 
\token_to_str:N A \exp_not:N \else: 
\exp_not:n { \fi: \fi: \fi: \fi: \fi: \fi: \fi: \fi: } 
\exp_after:wN \exp_not:N \_t_analysis_b_normals:ww 
\exp_not:N \int_value:w \exp_not:N \int_eval:w - 1 + 
}

(End definition for \_t_analysis_b_char:Nww.)

\_t_analysis_b_cs:Nww
\_t_analysis_b_cs_test:ww

If the token we grab is a control sequence, leave \{ \langle \text{catcode} \rangle \langle \text{charcode} \rangle \} followed by \s__tl in the input stream, and call \_t_analysis_b_normals:ww with updated arguments.

\cs_new:Npx \_t_analysis_b_cs:Nww \_t_analysis_b_cs_test:ww #1
{ 
\exp:N \if_meaning:w #1 \exp:N \tex_undefined:D 
\ifcatcode:w #1 \c_catcode_other_token \exp_not:N \else: 
\ifcatcode:w #1 \c_catcode_letter_token \exp_not:N \else: 
\ifcatcode:w #1 \c_math_toggle_token \exp_not:N \else: 
\ifcatcode:w #1 \c_math_superscript_token \exp_not:N \else: 
\ifcatcode:w #1 \c_math_subscript_token \exp_not:N \else: 
\ifcatcode:w #1 \c_space_token \exp_not:N \else: 
\exp_not:n { \fi: \fi: \fi: \fi: \fi: \fi: \fi: \fi: } 
\exp_after:wN \exp_not:N \_t_analysis_b_normals:ww 
\exp_not:N \int_value:w \exp_not:N \int_eval:w - 1 + 
}

\cs_new:Npx \_t_analysis_b_cs_test:ww #1 ; #2 ; #3 ; #4 ;
{ 
\exp_after:wN \_t_analysis_b_normals:ww 

438
Here, \#1 is the current index in the array built in the first pass. Check now whether we reached the end (we shouldn’t keep the trailing end-group character that marked the end of the token list in the first pass). Unpack the \toks register: when x-expanding again, we will get the special token. Then leave the category code in the input stream, followed by the character code, and call \_\_tl_analysis_b_loop:w with the next index.

\_\_tl_analysis_b_special:w
\_\_tl_analysis_b_special_char:w
\_\_tl_analysis_b_special_space:w
44.8 Mapping through the analysis

First obtain the analysis of the token list into \texttt{\g__tl_analysis_result_tl}. To allow
nested mappings, increase the nesting depth \texttt{\g__kernel_prg_map_int} (shared between
all modules), then define the payload macro, which runs the user code and has a name
specific to that nesting depth. The looping macro grabs the \texttt{\langle\text{tokens}\rangle}, \texttt{\langle\text{catcode}\rangle}
and \texttt{\langle\text{char code}\rangle}; it checks for the end of the loop with \texttt{\use_none:n \#2}, normally empty, but which
becomes \texttt{\tl_map_break:} at the end; it then calls the payload macro with the arguments
in the correct order (this is the reason why we cannot directly use the same macro for
looping and payload), and loops by calling itself. When the loop ends, remember to
decrease the nesting depth.

```latex
\cs_new_protected:Npn \tl_analysis_map_inline:Nn #1 #2 \s__tl #3 #4 \s__tl
\{ \use_none:n #3
\#1 \{#2\} \#4 \{#3\}
\__tl_analysis_map:NwNw #1 \}
\cs_new_protected:Npn \__tl_analysis_map:NwNw #1 #2 \s__tl #3 #4 \s__tl
\{ \__tl_analysis_map:Nn \_\_\_tl_analysis_map_inline_ \int_use:N \g__kernel_prg_map_int :wNw \}
```

(End definition for \texttt{\_\_\_tl_analysis_map_inline:Nn} and others. These functions are documented on page 45.)

44.9 Showing the results

Add to \texttt{\_\_\_tl_analysis:n} a third pass to display tokens to the terminal. If the token
list variable is not defined, throw the same error as \texttt{\tl_show:N} by simply calling that
function.

```latex
\cs_new_protected:Npn \__tl_analysis_log:N \_\_\_tl_analysis_log:NW
\cs_new_protected:Npn \__tl_analysis_show:N \_\_\_tl_analysis_show:N
\cs_new_protected:Npn \_\_\_tl_analysis_show:NW
\cs_new_protected:Npn \\_\_\_tl_analysis_show:NW
```

(End definition for \texttt{\_\_\_tl_analysis_map_inline:Nn} and others. These functions are documented on page 45.)
\cs_new_protected:Npn \__tl_analysis_show:NNN #1#2#3
\{
  \tl_if_exist:NTF #3
  \{
    \exp_args:No \__tl_analysis:n {#3}
    #1 \{ tl \} \{ show-analysis \}
    \{ \token_to_str:N #3 \} \{ \__tl_analysis_show: \} \{ \} \{ \}
  \}
  { \#2 \#3 }
\}
\}(End definition for \tl_analysis_show:N, \tl_analysis_log:N, and \__tl_analysis_show:NNN. These functions are documented on page 45.)

No existence test needed here.

\cs_new_protected:Npn \tl_analysis_show:n
{ \__tl_analysis_show:Nn \msg_show:nnxxxx }
\cs_new_protected:Npn \tl_analysis_log:n
{ \__tl_analysis_show:Nn \msg_log:nnxxxx }
\cs_new_protected:Npn \__tl_analysis_show:Nn #1#2
{ \__tl_analysis:n {#2}
  #1 \{ tl \} \{ show-analysis \} \{ \} \{ \} \{ \__tl_analysis_show: \} \{ \} \{ \}
}
\cs_new:Npn \__tl_analysis_show:
{ \__tl_analysis_show_loop:wNw \g__tl_analysis_result_tl \s__tl ? \prg_break: \} \s__tl \prg_break_point:
\}
\cs_new:Npn \__tl_analysis_show_loop:wNw #1 \s__tl #2 #3 \s__tl
{\use_none:n \#2
  \iow_newline: > \use:n { - } { - }
  \if_int_compare:w "\#2 = \c_zero_int \exp_after:wN \__tl_analysis_show_cs:n
  \else:
    \if_int_compare:w "\#2 = 13 \exp_stop_f:
      \exp_after:wN \__tl_analysis_show_active:n
    \else:
      \exp_after:wN \__tl_analysis_show_normal:n
    \fi:
    \fi:
  \fi:
}\use:n { #1}
\__tl_analysis_show_loop:wNw

(End definition for \tl_analysis_show:n, \tl_analysis_log:n, and \__tl_analysis_show:NNn. These functions are documented on page 45.)

Here, \#1 o- and x-expands to the token; \#2 is the category code (one uppercase hexadecimal digit), 0 for control sequences; \#3 is the character code, which we ignore. In the cases of control sequences and active characters, the meaning may overflow one line, and we want to truncate it. Those cases are thus separated out.

\cs_new:Npn \__tl_analysis_show:
{ \__tl_analysis_show_loop:wNw \g__tl_analysis_result_tl \s__tl { ? \prg_break: \} \s__tl \prg_break_point:
}
\cs_new:Npn \__tl_analysis_show_loop:wNw #1 \s__tl #2 \#3 \s__tl
{\use_none:n \#2
  \iow_newline: > \use:n { - } { - }
  \if_int_compare:w "\#2 = \c_zero_int \exp_after:wN \__tl_analysis_show_cs:n
  \else:
    \if_int_compare:w "\#2 = 13 \exp_stop_f:
      \exp_after:wN \__tl_analysis_show_active:n
    \else:
      \exp_after:wN \__tl_analysis_show_normal:n
    \fi:
    \fi:
}\use:n { #1}
\__tl_analysis_show_loop:wNw

441
Non-active characters are a simple matter of printing the character, and its meaning.
Our test suite checks that begin-group and end-group characters do not mess up \TeX{}'s alignment status.

\begin{verbatim}
\cs_new:Npn \tl_analysis_show_normal:n #1 #1 \token_to_str:N #1 \exp_after:wN \token_to_meaning:N #1 }
\end{verbatim}

This expands to the value of \texttt{#1} if it has any.

\begin{verbatim}
\cs_new:Npn \tl_analysis_show_value:N #1 \token_if_expandable:NF #1 { \token_if_chardef:NTF #1 \prg_break: { } \token_if_mathchardef:NTF #1 \prg_break: { } \token_if_dim_register:NTF #1 \prg_break: { } \token_if_int_register:NTF #1 \prg_break: { } \token_if_skip_register:NTF #1 \prg_break: { } \token_if_toks_register:NTF #1 \prg_break: { } \use_none:nnn \prg_break_point: \use:n { \exp_after:wN = \tex_the:D #1 } }
\end{verbatim}

Control sequences and active characters are printed in the same way, making sure not to go beyond the \texttt{l_iow_line_count_int}. In case of an overflow, we replace the last characters by \texttt{\__tl_analysis_show_etc_str}.

\begin{verbatim}
\cs_new:Npn \tl_analysis_show_cs:n \exp_after:wN \token_to_str:N \token_to_meaning:N \__tl_analysis_show_value:N #1 }
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \tl_analysis_show_active:n \exp_args:No \__tl_analysis_show_long:nn {control-sequence=} \__tl_analysis_show_etc_str
\cs_new:Npn \tl_analysis_show_long:nn #1 #2 #3 \__tl_analysis_show_etc_str
\int_compare:nNnTF { \str_count:n { \token_to_str:N \token_to_meaning:N #1 } } > { \l_iow_line_count_int - 3 } { \__tl_analysis_show_long_aux:nnnn #1#2#3#4 \use:n #1 \__tl_analysis_show_etc_str #1
\use:n { \exp_after:wN = \tex_the:D #1 } }
\end{verbatim}
44.10 Peeking ahead

The break statements use the general \prg_map_break:Nn.

\peek_analysis_map_break:
\peek_analysis_map_break:n

\l__tl_peek_charcode_int
\int_new:N \l__tl_peek_charcode_int
(End definition for \l__tl_peek_charcode_int.)

\__tl_analysis_char_arg:Nw
\__tl_analysis_char_arg_aux:Nw
After a call to \futurelet \l__tl_analysis_token followed by a stringified character token (either explicit space or catcode other character), grab the argument and pass it to \#1. We only need to do anything in the case of a space.

\peek_analysis_map_inline:n
\peek_analysis_map_inline:n
Save the user’s code in a control sequence that is suitable for nested maps. We may wish to pass to this function an \outer control sequence or active character; for this we will undefine potentially-\outer tokens within a group, closed after the function receives its arguments. This user’s code function also calls the loop auxiliary, and includes the trailing \prg_break_point:Nn for when the user wants to stop the loop. The loop auxiliary must remove that break point because it must look at the input stream.

\peek_analysis_map_inline:n
\peek_analysis_map_inline:n

443
The loop starts a group (closed by the user-code function defined above) with a normalized escape character, and checks if the next token is special or N-type.

Normal tokens are not too hard, but can be \outer, hence the \exp_not:N in the code above. If the token is expandable then it might be an \outer or a \TeX conditional, so to be safe we set it to \scan_stop: (the assignment is local and stopped by the \group_end: upon calling the user’s code). Then distinguish characters (including active ones and macro parameter characters) from control sequences (whose string representation is more than one character because the escape character is printable). For a control sequence call the user code with suitable arguments.
For special characters the idea is to eventually act with `\token_to_str:N`, then pick up one by one the characters of this string representation until hitting the token that follows. First determine the character code of (the meaning of) the ⟨token⟩ (which we know is a special token), make sure the escape character is different from it, normalize the meanings of two active characters and the empty control sequence, and filter out these cases in `\__tl_peek_analysis_retest`.

At this point we know the meaning of the ⟨token⟩ in the input stream is `\_\_tl_peek_analysis_special`, either a space (32, 10) or a begin-group or end-group token (catcode 1 or 2),
and we excluded a few cases that would be difficult later (empty control sequence, active character with the same character code as its meaning or as the escape character). Now look at the (next token) following it using a combination of \afterassignment and \futurelet. The syntax of this primitive is \futurelet \langle peek token ⟩ \langle first token ⟩ \langle next token ⟩, and it sets \langle peek token ⟩ equal to \langle next token ⟩. Traditionally, one takes \langle first token ⟩ to be some macro that regains control of the code and, e.g., analyses \langle peek token ⟩.

Here, both \langle first token ⟩ and \langle next token ⟩ are mostly unknown tokens in the input stream (but we know the \langle first token ⟩ has catcode 1, 2 or 10), where \langle first token ⟩ was already stored as \l__tl_peak_token, and we regain control using \afterassignment, which inserts its argument after the assignment, hence after \langle peek token ⟩ but before \langle first token ⟩.

We then hit the \langle first token ⟩ with \token_to_str:N and grab characters until finding \l__tl_analysis_next_token. More precisely, by looking at the first character in the string representation of the \langle first token ⟩ we distinguish three cases: a stringified control sequence starts with the escape character; for an explicit character we find that same character; for an explicit character we find anything else (we made sure to exclude the case of an active character whose string representation coincides with the other two cases).

When \#1 is a stringified active character we pass appropriate arguments to the user’s code; thankfully \char_generate:nn can make active characters.
When \#1 matches the character we had extracted from the meaning of \l\_peek_token, the token was an explicit character, which can be a standard space, or a begin-group or end-group character with some character code. In the latter two cases we call \char\_generate:nn with suitable arguments and put suitable \if\_false: \fi: constructions to make the result balanced and such that \alpha\-expanding or \kappa\-expanding gives back a single (unbalanced) begin-group or end-group character.

\cs\_new\_protected:Npn \__tl\_peek\_analysis\_explicit:n \#1
\{\tl\_put\_right:Nx \l\_tl\_peek\_code\_tl
\{\if\_meaning:w \l\_peek\_token \c\_space\_token
\{ - \} \{ 32 \} \token\_to\_str:N A
\else:\n\if\_catcode:w \l\_peek\_token \c\_group\_begin\_token
\{ \exp\_not:N \exp\_after:wN
\char\_generate:nn \{ \#1 \} \{ 1 \}
\exp\_not:N \if\_false:
\if\_false: \{ \fi: \}
\exp\_not:N \fi:
\}
\{ \int\_value:w \#1 \}
1
\else:
\{\exp\_not:N \if\_false:
\if\_false: \} \fi:
\exp\_not:N \fi:
\char\_generate:nn \{ \#1 \} \{ 2 \}
\}
\{ \int\_value:w \#1 \}
2
\fi:
\fi:
\}
\l\_tl\_peek\_code\_tl
\}

Finally there is the case of a special token whose string representation starts with an escape character, namely the token was a control sequence. In that case we could have grabbed the token directly as an \N\-type argument, but of course we couldn’t know that until we had run all the various tests including stringifying the token. We are thus left with the hard work of picking up one by one the characters in the csname (being careful about spaces), until finding a token that matches the \langle next token \rangle picked up earlier (which was not stringified), such that the control sequence that we found so far indeed has the expected meaning \l\_peek\_token. This comparison with \l\_peek\_token catches a reasonably common case like \c\_group\_begin\_token _ in which the trailing _ has category code other: without comparison of the constructed csname with \l\_peek\_token collection would stop at \c, which is wrong.
\cs_new_protected:Npn \__tl_peek_analysis_escape: 
\{ 
\tl_clear:N \l__tl_internal_a_tl 
\tex_futurelet:D \l__tl_analysis_token 
\__tl_peek_analysis_collect:w 
\} 
\cs_new_protected:Npn \__tl_peek_analysis_collect:w 
\{ \__tl_analysis_char_arg:Nw \__tl_peek_analysis_collect:n \} 
\cs_new_protected:Npn \__tl_peek_analysis_collect:n #1 
\{ 
\tl_put_right:Nn \l__tl_internal_a_tl {#1} 
\__tl_peek_analysis_collect_loop: 
\} 
\cs_new_protected:Npn \__tl_peek_analysis_collect_loop: 
\{ 
\tex_futurelet:D \l__tl_analysis_token 
\__tl_peek_analysis_collect_test: 
\} 
\cs_new_protected:Npn \__tl_peek_analysis_collect_test: 
\{ 
\if_meaning:w \l__tl_analysis_token \l__tl_analysis_next_token 
\exp_after:wN \if_meaning:w \cs:w \l__tl_internal_a_tl \cs_end: \l_peek_token 
\__tl_peek_analysis_collect_end:NNN 
\fi: 
\fi: 
\__tl_peek_analysis_collect:w 
\} 
End by calling the user code with suitable arguments (here \#1, \#2 are \fi:), which closes the group begun early on.
\cs_new_protected:Npn \__tl_peek_analysis_collect_end:NNN #1#2#3 
\{ 
\l__tl_put_right:Nx \l__tl_analysis_token \l__tl_analysis_next_token 
\exp_after:wN \if_meaning:w \cs:w \l__tl_internal_a_tl \cs_end: \l_peek_token 
\__tl_peek_analysis_collect_end:NNN 
\fi: 
\fi: 
\__tl_peek_analysis_collect:w 
\} 
(End definition for \peek_analysis_map_inline:n and others. This function is documented on page 194.)

\c__tl_analysis_show_etc_str
When a control sequence (or active character) and its meaning are too long to fit in one line of the terminal, the end is replaced by this token list.
\tl_const:Nx \c__tl_analysis_show_etc_str \{ \token_to_str:N \ETC. \} 
(End definition for \c__tl_analysis_show_etc_str.)

\section{Messages}

\c__tl_analysis_show_etc_str
When a control sequence (or active character) and its meaning are too long to fit in one line of the terminal, the end is replaced by this token list.
\tl_const:Nx \c__tl_analysis_show_etc_str \% ( 
\{ \token_to_str:N \ETC. \} 
(End definition for \c__tl_analysis_show_etc_str.)
\msg_new:nnn { tl } { show-analysis }
{
  \tl_if_empty:nF {#1} { #1 ~ }
  \tl_if_empty:nTF {#2}
    { is-empty }
    { contains-the-tokens: #2 }
}\
Chapter 45

l3regex implementation

45.1 Plan of attack

Most regex engines use backtracking. This allows to provide very powerful features (back-references come to mind first), but it is costly, and raises the problem of catastrophic backtracking. Since \TeX is not first and foremost a programming language, complicated code tends to run slowly, and we must use faster, albeit slightly more restrictive, techniques, coming from automata theory.

Given a regular expression of \( n \) characters, we do the following:

- (Compiling.) Analyse the regex, finding invalid input, and convert it to an internal representation.
- (Building.) Convert the compiled regex to a non-deterministic finite automaton (NFA) with \( O(n) \) states which accepts precisely token lists matching that regex.
- (Matching.) Loop through the query token list one token (one “position”) at a time, exploring in parallel every possible path (“active thread”) through the NFA, considering active threads in an order determined by the quantifiers’ greediness.

We use the following vocabulary in the code comments (and in variable names).

- **Group**: index of the capturing group, \(-1\) for non-capturing groups.
- **Position**: each token in the query is labelled by an integer \( \langle \text{position} \rangle \), with \( \text{min_pos} - 1 \leq \langle \text{position} \rangle \leq \text{max_pos} \). The lowest and highest positions \( \text{min_pos} - 1 \) and \( \text{max_pos} \) correspond to imaginary begin and end markers (with non-existent category code and character code). \( \text{max_pos} \) is only set quite late in the processing.
- **Query**: the token list to which we apply the regular expression.
- **State**: each state of the NFA is labelled by an integer \( \langle \text{state} \rangle \) with \( \text{min_state} \leq \langle \text{state} \rangle < \text{max_state} \).
- **Active thread**: state of the NFA that is reached when reading the query token list for the matching. Those threads are ordered according to the greediness of quantifiers.
Step: used when matching, starts at 0, incremented every time a character is read, and is not reset when searching for repeated matches. The integer \l__regex_step_int is a unique id for all the steps of the matching algorithm.

We use l3intarray to manipulate arrays of integers. We also abuse \TeX’s \toks registers, by accessing them directly by number rather than tying them to control sequence using the \nevtoks allocation functions. Specifically, these arrays and \toks are used as follows. When building, \toks\langle state\rangle holds the tests and actions to perform in the \langle state\rangle of the NFA. When matching,

- \g__regex_state_active_intarray holds the last \langle step\rangle in which each \langle state\rangle was active.

- \g__regex_thread_info_intarray consists of blocks for each \langle thread\rangle (with min_thread ≤ \langle thread\rangle < max_thread). Each block has 1+2\l__regex_capturing_group_int entries: the \langle state\rangle in which the \langle thread\rangle currently is, followed by the beginnings of all submatches, and then the ends of all submatches. The \langle threads\rangle are ordered starting from the best to the least preferred.

- \g__regex_submatch_prev_intarray, \g__regex_submatch_begin_intarray and \g__regex_submatch_end_intarray hold, for each submatch (as would be extracted by \regex_extract_all\langle\rangle\rangle), the place where the submatch started to be looked for and its two end-points. For historical reasons, the minimum index is twice max_state, and the used registers go up to \l__regex_submatch_int. They are organized in blocks of \l__regex_capturing_group_int entries, each block corresponding to one match with all its submatches stored in consecutive entries.

When actually building the result,

- \toks\langle position\rangle holds \langle tokens\rangle which o- and x-expand to the \langle position\rangle-th token in the query.

- \g__regex_balance_intarray holds the balance of begin-group and end-group character tokens which appear before that point in the token list.

The code is structured as follows. Variables are introduced in the relevant section. First we present some generic helper functions. Then comes the code for compiling a regular expression, and for showing the result of the compilation. The building phase converts a compiled regex to NFA states, and the automaton is run by the code in the following section. The only remaining brick is parsing the replacement text and performing the replacement. We are then ready for all the user functions. Finally, messages, and a little bit of tracing code.

### 45.2 Helpers

\__regex_int_eval:w Access the primitive: performance is key here, so we do not use the slower route via \int_eval:n.

\cs_new_eq:NN \__regex_int_eval:w \tex_numexpr:D

(End definition for \__regex_int_eval:w.)

\__regex_standard_escapechar: Make the \escapechar into the standard backslash.

\cs_new_protected:Npn \__regex_standard_escapechar:
\{ \int_set:Nn \tex_escapechar:D { ‘‘\ } \}

451
Unpack a \toks given its number.
\cs_new:Npn \__regex_toks_use:w { \tex_the:D \tex_toks:D }

Empty a \toks or set it to a value, given its number.
\cs_new_protected:Npn \__regex_toks_clear:N #1
\cs_new_eq:NN \__regex_toks_set:Nn \tex_toks:D
\cs_new_protected:Npn \__regex_toks_set:No #1
\cs_toks:D #1 \exp_after:wN }

Copy \#3 \toks registers from \#2 onwards to \#1 onwards, like C's memcpy.
\cs_new_protected:Npn \__regex_toks_memcpy:NNn #1#2#3
\prg_replicate:nn {#3}
\tex_toks:D #1 = \tex_toks:D #2
\int_incr:N #1
\int_incr:N #2
}

During the building phase we wish to add x-expanded material to \toks, either to the left or to the right. The expansion is done “by hand” for optimization (these operations are used quite a lot). The \NN version of \__regex_toks_put_right:Nx is provided because it is more efficient than x-expanding with \exp_not:n.
\cs_new_protected:Npn \__regex_toks_put_left:Nx #1#2
\cs_set_nopar:Npx \__regex_tmp:w { #2 }
\tex_toks:D #1 \exp_after:wN \exp_after:wN \exp_after:wN
\exp_after:wN \__regex_tmp:w \tex_the:D \tex_toks:D #1

\cs_new_protected:Npn \__regex_toks_put_right:Nx #1#2
\cs_set_nopar:Npx \__regex_tmp:w {#2}
\tex_toks:D #1 \exp_after:wN
\{ \tex_the:D \tex_toks:D \exp_after:wN #1 \__regex_tmp:w \}

Expands to the string representation of the token (known to be a control sequence) at the current position \l__regex_curr_pos_int. It should only be used in x-expansion to avoid losing a leading space.
\cs_new:Npn \__regex_curr_cs_to_str:
\__regex_intarray_item:NnF \__regex_intarray_item_aux:nNF
Item of intarray, with a default value.
\__regex_intarray_item:NnF \__regex_intarray_item_aux:nNF #1#2
{ \exp_args:Nf \__regex_intarray_item_aux:nNF { \int_eval:n {#2} } #1 }
\__regex_intarray_item_aux:nNF #1#2
{ \if_int_compare:w #1 > \c_zero_int \exp_after:wN \use_i:nn \else: \exp_after:wN \use_ii:nn \fi:
{ \__kernel_intarray_item:Nn #2 {#1} }
}
(End definition for \__regex_intarray_item:NnF and \__regex_intarray_item_aux:nNF.)
\__regex_maplike_break:
Analogous to \tl_map_break:, this correctly exits \tl_map_inline:nn and similar constructions and jumps to the matching \prg_break_point:Nn \__regex_maplike_break:
{ }
\__regex_maplike_break:
{ \prg_map_break:Nn \__regex_maplike_break: { } }}
(End definition for \__regex_maplike_break:)
\__regex_odd_items:n \__regex_even_items:n \__regex_even_items_loop:nn
Map through a token list one pair at a time, leaving the odd-numbered or even-numbered items (the first item is numbered 1).
\__regex_odd_items:n \__regex_even_items:n #1 \__regex_even_items:n { ? #1 } \__regex_even_items:n #1
{ \__regex_even_items_loop:nn #1 \q__regex_nil \q__regex_nil \prg_break_point:
\__regex_even_items_loop:nn #1 \q__regex_nil \q__regex_nil \prg_break_point:
\__regex_even_items_loop:nn #1 \q__regex_nil \q__regex_nil \prg_break_point:
\__regex_even_items_loop:nn #1 \q__regex_nil \q__regex_nil \prg_break_point:
}
(End definition for \__regex_odd_items:n, \__regex_even_items:n, and \__regex_even_items_loop:nn)

45.2.1 Constants and variables
\__regex_tmp:w
Temporary function used for various short-term purposes.
\__regex_tmp:w { }
(End definition for \__regex_tmp:w.)
Temporary variables used for various purposes.

\tl_new:N \l__regex_internal_a_tl
\tl_new:N \l__regex_internal_b_tl
\tl_new:N \l__regex_internal_c_tl
\int_new:N \l__regex_internal_a_int
\int_new:N \l__regex_internal_b_int
\int_new:N \l__regex_internal_c_int
\bool_new:N \l__regex_internal_bool
\seq_new:N \l__regex_internal_seq
\tl_new:N \g__regex_internal_tl

(End definition for \l__regex_internal_a_tl and others.)

\tl_new:N \l__regex_build_tl

This temporary variable is specifically for use with the \tl_build machinery.

(End definition for \l__regex_build_tl.)

\c__regex_no_match_regex

This regular expression matches nothing, but is still a valid regular expression. We could use a failing assertion, but I went for an empty class. It is used as the initial value for regular expressions declared using \regex_new:N.

\tl_const:Nn \c__regex_no_match_regex
{
__regex_branch:n
{ \__regex_class:NnnnN \c_true_bool {} { 1 } { 0 } \c_true_bool }
}

(End definition for \c__regex_no_match_regex.)

\l__regex_balance_int

During this phase, \l__regex_balance_int counts the balance of begin-group and end-group character tokens which appear before a given point in the token list. This variable is also used to keep track of the balance in the replacement text.

(End definition for \l__regex_balance_int.)

45.2.2 Testing characters

\c__regex_ascii_min_int
\c__regex_ascii_max_control_int
\c__regex_ascii_max_int

\int_const:Nn \c__regex_ascii_min_int { 0 }
\int_const:Nn \c__regex_ascii_max_control_int { 31 }
\int_const:Nn \c__regex_ascii_max_int { 127 }

(End definition for \c__regex_ascii_min_int, \c__regex_ascii_max_control_int, and \c__regex_ascii_max_int.)

\c__regex_ascii_lower_int

\int_const:Nn \c__regex_ascii_lower_int { 'a - 'A }

(End definition for \c__regex_ascii_lower_int.)
45.2.3 Internal auxiliaries

\_\_regex\_recursion\_stop

Internal recursion quarks.

\verbatimverbatim
\cs_new:N \_\_regex\_recursion\_stop
(End definition for \_\_regex\_recursion\_stop.)
\endverbatimverbatim

\_\_regex\_nil

Internal quarks.

\verbatimverbatim
\cs_new:N \_\_regex\_nil
(End definition for \_\_regex\_nil.)
\endverbatimverbatim

Functions to gobble up to a quark.

\verbatimverbatim
\cs_new:Npn \_\_regex\_use\_none\_delimit\_by_q\_recursion\_stop:w #1 \_\_regex\_recursion\_stop { } \_\_regex\_use\_none\_delimit\_by_q\_nil:w
\cs_new:Npn \_\_regex\_use\_i\_delimit\_by_q\_recursion\_stop:nw #1 #2 \_\_regex\_recursion\_stop {#1} \_\_regex\_use\_i\_delimit\_by_q\_nil:wnw
\cs_new:Npn \_\_regex\_use\_none\_delimit\_by_q\_nil:w #1 \_\_regex\_nil { } \_\_regex\_use\_none\_delimit\_by_q\_nil:w
(End definition for \_\_regex\_use\_none\_delimit\_by_q\_recursion\_stop:w, \_\_regex\_use\_i\_delimit\_by_q\_recursion\_stop:nw, and \_\_regex\_use\_none\_delimit\_by_q\_nil:w.)
\endverbatimverbatim

\_\_regex\_quark\_if\_nil\_p:n
\_\_regex\_quark\_if\_nil:nTF

Branching quark conditional.

\verbatimverbatim
\__kernel_quark_new\_conditional:Nn \_\_regex\_quark\_if\_nil:N { F }
(End definition for \_\_regex\_quark\_if\_nil:nTF.)
\endverbatimverbatim

\_\_regex\_break\_point:TF
\_\_regex\_break\_true:w

When testing whether a character of the query token list matches a given character class in the regular expression, we often have to test it against several ranges of characters, checking if any one of those matches. This is done with a structure like

\verbatimverbatim
\{ (test1) \ldots (test_n) \_\_regex\_break\_point:TF { (true code) } { (false code) } \}
\endverbatimverbatim

If any of the tests succeeds, it calls \_\_regex\_break\_true:w, which cleans up and leaves \{true code\} in the input stream. Otherwise, \_\_regex\_break\_point:TF leaves the \{false code\} in the input stream.

\verbatimverbatim
\cs_new\_protected:Npm \_\_regex\_break\_true:w
#1 \_\_regex\_break\_point:TF #2 #3 \{#2 \}
\cs_new\_protected:Npm \_\_regex\_break\_point:TF #1 #2 \{ #2 \}
(End definition for \_\_regex\_break\_point:TF and \_\_regex\_break\_true:w.)
\endverbatimverbatim

\_\_regex\_item\_reverse:n

This function makes showing regular expressions easier, and lets us define \texttt{\textbackslash D} in terms of \texttt{d} for instance. There is a subtlety: the end of the query is marked by \texttt{-2}, and thus matches \texttt{\textbackslash D} and other negated properties; this case is caught by another part of the code.

\verbatimverbatim
\cs_new\_protected:Npm \_\_regex\_item\_reverse:n #1 \{
#1 
\_\_regex\_break\_point:TF \{ \} \_\_regex\_break\_true:w
\}
(End definition for \_\_regex\_item\_reverse:n.)
\endverbatimverbatim
Simple comparisons triggering \_\_regex_break_true:w when true.
\cs_new_protected:Npn \_\_regex_item_caseful_equal:n #1
\begin{verbatim}
  \if_int_compare:w #1 = \l__regex_curr_char_int
  \exp_after:wN \_\_regex_break_true:w
  \fi:
\end{verbatim}
\cs_new_protected:Npn \_\_regex_item_caseful_range:nn #1 #2
\begin{verbatim}
  \reverse_if:N \if_int_compare:w #1 > \l__regex_curr_char_int
  \reverse_if:N \if_int_compare:w #2 < \l__regex_curr_char_int
  \exp_after:wN \exp_after:wN \exp_after:wN \_\_regex_break_true:w
  \fi:
\end{verbatim}
(End definition for \_\_regex_item_caseful_equal:n and \_\_regex_item_caseful_range:nn.)

For caseless matching, we perform the test both on the curr_char and on the case_changed_char. Before doing the second set of tests, we make sure that case_changed_char has been computed.
\cs_new_protected:Npn \_\_regex_item_caseless_equal:n #1
\begin{verbatim}
  \if_int_compare:w #1 = \l__regex_curr_char_int
  \exp_after:wN \_\_regex_break_true:w
  \fi:
  \_\_regex_maybe_compute_ccc:
  \if_int_compare:w #1 = \l__regex_case_changed_char_int
  \exp_after:wN \_\_regex_break_true:w
  \fi:
\end{verbatim}
\cs_new_protected:Npn \_\_regex_item_caseless_range:nn #1 #2
\begin{verbatim}
  \reverse_if:N \if_int_compare:w #1 > \l__regex_curr_char_int
  \reverse_if:N \if_int_compare:w #2 < \l__regex_curr_char_int
  \exp_after:wN \exp_after:wN \exp_after:wN \_\_regex_break_true:w
  \fi:
  \_\_regex_maybe_compute_ccc:
  \if_int_compare:w #1 > \l__regex_case_changed_char_int
  \exp_after:wN \exp_after:wN \exp_after:wN \_\_regex_break_true:w
  \fi:
\end{verbatim}
(End definition for \_\_regex_item_caseless_equal:n and \_\_regex_item_caseless_range:nn.)

This function is called when \_\_regex_case_changed_char_int has not yet been computed. If the current character code is in the range \([65,90]\) (upper-case), then add 32, making it lowercase. If it is in the lower-case letter range \([97,122]\), subtract 32.
\cs_new_protected:Npn \_\_regex_compute_case_changed_char:
\begin{verbatim}
  \int_set_eq:NN \l__regex_case_changed_char_int \l__regex_curr_char_int
\end{verbatim}

Those must always be defined to expand to a caseful (default) or caseless version, and not be protected: they must expand when compiling, to hard-code which tests are caseless or caseful.

\cs_new_eq:NN \__regex_item_equal:n \__regex_item_equal:nT #1 #2
\cs_new_eq:NN \__regex_item_range:nn \__regex_item_range:nnT #1 #2
\cs_new_protected:Npn \__regex_item_catcode:nT #1
\cs_new_protected:Npn \__regex_item_catcode_reverse:nT #1 #2
\cs_new_protected:Npm \__regex_item_catcode:

The argument is a sum of powers of 4 with exponents given by the allowed category codes (between 0 and 13). Dividing by a given power of 4 gives an odd result if and only if that category code is allowed. If the catcode does not match, then skip the character code tests which follow.

\cs_new_protected:Npm \__regex_item_catcode:nT \#1
\cs_new_protected:Npm \__regex_item_catcode_reverse:nT \#1 \#2

(End definition for \__regex_item_equal:n and \__regex_item_range:nn)

(End definition for \__regex_item_catcode:nT, \__regex_item_catcode_reverse:nT, and \__regex_item_catcode:)

457
This matches an exact ⟨category⟩-⟨character code⟩ pair, or an exact control sequence, more precisely one of several possible control sequences, separated by \scan_stop:.

\cs_new_protected:Npn \__regex_item_exact:nn #1#2
\{\if_int_compare:w #1 = \l__regex_curr_catcode_int
\if_int_compare:w #2 = \l__regex_curr_char_int
\exp_after:wN \exp_after:wN \exp_after:wN \__regex_break_true:w
\fi:\fi:\}
\cs_new_protected:Npn \__regex_item_exact_cs:n #1
\{\int_compare:nNnTF \l__regex_curr_catcode_int = 0
{\__kernel_tl_set:Nx \l__regex_internal_a_tl
{\scan_stop: \__regex_curr_cs_to_str: \scan_stop: }
\tl_if_in:noTF {\scan_stop: #1 \scan_stop: }
\l__regex_internal_a_tl
{\__regex_break_true:w } { }
}
\}

(End definition for \__regex_item_exact:nn and \__regex_item_exact_cs:n)

\__regex_item_cs:n

Match a control sequence (the argument is a compiled regex). First test the catcode of the current token to be zero. Then perform the matching test, and break if the csname indeed matches.

\cs_new_protected:Npn \__regex_item_cs:n #1
\{\int_compare:nNnT \l__regex_curr_catcode_int = 0
{\group_begin:
\__regex_single_match:
\__regex_disable_submatches:
\__regex_build_for_cs:n {#1}
\bool_set_eq:NN \l__regex_saved_success_bool \g__regex_success_bool
\exp_args:Nx \__regex_match_cs:n { \__regex_curr_cs_to_str: }
\if_meaning:w \c_true_bool \g__regex_success_bool
\group_insert_after:N \__regex_break_true:w
\fi:
\bool_gset_eq:NN \g__regex_success_bool \l__regex_saved_success_bool
\group_end:
}

(End definition for \__regex_item_cs:n)

45.2.4 Character property tests

\__regex_prop_d:
\__regex_prop_h:
\__regex_prop_s:
\__regex_prop_v:
\__regex_prop_w:
\__regex_prop_N:

Character property tests for \d, \W, etc. These character properties are not affected by the \(?i\) option. The characters recognized by each one are as follows: \d=[0-9],

458
\w=[0-9A-Z_a-z], \s=\[\n\v\l\r\m\], \h=\[\n\v\l\r\m\], the upper case counterparts match anything that the lower case does not match. The order in which the various tests appear is optimized for usual mostly lower case letter text.

4622 \cs_new_protected:Npn \__regex_prop_d:
4623 { \__regex_item_caseful_range:nn { '0' } { '9' } }
4624 \cs_new_protected:Npn \__regex_prop_h:
4625 { \__regex_item_caseful_equal:n { '\' } 
4626 \__regex_item_caseful_equal:n { '\^^I' } }
4627 \cs_new_protected:Npn \__regex_prop_s:
4628 { \__regex_item_caseful_equal:n { '\' } 
4629 \__regex_item_caseful_equal:n { '\^^I' } 
4630 \__regex_item_caseful_equal:n { '\^^J' } 
4631 \__regex_item_caseful_equal:n { '\^^L' } 
4632 \__regex_item_caseful_equal:n { '\^^M' } }
4633 \cs_new_protected:Npn \__regex_prop_v:
4634 { \__regex_item_caseful_range:nn { '\^^J' } { '\^^M' } } % lf, vtab, ff, cr
4635 \cs_new_protected:Npn \__regex_prop_w:
4636 { \__regex_item_caseful_range:nn { 'a' } { 'z' } 
4637 \__regex_item_caseful_range:nn { 'A' } { 'Z' } 
4638 \__regex_item_caseful_range:nn { '0' } { '9' } 
4639 \__regex_item_caseful_equal:n { '_' } }
4640 \cs_new_protected:Npn \__regex_prop_N:
4641 { \__regex_item_reverse:n { \__regex_item_caseful_equal:n { '\^^J' } } }

(End definition for \__regex_prop_d: and others.)

\__regex_posix_alnum: POSIX properties. No surprise.
\__regex_posix_alpha: \cs_new_protected:Npn \__regex_posix_alpha:
\__regex_posix_ascii: { \__regex_posix_alpha: \__regex_posix_digit: }
\__regex_posix_blank: \cs_new_protected:Npn \__regex_posix_alpha:
\__regex_posix_cntrl: { \__regex_posix_lower: \__regex_posix_upper: }
\__regex_posix_digit: \cs_new_protected:Npn \__regex_posix_asciib:
\__regex_posix_graph:
\__regex_posix_lower:
\__regex_posix_print:
\__regex_posix_punct:
\__regex_posix_space:
\__regex_posix_upper:
\__regex_posix_word:
\__regex_posix_xdigit:
\cs_new_eq:NN \__regex_posix_blank: \__regex_prop_h:
\cs_new_protected:Npn \__regex_posix_cntrl:
{ \__regex_item_caseful_equal:n \c__regex_ascii_max_int 
\__regex_item_caseful_equal:n \c__regex_ascii_max_control_int 
\__regex_item_caseful_equal:n \c__regex_ascii_max_int }
459
45.2.5 Simple character escape

Before actually parsing the regular expression or the replacement text, we go through
them once, converting \n to the character 10, etc. In this pass, we also convert any special
character (*, ?, {, etc.) or escaped alphanumeric character into a marker indicating that
this was a special sequence, and replace escaped special characters and non-escaped
alphanumeric characters by markers indicating that those were “raw” characters. The
rest of the code can then avoid caring about escaping issues (those can become quite
complex to handle in combination with ranges in character classes).

Usage: \_regex_escape_use:nnnn \langle \text{inline 1} \rangle \langle \text{inline 2} \rangle \langle \text{inline 3} \rangle \{\langle \text{token list} \rangle \}

The \text{token list} is converted to a string, then read from left to right, interpreting back-
slashes as escaping the next character. Unescaped characters are fed to the function
\langle \text{inline 1} \rangle, and escaped characters are fed to the function \langle \text{inline 2} \rangle within an x-expansion
context (typically those functions perform some tests on their argument to decide how
to output them). The escape sequences \a, \e, \f, \n, \r, \t and \x are recognized, and
those are replaced by the corresponding character, then fed to \langle \text{inline 3} \rangle. The result is
then left in the input stream. Spaces are ignored unless escaped.

The conversion is done within an x-expanding assignment.

\_regex_escape_use:nnnn

The result is built in \_regex_internal_a_tl, which is then left in the input stream.
Tracing code is added as appropriate inside this token list. Go through #4 once, applying
#1, #2, or #3 as relevant to each character (after de-escaping it).
\cs_new_protected:Npn \__regex_escape_use:nnnn \ #1\#2\#3\#4
{\group_begin: \tl_clear:N \l__regex_internal_a_tl
\cs_set:Npn \__regex_escape_unescaped:N ##1 { #1 }
\cs_set:Npn \__regex_escape_escaped:N ##1 { #2 }
\cs_set:Npn \__regex_escape_raw:N ##1 { #3 }
\__regex_standard_escapechar:
\__kernel_tl_gset:Nx \g__regex_internal_tl { \__kernel_str_to_other_fast:n { #4} }
\tl_put_right:Nx \l__regex_internal_a_tl { \exp_after:wN \__regex_escape_loop:N \g__regex_internal_tl \scan_stop: \prg_break_point: }
\exp_after:wN \group_end: \l__regex_internal_a_tl}

\__regex_escape_loop:N \__regex_escape_\:w \__regex_escape_loop:N reads one character: if it is special (space, backslash, or endmarker), perform the associated action, otherwise it is simply an unescaped character. After a backslash, the same is done, but unknown characters are “escaped”.

\cs_new:Npn \__regex_escape_loop:N #1
{\cs_if_exist_use:cF { \__regex_escape_\token_to_str:N #1:w } { \__regex_escape_unescaped:N #1 }
\__regex_escape_loop:N}
\cs_new:cpn { \__regex_escape_ \c_backslash_str :w }
\cs_new:cpn { \__regex_escape_/ \iow_char:N\scan_stop: :w }
\cs_new:cpn { \__regex_escape_/\iow_char:N\scan_stop: :w }
\cs_new:cpn { \__regex_escape_/a:w }
\cs_new:cpn { \__regex_escape_/e:w }
\cs_new:cpn { \__regex_escape_/f:w }
\cs_new:cpn { \__regex_escape_/n:w }
\cs_new:cpn { \__regex_escape_/r:w }
\cs_new:cpn { \__regex_escape_/t:w }
\cs_new:cpn { \__regex_escape_\:w }

\__regex_escape_unescaped:N \__regex_escape_escaped:N \__regex_escape_raw:N
Those functions are never called before being given a new meaning, so their definitions here don’t matter.

\cs_new_eq:NN \__regex_escape_unescaped:N ?
\cs_new_eq:NN \__regex_escape_escaped:N ?
\cs_new_eq:NN \__regex_escape_raw:N ?

(End definition for \__regex_escape_unescaped:N, \__regex_escape_escaped:N, and \__regex_escape_-
raw:N.)

\__regex_escape_\scan_stop:w \__regex_escape_/\scan_stop:w \__regex_escape_/a:w \__regex_escape_/e:w \__regex_escape_/r:w \__regex_escape_/t:w \__regex_escape_/w

The loop is ended upon seeing the end-marker “break”, with an error if the string ended in a backslash. Spaces are ignored, and \a, \e, \f, \n, \r, \t take their meaning here.
\msg\expandableerror{nn { regex } { trailing-backslash }}
\prg\break:

\cs\new:c { __regex_escape_-w } { }
\cs\new:c { __regex_escape_/a:w } { }
\cs\new:c { __regex_escape_/t:w } { }
\cs\new:c { __regex_escape_/n:w } { }
\cs\new:c { __regex_escape_/f:w } { }
\cs\new:c { __regex_escape_/r:w } { }
\cs\new:c { __regex_escape_/e:w } { }
\cs\new:c { __regex_escape_/x:w } { \__regex_escape_x_end:w }
\cs\new:c { __regex_escape_x_test:N } { }
\cs\new:c { __regex_escape_x_testii:N } { }
\cs\new:c { __regex_escape_x_large:n } { }

(End definition for \__regex_escape_-w and others.)

When \textbackslash x is encountered, \__regex_escape_x_test:N is responsible for grabbing some hexadecimal digits, and feeding the result to \__regex_escape_x_end:w. If the number is too big interrupt the assignment and produce an error, otherwise call \__regex_escape_raw:N on the corresponding character token.

\cs\new:Npn { \__regex_escape_x_test:N } #1 { }
\cs\new:Npn { \__regex_escape_x_testii:N } #1 { }
\if_charcode:w \c_space_token #1
  \exp_after:wN \__regex_escape_x_test:N
\else:
  \exp_after:wN \__regex_escape_x_testii:N
  \exp_after:wN #1
\fi:
\}
\cs_new:Npn \__regex_escape_x_testii:N #1
  \if_charcode:w \c_left_brace_str #1
  \exp_after:wN \__regex_escape_x_loop:N
  \else:
  \__regex_hexadecimal_use:NTF #1
  \{ \exp_after:wN \__regex_escape_x:N \}
  \{ ; \exp_after:wN \__regex_escape_loop:N \exp_after:wN #1 \}
\fi:
\}
\end{definition}

\__regex_escape_x:N
This looks for the second digit in the unbraced case.
\cs_new:Npn \__regex_escape_x:N #1
  \if_meaning:w \scan_stop: #1
  \exp_after:wN \use_i:nnn \exp_after:wN ;
  \fi:
  \use:n
  \{ \__regex_hexadecimal_use:NTF #1
  \{ \__regex_escape_x_loop:N \}
  \{ ; \__regex_escape_loop:N #1 \}
\}
\end{definition}

\__regex_escape_x_loop:N
\__regex_escape_x_loop_error:
Grab hexadecimal digits, skip spaces, and at the end, check that there is a right brace, otherwise raise an error outside the assignment.
\cs_new:Npn \__regex_escape_x_loop:N #1
  \if_meaning:w \scan_stop: #1
  \exp_after:wN \use_i:nnn \exp_after:wN \use_ii:nn
  \fi:
  \use_ii:nn
  \{ ; \__regex_escape_x_loop_error:n \{ \} {#1} \}
  \{ \__regex_hexadecimal_use:NTF \}
  \{ \__regex_escape_x_loop:N \}
  \{ \token_if_eq_charcode:NNTF \c_space_token #1
  \{ \__regex_escape_x_loop:N \}
  \}
TEX detects uppercase hexadecimal digits for us but not the lowercase letters, which we need to detect and replace by their uppercase counterpart.

These two tests are used in the first pass when parsing a regular expression. That pass is responsible for finding escaped and non-escaped characters, and recognizing which ones have special meanings and which should be interpreted as “raw” characters. Namely,

- alphanumerics are “raw” if they are not escaped, and may have a special meaning when escaped;
- non-alphanumeric printable ascii characters are “raw” if they are escaped, and may have a special meaning when not escaped;
- characters other than printable ascii are always “raw”.

(End definition for \_\_regex_hexadecimal_use:NTF.)
The code is ugly, and highly based on magic numbers and the ascii codes of characters. This is mostly unavoidable for performance reasons. Maybe the tests can be optimized a little bit more. Here, “alphanumeric” means 0–9, A–Z, a–z; “special” character means non-alphanumeric but printable ascii, from space (hex 20) to del (hex 7E).

45.3 Compiling

A regular expression starts its life as a string of characters. In this section, we convert it to internal instructions, resulting in a “compiled” regular expression. This compiled
expression is then turned into states of an automaton in the building phase. Compiled regular expressions consist of the following:

- \_\_regex_class:n\#nnnN \{boolean\} \{\(\text{tests}\)\} \{\(\text{min}\)\} \{\(\text{more}\)\} \(\text{lazyness}\)
- \_\_regex_group:n\#nnnN \{\(\text{branches}\)\} \{\(\text{min}\)\} \{\(\text{more}\)\} \(\text{lazyness}\), also \_\_regex_group_no_capture:n\#nnnN and \_\_regex_group_resetting:n\#nnnN with the same syntax.
- \_\_regex_branch:n \{\(\text{contents}\)\}
- \_\_regex_command_K:
- \_\_regex_assertion:n \{\(\text{assertion test}\)\}, where the \(\text{assertion test}\) is \_\_regex_b_test: or \_\_regex_Z_test: or \_\_regex_A_test: or \_\_regex_G_test:
- \_\_regex_item_caseful_equal:n \{\(\text{char code}\)\}
- \_\_regex_item_caseless_equal:n \{\(\text{char code}\)\}
- \_\_regex_item_caseful_range:nn \{\(\text{min}\)\} \{\(\text{max}\)\}
- \_\_regex_item_caseless_range:nn \{\(\text{min}\)\} \{\(\text{max}\)\}
- \_\_regex_item_catcode:nT \{\(\text{catcode bitmap}\)\} \{\(\text{tests}\)\}
- \_\_regex_item_catcode_reverse:nT \{\(\text{catcode bitmap}\)\} \{\(\text{tests}\)\}
- \_\_regex_item_reverse:n \{\(\text{tests}\)\}
- \_\_regex_item_exact:nn \{\(\text{catcode}\)\} \{\(\text{char code}\)\}
- \_\_regex_item_exact_cs:n \{\(\text{csnames}\)\}, more precisely given as \(\text{csname}\) \text{\scan stop:} \(\text{csname}\) \text{\scan_stop:} \(\text{csname}\) and so on in a brace group.
- \_\_regex_item_cs:n \{\(\text{compiled regex}\)\}

45.3.1 Variables used when compiling

We make sure to open the same number of groups as we close.

\int_new:N \l__regex_group_level_int
\int_new:N \l__regex_mode_int
\int_const:Nn \c__regex_cs_in_class_mode_int { -6 }
\int_const:Nn \c__regex_cs_mode_int { -2 }
\int_const:Nn \c__regex_outer_mode_int { 0 }
\int_const:Nn \c__regex_catcode_mode_int { 2 }
\int_const:Nn \c__regex_catcode_in_class_mode_int { 3 }
\int_const:Nn \c__regex_class_mode_int { 6 }

While compiling, ten modes are recognized, labelled \(-63, -23, -6, -2, 0, 2, 3, 6, 23, 63\). See section 45.3.3. We only define some of these as constants.
We wish to allow constructions such as `\c[^BE] (..\cL[a-z]..)`, where the outer catcode test applies to the whole group, but is superseded by the inner catcode test. For this to work, we need to keep track of lists of allowed category codes: `\l__regex_catcodes_int` and `\l__regex_default_catcodes_int` are bitmaps, sums of $4^c$, for all allowed catcodes $c$. The latter is local to each capturing group, and we reset `\l__regex_catcodes_int` to that value after each character or class, changing it only when encountering a `\c` escape. The boolean records whether the list of categories of a catcode test has to be inverted: compare `\c[^[BE]]` and `\c[BE]`.

\begin{verbatim}
\int_new:N \l__regex_catcodes_int
\int_new:N \l__regex_default_catcodes_int
\bool_new:N \l__regex_catcodes_bool
\end{verbatim}

Constants: $4^c$ for each category, and the sum of all powers of 4.

\begin{verbatim}
\int_const:Nn \c__regex_catcode_C_int { "1 } \int_const:Nn \c__regex_catcode_B_int { "4 } \int_const:Nn \c__regex_catcode_E_int { "10 } \int_const:Nn \c__regex_catcode_M_int { "40 } \int_const:Nn \c__regex_catcode_T_int { "100 } \int_const:Nn \c__regex_catcode_P_int { "1000 } \int_const:Nn \c__regex_catcode_U_int { "4000 } \int_const:Nn \c__regex_catcode_D_int { "10000 } \int_const:Nn \c__regex_catcode_S_int { "100000 } \int_const:Nn \c__regex_catcode_L_int { "400000 } \int_const:Nn \c__regex_catcode_O_int { "1000000 } \int_const:Nn \c__regex_catcode_A_int { "4000000 } \int_const:Nn \c__regex_all_catcodes_int { "5515155 }
\end{verbatim}

The compilation step stores its result in this variable.

\begin{verbatim}
\cs_new_eq:NN \l__regex_internal_regex \c__regex_no_match_regex
\end{verbatim}

This sequence holds the prefix that makes up the line displayed to the user. The various items must be removed from the right, which is tricky with a token list, hence we use a sequence.

\begin{verbatim}
\seq_new:N \l__regex_show_prefix_seq
\end{verbatim}

A hack. To know whether a given class has a single item in it or not, we count the number of lines when showing the class.

\begin{verbatim}
\int_new:N \l__regex_show_lines_int
\end{verbatim}
45.3.2 Generic helpers used when compiling

\_\_regex\_two\_if\_eq:NNNNTF

Used to compare pairs of things like \_\_regex\_compile\_special:N ? together. It’s often inconvenient to get the catcodes of the character to match so we just compare the character code. Besides, the expanding behaviour of \if:w is very useful as that means we can use \c\_left\_brace\_str and the like.

4921 \prg\_new\_conditional:Npnn \_\_regex\_two\_if\_eq:NNN #1#2#3#4 { TF }
4922 {
4923 \if\_meaning:w #1 #3
4924 \if:w #2 #4
4925 \prg\_return\_true:
4926 \else:
4927 \prg\_return\_false:
4928 \fi:
4929 \else:
4930 \prg\_return\_false:
4931 \fi:
4932 }

(End definition for \_\_regex\_two\_if\_eq:NNNNTF.)

\_\_regex\_get\_digits:NTFw \_\_regex\_get\_digits\_loop:w

If followed by some raw digits, collect them one by one in the integer variable #1, and take the true branch. Otherwise, take the false branch.

4933 \cs\_new\_protected:Npn \_\_regex\_get\_digits:NTFw #1#2#3#4#5
4934 {
4935 \_\_regex\_if\_raw\_digit:NNTF #4 #5
4936 { #1 = #5 \_\_regex\_get\_digits\_loop:nw {#2} }
4937 { #3 #4 #5 }
4938 }
4939 \cs\_new:Npn \_\_regex\_get\_digits\_loop:nw #1#2#3
4940 {
4941 \_\_regex\_if\_raw\_digit:NNTF #2 #3
4942 { #3 \_\_regex\_get\_digits\_loop:nw {#1} }
4943 { \scan\_stop: #1 #2 #3 }
4944 }

(End definition for \_\_regex\_get\_digits:NTFw and \_\_regex\_get\_digits\_loop:w.)

\_\_regex\_if\_raw\_digit:NNTF

Test used when grabbing digits for the \{m,n\} quantifier. It only accepts non-escaped digits.

4945 \prg\_new\_conditional:Npnn \_\_regex\_if\_raw\_digit:NN #1#2 { TF }
4946 {
4947 \if\_meaning:w \_\_regex\_compile\_raw:N #1
4948 \if\_int\_compare:w 1 < 1 #2 \exp\_stop\_f:
4949 \prg\_return\_true:
4950 \else:
4951 \prg\_return\_false:
4952 \fi:
4953 \else:
4954 \prg\_return\_false:
4955 \fi:
4956 }

(End definition for \_\_regex\_if\_raw\_digit:NNTF.)
45.3.3 Mode

When compiling the NFA corresponding to a given regex string, we can be in ten distinct modes, which we label by some magic numbers:

-6 \[\text{\textbackslash{}c\{...\}}\] control sequence in a class,
-2 \[\text{\textbackslash{}c\{...\}}\] control sequence,
0 ... outer,
2 \text{\textbackslash{...}} catcode test,
6 \[\text{\textbackslash{...}}\] catcode test in a class,
-63 \[\text{\textbackslash{\{...\}}}\] class inside mode −6,
-23 \text{\textbackslash{\{...\}}} class inside mode −2,
3 [...] class inside mode 0,
23 \text{\textbackslash{[...}\}} class inside mode 2,
63 \text{\textbackslash{[{...}\}}} class inside mode 6.

This list is exhaustive, because \textbackslash{c} escape sequences cannot be nested, and character classes cannot be nested directly. The choice of numbers is such as to optimize the most useful tests, and make transitions from one mode to another as simple as possible.

- Even modes mean that we are not directly in a character class. In this case, a left bracket appends 3 to the mode. In a character class, a right bracket changes the mode as \( m \to (m - 15)/13 \), truncated.

- Grouping, assertion, and anchors are allowed in non-positive even modes (0, −2, −6), and do not change the mode. Otherwise, they trigger an error.

- A left bracket is special in even modes, appending 3 to the mode; in those modes, quantifiers and the dot are recognized, and the right bracket is normal. In odd modes (within classes), the left bracket is normal, but the right bracket ends the class, changing the mode from \( m \) to \( (m - 15)/13 \), truncated; also, ranges are recognized.

- In non-negative modes, left and right braces are normal. In negative modes, however, left braces trigger a warning; right braces end the control sequence, going from −2 to 0 or −6 to 3, with error recovery for odd modes.

- Properties (such as the \textbackslash{d} character class) can appear in any mode.

\texttt{}\texttt{__regex\_if\_in\_class:TF} Test whether we are directly in a character class (at the innermost level of nesting). There, many escape sequences are not recognized, and special characters are normal. Also, for every raw character, we must look ahead for a possible raw dash.
The \texttt{\textbackslash c} escape sequence is only allowed in modes 0 and 3, \textit{i.e.}, not within any other \texttt{\textbackslash c} escape sequence.
\__regex_mode_quit_c:  This function changes the mode as it is needed just after a catcode test.

\cs_new_protected:Npn \__regex_mode_quit_c:  \{  \if_int_compare:w \l__regex_mode_int = \c__regex_catcode_mode_int  \int_set_eq:NN \l__regex_mode_int \c__regex_outer_mode_int  \else:  \if_int_compare:w \l__regex_mode_int = \c__regex_catcode_in_class_mode_int  \int_set_eq:NN \l__regex_mode_int \c__regex_class_mode_int  \fi:  \fi:  \}

(End definition for \__regex_mode_quit_c:)

45.3.4 Framework

\__regex_compile:w  \__regex_compile_end:  Used when compiling a user regex or a regex for the \texttt{\{...\}} escape sequence within another regex. Start building a token list within a group (with x-expansion at the outset), and set a few variables (group level, catcodes), then start the first branch. At the end, make sure there are no dangling classes nor groups, close the last branch: we are done building \l__regex_internal_regex.

\cs_new_protected:Npn \__regex_compile:w  \{  \group_begin:  \tl_build_begin:N \l__regex_build_tl  \int_zero:N \l__regex_group_level_int  \int_set_eq:NN \l__regex_default_catcodes_int \c__regex_all_catcodes_int  \cs_set:Npn \__regex_item_equal:n { \__regex_item_caseful_equal:n }  \cs_set:Npn \__regex_item_range:nn { \__regex_item_caseful_range:nn }  \tl_build_put_right:Nn \l__regex_build_tl { \__regex_branch:n { \if_false: } \fi: }  \}

\cs_new_protected:Npn \__regex_compile_end:  \{  \__regex_if_in_class:TF  \{  \msg_error:nn { regex } { missing-rbrack }  \use:c { \__regex_compile_}:  \}

471
The compilation is done between `\__regex_compile:w` and `\__regex_compile_end:`, starting in mode 0. Then `\__regex_escape_use:nnnn` distinguishes special characters, escaped alphanumerics, and raw characters, interpreting `\a`, `\x` and other sequences. The 4 trailing `\prg_do_nothing:` are needed because some functions defined later look up to 4 tokens ahead. Before ending, make sure that any `\c{...}` is properly closed. No need to check that brackets are closed properly since `\__regex_compile_end:` does that. However, catch the case of a trailing `\cL` construction.
\_\_regex_compile_use:n
Use a regex, regardless of whether it is given as a string (in which case we need to compile)
or as a regex variable. This is used for \regex_match_case:nn and related functions to
allow a mixture of explicit regex and regex variables.
\cs_new_protected:Npn \_\_regex_compile_use:n #1
\tl_if_single_token:nT {#1}
\exp_after:wN \_\_regex_compile_use_aux:w 
\token_to_meaning:N #1 ~ \q__regex_nil 
\_\_regex_compile:n {#1} \l__regex_internal_regex
\cs_new_protected:Npm \_\_regex_compile_use:n #1
\cs_if_exist_use:cF { __regex_compile_#1: } 
\{ \_\_regex_compile_raw:N #1 \}
\cs_new_protected:Npm \_\_regex_compile_escaped:N #1
\cs_if_exist_use:cF { __regex_compile_/#1: } 
\{ \_\_regex_compile_raw:N #1 \}
(End definition for \_\_regex_compile_use:n)

\_\_regex_compile_escaped:N
\_\_regex_compile_special:N
If the special character or escaped alphanumeric has a particular meaning in regexes,
the corresponding function is used. Otherwise, it is interpreted as a raw character. We
distinguish special characters from escaped alphanumeric characters because they behave
differently when appearing as an end-point of a range.
\cs_new_protected:Npm \_\_regex_compile_special:N #1
\cs_if_exist_use:cF { __regex_compile_#1: } 
\{ \_\_regex_compile_raw:N #1 \}
\cs_new_protected:Npm \_\_regex_compile_escaped:N #1
\cs_if_exist_use:cF { __regex_compile_/#1: } 
\{ \_\_regex_compile_raw:N #1 \}
(End definition for \_\_regex_compile_escaped:N and \_\_regex_compile_special:N)

\_\_regex_compile_one:n
This is used after finding one “test”, such as \d, or a raw character. If that followed a
catcode test (e.g., \cL), then restore the mode. If we are not in a class, then the test is

(End definition for \_\_regex_compile_one:n)
“standalone”, and we need to add \_\_regex\_class:NNNN and search for quantifiers. In any case, insert the test, possibly together with a catcode test if appropriate.

\cs_new_protected:Npn \_\_regex\_compile\_one:n #1
{ \_\_regex\_mode\_quit_c:
   \_\_regex\_if\_in\_class:TF { }
   \tl_build\_put\_right:Nn \l\_\_regex\_build\_tl
   \{ \_\_regex\_class:NNNN \c\_true_bool { \_if\_false: } \_fi: \}
\tl_build\_put\_right:Nx \l\_\_regex\_build\_tl
{ \_\_regex\_if\_in\_class:TF { \_\_regex\_compile\_quantifier:w } }
\int\_set\_eq:NN \l\_\_regex\_catcodes\_int \l\_\_regex\_default\_catcodes\_int
\_\_regex\_if\_in\_class:TF { } { \_\_regex\_compile\_quantifier:w }

\_\_regex\_compile\_abort\_tokens:n
This function places the collected tokens back in the input stream, each as a raw character. Spaces are not preserved.
\cs_new_protected:Npn \_\_regex\_compile\_abort\_tokens:x #1
{ \_\_regex\_if\_in\_class:TFw #1#2#3#4
   \token\_if\_eq\_meaning:NNTF #3 \_\_regex\_compile\_special:N
   \{ \_\_regex\_compile\_quantifier:w \} \{ #1 \#2 \#3 \#4
\cs\_generate\_variant:Nn \_\_regex\_compile\_abort\_tokens:n { x }

45.3.5 Quantifiers
This looks ahead and checks whether there are any quantifier (special character equal to either of \?\+\{}). This is useful for the \ua and \ur escape sequences.
\cs_new_protected:Npn \_\_regex\_compile\_if\_quantifier:TFw
{ \_\_regex\_if\_eq\_meaning:NNTF #3 \_\_regex\_compile\_special:N
   \{ \_\_regex\_compile\_quantifier:w \} \{ #1 \#2 \#3 \#4
\cs\_generate\_variant:Nn \_\_regex\_compile\_if\_quantifier:TFw

474
This looks ahead and finds any quantifier (special character equal to either of `?`+`*`).

Those functions are called whenever there is no quantifier, or a braced construction is invalid (equivalent to no quantifier, and whatever characters were grabbed are left raw).

Once the “main” quantifier (`?`, `*`, `+` or a braced construction) is found, we check whether it is lazy (followed by a question mark). We then add to the compiled regex a closing brace (ending `\__regex_class:NnnnN` and friends), the start-point of the range, its end-point, and a boolean, `true` for lazy and `false` for greedy operators.

For each “basic” quantifier, `?`, `*`, `+`, feed the correct arguments to `\__regex_compile_quantifier_lazyness:nnNN`. 

−1 means that there is no upper bound on the number of repetitions.
Three possible syntaxes: \{\langle\text{int}\rangle\}, \langle\text{int}\rangle, or \{\langle\text{int}\rangle,\langle\text{int}\rangle\}. Any other syntax causes us to abort and put whatever we collected back in the input stream, as raw characters, including the opening brace. Grab a number into \l__regex_internal_a_int. If the number is followed by a right brace, the range is \[a,a\]. If followed by a comma, grab one more number, and call the \_ii or \_iii auxiliary. Those auxiliaries check for a closing brace, leading to the range \[a,\infty\] or \[a,b\], encoded as \{a\}{−1} and \{a\}{b−a}.
\__regex_compile_quantifier_braced_auxiii:w #1#2
\__regex_compile_quantifier_abort:xNN
\__regex_compile_quantifier_bracedAuxiii:w #1#2
\__regex_compile_quantifier_abort:xNN

(End definition for \__regex_compile_quantifier_{:w and others.}

5.3.6 Raw characters
\__regex_compile_raw_error:N
\__regex_compile_raw:N

\__regex_compile_raw_error:N
Within character classes, and following catcode tests, some escaped alphanumeric
sequences such as \b do not have any meaning. They are replaced by a raw character, after
spitting out an error.

(End definition for \__regex_compile_raw_error:N.)

\__regex_compile_raw:N
If we are in a character class and the next character is an unescaped dash, this denotes
a range. Otherwise, the current character #1 matches itself.

(End definition for \__regex_compile_raw:N.)
We have just read a raw character followed by a dash; this should be followed by an end-point for the range. Valid end-points are: any raw character; any special character, except a right bracket. In particular, escaped characters are forbidden.

```latex
\__regex_if_end_range:NNTF
\__regex_compile_range:Nw

\__regex_compile_raw:N
\__regex_item_equal:n
\__regex_item_range:nn
\__regex_compile_one:n
```

(End definition for \__regex_compile_raw:Nw.)
5325 \tl_build_put_right:Nx \l__regex_build_tl
5326 {
5327 \__regex_item_equal:n \{ \int_value:w '#1 \exp_stop_f: \}
5328 \__regex_item_equal:n \{ \int_value:w '-' \exp_stop_f: \}
5330 \}
5331 \#2\#3
5332 }

(End definition for \__regex_compile_range:Nw and \__regex_if_end_range:NNTF.)

45.3.7 Character properties

\__regex_compile_.: In a class, the dot has no special meaning. Outside, insert \__regex_prop_.:, which
matches any character or control sequence, and refuses −2 (end-marker).

5333 \cs_new_protected:cpx \{ \__regex_compile_.: \}
5334 \{ \exp_not:N \__regex_if_in_class:TF
5335 \{ \__regex_compile_raw:N . \}
5336 \}
5337 \cs_new_protected:cpn \{ \__regex_prop_.: \}
5338 \if_int_compare:w \l__regex_curr_char_int > - 2 \exp_stop_f:
5339 \exp_after:wN \__regex_break_true:w
5340 \fi:
5341 \}

(End definition for \__regex_compile_.: and \__regex_prop_.:)

\__regex_compile_/d: The constants \__regex_prop_d:, etc. hold a list of tests which match the corresponding
character class, and jump to the \__regex_break_point:TF marker. As for a normal
character, we check for quantifiers.

5346 \cs_set_protected:Npn \__regex_tmp:w #1#2
5347 \cs_new_protected:cpx \{ \__regex_compile_/#1: \}
5348 \cs_new_protected:cpx \{ \__regex_compile_/#1: \}
5349 \cs_new_protected:cpx \{ \__regex_compile_/#2: \}
5350 \}
5351 \cs_new_protected:cpn \{ \__regex_compile_/#1: \}
5352 \}
5353 \}
5354 \}
5355 \}
5356 \}
5357 \}
5358 \}
5359 \}
5360 \}
5361 \}
5362 \}

(End definition for \__regex_compile_/d: and others.)
45.3.8 Anchoring and simple assertions

In modes where assertions are forbidden, anchors such as \A produce an error (\A is invalid in classes); otherwise they add an \__regex_assertion:Nn test as appropriate (the only negative assertion is \B). The test functions are defined later. The implementation for $ and ^ is only different from \A etc because these are valid in a class.

\cs_new_protected:Npn \__regex_compile_anchor_letter:NNN #1#2#3
\__regex_if_in_class_or_catcode:TF { \__regex_compile_raw_error:N #1 }
\tl_build_put_right:Nn \l__regex_build_tl { \__regex_assertion:Nn #2 {#3} }
\cs_set_protected:Npn \__regex_tmp:w #1#2
\cs_new_protected:cpn { __regex_compile_#1: }
\exp_args:Nx \__regex_tmp:w { \iow_char:N \^ } { \c_true_bool }
\exp_args:Nx \__regex_tmp:w { \iow_char:N \$ } { \c_true_bool }
(End definition for \__regex_compile_anchor_letter:NNN and others.)

45.3.9 Character classes

Outside a class, right brackets have no meaning. In a class, change the mode (m \rightarrow (m - 15)/13, truncated) to reflect the fact that we are leaving the class. Look for quantifiers, unless we are still in a class after leaving one (the case of [...\cL[...]. . . ]). quantifiers.

\cs_new_protected:cpn { __regex_compile_]: }
\__regex_if_in_class_or_catcode:TF { \__regex_compile_raw:N #1 }
\tl_build_put_right:Nn \l__regex_build_tl { \__regex_assertion:Nn \c_true_bool {#2} }
\cs_set_protected:Npn \__regex_tmp:w #1#2
\exp_args:Nx \__regex_tmp:w { \iow_char:N \^ } { \c_true_bool }
\exp_args:Nx \__regex_tmp:w { \iow_char:N \$ } { \c_true_bool }

(End definition for \__regex_compile_anchor_letter:NNN and others.)

45.3.9 Character classes

Outside a class, right brackets have no meaning. In a class, change the mode (m \rightarrow (m - 15)/13, truncated) to reflect the fact that we are leaving the class. Look for quantifiers, unless we are still in a class after leaving one (the case of [...\cL[...]. . . ]). quantifiers.
In a class, left brackets might introduce a POSIX character class, or mean nothing. Immediately following \c{category}, we must insert the appropriate catcode test, then parse the class; we pre-expand the catcode as an optimization. Otherwise (modes 0, −2 and −6) just parse the class. The mode is updated later.

\_regex_compile\[:

\_regex_compile_class_normal\: In the “normal” case, we insert \_regex_class\:NnnnN \langle boolean \rangle in the compiled code. The \langle boolean \rangle is true for positive classes, and false for negative classes, characterized by a leading \^\textasciitilde. The auxiliary \_regex_compile_class\:TFNN also checks for a leading ] which has a special meaning.

\_regex_compile_class\:catcode\:w

This function is called for a left bracket in modes 2 or 6 (catcode test, and catcode test within a class). In mode 2 the whole construction needs to be put in a class (like single character). Then determine if the class is positive or negative, inserting \_regex\:-item\:catcode\:nT or the reverse variant as appropriate, each with the current catcodes bitmap \#1 as an argument, and reset the catcodes.
The first character is -, then the class is negative (use #2), otherwise it is positive (use #1). If the next character is a right bracket, then it should be changed to a raw one.

\cs_new_protected:Npn \__regex_compile_class:TFNN #1#2#3#4
\l__regex_mode_int = \int_value:w \l__regex_mode_int 3 \exp_stop_f:
\__regex_two_if_eq:NNNNTF #3 #4 \__regex_compile_special:N ^
\tl_build_put_right:Nn \l__regex_build_tl { #2 { \if_false: } \fi: }
\__regex_compile_class:NN
\tl_build_put_right:Nn \l__regex_build_tl { #1 { \if_false: } \fi: }
\__regex_compile_class:NN #3 #4
\}
\cs_new_protected:Npn \__regex_compile_class:NN #1#2
\token_if_eq_meaning:NNT \__regex_compile_special:N #1
\str_case:nn { #2 }
\msg_warning:nnx { regex } posix-unsupported
\msg_warning:nnx { regex } posix-unsupported
\}

Here we check for a syntax such as [:alpha:]. We also detect [= and ., which have a meaning in POSIX regular expressions, but are not implemented in l3regex. In case we see [:, grab raw characters until hopefully reaching :]. If that’s missing, or the POSIX class is unknown, abort. If all is right, add the test to the current class, with an extra \__regex_item_reverse:n for negative classes.

\cs_new_protected:Npn \__regex_compile_class_posix_test:w #1
\token_if_eq_meaning:NNT \__regex_compile_special:N #1
\str_case:nn { #2 }
{ \msg_warning:nxx { regex } POSIX-unsupported
{ posix-unsupported } { = } }
{ posix-unsupported } { . }
45.3.10 Groups and alternations

The contents of a regex group are turned into compiled code in \_\_\_regex_build_tl, which ends up with items of the form \_\_\_regex_branch:n \{(concatenation)\}. This construction is done using \tl_build... functions within a \TeX group, which automatically makes sure that options (case-sensitivity and default catcode) are reset at the end of the group. The argument \#1 is \_\_\_regex_group:nnnN or a variant thereof. A small subtlety to support \cL(abc) as a shorthand for \La\Lb\Lc: exit any pending catcode test, save the category code at the start of the group as the default catcode for that group, and make sure that the catcode is restored to the default outside the group.

\cs_new_protected:Npn \_\_\_regex_compile_group_begin:N #1
\tl_build_put_right:Nn \l__regex_build_tl { #1 { \if_false: } \fi: }
\__regex_mode_quit_c:
\tl_build_begin:N \l__regex_build_tl
\int_set_eq:NN \l__regex_default_catcodes_int \l__regex_catcodes_int
\int_incr:N \l__regex_group_level_int
\tl_build_put_right:Nn \l__regex_build_tl { \__regex_branch:n { \if_false: } \fi: }
\cs_new_protected:Npn \_\_\_regex_compile_group_end:
\if_int_compare:w \l__regex_group_level_int > \c_zero_int
\tl_build_put_right:Nn \l__regex_build_tl \{ \l__regex_build_tl \}
\tl_build_end:N \l__regex_build_tl
\exp_args:NNNx \group_end:
\tl_build_put_right:Nn \l__regex_build_tl \{ \l__regex_build_tl \}
\int_set_eq:NN \l__regex_catcodes_int \l__regex_default_catcodes_int
\exp_after:wN \_\_\_regex_compile_quantifier:w
\else:
\msg_warning:nn { regex } { extra-rparen }
\exp_after:wN \_\_\_regex_compile_raw:N \exp_after:wN ) \fi:
\}

(End definition for \_\_\_regex_compile_group_begin:N and \_\_\_regex_compile_group_end:.)

\_\_\_regex_compile_: In a class, parentheses are not special. In a catcode test inside a class, a left parenthesis gives an error, to catch [a\cL(bcd)e]. Otherwise check for a ?, denoting special groups, and run the code for the corresponding special group.

\cs_new_protected:cpn { \_\_\_regex_compile_: }
\__regex_if_in_class:TF { \_\_\_regex_compile_raw:N ( }
\if_int_compare:w \l__regex_mode_int > \c_zero_int
\c__regex_catcode_in_class_mode_int = \c__regex_catcode_in_class_mode_int
\}
\msg_error:nn { regex } { c-lparen-in-class }
\exp_after:wN \_regex_compile_raw:N \exp_after:wN ( \else:
\exp_after:wN \_regex_compile_lparen:w \fi:
\)
\cs_new_protected:Npn \_regex_compile_lparen:w #1#2#3#4
{ \_regex_two_if_eq:NNNNTF #1 #2 \_regex_compile_special:N ?
  { \cs_if_exist_use:cF
    { \_regex_compile_special_group_\token_to_str:N #4 :w }
   { \msg_warning:nnx { regex } { special-group-unknown }
     { (? #4 )
    \_regex_compile_group_begin:N \_regex_group:nnnN \_regex_compile_raw:N ? #3 #4
    \_regex_compile_group_end: }
  }
 { \_regex_compile_group_begin:N \_regex_group:nnnN #1 #2 #3 #4
  }
}
(End definition for \_regex_compile_(::)

\_regex_compile_|: In a class, the pipe is not special. Otherwise, end the current branch and open another one.
\cs_new_protected:cpn { \_regex_compile_|: }
{ \_regex_if_in_class:TF { \_regex_compile_raw:N | }
  { \tl_build_put_right:Nn \l__regex_build_tl \_regex_build_tl
    { \if_false: { \fi: } \_regex_branch:n { \if_false: } \fi: }
  }
(End definition for \_regex_compile_|::)

\_regex_compile_): Within a class, parentheses are not special. Outside, close a group.
\cs_new_protected:cpn { \_regex_compile_): }
{ \_regex_if_in_class:TF { \_regex_compile_raw:N ) }
  { \_regex_group_end: }
(End definition for \_regex_compile_):::)

\_regex_compile_special_group_::w \_regex_compile_special_group_|:w Non-capturing, and resetting groups are easy to take care of during compilation; for those groups, the harder parts come when building.
\cs_new_protected:cpn { \_regex_compile_special_group_::w }
{ \_regex_compile_group_begin:N \_regex_group_no_capture:nnnN }
\cs_new_protected:cpn { \_regex_compile_special_group_|:w }
{ \_regex_compile_group_begin:N \_regex_group_resetting:nnnN }

485
The match can be made case-insensitive by setting the option with (?i); the original behaviour is restored by (?-i). This is the only supported option.

\cs_new_protected:Npn \__regex_compile_special_group_i:w \#1\#2
{ \__regex_two_if_eq:NNNNTF \#1 \#2 \__regex_compile_special:N )

{ \__regex_item_equal:n { \__regex_item_caseless_equal:n }
\cs_set:Npn \__regex_item_range:nn { \__regex_item_caseless_range:nn }
}

{ \msg_warning:nnx { regex } { unknown-option } { (?i \#2 }
\__regex_compile_raw:N ( \__regex_compile_raw:N ? \__regex_compile_raw:N i
\#1 \#2
}

\cs_new_protected:cpn { __regex_compile_special_group_-:w } \#1\#2\#3\#4
{ \__regex_two_if_eq:NNNNTF \#1 \#2 \__regex_compile_raw:N )
\use_ii:nn

{ \__regex_item_equal:n { \__regex_item_caseful_equal:n }
\cs_set:Npn \__regex_item_range:nn { \__regex_item_caseful_range:nn }
}

{ \msg_warning:nnx { regex } { unknown-option } { (?-\#2\#4 }
\__regex_compile_raw:N ( \__regex_compile_raw:N ? \__regex_compile_raw:N -
\#1 \#2 \#3 \#4
}

\cs_new_protected:cpn { __regex_compile_special_group_-:w } \#1\#2\#3\#4
{ \__regex_two_if_eq:NNNNTF \#1 \#2 \__regex_compile_raw:N )
\use_ii:nn

{ \__regex_item_equal:n { \__regex_item_caseful_equal:n }
\cs_set:Npn \__regex_item_range:nn { \__regex_item_caseful_range:nn }
}

{ \msg_warning:nnx { regex } { unknown-option } { (?-\#2\#4 }
\__regex_compile_raw:N ( \__regex_compile_raw:N ? \__regex_compile_raw:N -
\#1 \#2 \#3 \#4
}

(End definition for \__regex_compile_special_group_i:w and \__regex_compile_special_group_-:w.)

45.3.11 Catcodes and csnames

The \c escape sequence can be followed by a capital letter representing a character category, by a left bracket which starts a list of categories, or by a brace group holding a regular expression for a control sequence name. Otherwise, raise an error.

\cs_new_protected:cpn { __regex_compile_c:c: } \{ \__regex_chk_c_allowed:T \{ \__regex_compile_c_test:NN \} \}
\cs_new_protected:Npn \__regex_compile_c_test:NN \#1\#2
{ \token_if_eq_meaning:NNTF \#1 \__regex_compile_raw:N ...

486
\int_if_exist:cTF { \c__regex_catcode_#2_int } \\
\int_set_eq:Nc \l__regex_catcodes_int \\
\l__regex_mode_int \\
= \if_case:w \l__regex_mode_int \\
\c__regex_catcode_mode_int \\
\else: \\
\c__regex_catcode_in_class_mode_int \\
\fi:
\token_if_eq_charcode:NNT C #2 { \__regex_compile_c_C:NN }
\}
\cs_if_exist_use:cF { __regex_compile_c_#2:w } 
{ 
\msg_error:nnx { regex } { c-missing-category } {#2} 
#1 #2 
}
\}
\end{definition} \__.\regex_compile_/c: and \__.\regex_compile_c_test:NN.
\\__.\regex_compile_c_C:NN
If \cC is not followed by . or (...) then complain because that construction cannot match anything, except in cases like \cC[\cCf...], where it has no effect.
\cs_new_protected:Npn \__regex_compile_c_C:NN #1#2 \\
{ 
\token_if_eq_meaning:NNTF #1 \__regex_compile_special:N \\
\token_if_eq_charcode:NNTF #2 . \\
{ \use_none:n } \\
{ \token_if_eq_charcode:NNF #2 ( ) % } \\
{ \use:n } \\
{ \msg_error:nnn { regex } { c-C-invalid } {#2} } 
#1 #2 
}\end{definition} \__.\regex_compile_c_C:NN.
\\__.\regex_compile_c_[::w
\__regex_compile_c_lbrack_loop:NN \__regex_compile_c_lbrack_add:N \__regex_compile_c_lbrack_end:
\When encountering \c[, the task is to collect uppercase letters representing character categories. First check for ^ which negates the list of category codes.
\cs_new_protected:cpn { \__regex_compile_c_[::w } #1#2 \\
{ 
\l__regex_mode_int \\
= \if_case:w \l__regex_mode_int \\
\c__regex_catcode_mode_int \\
\else: \\
\c__regex_catcode_in_class_mode_int \\
\fi:
\int_zero:N \l__regex_catcodes_int \\
\__regex_two_if_eq:NNNTF #1 #2 \__regex_compile_special:N ^ \\
{ 
\bool_set_false:N \l__regex_catcodes_bool \\
\__regex_compile_c_lbrack_loop:NN
The case of a left brace is easy, based on what we have done so far: in a group, compile the regular expression, after changing the mode to forbid nesting \c. Additionally, disable submatch tracking since groups don’t escape the scope of \c{...}.
\_\_regex_compile_{:}

We forbid unescaped left braces inside a \c{...} escape because they otherwise lead to the confusing question of whether the first right brace in \c{{}x} should end \c or whether one should match braces.

\cs_new_protected:cpn \{ __regex_compile_ \c_left_brace_str : \}
\__regex_if_in_cs:TF
\{ \msg_error:nnn { regex } { cu-lbrace } { c } \}
\{ \exp_after:wN \__regex_compile_raw:N \c_left_brace_str \}
\}

(End definition for \_\_regex_compile_c{}:)
\__regex_compile_/u:

The \u escape is invalid in classes and directly following a catcode test. Otherwise test
for a following \r (for \ur), and call an auxiliary responsible for finding the variable name.

\cs_new_protected:cpn { \__regex_compile_/u: } #1#2

\__regex_if_in_class_or_catcode:TF
\__regex_compile_raw_error:N u #1 \#2
\__regex_two_if_eq:NNNNTF #1 \#2 \__regex_compile_raw:N r
\__regex_compile_u_brace:NNN \__regex_compile_ur_end: \}
\__regex_compile_u_brace:NNN \__regex_compile_u_end: #1 \#2

(End definition for \__regex_compile_: and others.)

45.3.12 Raw token lists with \u

The \u escape is invalid in classes and directly following a catcode test. Otherwise test
for a following \r (for \ur), and call an auxiliary responsible for finding the variable name.

\cs_new_protected:cpn { \__regex_compile_/u: } #1#2

\__regex_if_in_class_or_catcode:TF
\__regex_compile_raw_error:N u #1 \#2
\__regex_two_if_eq:NNNNTF #1 \#2 \__regex_compile_raw:N r
\__regex_compile_u_brace:NNN \__regex_compile_ur_end: \}
\__regex_compile_u_brace:NNN \__regex_compile_u_end: #1 \#2

490
\__regex_compile_u_brace:NNN

This enforces the presence of a left brace, then starts a loop to find the variable name.

\cs_new:Npn \__regex_compile_u_brace:NNN #1#2#3
  {\__regex_two_if_eq:NNNNTF #2 #3 \__regex_compile_special:N \c_left_brace_str
  {\tl_set:Nn \l__regex_internal_b_tl {#1}\_kernel_tl_set:Nx \l__regex_internal_a_tl \if_false: \fi:
  \_regex_compile_u_loop:NN}
  {\msg_error:nn { regex } { u-missing-lbrace }
  \token_if_eq_meaning:NNTF #1 \__regex_compile_ur_end:
  {\_regex_compile_raw:N u \_regex_compile_raw:N r }
  {\_regex_compile_raw:N u }
  #2 #3}
}

(End definition for \_regex_compile_u_brace:NNN.)

\__regex_compile_u_loop:NN

We collect the characters for the argument of \u within an x-expanding assignment. In principle we could just wait to encounter a right brace, but this is unsafe: if the right brace was missing, then we would reach the end-markers of the regex, and continue, leading to obscure fatal errors. Instead, we only allow raw and special characters, and stop when encountering a special right brace, any escaped character, or the end-marker.

\cs_new:Npn \__regex_compile_u_loop:NN #1#2
  {\token_if_eq_meaning:NNTF #1 \_regex_compile_raw:N
  { #2 \_regex_compile_u_loop:NN }
  {\token_if_eq_meaning:NNTF #1 \_regex_compile_special:N
  {\exp_after:wN \token_if_eq_charcode:NNTF \c_right_brace_str #2
  { \if_false: \fi: \l__regex_internal_b_tl }
  { \if_charcode:w \c_left_brace_str #2
  \msg_expandable_error:nnn { regex } { cu-lbrace } { u }
  \else:
  #2
  \fi:
  \_regex_compile_u_loop:NN
  }
  }
  { \if_false: \fi: }
  \msg_error:nnx { regex } { u-missing-rbrace } {#2}
  \l__regex_internal_b_tl
  #1 #2}
}

(End definition for \_regex_compile_u_loop:NN.)
For the \texttt{$\ur{\ldots}$} construction, once we have extracted the variable’s name, we replace all groups by non-capturing groups in the compiled regex (passed as the argument of \_\_regex_compile_u:n). If that has a single branch (namely \texttt{\tl_if_empty:oTF} is false) and there is no quantifier, then simply insert the contents of this branch (obtained by \texttt{\use_ii:nn}, which is expanded later). In all other cases, insert a non-capturing group and look for quantifiers to determine the number of repetition etc.

\begin{verbatim}
\cs_new_protected:Npn \_regex_compile_u_end:
{ \group_begin:
  \cs_set:Npn \_regex_group:nnnN { \_regex_group_no_capture:nnnN } 
  \exp_args:NNx
  \group_end:
  \_regex_compile_u:n { \use:c { \l__regex_internal_a_tl } } }
\cs_new_protected:Npn \_regex_compile_u:n #1
{ \tl_if_empty:oTF \_regex_compile_u_aux:w #1 {} ? ? \q__regex_nil }
{ \_regex_compile_if_quantifier:TFw }{ \use_i:nn }
{ \tl_build_put_right:Nn \l__regex_build_tl
  { \_regex_group_no_capture:nnnN { \if_false: } \fi: #1 } \_regex_compile_quantifier:w }
\cs_new:Npn \_regex_compile_u_payload:
{ \tl_build_put_right:Nn \l__regex_build_tl { \if_false: { \fi: } } \_regex_compile_quantifier:w }
\end{verbatim}
\begin{verbatim}
{ \__regex_compile_u_payload: }
\cs_new_protected:Npn \__regex_compile_u_payload:
{
  \tl_set:Nv \l__regex_internal_a_tl { \l__regex_internal_a_tl }
  \if_int_compare:w \l__regex_mode_int = \c__regex_outer_mode_int
  \__regex_compile_u_not_cs:
  \else:
  \__regex_compile_u_in_cs:
  \fi:
}

(End definition for \__regex_compile_u_end: and \__regex_compile_u_payload:.)

\__regex_compile_u_in_cs:
When \u appears within a control sequence, we convert the variable to a string with escaped spaces. Then for each character insert a class matching exactly that character, once.
\begin{verbatim}
\cs_new_protected:Npn \__regex_compile_u_in_cs:
{
  \__kernel_tl_gset:Nx \g__regex_internal_tl
    { \exp_args:No \__kernel_str_to_other_fast:n \l__regex_internal_a_tl }
  \tl_build_put_right:Nx \l__regex_build_tl
    { \tl_map_function:NN \g__regex_internal_tl \__regex_compile_u_in_cs_aux:n }
}
\cs_new:Npn \__regex_compile_u_in_cs_aux:n #1
{
  \__regex_class:NnnnN \c_true_bool
    { \__regex_item_caseful_equal:n { \int_value:w '#1 } } { 1 } { 0 } \c_false_bool
}

(End definition for \__regex_compile_u_in_cs:)
\end{verbatim}

\__regex_compile_u_not_cs:
In mode 0, the \u escape adds one state to the NFA for each token in \l__regex_internal_a_tl. If a given \textit{token} is a control sequence, then insert a string comparison test, otherwise, \__regex_item_exact:nn which compares catcode and character code.
\begin{verbatim}
\cs_new_protected:Npn \__regex_compile_u_not_cs:
{
  \tl_analysis_map_inline:Nn \l__regex_internal_a_tl
  { \tl_build_put_right:Nx \l__regex_build_tl
    { \tl_map_function:NN \g__regex_internal_tl \__regex_compile_u_not_cs_aux:n \#1 }
  }
}
\cs_new:Npn \__regex_compile_u_not_cs_aux:n #1
{
  \__regex_class:NnnnN \c_true_bool
    { \__regex_item_exact_cs:n \exp_after:wN \cs_to_str:N ##1 } \else:
    \__regex_item_exact:nn \exp_after:wN \cs_to_str:N \#1 }
}

(End definition for \__regex_compile_u_not_cs:)
\end{verbatim}
\end{verbatim}
\_\_regex_compile_/K:

The \texttt{/K} control sequence is currently the only “command”, which performs some action, rather than matching something. It is allowed in the same contexts as \texttt{\b}. At the compilation stage, we leave it as a single control sequence, defined later.

Before showing a regex we check that it is “clean” in the sense that it has the correct internal structure. We do this (in the implementation of \texttt{\regex_show:N} and \texttt{\regex_\-log:N}) by comparing it with a cleaned-up version of the same regex. Along the way we also need similar functions for other types: all \texttt{\_\_regex_clean\_⟨type⟩:n} functions produce valid \langle\texttt{type}\rangle tokens (bool, explicit integer, etc.) from arbitrary input, and the output coincides with the input if that was valid.
\cs_new:Npn \__regex_clean_regex_loop:w #1 \__regex_branch:n #2
{
\quark_if_recursion_tail_stop:n {#2}
\__regex_branch:n { \__regex_clean_branch:n {#2} }
\__regex_clean_regex_loop:w
}
\cs_new:Npn \__regex_clean_branch:n #1
{
\__regex_clean_branch_loop:n #1

? ? ? ? ? \prg_break_point:
}
\cs_new:Npn \__regex_clean_branch_loop:n #1
{
\tl_if_single:nF {#1} { \prg_break: }
\token_case_meaning:NnF #1
{
\__regex_command_K: { #1 \__regex_clean_branch_loop:n }
\__regex_assertion:Nn { #1 \__regex_clean_assertion:Nn }
\__regex_class:NnnnN { #1 \__regex_clean_class:NnnnN }
\__regex_group:nnnN { #1 \__regex_clean_group:nnnN }
\__regex_group_no_capture:nnnN { #1 \__regex_clean_group:nnnN }
\__regex_group_resetting:nnnN { #1 \__regex_clean_group:nnnN }
}
{ \prg_break: }
}
\cs_new:Npn \__regex_clean_assertion:Nn #1#2
{
\__regex_clean_bool:n {#1}
\tl_if_single:nF {#2} { { \__regex_A_test: } \prg_break: }
\token_case_meaning:NnTF #2
{
\__regex_A_test: { }
\__regex_G_test: { }
\__regex_Z_test: { }
\__regex_b_test: { }
}
{ \prg_break: }
}
\cs_new:Npn \__regex_clean_class:NnnnN #1#2#3#4#5
{
\__regex_clean_bool:n {#1}
\__regex_clean_class:n {#2}
{ \int_max:nn { 0 } { \__regex_clean_int:n {#3} } }
{ \int_max:nn { -1 } { \__regex_clean_int:n {#4} } }
\__regex_clean_bool:n {#5}
\__regex_clean_branch_loop:n
}
\cs_new:Npn \__regex_clean_group:nnnN #1#2#3#4

495
\__regex_clean_regex:n {\#1} \\
\__regex_clean_int:n {\#2} \\
\__regex_clean_int:n {\#3} \\
\__regex_clean_bool:n {\#4} \\
\__regex_clean_branch_loop:n \\
\__regex_clean_class:n #1 \\
\__regex_clean_class_loop:nnn ????

\cs_new:Npn \__regex_clean_class_loop:nnn #1 #2 #3 
\tl_if_single:nF {#1} { \prg_break: } 
\token_case_meaning:NnTF #1 
\__regex_item_cs:n { #1 { \__regex_clean_regex:n {#2} } } 
\__regex_item_exact_cs:n { #1 { \__regex_clean_exact_cs:n {#2} } } 
\__regex_item_caseful_equal:n { #1 { \__regex_clean_int:n {#2} } } 
\__regex_item_caseless_equal:n { #1 { \__regex_clean_int:n {#2} } } 
\__regex_item_reverse:n { #1 { \__regex_clean_class:n {#2} } } 
\__regex_clean_class_loop:nnn {#3} 
\token_case_meaning:NnTF #1 
\__regex_item_exact:nn { } 
\__regex_item_caseful_range:nn { } 
\__regex_item_caseless_range:nn { } 
\__regex_clean_int:n {#2} 
\__regex_clean_int:n {#3} 
\prg_break: 

\cs_new:Npn \__regex_clean_exact_cs:n #1 
\exp_last_unbraced:Nf \use_none:n 
\__regex_clean_exact_cs:w #1 
\scan_stop: \q_recursion_tail \scan_stop: \q_recursion_stop 
\prg_break: 

\cs_new:Npn \__regex_clean_exact_cs:w #1 \scan_stop: 

Within a group and within \_tl_build_begin:N ... \_tl_build_end:N we redefine all the function that can appear in a compiled regex, then run the regex. The result stored in \_l__regex_internal_a_tl is then meant to be shown.

\cs_new_protected:Npn \_regex_show:N #1
{\group_begin: \_regex_branch:n {#1}
  \tl_build_begin:N \l__regex_build_tl
  \cs_set_protected:Npn \_regex_branch:n { #1 }
  \__regex_clean_exact_cs:w
  \__regex_clean_bool:n {#1}
  \scan_stop: \tl_to_str:n {#1}
}

(End definition for \_regex_clean_bool:n and others.)

\_regex_show:N
\cs_set_protected:Npn \_regex_item_catcode:nT
\{ \_regex_show_item_catcode:NnT \c_true_bool \}
\cs_set_protected:Npn \_regex_item_catcode_reverse:nT
\{ \_regex_show_item_catcode:NnT \c_false_bool \}
\cs_set_protected:Npn \_regex_item_reverse:n
\{ \_regex_show_scope:nn { Reversed~match } \}
\cs_set_protected:Npn \_regex_item_exact:nn ##1##2
\{ \_regex_show_one:n { char~\_regex_show_char:n{##2},~catcode~##1 } \}
\cs_set_eq:NN \_regex_item_exact_cs:n \_regex_show_item_exact_cs:n
\cs_set_protected:Npn \_regex_item_cs:n
\{ \_regex_show_scope:nn { control~sequence } \}
\cs_set:cpn { __regex_prop_.: } { \_regex_show_one:n { any~token } }
\seq_clear:N \l__regex_show_prefix_seq
\__regex_show_push:n { ~ }
\cs_if_exist_use:N #1
\tl_build_end:N \l__regex_build_tl
\exp_args:NNNo
\group_end:
\tl_set:Nn \l__regex_internal_a_tl { \l__regex_build_tl }

(End definition for \_regex_show:N.)

\_regex_show_char:n  Show a single character, together with its ascii representation if available. This could be
extended to beyond ascii. It is not ideal for parentheses themselves.
\cs_new:Npn \_regex_show_char:n #1
\{ \int_eval:n {#1}
\int_compare:nT { 32 <= #1 <= 126 }
\{ ~ ( \char_generate:nn {#1} {12} ) \}
\}

(End definition for \_regex_show_char:n.)

\_regex_show_one:n  Every part of the final message go through this function, which adds one line to the
output, with the appropriate prefix.
\cs_new_protected:Npn \_regex_show_one:n #1
\{ \int_incr:N \l__regex_show_lines_int
\tl_build_put_right:Nx \l__regex_build_tl
\{ \exp_not:N \iow_newline:
\seq_map_function:NN \l__regex_show_prefix_seq \use:n #1
\}
\}

(End definition for \_regex_show_one:n.)

\_regex_show_push:n  Enter and exit levels of nesting. The scope function prints its first argument as an
“introduction”, then performs its second argument in a deeper level of nesting.
\cs_new_protected:Npn \_regex_show_push:n #1
\{ \seq_put_right:Nx \l__regex_show_prefix_seq { #1 - } \}

(End definition for \_regex_show_one:n.)
\cs_new_protected:Npn \_\regex_show_pop: 
  \seq_pop_right:NN \l_\regex_show_prefix_seq \l_\regex_internal_a_tl 
\cs_new_protected:Npn \_\regex_show_scope:nn #1 #2 
  \__regex_show_one:n {#1} \_\regex_show_push:n { ~ } \__regex_show_pop: 
\__regex_show_group_aux:nnnnN

We display all groups in the same way, simply adding a message, (no capture) or (resetting), to special groups. The odd \use_ii:nn avoids printing a spurious +-branch for the first branch.

\cs_new_protected:Npn \_\regex_show_group_aux:nnnnN #1#2#3#4#5 
  \__regex_show_one:n { ,-group~begin #1 } \_\regex_show_push:n { | } \use_ii:nn #2 \_\regex_show_pop: 
  \__regex_show_one:n \_\regex_msg_repeated:nnN {#3} {#4} #5 
\__regex_show_class:NnnnN

I’m entirely unhappy about this function: I couldn’t find a way to test if a class is a single test. Instead, collect the representation of the tests in the class. If that had more than one line, write Match or Don’t match on its own line, with the repeating information if any. Then the various tests on lines of their own, and finally a line. Otherwise, we need to evaluate the representation of the tests again (since the prefix is incorrect). That’s clunky, but not too expensive, since it’s only one test.

\cs_set:Npn \_\regex_show_class:NnnnN #1#2#3#4#5 
  \group_begin: \tl_build_begin:N \l_\regex_build_tl \int_zero:N \l_\regex_show_lines_int \_\regex_show_push:n {~} \__\regex_show_one:n { \bool_if:NTF #1 { Fail } { Pass } } \group_end: 
  \__\regex_show_one:n \bool_if:nTF { \int_compare:nTF { \l_\regex_show_lines_int = 0 } { \__\regex_show_push:n {~} \__\regex_show_one:n { \bool_if:nTF { #1 \&\& \int_compare_p:n { \l_\regex_show_lines_int = 1 } } { \group_end: \_\regex_show_one:n \tl_build_put_right:Nn \l_\regex_build_tl { \_\regex_msg_repeated:mnN {#3} {#4} #5 } } \group_end: \tl_build_put_right:Nn \l_\regex_build_tl { \_\regex_msg_repeated:mnN {#3} {#4} #5 } } } 
\_\regex_show_group_class:NnnnN
Building

45.4.1 Variables used while building

The last state that was allocated is \l__regex_max_state_int - 1, so that \l__regex_max_state_int always points to a free state. The min_state variable is 1 to begin with, but gets shifted in nested calls to the matching code, namely in \c{...} constructions.
Alternatives are implemented by branching from a left state into the various choices, then merging those into a right state. We store information about those states in two sequences. Those states are also used to implement group quantifiers. Most often, the left and right pointers only differ by 1.

\l__regex_left_state_int \l__regex_right_state_int

\l__regex_left_state_seq \l__regex_right_state_seq

\l__regex_capturing_group_int

\l__regex_capturing_group_int is the next ID number to be assigned to a capturing group. This starts at 0 for the group enclosing the full regular expression, and groups are counted in the order of their left parenthesis, except when encountering resetting groups.

\l__regex_capturing_group_int

\l__regex_capturing_group_int

45.4.2 Framework

This phase is about going from a compiled regex to an NFA. Each state of the NFA is stored in a \toks. The operations which can appear in the \toks are

- \__regex_action_start_wildcard:{boolean} inserted at the start of the regular expression, where a true {boolean} makes it unanchored.
- \__regex_action_success: marks the exit state of the NFA.
- \__regex_action_cost:n {⟨shift⟩} is a transition from the current {⟨state⟩} to {⟨state⟩ + ⟨shift⟩}, which consumes the current character: the target state is saved and will be considered again when matching at the next position.
- \__regex_action_free:n{⟨shift⟩}, and \__regex_action_free_group:n{⟨shift⟩} are free transitions, which immediately perform the actions for the state {⟨state⟩ + ⟨shift⟩} of the NFA. They differ in how they detect and avoid infinite loops. For now, we just need to know that the group variant must be used for transitions back to the start of a group.
- \__regex_action_submatch:n{⟨group⟩} {⟨key⟩} where the {⟨key⟩} is < or > for the beginning or end of group numbered {⟨group⟩}. This causes the current position in the query to be stored as the {⟨key⟩} submatch boundary.
- One of these actions, within a conditional.

We strive to preserve the following properties while building.

- The current capturing group is capturing_group − 1, and if a group opened now it would be labelled capturing_group.
The last allocated state is $$\text{max}\_\text{state} - 1$$, so $$\text{max}\_\text{state}$$ is a free state.

- The left state points to a state to the left of the current group or of the last class.

- The right state points to a newly created, empty state, with some transitions leading to it.

- The left/right sequences hold a list of the corresponding end-points of nested groups.

The n-type function first compiles its argument. Reset some variables. Allocate two states, and put a wildcard in state 0 (transitions to state 1 and 0 state). Then build the regex within a (capturing) group numbered 0 (current value of $$\text{capturing}\_\text{group}$$). Finally, if the match reaches the last state, it is successful. A false boolean for argument #1 for the auxiliaries will suppress the wildcard and make the match anchored: used for $$\text{peek}\_\text{regex}\_\text{NTF}$$ and similar.

```latex
\begin{verbatim}
\cs_new_protected:Npn \__regex_build:n { \__regex_build_aux:Nn \c_true_bool }
\cs_new_protected:Npn \__regex_build:N { \__regex_build_aux:NN \c_true_bool }
\cs_new_protected:Npn \__regex_build_aux:Nn #1#2 { \__regex_compile:n {#2} \__regex_build_aux:NN #1 \l__regex_internal_regex }
\cs_new_protected:Npn \__regex_build_aux:NN #1#2 { \__regex_standard_escapechar: \int_zero:N \l__regex_capturing_group_int \int_set_eq:NN \l__regex_max_state_int \l__regex_min_state_int \__regex_build_new_state: \__regex_build_new_state: \__regex_toks_put_right:Nn \l__regex_left_state_int { \__regex_action_start_wildcard:N #1 } \__regex_group:nnnN {#2} { 1 } { 0 } \c_false_bool \__regex_toks_put_right:Nn \l__regex_right_state_int { \__regex_action_success: } }
\end{verbatim}
```

(The definition for \__regex_build:n and others.)

\begin{verbatim}
\g__regex_case_int \int_new:N \g__regex_case_int
\l__regex_case_max_group_int \int_new:N \l__regex_case_max_group_int
\end{verbatim}

(The definition for \g__regex_case_int.)
See \_regex_build:n, but with a loop.

\cs_new_protected:Npn \_regex_case_build:n \#1
\{ \_regex_case_build_aux:Nn \c_true_bool {\#1}
\int_gzero:N \g__regex_case_int
\}
\cs_generate_variant:Nn \_regex_case_build:n { x }
\cs_new_protected:Npn \_regex_case_build_aux:Nn #1#2
{ \_regex_standard_escapechar:
\int_set_eq:NN \l__regex_max_state_int \l__regex_min_state_int
\_regex_build_new_state:
\_regex_build_new_state:
\_regex_toks_put_right:Nn \l__regex_left_state_int
{ \_regex_action_start_wildcard:N #1 }
\%
\_regex_build_new_state:
\_regex_toks_put_left:Nx \l__regex_left_state_int
{ \_regex_action_submatch:nN { 0 } < }
\_regex_push_lr_states:
\int_zero:N \l__regex_case_max_group_int
\int_gzero:N \g__regex_case_int
\tl_map_inline:nn {#2}
{ \int_gincr:N \g__regex_case_int
\_regex_case_build_loop:n {##1}
}
\int_set_eq:NN \l__regex_capturing_group_int \l__regex_case_max_group_int
\_regex_pop_lr_states:
\}
\cs_new_protected:Npn \_regex_case_build_loop:n \#1
{ \int_set:Nn \l__regex_capturing_group_int { 1 }
\_regex_compile_use:n {#1}
\int_set:Nn \l__regex_case_max_group_int
{ \int_max:nn { \l__regex_case_max_group_int } { \l__regex_capturing_group_int } }
\seq_pop:NN \l__regex_right_state_seq \l__regex_internal_a_tl
\int_set:Nn \l__regex_right_state_int \l__regex_internal_a_tl
\_regex_toks_put_left:Nx \l__regex_left_state_int
\_regex_action_submatch:nN \l__regex_right_state_int
\{ \_regex_action_submatch:nN \l__regex_right_state_int
\int_gset:Nn \g__regex_case_int \l__regex_case_max_group_int
\_regex_action_success:
\}
\_regex_toks_clear:N \l__regex_max_state_int
\seq_pop:No \l__regex_right_state_seq
\{ \int_use:N \g__regex_case_int \}
\_regex_action_success:
\}
\_regex_toks_clear:N \l__regex_max_state_int
\seq_pop:No \l__regex_right_state_seq
\{ \int_use:N \g__regex_case_int \}
\int_incr:N \l__regex_max_state_int
\}

503
The matching code relies on some global intarray variables, but only uses a range of their entries. Specifically,

- \texttt{\_\_regex_state_active_intarray} from \texttt{\_\_regex_min_state_int} to \texttt{\_\_regex_max_state_int}.

Here, in this nested call to the matching code, we need the new versions of this range to involve completely new entries of the intarray variables, so we begin by setting (the new) \texttt{\_\_regex_min_state_int} to (the old) \texttt{\_\_regex_max_state_int} to use higher entries.

When using a regex to match a \texttt{cs}, we don’t insert a wildcard, we anchor at the end, and since we ignore submatches, there is no need to surround the expression with a group. However, for branches to work properly at the outer level, we need to put the appropriate left and right states in their sequence.

```latex
\cs_new_protected:Npn \_\_regex_build_for_cs:n #1
\int_set_eq:NN \l__regex_min_state_int \l__regex_max_state_int
\_\_regex_build_new_state:
\_\_regex_push_lr_states:
#1
\_\_regex_pop_lr_states:
\_\_regex_toks_put_right:Nn \l__regex_right_state_int
\if_int_compare:w -2 = \l__regex_curr_char_int
\exp_after:wN \_\_regex_action_success:
\fi:
```

(End definition for \_\_regex_build_for_cs:n.)

45.4.3 Helpers for building an nfa

When building the regular expression, we keep track of pointers to the left-end and right-end of each group without help from \LaTeX’s grouping.

```latex
\cs_new_protected:Npn \_\_regex_push_lr_states:
\seq_push:No \l__regex_left_state_seq
\seq_push:No \l__regex_left_state_int
\seq_push:No \l__regex_right_state_seq
\seq_push:No \l__regex_right_state_int
```

(End definition for \_\_regex_push_lr_states.)
Add a transition from `#2` to `#3` using the function `#1`. The left function is used for higher priority transitions, and the right function for lower priority transitions (which should be performed later). The signatures differ to reflect the differing usage later on. Both functions could be optimized.

```latex
\cs_new_protected:Npn \__regex_build_transition_left:NNN #1#2#3
\{ \__regex_toks_put_left:Nx #2 { #1 \{ \int_eval:n \{ #3 - #2 \} \} }\}
\cs_new_protected:Npn \__regex_build_transition_right:nNn #1#2#3
\{ \__regex_toks_put_right:Nx #2 { #1 \{ \int_eval:n \{ #3 - #2 \} \} }\}
```

Add a new empty state to the NFA. Then update the left, right, and max states, so that the right state is the new empty state, and the left state points to the previously “current” state.

```latex
\cs_new_protected:Npn \__regex_build_new_state:
\{ \__regex_toks_clear:N \l__regex_max_state_int
\int_set_eq:NN \l__regex_left_state_int \l__regex_right_state_int
\int_set_eq:NN \l__regex_right_state_int \l__regex_max_state_int
\int_incr:N \l__regex_max_state_int
\}
```

This function creates a new state, and puts two transitions starting from the old current state. The order of the transitions is controlled by `#1`, true for lazy quantifiers, and false for greedy quantifiers.

```latex
\cs_new_protected:Npn \__regex_build_transitions_lazyness:NNNNN #1#2#3#4#5
\{ \__regex_build_new_state:
\__regex_toks_put_right:Nx \l__regex_left_state_int
\if_meaning:w \c_true_bool #1
\#2 \{ \int_eval:n \{ #3 - \l__regex_left_state_int \} \}
\#4 \{ \int_eval:n \{ #4 - \l__regex_left_state_int \} \}
\else:\n\#4 \{ \int_eval:n \{ #5 - \l__regex_left_state_int \} \}
\#2 \{ \int_eval:n \{ #3 - \l__regex_left_state_int \} \}
\fi:\n\}
```

### 45.4.4 Building classes

The arguments are: `(boolean) ` `{(tests)}` `{(min)}` `{(more)}` `{(lazyness)}`. First store the tests with a trailing `\__regex_action_cost:n`, in the true branch of `\__regex_break_point:TF` for positive classes, or the false branch for negative classes. The integer `(more)` is 0 for fixed repetitions, −1 for unbounded repetitions, and `(max) – (min)` for a range of repetitions.
\cs_new_protected:Npn \__regex_class:NnnnN #1#2#3#4#5
{
\cs_set:Npx \__regex_tests_action_cost:n ##1
{
\exp_not:n \exp_not:n {#2}
\bool_if:NTF #1
{
{ \__regex_break_point:TF { \__regex_action_cost:n {##1} } { } }
{ \__regex_break_point:TF { } { \__regex_action_cost:n {##1} } }
}
\if_case:w - #4 \exp_stop_f:
\__regex_class_repeat:n {#3}
or: \__regex_class_repeat:nN {#3} #5
\else: \__regex_class_repeat:nnN {#3} {#4} #5
\fi:
}
\cs_new:Npn \__regex_tests_action_cost:n { \__regex_action_cost:n }

(End definition for \__regex_class:NnnnN and \__regex_tests_action_cost:n.)

\__regex_class_repeat:n This is used for a fixed number of repetitions. Build one state for each repetition, with a
transition controlled by the tests that we have collected. That works just fine for \#1 = 0
repetitions: nothing is built.
\cs_new_protected:Npn \__regex_class_repeat:n #1
{
\prg_replicate:nn {#1}
{\__regex_build_new_state:
\__regex_build_transition_right:nNn \__regex_tests_action_cost:n
\l__regex_left_state_int \l__regex_right_state_int}
}

(End definition for \__regex_class_repeat:n.)

\__regex_class_repeat:nN This implements unbounded repetitions of a single class (e.g. the * and + quantifiers). If
the minimum number \#1 of repetitions is 0, then build a transition from the current state
to itself governed by the tests, and a free transition to a new state (hence skipping the
tests). Otherwise, call \__regex_class_repeat:n for the code to match \#1 repetitions,
and add free transitions from the last state to the previous one, and to a new one. In
both cases, the order of transitions is controlled by the lazyness boolean \#2.
\cs_new_protected:Npn \__regex_class_repeat:nN #1#2
{
\if_int_compare:w \#1 = \c_zero_int
\__regex_build_transitions_lazyness:NNNNN #2
\__regex_action_free:n \l__regex_right_state_int
\__regex_tests_action_cost:n \l__regex_left_state_int
\else:
\__regex_class_repeat:n {#1}
\int_set_eq:NN \l__regex_internal_a_int \l__regex_left_state_int
\__regex_build_transitions_lazyness:NNNNN #2
\__regex_action_free:n \l__regex_right_state_int
\__regex_action_free:n \l__regex_internal_a_int
\fi:
}
We want to build the code to match from \#1 to \#1 + \#2 repetitions. Match \#1 repetitions (can be 0). Compute the final state of the next construction as \(a\). Build \#2 > 0 states, each with a transition to the next state governed by the tests, and a transition to the final state \(a\). The computation of \(a\) is safe because states are allocated in order, starting from \texttt{max_state}.

\begin{verbatim}
\cs_new_protected:Npn \__regex_class_repeat:nnN \#1\#2\#3
\{
\__regex_class_repeat:n \#1\n\int_set:Nn \l__regex_internal_a_int\{ \l__regex_max_state_int + \#2 - 1 \}
\prg_replicate:nn { \#2 }
\{
\__regex_build_transitions_lazyness:NNNNN \#3
\__regex_action_free:n \l__regex_internal_a_int
\__regex_tests_action_cost:n \l__regex_right_state_int
\}
\}
\end{verbatim}

(End definition for \__regex_class_repeat:nnN.)

\begin{verbatim}
\__regex_group_aux:nnnnN \Arguments:\{\texttt{\langle label\rangle}\}\{\texttt{\langle contents\rangle}\}\{\texttt{\langle min\rangle}\}\{\texttt{\langle more\rangle}\}\{\texttt{\langle lazyness\rangle}\}. If \texttt{\langle min\rangle} is 0, we need to add a state before building the group, so that the thread which skips the group does not also set the start-point of the submatch. After adding one more state, the \texttt{left_state} is the left end of the group, from which all branches stem, and the \texttt{right_state} is the right end of the group, and all branches end their course in that state. We store those two integers to be queried for each branch, we build the NFA states for the contents \#2 of the group, and we forget about the two integers. Once this is done, perform the repetition: either exactly \#3 times, or \#3 or more times, or between \#3 and \#3 + \#4 times, with \texttt{lazyness} \#5. The \texttt{\langle label\rangle} \#1 is used for submatch tracking. Each of the three auxiliaries expects \texttt{left_state} and \texttt{right_state} to be set properly.

\begin{verbatim}
\cs_new_protected:Npn \__regex_group_aux:nnnnN \#1\#2\#3\#4\#5
\{
\if_int_compare:w \#3 = \c_zero_int
\__regex_build_new_state:\n\__regex_build_transition_right:nNn \__regex_action_free_group:n \l__regex_left_state_int \l__regex_right_state_int
\fi:\n\__regex_build_new_state:\n\__regex_push_lr_states:\n\#2\n\__regex_pop_lr_states:\n\if_case:w \#4 \exp_stop_f:\n\__regex_group_repeat:nn \#1 \#3
\or: \__regex_group_repeat:nnnN \#1 \#3 \#5
\else: \__regex_group_repeat:nnnnN \#1 \#3 \#4 \#5
\fi:\n\}
\end{verbatim}

(End definition for \__regex_group_aux:nnnnN.)

45.4.5 Building groups

Arguments: \{\texttt{\langle label\rangle}\}\{\texttt{\langle contents\rangle}\}\{\texttt{\langle min\rangle}\}\{\texttt{\langle more\rangle}\}\{\texttt{\langle lazyness\rangle}\}. If \texttt{\langle min\rangle} is 0, we need to add a state before building the group, so that the thread which skips the group does not also set the start-point of the submatch. After adding one more state, the \texttt{left_state} is the left end of the group, from which all branches stem, and the \texttt{right_state} is the right end of the group, and all branches end their course in that state. We store those two integers to be queried for each branch, we build the NFA states for the contents \#2 of the group, and we forget about the two integers. Once this is done, perform the repetition: either exactly \#3 times, or \#3 or more times, or between \#3 and \#3 + \#4 times, with \texttt{lazyness} \#5. The \texttt{\langle label\rangle} \#1 is used for submatch tracking. Each of the three auxiliaries expects \texttt{left_state} and \texttt{right_state} to be set properly.
Hand to \cs{__regex_group:nnnN} the label of that group (expanded), and the group itself, with some extra commands to perform.

```latex
\cs_new_protected:Npn \__regex_group:nnnN #1
\exp_args:No \__regex_group_aux:nnnnN
\{ \int_use:N \l__regex_capturing_group_int \}
\int_incr:N \l__regex_capturing_group_int
\#1
\}
\cs_new_protected:Npn \__regex_group_no_capture:nnnN
\{ \__regex_group_aux:nnnnN \{-1\} \}
```

(End definition for \cs{__regex_group:nnnN} and \cs{__regex_group_no_capture:nnnN}.)

Again, hand the label −1 to \cs{__regex_group_aux:nnnnN}, but this time we work a little bit harder to keep track of the maximum group label at the end of any branch, and to reset the group number at each branch. This relies on the fact that a compiled regex always is a sequence of items of the form \cs{__regex_branch:n} \{\textit{branch}\}.

```latex
\cs_new_protected:Npn \__regex_branch:n #1\__regex_group_reseting:nnnN \#1
\{ \__regex_group_aux:nnnnN \{-1\}
\exp_args:Noo \__regex_group_reseting_loop:nnNn
\{ \int_use:N \l__regex_capturing_group_int \}
\int_set:Nn \l__regex_capturing_group_int {#1}
\use_none:nn \{ \int_set:Nn \l__regex_capturing_group_int {#1} \}
\{ \int_max:nn {#1} \{ \int_use:N \l__regex_capturing_group_int \} \}
\{ \int_use:N \l__regex_capturing_group_int \}
\exp_args:Nf \__regex_group_reseting_loop:nnNn
\{ \int_max:nn {#1} \{ \int_use:N \l__regex_capturing_group_int \} \}
\}
```

(End definition for \cs{__regex_group_reseting:nnnN} and \cs{__regex_group_reseting_loop:nnNn}.)

Add a free transition from the left state of the current group to a brand new state, starting point of this branch. Once the branch is built, add a transition from its last state to the right state of the group. The left and right states of the group are extracted from the relevant sequences.

```latex
\cs_new_protected:Npn \__regex_branch:n \#1\__regex_build_new_state:
\seq_get:NN \l__regex_left_state_seq \l__regex_internal_a_tl
\int_set:Nn \l__regex_left_state_int \l__regex_internal_a_tl
\\int_set:Nn \l__regex_left_state_seq \l__regex_internal_a_tl
\\int_max:nn \{\#1\} \{ \l__regex_capturing_group_int \}
\}
```

508
This function is called to repeat a group a fixed number of times \#2; if this is 0 we remove the group altogether (but don’t reset the capturing_group label). Otherwise, the auxiliary \__regex_group_repeat_aux:n copies \#2 times the \toks for the group, and leaves internal_a pointing to the left end of the last repetition. We only record the submatch information at the last repetition. Finally, add a state at the end (the transition to it has been taken care of by the replicating auxiliary).

\__regex_group_repeat:nn
\__regex_group_submatches:nNN
\__regex_group_repeat_aux:n

This inserts in states \#2 and \#3 the code for tracking submatches of the group \#1, unless inhibited by a label of −1.

Here we repeat \toks ranging from left_state to max_state, \#1 > 0 times. First add a transition so that the copies “chain” properly. Compute the shift c between the original copy and the last copy we want. Shift the right_state and max_state to their final values. We then want to perform c copy operations. At the end, b is equal to the max_state, and a points to the left of the last copy of the group.
This function is called to repeat a group at least \( n \) times; the case \( n = 0 \) is very different from \( n > 0 \). Assume first that \( n = 0 \). Insert submatch tracking information at the start and end of the group, add a free transition from the right end to the “true” left state \( a \) (remember: in this case we had added an extra state before the left state). This forms the loop, which we break away from by adding a free transition from \( a \) to a new state.

Now consider the case \( n > 0 \). Repeat the group \( n \) times, chaining various copies with a free transition. Add submatch tracking only to the last copy, then add a free transition from the right end back to the left end of the last copy, either before or after the transition to move on towards the rest of the NFA. This transition can end up before submatch tracking, but that is irrelevant since it only does so when going again through the group, recording new matches. Finally, add a state; we already have a transition pointing to it from \( \text{\texttt{\_\_regex_group_repeat\_aux:n}} \).
We wish to repeat the group between #2 and #2 + #3 times, with a laziness controlled by #4. We insert submatch tracking up front: in principle, we could avoid recording submatches for the first #2 copies of the group, but that forces us to treat specially the case #2 = 0. Repeat that group with submatch tracking #2 + #3 times (the maximum number of repetitions). Then our goal is to add #3 transitions from the end of the #2-th group, and each subsequent groups, to the end. For a lazy quantifier, we add those transitions to the left states, before submatch tracking. For the greedy case, we add the transitions to the right states, after submatch tracking and the transitions which go on with more repetitions. In the greedy case with #2 = 0, the transition which skips over all copies of the group must be added separately, because its starting state does not follow the normal pattern: we had to add it “by hand” earlier.

\cs_new_protected:Npn \__regex_group_repeat:nnnN #1#2#3#4 
{
  \__regex_group_submatches:nNN {#1}
  \l__regex_left_state_int \l__regex_right_state_int
  \__regex_group_repeat_aux:n { #2 + #3 }
  \if_meaning:w \c_true_bool #4
    \int_set_eq:NN \l__regex_left_state_int \l__regex_max_state_int
    \prg_replicate:nn { #3 }
    { \int_sub:Nn \l__regex_left_state_int
      { \l__regex_internal_b_int - \l__regex_internal_a_int } \__regex_build_transition_left:NNN \__regex_action_free:n \l__regex_left_state_int \l__regex_max_state_int
    }
  \else:
    \prg_replicate:nn { #3 - 1 }
    { \int_sub:Nn \l__regex_right_state_int
      { \l__regex_internal_b_int - \l__regex_internal_a_int } \__regex_build_transition_right:nNn \__regex_action_free:n \l__regex_right_state_int \l__regex_max_state_int
    }
  \fi:
  \if_int_compare:w #2 = \c_zero_int
    \int_set:Nn \l__regex_right_state_int
    \l__regex_left_state_int - 1
  \else:
    \int_sub:Nn \l__regex_right_state_int
    { \l__regex_internal_b_int - \l__regex_internal_a_int } \__regex_build_transition_right:nNn \__regex_action_free:n \l__regex_right_state_int \l__regex_max_state_int
  \fi:
}\end{definition}
45.4.6 Others

Usage: \_\_regex\_assertion:NN (\textit{boolean}) \{\textit{test}\}, where the \textit{test} is either of the two other functions. Add a free transition to a new state, conditionally to the assertion test. The \_\_regex\_b\_test: test is used by the \texttt{\}\texttt{\}} and \texttt{\}\texttt{\}} escape: check if the last character was a word character or not, and do the same to the current character. The boundary-markers of the string are non-word characters for this purpose.

\begin{verbatim}
\cs_new_protected:Npn \_\_regex\_assertion:NN #1#2
\{\_\_regex\_build\_new\_state: \_\_regex\_toks\_put\_right:Nx \l__regex\_left\_state\_int
\{ \exp_not:n {#2} \_\_regex\_break\_point:TF \bool_if:NF #1 { { } } \{ \_\_regex\_action\_free:n \{ \int_eval:n \{ \l\_\_regex\_right\_state\_int - \l\_\_regex\_left\_state\_int \} \} } \bool_if:NT #1 { { } } \}
\}
\cs_new_protected:Npn \_\_regex\_b\_test:
\{ \group_begin: \_\_int\_set\_eq:NN \_\_regex\_curr\_char\_int \_\_regex\_last\_char\_int \_\_regex\_prop\_w:\ 
\_\_regex\_break\_point:TF \{ \group_end: \_\_regex\_item\_reverse:n \_\_regex\_prop\_w: \} \{ \group_end: \_\_regex\_prop\_w: \} \}
\cs_new_protected:Npn \_\_regex\_Z\_test:
\{ \_\_if\_int\_compare:w -2 = \_\_regex\_curr\_char\_int \_\_exp\_after:wN \_\_regex\_break\_true:w \_\_fi: \}
\cs_new_protected:Npn \_\_regex\_A\_test:
\{ \_\_if\_int\_compare:w -2 = \_\_regex\_last\_char\_int \_\_exp\_after:wN \_\_regex\_break\_true:w \_\_fi: \}
\cs_new_protected:Npn \_\_regex\_G\_test:
\{ \_\_if\_int\_compare:w \_\_regex\_curr\_pos\_int = \_\_regex\_start\_pos\_int
\end{verbatim}
45.5 Matching

We search for matches by running all the execution threads through the NFA in parallel, reading one token of the query at each step. The NFA contains “free” transitions to other states, and transitions which “consume” the current token. For free transitions, the instruction at the new state of the NFA is performed immediately. When a transition consumes a character, the new state is appended to a list of “active states”, stored in \texttt{\g__regex_thread_info_intarray} (together with submatch information): this thread is made active again when the next token is read from the query. At every step (for each token in the query), we unpack that list of active states and the corresponding submatch props, and empty those.

If two paths through the NFA “collide” in the sense that they reach the same state after reading a given token, then they only differ in how they previously matched, and any future execution would be identical for both. (Note that this would be wrong in the presence of back-references.) Hence, we only need to keep one of the two threads: the thread with the highest priority. Our NFA is built in such a way that higher priority actions always come before lower priority actions, which makes things work.

The explanation in the previous paragraph may make us think that we simply need to keep track of which states were visited at a given step: after all, the loop generated when matching \texttt{(a?)*} against \texttt{a} is broken, isn’t it? No. The group first matches \texttt{a}, as it should, then repeats; it attempts to match \texttt{a} again but fails; it skips \texttt{a}, and finds out that this state has already been seen at this position in the query: the match stops. The capturing group is (wrongly) \texttt{a}. What went wrong is that a thread collided with itself, and the later version, which has gone through the group one more times with an empty match, should have a higher priority than not going through the group.
We solve this by distinguishing “normal” free transitions \_\_regex_action_free:n from transitions \_\_regex_action_free_group:n which go back to the start of the group. The former keeps threads unless they have been visited by a “completed” thread, while the latter kind of transition also prevents going back to a state visited by the current thread.

45.5.1 Variables used when matching

The tokens in the query are indexed from min_pos for the first to max_pos – 1 for the last, and their information is stored in several arrays and \toks registers with those numbers. We match without backtracking, keeping all threads in lockstep at the curr_pos in the query. The starting point of the current match attempt is start_pos, and success_pos, updated whenever a thread succeeds, is used as the next starting position.

The character and category codes of the token at the current position and a token list expanding to that token; the character code of the token at the previous position; the character code of the token just before a successful match; and the character code of the result of changing the case of the current token (A-Z↔a-z). This last integer is only computed when necessary, and is otherwise \c_max_int. The curr_char variable is also used in various other phases to hold a character code.

For every character in the token list, each of the active states is considered in turn. The variable \_\_regex_curr_state_int holds the state of the NFA which is currently considered: transitions are then given as shifts relative to the current state.

The submatches for the thread which is currently active are stored in the curr_submatches list, which is almost a comma list, but ends with a comma. This list is stored by \_\_regex_store_state:n into an intarray variable, to be retrieved when matching at the next position. When a thread succeeds, this list is copied to \_\_regex_success_submatches_t1: only the last successful thread remains there.
This integer, always even, is increased every time a character in the query is read, and not reset when doing multiple matches. We store in \g__regex_state_active_intarray the last step in which each \langle state \rangle in the NFA was encountered. This lets us break infinite loops by not visiting the same state twice in the same step. In fact, the step we store is equal to step when we have started performing the operations of \toks \langle state \rangle, but not finished yet. However, once we finish, we store step + 1 in \g__regex_state_active_intarray. This is needed to track submatches properly (see building phase). The step is also used to attach each set of submatch information to a given iteration (and automatically discard it when it corresponds to a past step).

\int_new:N \l__regex_step_int
(End definition for \l__regex_step_int.)

All the currently active threads are kept in order of precedence in \g__regex_thread_info_intarray together with the corresponding submatch information. Data in this intarray is organized as blocks from min_thread (included) to max_thread (excluded). At the start of every step, the whole array is unpacked, so that the space can immediately be reused, and max_thread is reset to min_thread, effectively clearing the array.

\int_new:N \l__regex_min_thread_int
\int_new:N \l__regex_max_thread_int
(End definition for \l__regex_min_thread_int and \l__regex_max_thread_int.)

\g__regex_state_active_intarray
\g__regex_thread_info_intarray
\g__regex_state_active_intarray stores the last \langle step \rangle in which each \langle state \rangle was active. \g__regex_thread_info_intarray stores threads to be considered in the next step, more precisely the states in which these threads are.

\intarray_new:Nn \g__regex_state_active_intarray { 65536 }
\intarray_new:Nn \g__regex_thread_info_intarray { 65536 }
(End definition for \g__regex_state_active_intarray and \g__regex_thread_info_intarray.)

\l__regex_matched_analysis_tl
\l__regex_curr_analysis_tl
The list \l__regex_curr_analysis_tl consists of a brace group containing three brace groups corresponding to the current token, with the same syntax as \tl_analysis_map_inline:nn. The list \l__regex_matched_analysis_tl (constructed under the \tl_build machinery) has one item for each token that has already been treated so far in a given match attempt: each item consists of three brace groups with the same syntax as \tl_analysis_map_inline:nn.

\tl_new:N \l__regex_matched_analysis_tl
\tl_new:N \l__regex_curr_analysis_tl
(End definition for \l__regex_matched_analysis_tl and \l__regex_curr_analysis_tl.)

\l__regex_every_match_tl
Every time a match is found, this token list is used. For single matching, the token list is empty. For multiple matching, the token list is set to repeat the matching, after performing some operation which depends on the user function. See \__regex_single_match: and \__regex_multi_match:n.

\tl_new:N \l__regex_every_match_tl
(End definition for \l__regex_every_match_tl.)
When doing multiple matches, we need to avoid infinite loops where each iteration matches the same empty token list. When an empty token list is matched, the next successful match of the same empty token list is suppressed. We detect empty matches by setting `\_regex_fresh_thread_bool` to `true` for threads which directly come from the start of the regex or from the `\K` command, and testing that boolean whenever a thread succeeds. The function `\_regex_if_two_empty_matches:F` is redefined at every match attempt, depending on whether the previous match was empty or not: if it was, then the function must cancel a purported success if it is empty and at the same spot as the previous match; otherwise, we definitely don’t have two identical empty matches, so the function is `\use:n`.

```latex
\bool_new:N \l__regex_fresh_thread_bool
\bool_new:N \l__regex_empty_success_bool
\cs_new_eq:NN \__regex_if_two_empty_matches:F \use:n
```

(End definition for `\_regex_fresh_thread_bool`, `\_regex_empty_success_bool`, and `\_regex_if_two_empty_matches:F`.)

The boolean `\_regex_match_success_bool` is true if the current match attempt was successful, and `\g__regex_success_bool` is true if there was at least one successful match. This is the only global variable in this whole module, but we would need it to be local when matching a control sequence with `\c{...}`. This is done by saving the global variable into `\_regex_saved_success_bool`, which is local, hence not affected by the changes due to inner regex functions.

```latex
\bool_new:N \g__regex_success_bool
\bool_new:N \l__regex_saved_success_bool
\bool_new:N \l__regex_match_success_bool
```

(End definition for `\g__regex_success_bool`, `\_regex_saved_success_bool`, and `\_regex_match_success_bool`.)

### 45.5.2 Matching: framework

Initialize the variables that should be set once for each user function (even for multiple matches). Namely, the overall matching is not yet successful; none of the states should be marked as visited (`\g__regex_state_active_intarray`), and we start at step 0; we pretend that there was a previous match ending at the start of the query, which was not empty (to avoid smothering an empty match at the start). Once all this is set up, we are ready for the ride. Find the first match.

```latex
\cs_new_protected:Npn \__regex_match:n #1
\cs_new_protected:Npn \__regex_match_cs:n #1
\cs_new_protected:Npn \__regex_match_init:
```

516
\__regex_match_once_init: This function resets various variables used when finding one match. It is called before the loop through characters, and every time we find a match, before searching for another match (this is controlled by the every_match token list).

First initialize some variables: set the conditional which detects identical empty matches; this match attempt starts at the previous success_pos, is not yet successful, and has no submatches yet; clear the array of active threads, and put the starting state 0 in it. We are then almost ready to read our first token in the query, but we actually start one position earlier than the start because \__regex_match_one_token:nnN increments \l__regex_curr_pos_int and saves \l__regex_curr_char_int as the last_char so that word boundaries can be correctly identified.

(End definition for \__regex_match:n, \__regex_match_cs:n, and \__regex_match_init:)

\__regex_match_once_init: This function resets various variables used when finding one match. It is called before the loop through characters, and every time we find a match, before searching for another match (this is controlled by the every_match token list).

First initialize some variables: set the conditional which detects identical empty matches; this match attempt starts at the previous success_pos, is not yet successful, and has no submatches yet; clear the array of active threads, and put the starting state 0 in it. We are then almost ready to read our first token in the query, but we actually start one position earlier than the start because \__regex_match_one_token:nnN increments \l__regex_curr_pos_int and saves \l__regex_curr_char_int as the last_char so that word boundaries can be correctly identified.
\_\texttt{\_\_regex\_single\_match:} For a single match, the overall success is determined by whether the only match attempt is a success. When doing multiple matches, the overall matching is successful as soon as any match succeeds. Perform the action \#1, then find the next match.

\begin{verbatim}
\cs_new_protected:Npn \_\_regex\_single\_match:
\{\tl_set:Nn \l__regex\_every\_match\_tl
\brace_gset_eq:NN \g__regex\_success\_bool \l__regex\_match\_success\_bool
\_\_regex\_maplike\_break:
\}
\cs_new_protected:Npn \_\_regex\_multi\_match:n #1
\{\tl_set:Nn \l__regex\_every\_match\_tl
\if_meaning:w \c false_bool \l__regex\_match\_success\_bool
\exp_after:wN \_\_regex\_maplike\_break:
\fi:
\bool_gset_true:N \g__regex\_success\_bool #1
\_\_regex\_match\_once\_init:
\}
\end{verbatim}

\textit{(End definition for \_\_regex\_match\_once\_init:.)}

\_\_\_regex\_multi\_match:n At each new position, set some variables and get the new character and category from the query. Then unpack the array of active threads, and clear it by resetting its length (\texttt{max\_thread}). This results in a sequence of \_\_\_regex\_use\_state\_and\_submatches:w \langle\texttt{state}\rangle,\langle\texttt{submatch\_clist}\rangle; and we consider those states one by one in order. As soon
as a thread succeeds, exit the step, and, if there are threads to consider at the next position, and we have not reached the end of the string, repeat the loop. Otherwise, the last thread that succeeded is the match. We explain the fresh_thread business when describing \_\_regex\_action\_wildcard:

\begin{verbatim}
\cs_new_protected:Npn \_\_regex_match_one_token:nnN #1#2#3
\int_add:Nn \l__regex_step_int { 2 }
\int_incr:N \l__regex_curr_pos_int
\int_set_eq:NN \l__regex_last_char_int \l__regex_curr_char_int
\cs_set_eq:NN \__regex_maybe_compute_ccc: \__regex_compute_case_changed_char:
\tl_set:Nn \l__regex_curr_token_tl {#1}
\int_set:Nn \l__regex_curr_char_int {#2}
\int_set:Nn \l__regex_curr_catcode_int { #3 }
\tl_build_put_right:Nx \l__regex_matched_analysis_tl { \exp_not:o \l__regex_curr_analysis_tl }
\tl_set:Nn \l__regex_curr_analysis_tl { { {#1} {#2} #3 } }
\use:x
\int_set_eq:NN \l__regex_max_thread_int \l__regex_min_thread_int
\int_step_function:nnN { \l__regex_min_thread_int }
\int_step_function:nnN { \l__regex_max_thread_int - 1 }\_\_regex_match_one_active:n
\prg_break_point:
\bool_set_false:N \l__regex_fresh_thread_bool
\if_int_compare:w \l__regex_max_thread_int > \l__regex_min_thread_int\if_int_compare:w -2 < \l__regex_curr_char_int\exp_after:wN \exp_after:wN \exp_after:wN \use_none:n\fi:
\fi:
\l__regex_every_match_tl
\cs_new:Npn \_\_regex_match_one_active:n #1
\__regex_use_state_and_submatches:w
\__kernel_intarray_range_to_clist:Nnn \g__regex_thread_info_intarray
{ 1 + #1 * (\l\_regex\_capturing\_group\_int * 2 + 1) }
{ (1 + #1) * (\l\_regex\_capturing\_group\_int * 2 + 1) }
\end{verbatim}

(End definition for \_\_regex\_match\_one\_token:nnN and \_\_regex\_match\_one\_active:n.)

45.5.3 Using states of the nfa

\_\_regex\_use\_state: Use the current NFA instruction. The state is initially marked as belonging to the current step: this allows normal free transition to repeat, but group-repeating transitions won’t. Once we are done exploring all the branches it spawned, the state is marked as step + 1: any thread hitting it at that point will be terminated.
\_\_regex\_use\_state\_and\_submatches:w
This function is called as one item in the array of active threads after that array has
been unpacked for a new step. Update the curr\_state and curr\_submatches and use
the state if it has not yet been encountered at this step.
\cs\_new\_protected:Npn \_\_regex\_use\_state\_and\_submatches:w \#1 , \#2 ;
\{ \int\_set:Nn \l\_\_regex\_curr\_state\_int \{\#1\}
\if\_int\_compare:w
\\_\_kernel\_intarray\_item:Nn \g\_\_regex\_state\_active\_intarray
\{ \l\_\_regex\_curr\_state\_int \}
< \l\_\_regex\_step\_int
\tl\_set:Nn \l\_\_regex\_curr\_submatches\_tl \{ \#2 , \}
\exp\_afte:r:wN \_\_regex\_use\_state:
\fi:
\scan\_stop:
\}
(End definition for \_\_regex\_use\_state\_and\_submatches:w.)

\_\_regex\_action\_start\_wildcard:N
For an unanchored match, state 0 has a free transition to the next and a costly one to
itself, to repeat at the next position. To catch repeated identical empty matches, we need
to know if a successful thread corresponds to an empty match. The instruction resetting
\l\_\_regex\_fresh\_thread\_bool may be skipped by a successful thread, hence we had to
add it to \_\_regex\_match\_one\_token:nnN too.
\cs\_new\_protected:Npm \_\_regex\_action\_start\_wildcard:N \#1
\{ \bool\_set\_true:N \l\_\_regex\_fresh\_thread\_bool
\_\_regex\_action\_free:n \{\#1\}
\bool\_set\_false:N \l\_\_regex\_fresh\_thread\_bool
\bool\_if:NT \#1 { \_\_regex\_action\_cost:n \{0\} }
\}
(End definition for \_\_regex\_action\_start\_wildcard:N.)

\_\_regex\_action\_free:n
\_\_regex\_action\_free\_group:n
\_\_regex\_action\_free\_aux:nn
These functions copy a thread after checking that the NFA state has not already been used
at this position. If not, store submatches in the new state, and insert the instructions for
that state in the input stream. Then restore the old value of \l\_\_regex\_curr\_state\_int and of the current submatches. The two types of free transitions differ by how they
test that the state has not been encountered yet: the group version is stricter, and will
not use a state if it was used earlier in the current thread, hence forcefully breaking the
loop, while the “normal” version will revisit a state even within the thread itself.
\cs\_new\_protected:Npm \_\_regex\_action\_free:n
\_\_regex\_action\_free\_aux:nn { > \_\_regex\_step\_int \else: } }
\cs_new_protected:Npn \_\_regex\_action\_free\_group:n
  { \_\_regex\_action\_free\_aux:nn { < \_\_regex\_step\_int } }
\cs_new_protected:Npn \_\_regex\_action\_free\_aux:nn \#1\#2
  {
    \use:x
      {
        \int_add:Nn \l__regex_curr_state_int {#2}
        \exp_not:n
          {
            \if_int_compare:w
              \_\_kernel_intarray_item:Nn \g__regex_state_active_intarray
              { \_\_regex_curr_state_int }
              #1
            \exp_after:wN \_\_regex_use_state:
            \fi:
          }
        \int_set:Nn \l__regex_curr_state_int
          { \int_use:N \l__regex_curr_state_int }
        \tl_set:Nn \exp_not:N \l__regex_curr_submatches_tl
          \exp_not:o \l__regex_curr_submatches_tl
      }
  }
(End definition for \_\_regex\_action\_free:n, \_\_regex\_action\_free\_group:n, and \_\_regex\_action\_free\_aux:nn.)

\_\_regex\_action\_cost:n
A transition which consumes the current character and shifts the state by \#1. The resulting state is stored in the appropriate array for use at the next position, and we also store the current submatches.
\cs_new_protected:Npn \_\_regex\_action\_cost:n \#1
  {
    \exp_args:Nx \_\_regex\_store\_state:n
      { \int_eval:n { \l__regex_curr_state_int + \#1 } }
  }
(End definition for \_\_regex\_action\_cost:n.)

\_\_regex\_store\_state:n
\_\_regex\_store\_submatches:
Put the given state and current submatch information in \g__regex\_thread\_info\_intarray, and increment the length of the array.
\cs_new_protected:Npn \_\_regex\_store\_state:n \#1
  {
    \exp_args:No \_\_regex\_store\_submatches:nn
      \l__regex\_curr\_submatches\_tl \#1
      \int_incr:N \l__regex\_max\_thread\_int
  }
\cs_new_protected:Npn \_\_regex\_store\_submatches:nn \#1\#2
  {
    \_\_kernel_intarray_gset_range_from_clist:Nnn
      \g__regex\_thread\_info\_intarray
      \_\_regex\_int_eval:w
      1 + \l__regex\_max\_thread\_int *
      (\l__regex\_capturing\_group\_int + 2 + 1)
Some user functions don’t require tracking submatches. We get a performance improvement by simply defining the relevant functions to remove their argument and do nothing with it.

Update the current submatches with the information from the current position. Maybe a bottleneck.

There is a successful match when an execution path reaches the last state in the NFA, unless this marks a second identical empty match. Then mark that there was a successful match; it is empty if it is “fresh”; and we store the current position and submatches. The current step is then interrupted with \texttt{prg\_break}; and only paths with higher precedence are pursued further. The values stored here may be overwritten by a later success of a path with higher precedence.
Replacement

45.6.1 Variables and helpers used in replacement

\l__regex_replacement_csnames_int

The behaviour of closing braces inside a replacement text depends on whether a sequences \c{ or \u{ has been encountered. The number of “open” such sequences that should be closed by } is stored in \l__regex_replacement_csnames_int, and decreased by 1 by each }.

\int_new:N \l__regex_replacement_csnames_int

(End definition for \l__regex_replacement_csnames_int.)

\l__regex_replacement_category_tl

\l__regex_replacement_category_seq

This sequence of letters is used to correctly restore categories in nested constructions such as \c{abc\cD(_d)}.

\tl_new:N \l__regex_replacement_category_tl
\seq_new:N \l__regex_replacement_category_seq

(End definition for \l__regex_replacement_category_tl and \l__regex_replacement_category_seq.)

\g__regex_balance_tl

This token list holds the replacement text for \_\_regex_replacement_balance_one_match:n while it is being built incrementally.

\tl_new:N \g__regex_balance_tl

(End definition for \g__regex_balance_tl.)

\_\_regex_replacement_balance_one_match:n

This expects as an argument the first index of a set of entries in \g__regex_submatch_begin_intarray (and related arrays) which hold the submatch information for a given match. It can be used within an integer expression to obtain the brace balance incurred by performing the replacement on that match. This combines the braces lost by removing the match, braces added by all the submatches appearing in the replacement, and braces appearing explicitly in the replacement. Even though it is always redefined before use, we initialize it as for an empty replacement. An important property is that concatenating several calls to that function must result in a valid integer expression (hence a leading + in the actual definition).

\cs_new:Npn \_\_regex_replacement_balance_one_match:n #1
{ - \_\_regex_submatch_balance:n #1 }

(End definition for \_\_regex_replacement_balance_one_match:n.)
The input is the same as \_\_regex_replacement_balance_one_match:n. This function is redefined to expand to the part of the token list from the end of the previous match to a given match, followed by the replacement text. Hence concatenating the result of this function with all possible arguments (one call for each match), as well as the range from the end of the last match to the end of the string, produces the fully replaced token list. The initialization does not matter, but (as an example) we set it as for an empty replacement.

\cs_new:Npn \_\_\_regex_replacement_do_one_match:n #1
\{\_\_regex_query_range:nn {\_kernel_intarray_item:Nn \g__regex_submatch_prev_intarray \{#1\} } {\_kernel_intarray_item:Nn \g__regex_submatch_begin_intarray \{#1\} } \}

(End definition for \_\_\_regex_replacement_do_one_match:n.)

This function lets us navigate around the fact that the primitive \exp_not:n requires a braced argument. As far as I can tell, it is only needed if the user tries to include in the replacement text a control sequence set equal to a macro parameter character, such as \c_parameter_token. Indeed, within an x-expanding assignment, \exp_not:N # behaves as a single #, whereas \exp_not:n {#} behaves as a doubled ##.

\cs_new:Npn \_\_\_regex_replacement_exp_not:N #1 {\exp_not:n {#1}}

(End definition for \_\_\_regex_replacement_exp_not:N.)

This is used for the implementation of \u, and it gets redefined for \peek_regex_replace_once:nnTF.

\cs_new_eq:NN \_\_\_regex_replacement_exp_not:V \exp_not:V

(End definition for \_\_\_regex_replacement_exp_not:V.)

45.6.2 Query and brace balance

When it is time to extract submatches from the token list, the various tokens are stored in \toks registers numbered from \_\_\_regex_min_pos_int inclusive to \_\_\_regex_max_pos_int exclusive. The function \_\_\_regex_query_range:nn \{\langle min\rangle\} \{\langle max\rangle\} unpacks registers from the position \langle min\rangle to the position \langle max\rangle − 1 included. Once this is expanded, a second x-expansion results in the actual tokens from the query. That second expansion is only done by user functions at the very end of their operation, after checking (and correcting) the brace balance first.

\cs_new:Npn \_\_\_regex_query_range:nn #1#2
\{\exp_after:wN \_\_\_regex_query_range_loop:ww \int_value:w \_\_\_regex_int_eval:w #1 \exp_after:wN ; \int_value:w \_\_\_regex_int_eval:w #2 ; \prg_break_point: \}

\cs_new:Npn \_\_\_regex_query_range_loop:ww #1 ; #2 ;
\{\if_int_compare:w #1 < #2 \exp_stop_f:
\else:
\exp_after:wN \prg_break:
\fi:
\__regex_toks_use:w \#1 \exp_stop_f:
\exp_after:wN \__regex_query_range_loop:ww
\int_value:w \__regex_int_eval:w \#1 + 1 ; \#2 ;
}

(End definition for \__regex_query_range:nn and \__regex_query_range_loop:ww.)

\__regex_query_submatch:n Find the start and end positions for a given submatch (of a given match).
\cs_new:Npn \__regex_query_submatch:n \#1
{ \__regex_query_range:nn \{ \__kernel_intarray_item:Nn \g__regex_submatch_begin_intarray {\#1} \} \{ \__kernel_intarray_item:Nn \g__regex_submatch_end_intarray {\#1} \} }

(End definition for \__regex_query_submatch:n.)

\__regex_submatch_balance:n Every user function must result in a balanced token list (unbalanced token lists cannot be stored by TeX). When we unpacked the query, we kept track of the brace balance, hence the contribution from a given range is the difference between the brace balances at the \( \langle \text{max pos} \rangle \) and \( \langle \text{min pos} \rangle \). These two positions are found in the corresponding "submatch" arrays.
\cs_new_protected:Npn \__regex_submatch_balance:n \#1
{ \int_eval:n
{ \__regex_intarray_item:NnF \g__regex_balance_intarray \{ \__kernel_intarray_item:Nn \g__regex_submatch_end_intarray {\#1} \} 0
- \__regex_intarray_item:NnF \g__regex_balance_intarray \{ \__kernel_intarray_item:Nn \g__regex_submatch_begin_intarray {\#1} \} 0 }
}

(End definition for \__regex_submatch_balance:n.)

45.6.3 Framework

\__regex_replacement:n \__regex_replacement:x \__regex_replacement_apply:Nn \__regex_replacement_set:n
The replacement text is built incrementally. We keep track in \l\_\_regex_balance_int of the balance of explicit begin- and end-group tokens and we store in \g\_\_regex_balance_int some code to compute the brace balance from submatches (see its description). Detect unescaped right braces, and escaped characters, with trailing \prg_do_nothing: because some of the later function look-ahead. Once the whole replacement text has been parsed, make sure that there is no open csname. Finally, define the balance_one_match and do_one_match functions.
\cs_new_protected:Npn \__regex_replacement:n
{ \__regex_replacement_apply:Nn \__regex_replacement_set:n }
\cs_new_protected:Npn \__regex_replacement_set:n #1
{
\group_begin:
\tl_build_begin:N \l__regex_build_tl
\int_zero:N \l__regex_balance_int
\tl_gclear:N \g__regex_balance_tl
\__regex_escape_use:nnnn
{ \if_charcode:w \c_right_brace_str ##1
  \__regex_replacement_rbrace:N
\else:
  \if_charcode:w \c_left_brace_str ##1
  \__regex_replacement_lbrace:N
\else:
  \__regex_replacement_normal:n
\fi:
\fi:
  #1
}
{ \__regex_replacement_escaped:N #1 }
{ \__regex_replacement_normal:n #1 }#2
\prg_do_nothing: \prg_do_nothing:
\if_int_compare:w \l__regex_replacement_csnames_int > \c_zero_int
\msg_error:nnx { regex } { replacement-missing-rbrace }
{ \int_use:N \l__regex_replacement_csnames_int }
\tl_build_put_right:Nx \l__regex_build_tl
{ \prg_replicate:nn \l__regex_replacement_csnames_int \cs_end: }
\fi:
\seq_if_empty:NF \l__regex_replacement_category_seq
{ \msg_error:nnx { regex } { replacement-missing-rparen }
  \{ \seq_count:N \l__regex_replacement_category_seq \}
\seq_clear:N \l__regex_replacement_category_seq }
\tl_gput_right:Nx \g__regex_balance_tl
{ + \int_use:N \l__regex_balance_int }
\tl_build_end:N \l__regex_build_tl
\exp_args:NNo
\group_end:
#1 \l__regex_build_tl
\cs_generate_variant:Nn \__regex_replacement:n { x }
\cs_new_protected:Npn \__regex_replacement_set:n #1
{ \cs_set:Npn \__regex_query_range:nn
  { \__kernel_intarray_item:Nn \g__regex_submatch_prev_intarray {##1} } }
\__regex_case_replacement:n
\__regex_case_replacement:x

\tl_new:N \l__regex_case_replacement_tl
\tl_new:N \l__regex_case_balance_tl
\cs_new_protected:Npn \__regex_case_replacement:n \#1
\tl_gset:Nn \l__regex_case_balance_tl
{\if_case:w
 \__kernel_intarray_item:Nn
 \g__regex_submatch_case_intarray {\#1}
 \tl_gset_eq:NN \l__regex_case_replacement_tl \l__regex_case_balance_tl
 \tl_map_tokens:nn {#1}
{ \__regex_replacement_apply:Nn \__regex_case_replacement_aux:n }
\tl_gset:No \l__regex_case_balance_tl \fi: }
\exp_args:No \__regex_case_replacement_set:n
{ \l__regex_case_replacement_tl \fi: }
\cs_generate_variant:Nn \__regex_case_replacement:n { x }
\cs_new_protected:Npn \__regex_case_replacement:n \#1
{ \tl_gput_right:Nn \l__regex_case_replacement_tl
 \tl_gput_right:No \l__regex_case_balance_tl
 { \exp_after:wN \or: \l__regex_balance_tl }
\exp_after:wN \or: \l__regex_case_balance_tl }
\cs_generate_variant:Nn \__regex_case_replacement:n { x }

(End definition for \__regex_case_replacement:n.)
\__regex_replacement_put:n

This gets redefined for \peek_regex_replace_once:nnTF.
\cs_new_protected:Npn \__regex_replacement_put:n
{ \tl_build_put_right:Nn \l__regex_build_tl }

(End definition for \__regex_replacement_put:n.)
\__regex_replacement_normal:n
\__regex_replacement_normal_aux:n

Most characters are simply sent to the output by \tl_build_put_right:Nn, unless a particular category code has been requested: then \__regex_replacement_c_A:w or a similar auxiliary is called. One exception is right parentheses, which restore the category
code in place before the group started. Note that the sequence is non-empty there: it contains an empty entry corresponding to the initial value of \l__regex_replacement_category_tl. The argument #1 is a single character (including the case of a catcode-other space). In case no specific catcode is requested, we take into account the current catcode regime (at the time the replacement is performed) as much as reasonable, with all impossible catcodes (escape, newline, etc.) being mapped to “other”.

\cs_new_protected:Npn \__regex_replacement_normal:n #1
{\int_compare:nNnTF \l__regex_replacement_csnames_int > 0
{\exp_args:No \__regex_replacement_put:n \token_to_str:N #1 }
{\tl_if_empty:NTF \l__regex_replacement_category_tl
{\__regex_replacement_normal_aux:N #1 }
{ % ( \token_if_eq_charcode:NNTF #1 )
\seq_pop:NN \l__regex_replacement_category_seq
\l__regex_replacement_category_tl
}
{\use:c \__regex_replacement_c_\l__regex_replacement_category_tl:w }
? #1}
}
\cs_new_protected:Npn \__regex_replacement_normal_aux:N #1
{\token_if_eq_charcode:NNTF #1 \c_space_token
\exp_after:wN \exp_after:wN
\if_case:w \tex_catcode:D '#1 \exp_stop_f:
\__regex_replacement_c_0:w
\or: \__regex_replacement_c_B:w
\or: \__regex_replacement_c_E:w
\or: \__regex_replacement_c_N:w
\or: \__regex_replacement_c_T:w
\or: \__regex_replacement_c_O:w
\or: \__regex_replacement_c_P:w
\or: \__regex_replacement_c_U:w
\or: \__regex_replacement_c_D:w
\or: \__regex_replacement_c_C:w
\or: \__regex_replacement_c_S:w
\or: \__regex_replacement_c_L:w
\or: \__regex_replacement_c_A:w
\else: \__regex_replacement_c_0:w
\fi:
? #1}
(End definition for \__regex_replacement_normal:n and \__regex_replacement_normal_aux:N)

528
As in parsing a regular expression, we use an auxiliary built from \#1 if defined. Otherwise, check for escaped digits (standing from submatches from 0 to 9): anything else is a raw character.

\cs_new_protected:Npn \__regex_replacement_escaped:N \#1
\cs_if_exist_use:cF { __regex_replacement_#1:w }
\__regex_replacement_put_submatch:n \{#1\}
\else:
\__regex_replacement_normal:n \{#1\}
\fi:
\}

(End definition for \__regex_replacement_escaped:N.)

45.6.4 Submatches

Insert a submatch in the replacement text. This is dropped if the submatch number is larger than the number of capturing groups. Unless the submatch appears inside a \c{...} or \u{...} construction, it must be taken into account in the brace balance. Later on, \#1 will be replaced by a pointer to the 0-th submatch for a given match. There is an \exp_not:N here as at the point-of-use of \g__regex_balance_tl there is an x-type expansion which is needed to get \#1 in correctly.

\cs_new_protected:Npn \__regex_replacement_put_submatch:n \#1
\if_int_compare:w \#1 < \l__regex_capturing_group_int
\__regex_replacement_put_submatch_aux:n \{#1\}
\else:
\msg_expandable_error:nnff { regex } { submatch-too-big }
\{#1\} \{ \int_eval:n \{ \l__regex_capturing_group_int - 1 \} \}
\fi:
\}
\cs_new_protected:Npn \__regex_replacement_put_submatch_aux:n \#1
\tl_build_put_right:Nn \l__regex_build_tl
\{ \__regex_query_submatch:n \{ \int_eval:n \{ \#1 + \#1 \} \} \}
\tl_gput_right:Nn \g__regex_balance_tl \{ \int_eval:n \{ \#1 + \#1 \} \}
\fi:
\}

(End definition for \__regex_replacement_put_submatch:n and \__regex_replacement_put_submatch_aux:n.)

\__regex_replacement_g:w \__regex_replacement_g_digits:NN

Grab digits for the \g escape sequence in a primitive assignment to the integer \l__regex_internal_a_int. At the end of the run of digits, check that it ends with a right brace.

\cs_new_protected:Npn \__regex_replacement_g:w \#1\#2
\token_if_eq_meaning:NNTF \#1 \__regex_replacement_lbrace:N
\cs_new:Npn \__regex_replacement_g_digits:NN #1#2
\{ \token_if_eq_meaning:NNTF \#1 \__regex_replacement_normal:n
{ \if_int_compare:w 1 < \numexpr#2\relax
\exp_stop_f:
#2
\exp_after:wN \use_i:nn
\exp_after:wN \__regex_replacement_g_digits:NN
\else:
\exp_stop_f:
\exp_after:wN \__regex_replacement_error:NNN
g
\fi:
}
\}
\{ \exp_stop_f:
\cs_if_exist:cTF \__regex_replacement_c_#2:w
\exp_args:No \__regex_replacement_put_submatch:n
\{ \use:N \l__regex_internal_a_int \use_none:nn
\else:
\exp_after:wN \__regex_replacement_error:NNN
g
\fi:
\}
\}
\{ #1 #2
\}

(End definition for \__regex_replacement_g:w and \__regex_replacement_g_digits:NN.)

\subsection{Csnames in replacement}

\__regex_replacement_c:w \textbackslash c may only be followed by an unescaped character. If followed by a left brace, start a control sequence by calling an auxiliary common with \textbackslash u. Otherwise test whether the category is known; if it is not, complain.

\cs_new_protected:Npm \__regex_replacement_c:w #1#2
\{ \token_if_eq_meaning:NNTTF \#1 \__regex_replacement_normal:n
\{ \cs_if_exist:cTF \__regex_replacement_c_#2:w
\exp_args:No \__regex_replacement_put_submatch:n
\{ \use:N \l__regex_internal_a_int \use_none:nn
\else:
\exp_after:wN \__regex_replacement_error:NNN
g
\fi:
\}
\}
\{ \exp_stop_f:
\\textbackslash \textbackslash u \__regex_replacement_rbrace:N \#1
\exp_args:No \__regex_replacement_put_submatch:n
\{ \use:N \l__regex_internal_a_int \use_none:nn
\else:
\exp_after:wN \__regex_replacement_error:NNN
g
\fi:
\}
\}
\}

(End definition for \__regex_replacement_c:w.)

530
Start a control sequence with `\cs:w`, protected from expansion by \#1 (either \__regex_replacement_exp_not:N or \exp_not:V), or turned to a string by \tl_to_str:V if inside another csname construction \c{...} or \u{...}. We use \tl_to_str:V rather than \tl_to_str:N to deal with integers and other registers.

```
\cs_new_protected:Npn \__regex_replacement_cu_aux:Nw \#1
    \{ \if_case:w \l__regex_replacement_csnames_int
        \tl_build_put_right:Nn \l__regex_build_tl
            \{ \exp_not:n \{ \exp_after:wN \#1 \cs:w \} \}
    \else:
        \tl_build_put_right:Nn \l__regex_build_tl
            \{ \exp_not:n \{ \exp_after:wN \tl_to_str:V \cs:w \} \}
    \fi:
    \int_incr:N \l__regex_replacement_csnames_int
\}
```

(End definition for \__regex_replacement_cu_aux:Nw.)

Check that \u{...} is followed by a left brace. If so, start a control sequence with \cs:w, which is then unpacked either with \exp_not:V or \tl_to_str:V depending on the current context.

```
\cs_new_protected:Npn \__regex_replacement_u:w \#1#2
    \{ \token_if_eq_meaning:NNTF \#1 \__regex_replacement_lbrace:N
        \__regex_replacement_cu_aux:Nw \__regex_replacement_exp_not:V \}
    \{ \__regex_replacement_error:NNN u \#1#2 \}
\}
```

(End definition for \__regex_replacement_u:w.)

Within a \c{...} or \u{...} construction, end the control sequence, and decrease the brace count. Otherwise, this is a raw right brace.

```
\cs_new_protected:Npn \__regex_replacement_rbrace:N \#1
    \{ \if_int_compare:w \l__regex_replacement_csnames_int > \c_zero_int
        \tl_build_put_right:Nn \l__regex_build_tl \{ \cs_end: \}
    \else:
        \__regex_replacement_normal:n \#1
    \fi:
\}
```

(End definition for \__regex_replacement_rbrace:N.)

Within a \c{...} or \u{...} construction, this is forbidden. Otherwise, this is a raw left brace.

```
\cs_new_protected:Npn \__regex_replacement_lbrace:N \#1
    \{ \if_int_compare:w \l__regex_replacement_csnames_int > \c_zero_int
        \msg_error:nnn { regex } { cu-lbrace } { u } \}
    \else:
        \__regex_replacement_normal:n \#1
    \fi:
\}
```

(End definition for \__regex_replacement_lbrace:N.)
45.6.6 Characters in replacement

Here, \#1 is a letter among BEMTPUDSLOA and \#2\#3 denote the next character. Complain if we reach the end of the replacement or if the construction appears inside \c{...} or \u{...}, and detect the case of a parenthesis. In that case, store the current category in a sequence and switch to a new one.

\texttt{\cs_new_protected:Npn \__regex_replacement_cat:NNN \#1\#2\#3
{
  \token_if_eq_meaning:NNTF \prg_do_nothing: \#3
  {
    \msg_error:nn \{ regex \} \{ replacement-catcode-end \ }
  }
  \int_compare:nNnTF { \l__regex_replacement_csnames_int } > 0
  {
    \msg_error:nnnn \{ regex \} \{ replacement-catcode-in-cs \} \{\#1\} \{\#3\}
    \#2 \#3
  }
  {
    \__regex_two_if_eq:NNNNTF \#2 \#3 \__regex_replacement_normal:n \(
      \seq_push:NV \l__regex_replacement_category_seq \l__regex_replacement_category_tl \tl_set:Nn \l__regex_replacement_category_tl \{\#1\}
    \)
  }
}
\token_if_eq_meaning:NNT \__regex_replacement_escaped:N \{ \__regex_char_if_alphanumeric:NTF \#3
  \}
\use:c \{ \__regex_replacement_c_\#1:w \} \#2 \#3
\)

(End definition for \__regex_replacement_cat:NNN.)

We now need to change the category code of the null character many times, hence work in a group. The catcode-specific macros below are defined in alphabetical order; if you are trying to understand the code, start from the end of the alphabet as those categories are simpler than active or begin-group.

\texttt{\group_begin:

\__regex_replacement_char:nNN

The only way to produce an arbitrary character–catcode pair is to use the \texttt{\lowercase} or \texttt{\uppercase} primitives. This is a wrapper for our purposes. The first argument is the null character with various catcodes. The second and third arguments are grabbed from the input stream: \#3 is the character whose character code to reproduce. We could use
\char_generate:nn but only for some catcodes (active characters and spaces are not supported).
\cs_new_protected:Npn \_regex_replacement_char:nNN #1#2#3
\tex_lccode:D 0 = '#3 \scan_stop:
\tex_lowercase:D { \_regex_replacement_put:n {#1} }
\end defn for \_regex_replacement_char:nNN
\_regex_replacement_c_A:w
For an active character, expansion must be avoided, twice because we later do two \texttt{x}-
expansions, to unpack \texttt{\toks} for the query, and to expand their contents to tokens of the
query.
\char_set_catcode_active:N \^\^@
\cs_new_protected:Npn \_regex_replacement_c_A:w
{ \_regex_replacement_char:nNN { \exp_not:n { \exp_not:N ^\^@ } } }
\end defn for \_regex_replacement_c_A:w
\_regex_replacement_c_B:w
An explicit begin-group token increases the balance, unless within a \texttt{\c{...}} or \texttt{\u{...}}
construction. Add the desired begin-group character, using the standard \texttt{\if_false:}
trick. We eventually \texttt{x}-expand twice. The first time must yield a balanced token list,
and the second one gives the bare begin-group token. The \texttt{\exp_after:wN} is not strictly
needed, but is more consistent with \texttt{l3tl-analysis}.
\char_set_catcode_group_begin:N \^\^@
\cs_new_protected:Npn \_regex_replacement_c_B:w
{ \_regex_replacement_char:nNN { \exp_not:n { \exp_after:wN \^\^@ \if_false: } \fi: } }
\end defn for \_regex_replacement_c_B:w
\_regex_replacement_c_C:w
This is not quite catcode-related: when the user requests a character with category “con-
trol sequence”, the one-character control symbol is returned. As for the active character, we
prepare for two \texttt{x}-expansions.
\cs_new_protected:Npn \_regex_replacement_c_C:w #1#2
\tl_build_put_right:Nn \_regex_build_tl
\exp_not:N \_regex_replacement_exp_not:N \exp_not:c (#2)
\end defn for \_regex_replacement_c_C:w
\_regex_replacement_c_D:w
Subscripts fit the mould: \texttt{\lowercase} the null byte with the correct category.
\char_set_catcode_math_subscript:N \^\^@
\cs_new_protected:Npn \_regex_replacement_c_D:w
\_regex_replacement_char:nNN { ^\^@ }
\end defn for \_regex_replacement_c_D:w

533
Similar to the begin-group case, the second x-expansion produces the bare end-group token.

\begin{verbatim}
\char_set_catcode_group_end:N \^^@
\cs_new_protected:Npn \__regex_replacement_c_E:w
{ \if_int_compare:w \l__regex_replacement_csnames_int = \c_zero_int
\int_decr:N \l__regex_balance_int
\fi:
\__regex_replacement_char:nNN
\exp_not:n \{ \if_false: \{ \fi: \^^@ \}
\}
\end{verbatim}

(End definition for \__regex_replacement_c_E:w.)

\__regex_replacement_c_L:w
Simply \lowercase a letter null byte to produce an arbitrary letter.

\begin{verbatim}
\char_set_catcode_letter:N \^^@
\cs_new_protected:Npn \__regex_replacement_c_L:w
{ \__regex_replacement_char:nNN \{ \^^@ \}
\}
\end{verbatim}

(End definition for \__regex_replacement_c_L:w.)

\__regex_replacement_c_M:w
No surprise here, we lowercase the null math toggle.

\begin{verbatim}
\char_set_catcode_math_toggle:N \^^@
\cs_new_protected:Npn \__regex_replacement_c_M:w
\{
\__regex_replacement_char:nNN \{ \^^@ \}
\}
\end{verbatim}

(End definition for \__regex_replacement_c_M:w.)

\__regex_replacement_c_O:w
Lowercase an other null byte.

\begin{verbatim}
\char_set_catcode_other:N \^^@
\cs_new_protected:Npn \__regex_replacement_c_O:w
\{
\__regex_replacement_char:nNN \{ \^^@ \}
\}
\end{verbatim}

(End definition for \__regex_replacement_c_O:w.)

\__regex_replacement_c_P:w
For macro parameters, expansion is a tricky issue. We need to prepare for two x-expansions and passing through various macro definitions. Note that we cannot replace one \exp_not:n by doubling the macro parameter characters because this would misbehave if a mischievous user asks for \c{\cP\#}, since that macro parameter character would be doubled.

\begin{verbatim}
\char_set_catcode_parameter:N \^^@
\cs_new_protected:Npn \__regex_replacement_c_P:w
\{
\__regex_replacement_char:nNN
\exp_not:n \{ \exp_not:n \{ \^^@\^^@\^^@\^^@ \} \}
\}
\end{verbatim}

(End definition for \__regex_replacement_c_P:w.)

\__regex_replacement_c_S:w
Spaces are normalized on input by \TeX{} to have character code 32. It is in fact impossible to get a token with character code 0 and category code 10. Hence we use 32 instead of 0 as our base character.

\begin{verbatim}
\cs_new_protected:Npn \__regex_replacement_c_S:w \#1#2
\{
\end{verbatim}

534
No surprise for alignment tabs here. Those are surrounded by the appropriate braces whenever necessary, hence they don’t cause trouble in alignment settings.

Simple call to \_\_regex\_replacement\_char:nNN which lowercases the math superscript \textsuperscript{^@}.

Restore the catcode of the null byte.

45.6.7 An error

Simple error reporting by calling one of the messages replacement-c, replacement-g, or replacement-u.

45.7 User functions

Before being assigned a sensible value, a regex variable matches nothing.

The usual scratch space.
Compile, then store the result in the user variable with the appropriate assignment function.

\cs_new_protected:Npn \regex_set:Nn #1#2
{ \_regex_compile:n {#2} \tl_set_eq:NN #1 \l__regex_internal_regex }

\cs_new_protected:Npn \regex_gset:Nn #1#2
{ \_regex_compile:n {#2} \tl_gset_eq:NN #1 \l__regex_internal_regex }

\cs_new_protected:Npn \regex_const:Nn #1#2
{ \_regex_compile:n {#2} \tl_const:Nx #1 { \exp_not:o \l__regex_internal_regex } }

User functions: the n variant requires compilation first. Then show the variable with some appropriate text. The auxiliary \_\regex_show:N is defined in a different section.

\cs_new_protected:Npn \regex_show:n { \__regex_show:Nn \msg_show:nnxxxx }
\cs_new_protected:Npn \regex_log:n { \__regex_show:Nn \msg_log:nnxxxx }
\cs_new_protected:Npn \__regex_show:Nn #1#2
{ \_regex_compile:n {#2} \__regex_show:N \l__regex_internal_regex #1 { regex } { show } { \token_to_str:N #2 } { } { \l__regex_internal_a_tl } { } }
\cs_new_protected:Npn \regex_show:N { \__regex_show:NN \msg_show:nnxxxx }
\cs_new_protected:Npn \regex_log:N { \__regex_show:NN \msg_log:nnxxxx }
\cs_new_protected:Npn \__regex_show:NN #1#2
{ \kernel_chk_tl_type:NnnT #2 { regex } \__regex_clean_regex:n {#2} }

Those conditionals are based on a common auxiliary defined later. Its first argument builds the NFA corresponding to the regex, and the second argument is the query token
Once we have performed the match, convert the resulting boolean to `'true` or `'false`.

The auxiliary errors if `#1` has an odd number of items, and otherwise it sets \_g\_regex\_case\_int according to which case was found (zero if not found). The true branch leaves the corresponding code in the input stream.

We define here 40 user functions, following a common pattern in terms of \_\_regex\_build\_n or \_\_regex\_build\_N with the appropriate regex argument, then all other necessary arguments (replacement text, token list, etc. The conditionals call \_\_regex\_return\_ to return either `true` or `false` once matching has been performed.
If the input is bad (odd number of items) then take the false branch. Otherwise, use the same auxiliary as \regex_replace_all:nN, but with more complicated code to build the automaton, and to find what replacement text to use. The \tl_item:n is only expanded once we know the value of \g__regex_case_int, namely which case matched.

\cs_new_protected:Npn \regex_replace_case_all:nNTF #1#2#3
\{ \regex_replace_case_once:nNTF {#1} {#2} {#3} \}
\cs_new_protected:Npn \regex_replace_case_all:nN #1#2
\{ \regex_replace_case_once:nNTF {#1} {#2} \}
\cs_new_protected:Npn \regex_replace_case_all:nF #1#2
\{ \regex_replace_case_once:nNTF {#1} {#2} \}

(End definition for \regex_replace_case_once:nNTF and others. This function is documented on page 58.)
45.7.1 Variables and helpers for user functions

\l__regex_match_count_int

The number of matches found so far is stored in \l__regex_match_count_int. This is only used in the \regex_count:nnN functions.

\int_new:N \l__regex_match_count_int

(End definition for \l__regex_match_count_int.)

__regex_begin
__regex_end

Those flags are raised to indicate begin-group or end-group tokens that had to be added when extracting submatches.

\flag_new:n { __regex_begin }
\flag_new:n { __regex_end }

(End definition for __regex_begin and __regex_end.)

\l__regex_min_submatch_int
\l__regex_submatch_int
\l__regex_zeroth_submatch_int

The end-points of each submatch are stored in two arrays whose index (submatch) ranges from \l__regex_min_submatch_int (inclusive) to \l__regex_submatch_int (exclusive). Each successful match comes with a 0-th submatch (the full match), and one match for each capturing group: submatches corresponding to the last successful match are labelled starting at zeroth_submatch. The entry \l__regex_zeroth_submatch_int in \g__regex_submatch_prev_intarray holds the position at which that match attempt started: this is used for splitting and replacements.

\int_new:N \l__regex_min_submatch_int
\int_new:N \l__regex_submatch_int
\int_new:N \l__regex_zeroth_submatch_int

(End definition for \l__regex_min_submatch_int, \l__regex_submatch_int, and \l__regex_zeroth_submatch_int.)
Hold the place where the match attempt begun, the end-points of each submatch, and which regex case the match corresponds to, respectively.

\texttt{\intarray_new:Nn \g__regex_submatch_prev_intarray { 65536 } \intarray_new:Nn \g__regex_submatch_begin_intarray { 65536 } \intarray_new:Nn \g__regex_submatch_end_intarray { 65536 } \intarray_new:Nn \g__regex_submatch_case_intarray { 65536 }}

(End definition for \texttt{\g__regex_submatch_prev_intarray} and others.)

The first thing we do when matching is to store the balance of begin-group/end-group characters into \texttt{\g__regex_balance_intarray}.

\texttt{\intarray_new:Nn \g__regex_balance_intarray { 65536 }}

(End definition for \texttt{\g__regex_balance_intarray}.)

Keep track of the number of left/right braces to add when performing a regex operation such as a replacement.

\texttt{\int_new:N \l__regex_added_begin_int \int_new:N \l__regex_added_end_int}

(End definition for \texttt{\l__regex_added_begin_int} and \texttt{\l__regex_added_end_int}.)

This function triggers either \texttt{\prg_return_false} or \texttt{\prg_return_true} as appropriate to whether a match was found or not. It is used by all user conditionals.

\texttt{\cs_new_protected:Npn \__regex_return: { \if_meaning:w \c_true_bool \g__regex_success_bool \prg_return_true: \else: \prg_return_false: \fi: }}

(End definition for \texttt{\__regex_return:})

To easily extract subsets of the input once we found the positions at which to cut, store the input tokens one by one into successive \texttt{\toks} registers. Also store the brace balance (used to check for overall brace balance) in an array.

\texttt{\cs_new_protected:Npn \__regex_query_set:n \__regex_query_set:nN}

540
45.7.2 Matching

\_\_regex_if_match:nn

We don’t track submatches, and stop after a single match. Build the NFA with \#1, and perform the match on the query \#2.

\cs_new_protected:Npn \_\_regex_if_match:nn \#1\#2
{
\group_begin:
\_\_regex_disable_submatches:
\_\_regex_single_match:
\#1
\_\_regex_match:n \{\#2\}
\group_end:
}

(End definition for \_\_regex_if_match:nn and \_\_regex_query_set:n and \_\_regex_query_set_aux:n.)

\_\_regex_match_case:nnTF
\_\_regex_match_case_aux:nn

The code would get badly messed up if the number of items in \#1 were not even, so we catch this case, then follow the same code as \regex_match:nnTF but using \_\_regex_case_build:n and without returning a result.

\cs_new_protected:Npn \_\_regex_match_case:nnTF \#1\#2
{
\group_begin:
\int_if_odd:nTF { \tl_count:n {\#1} } {mag_error:nxxxx \{ \token_to_str:n \_\_regex_match_case:nn(TF) \} \{ code \} \use_ii:nn}
\_\_regex_if_match:nn
{ \_\_regex_case_build:x { \_\_regex_tl_odd_items:n \{\#1\} } \{\#2\} \bool_if:NTF \g\_\_regex_success_bool}
\group_end:
}

(End definition for \_\_regex_match_case:nnTF and \_\_regex_match_case_aux:nn.)

\_\_regex_count:nnN

Again, we don’t care about submatches. Instead of aborting after the first “longest match” is found, we search for multiple matches, incrementing \_\_regex_match_count_int every time to record the number of matches. Build the NFA and match. At the end, store the result in the user’s variable.

\cs_new_protected:Npn \_\_regex_count:nnN \#1\#2\#3
{
\group_begin:

\_regex\_extract\_once:nnN
\_regex\_extract\_all:nnN

Match once or multiple times. After each match (or after the only match), extract the submatches using \_regex\_extract:. At the end, store the sequence containing all the submatches into the user variable #3 after closing the group.

\cs\_new\_protected:Npn \_regex\_extract\_once:nnN #1#2#3

\group\_begin:
\_regex\_single\_match:
#1
\_regex\_match:n {#2}
\_regex\_extract:
\_regex\_query\_set:n {#2}
\_regex\_group\_end\_extract\_seq:N #3
\group\_end:
\cs\_new\_protected:Npn \_regex\_extract\_all:nnN #1#2#3

\group\_begin:
\_regex\_multi\_match:n { \_regex\_extract: }
#1
\_regex\_match:n {#2}
\_regex\_query\_set:n {#2}
\_regex\_group\_end\_extract\_seq:N #3
\group\_end:

(End definition for \_regex\_count:nnN.)

\_regex\_split:nnN

Splitting at submatches is a bit more tricky. For each match, extract all submatches, and replace the zeroth submatch by the part of the query between the start of the match attempt and the start of the zeroth submatch. This is inhibited if the delimiter matched an empty token list at the start of this match attempt. After the last match, store the last part of the token list, which ranges from the start of the match attempt to the end of the query. This step is inhibited if the last match was empty and at the very end: decrement \_\_regex\_submatch\_int, which controls which matches will be used.

\cs\_new\_protected:Npn \_regex\_split:nnN #1#2#3

\group\_begin:
\_regex\_multi\_match:n
{
  \if_int_compare:w
  \_\_regex\_start\_pos\_int < \_\_regex\_success\_pos\_int

\_\_regex_extract:
\_\_kernel_intarray_gset:Nnn \g__regex_submatch_prev_intarray
{ \l__regex_zeroth_submatch_int } { 0 }
\_\_kernel_intarray_gset:Nnn \g__regex_submatch_end_intarray
{ \l__regex_zeroth_submatch_int }
{ \_\_kernel_intarray_item:Nn \g__regex_submatch_begin_intarray
{ \l__regex_zeroth_submatch_int }
}
\_\_kernel_intarray_gset:Nnn \g__regex_submatch_begin_intarray
{ \l__regex_zeroth_submatch_int }
{ \l__regex_start_pos_int }
\fi:
\_\_regex_match:n {#2}
\_\_regex_query_set:n {#2}
\_\_kernel_intarray_gset:Nnn \g__regex_submatch_prev_intarray
{ \l__regex_submatch_int } { 0 }
\_\_kernel_intarray_gset:Nnn \g__regex_submatch_end_intarray
{ \l__regex_submatch_int }
\_\_kernel_intarray_gset:Nnn \g__regex_submatch_begin_intarray
{ \l__regex_submatch_int }
{ \l__regex_start_pos_int }
\int_incr:N \l__regex_submatch_int
\if_meaning:w \c_true_bool \l__regex_empty_success_bool
\if_int_compare:w \l__regex_start_pos_int = \l__regex_max_pos_int
\int_decr:N \l__regex_submatch_int
\fi:
\fi:
\_\_regex_group_end_extract_seq:N #3
\}

The end-points of submatches are stored as entries of two arrays from \_\_\_regex_min_submatch_int to \_\_\_regex_submatch_int (exclusive). Extract the relevant ranges into \g__regex_internal_tl, separated by \__regex_tmp:w \{ \}. We keep track in the two flags \_\_regex_begin and \_\_regex_end of the number of begin-group or end-group tokens added to make each of these items overall balanced. At this step, \{ \} is counted as being balanced (same number of begin-group and end-group tokens). This problem is caught by \_\_\_regex_extract_check:w, explained later. After complaining about any begin-group or end-group tokens we had to add, we are ready to construct the user’s sequence outside the group.
\cs_new_protected:Npn \_\_regex_group_end_extract_seq:N #1
\{ \flag_clear:n \{ \_\_regex_begin \}
\flag_clear:n \{ \_\_regex_end \}
\cs_set_eq:NN \_\_regex_tmp:w \scan_stop:
\_\_kernel_tl_gset:Nx \g__regex_internal_tl
\{ \int_step_function:nnN \{ \_\_\_regex_min_submatch_int \}
{ \l__regex_submatch_int - 1 } \_\_\_regex_extract_seq_aux:n

(End definition for \_\_\_regex_split:nnN.)
\_\_regex_extract_seq:nn
\_\_regex_extract_seq:ww

The \_\_regex_extract_seq:nn auxiliary builds one item of the sequence of submatches. First compute the brace balance of the submatch, then extract the submatch from the query, adding the appropriate braces and raising a flag if the submatch is not balanced.
In \_regex_group_end_extract_seq:N we had to expand \_g_regex_internal_tl to turn \if_false: constructions into actual begin-group and end-group tokens. This is done with a \_kernel_tl_gset:Nx assignment, and \_regex_extract_check:w is run immediately after this assignment ends, thanks to the \afterassignment primitive. If all of the items were properly balanced (enough begin-group tokens before end-group tokens, so \{| is not) then \_regex_extract_check:w is called just before the closing brace of the \_kernel_tl_gset:Nx (thanks to our sneaky \if_false: { \fi: } construction), and finds that there is nothing left to expand. If any of the items is unbalanced, the assignment gets ended early by an extra end-group token, and our check finds more tokens needing to be expanded in a new \_kernel_tl_gset:Nx assignment. We need to add a begin-group and an end-group tokens to the unbalanced item, namely to the last item found so far, which we reach through a loop.

\cs_new_protected:Npn \_regex_extract_check:w
\begin{verbatim}
\exp_after:wN \_regex_extract_check:n
\exp_after:wN { \if_false: } \fi:
\end{verbatim}
\cs_new_protected:Npn \_regex_extract_check:n #1
\begin{verbatim}
\tl_if_empty:nF {#1}
\int_incr:N \l__regex_added_begin_int
\int_incr:N \l__regex_added_end_int
\tex_afterassignment:D \_regex_extract_check:w
\_kernel_tl_gset:Nx \g__regex_internal_tl
\begin{verbatim}
\exp_after:wN \_regex_extract_check_loop:w
\g__regex_internal_tl
\_regex_tmp:w \_regex_extract_check_end:w
#1
\end{verbatim}
\end{verbatim}
\end{verbatim}
\cs_new:Npn \_regex_extract_check_loop:w #1 \_regex_tmp:w #2 \_regex_extract_check_end:w
\begin{verbatim}
{ \exp_not:o {#1} }
#2
\_regex_tmp:w { }
\_regex_extract_check_loop:w \prg_do_nothing:
\end{verbatim}

Arguments of \_regex_extract_check_end:w are: #1 is the part of the item before the extra end-group token; #2 is junk; #3 is \prg_do_nothing: followed by the not-yet-expanded part of the item after the extra end-group token. In the replacement text, the first brace and the \if_false: { \fi: } construction are the added begin-group and end-group tokens (the latter being not-yet expanded, just like #3), while the closing brace after \exp_not:o {#1} replaces the extra end-group token that had ended the assignment early. In particular this means that the character code of that end-group token is lost.

\cs_new:Npn \_regex_extract_check_end:w
\begin{verbatim}
\exp_not:o #1\#2 \_regex_extract_check_loop:w #3 \_regex_tmp:w
\end{verbatim}
\begin{verbatim}
{ \exp_not:o {#1} }
\end{verbatim}

#3

545
\_\_regex_extract_map:N
\_\_regex_extract_map_aux:NNn
\_\_regex_extract_map_loop:w

This receives a seq internal function and maps it over all items in \_g\_\_regex_internal_tl. This token list takes the form \_\_regex_tmp:w {} ⟨item1⟩ \_\_regex_tmp:w {} ⟨item2⟩ ... \_\_regex_tmp:w, and the calling code has set \_\_regex_tmp:w equal to \_\_regex_extract_map_loop:w. The loop is otherwise pretty standard, with \prg_do_nothing: to avoid losing braces.

\cs_new:Npn \_\_regex_extract_map:N #1
{\exp_after:wN \_\_regex_extract_map_aux:NNn \exp_after:wN #1 \g\_\_regex_internal_tl \use_none:nnn}
\cs_new:Npn \_\_regex_extract_map_aux:NNn #1#2#3
{ #3 #2 #1 \prg_do_nothing: }
\cs_new:Npn \_\_regex_extract_map_loop:w #1#2 \_\_regex_tmp:w #3
{\exp_after:wN #1 \exp_after:wN {#2} \_\_regex_tmp:w \#1 \prg_do_nothing: }

(End definition for \_\_regex_extract_map:N, \_\_regex_extract_map_aux:NNn, and \_\_regex_extract_map_loop:w.)

\_\_regex_extract: \_\_regex_extract_aux:w

Our task here is to store the list of end-points of submatches, and store them in appropriate array entries, from \_\_regex_zeroth_submatch_int upwards. First, we store in \_g\_\_regex_submatch_prev_intarray the position at which the match attempt started. We extract the rest from the comma list \_l\_\_regex_success_submatches_tl, which starts with entries to be stored in \_g\_\_regex_submatch_begin_intarray and continues with entries for \_g\_\_regex_submatch_end_intarray.

\cs_new_protected:Npn \_\_regex_extract:
{\if_meaning:w \c_true_bool \_\_regex_success_bool
 \int_set_eq:NN \l\_\_regex_zeroth_submatch_int \l\_\_regex_submatch_int
 \prg_replicate:nn \l\_\_regex_capturing_group_int
 \{ \_\_kernel_intarray_gset:Nnn \g\_\_regex_submatch_prev_intarray
 \{ \_\_regex_submatch_int \} \{ 0 \} \_\_kernel_intarray_gset:Nnn \g\_\_regex_submatch_case_intarray
 \{ \_\_regex_submatch_int \} \{ 0 \} \int_incr:N \_\_regex_submatch_int
 \} \_\_kernel_intarray_gset:Nnn \g\_\_regex_submatch_prev_intarray
 \{ \_\_regex_zeroth_submatch_int \} \{ \_\_regex_start_pos_int \} \_\_kernel_intarray_gset:Nnn \g\_\_regex_submatch_case_intarray
 \{ \_\_regex_zeroth_submatch_int \} \{ \g\_\_regex_case_int \} \int_zero:N \_\_regex_internal_a_int
 \exp_after:wN \_\_regex_extract_aux:w \_\_regex_success_submatches_tl
 \prg_break_point: \_\_regex_use_none_delimit_by_q_recursion_stop:w , \q\_\_regex_recursion_stop

546
45.7.4 Replacement

Build the NFA and the replacement functions, then find a single match. If the match failed, simply exit the group. Otherwise, we do the replacement. Extract submatches. Compute the brace balance corresponding to replacing this match by the replacement (this depends on submatches). Prepare the replaced token list: the replacement function produces the tokens from the start of the query to the start of the match and the replacement text for this match; we need to add the tokens from the end of the match to the end of the query. Finally, store the result in the user's variable after closing the group: this step involves an additional x-expansion, and checks that braces are balanced in the final result.
\__regex_replace_all:nnN

Match multiple times, and for every match, extract submatches and additionally store the position at which the match attempt started. The entries from \l__regex_min_submatch_int to \l__regex_submatch_int hold information about submatches of every match in order; each match corresponds to \l__regex_capturing_group_int consecutive entries. Compute the brace balance corresponding to doing all the replacements: this is the sum of brace balances for replacing each match. Join together the replacement texts for each match (including the part of the query before the match), and the end of the query.

At this stage \l__regex_internal_a_tl (x-expands to the desired result). Guess from \l__regex_balance_int the number of braces to add before or after the result then
try expanding. The simplest case is when \l__regex_internal_a_tl together with the
braces we insert via \prg_replicate:nn give a balanced result, and the assignment
ends at the \if_false: \{ \fi: \} construction: then \__regex_group_end_replace_check:w
sees that there is no material left and we successfully found the result. The harder case is
that expanding \l__regex_internal_a_tl may produce extra closing braces and end the
assignment early. Then we grab the remaining code using; importantly, what follows has
not yet been expanded so that \__regex_group_end_replace_check:n grabs everything
until the last brace in \__regex_group_end_replace_try:, letting us try again with an
extra surrounding pair of braces.

\cs_new_protected:Npn \__regex_group_end_replace:N \#1
\int_set:Nn \l__regex_added_begin_int
\int_max:nn { - \l__regex_balance_int } { 0 }
\int_set:Nn \l__regex_added_end_int
\int_max:nn { \l__regex_balance_int } { 0 }
\__regex_group_end_replace_try:
\int_compare:nNnT { \l__regex_added_begin_int + \l__regex_added_end_int } > 0
{ \msg_error:nnnn \regex \result-unbalanced
{ replacing } \int_use:N \l__regex_added_begin_int
\int_use:N \l__regex_added_end_int }
\group_end:
\tl_set_eq:NN \#1 \g__regex_internal_tl
\cs_new_protected:Npn \__regex_group_end_replace_try:
\tex_afterassignment:D \__regex_group_end_replace_check:w
\__kernel_tl_gset:Nx \g__regex_internal_tl
\prg_replicate:nn { \l__regex_added_begin_int } \{ \{ \if_false: \} \fi: \}
\l__regex_internal_a_tl
\prg_replicate:nn { \l__regex_added_end_int } \{ \if_false: \{ \fi: \} \}
\if_false: \{ \fi: \}
\cs_new_protected:Npn \__regex_group_end_replace_check:w
\exp_after:wN \__regex_group_end_replace_check:n
\exp_after:wN \if_false: \fi:
\cs_new_protected:Npn \__regex_group_end_replace_check:n \#1
\tl_if_empty:nF {\#1}
\int_incr:N \l__regex_added_begin_int
\int_incr:N \l__regex_added_end_int
\__regex_group_end_replace_try:
\cs_new_protected:Npn \__regex_group_end_replace:n #1
\tl_if_empty:nF {\#1}
\l__regex_internal_a_tl
\l__regex_internal_a_tl
\__regex_group_end_replace_try:
\cs_new_protected:Npn \__regex_group_end_replace:N and others.

(End definition for \__regex_group_end_replace:N and others.)
Peeking ahead

\__regex_peek_true_tl
\__regex_peek_false_tl

True/false code arguments of \peek_regex:nTF or similar.

\__regex_replacement_tl

When peeking in \peek_regex_replace_once:nnTF we need to store the replacement text.

\__regex_input_tl \__regex_input_item:n

Stores each token found as \__regex_input_item:n \{\langle tokens\rangle\}, where the \langle tokens\rangle o-expand to the token found, as for tl_analysis_map_inline:nn.

\peek_regex:nTF \peek_regex:NTF \peek_regex:nF \peek_regex:NTF \peek_regex_remove_once:nTF \peek_regex_remove_once:NTF

The T and F functions just call the corresponding TF function. The four TF functions differ along two axes: whether to remove the token or not, distinguished by using \__regex_peek_end: or \__regex_peek_remove_end:n (the latter case needs an argument, as we will see), and whether the regex has to be compiled or is already in an N-type variable, distinguished by calling \__regex_build_aux:Nn or \__regex_build_aux:NN. The first argument of these functions is \c_false_bool to indicate that there should be no implicit insertion of a wildcard at the start of the pattern: otherwise the code would keep looking further into the input stream until matching the regex.

\cs_new_protected:Npn \peek_regex:nTF #1
\cs_new_protected:Npn \peek_regex:NT #1#2
\cs_new_protected:Npn \peek_regex:nF #1
\cs_new_protected:Npn \peek_regex:NTF #1
\__regex_peek:nnTF
\__regex_peek_aux:nnTF

Store the user’s true/false codes (plus \texttt{\textbackslash group\_end:}) into two token lists. Then build the automaton with \#1, without submatch tracking, and aiming for a single match. Then start matching by setting up a few variables like for any regex matching like \texttt{\regex\_match:nTF}, with the addition of \texttt{\__regex\_input\_tl} that keeps track of the tokens seen, to reinsert them at the end. Instead of \texttt{\tl\_analysis\_map\_inline:nn} on the input, we call \texttt{\peek\_analysis\_map\_inline:n} to go through tokens in the input stream. Since \texttt{\__regex\_match\_one\_token:nN} calls \texttt{\__regex\_map\_like\_break}: we need to catch that and break the \texttt{\peek\_analysis\_map\_inline:n} loop instead.

\cs_new_protected:Npn \\__regex\_peek\_aux:nnTF \#1\#2\#3\#4
\{\group\_begin:
\tl\_set:Nn \\__regex\_peek\_true\_tl \{\texttt{\group\_end:} \#3\}
\tl\_set:Nn \\__regex\_peek\_false\_tl \{\texttt{\group\_end:} \#4\}
\__regex\_single\_match:
\#1
\__regex\_match\_init:
\tl\_build\_clear:N \\__regex\_input\_tl
\__regex\_match\_once\_init:
\peek\_analysis\_map\_inline:n
\{\tl\_build\_put\_right:Nn \\__regex\_input\_tl
\{\\__regex\_input\_item:n \{\#1\}\}
\__regex\_match\_one\_token:nN \{\#1\} \{\#2\} \#3
\use\_none:nN
\prg\_break\_point:Nn \\__regex\_map\_like\_break:
\{\\peek\_analysis\_map\_break:n \{\#2\}\}
\}
\}

(End definition for \texttt{\__regex\_peek:nTF} and \texttt{\__regex\_peek\_aux:nTF}.)
Once the regex matches (or permanently fails to match) we call \texttt{\_\_regex\_peek\_end:}, or \texttt{\_\_regex\_peek\_remove\_end:n} with argument the last token seen. For \texttt{\_\_peek\_regex:nTF} we reinsert tokens seen by calling \texttt{\_\_regex\_peek\_reinsert:N} regardless of the result of the match. For \texttt{\_\_peek\_regex\_remove\_once:nTF} we reinsert the tokens seen only if the match failed; otherwise we just reinsert the tokens \#1, with one expansion. To be more precise, \#1 consists of tokens that \texttt{o-expand} and \texttt{x-expand} to the last token seen, for example it is \texttt{\exp_not:N \langle cs \rangle} for a control sequence. This means that just doing \texttt{\exp_after:wN \l__regex\_peek\_true\_tl \#1} would be unsafe because the expansion of \texttt{\langle cs \rangle} would be suppressed.

\begin{verbatim}
\cs_new_protected:Npn \_\_regex\_peek\_end: \l__regex\_peek\_true\_tl \exp_end:
\cs_set_eq:NN \__regex\_input\_item:n \__regex\_reinsert\_item:n
\exp_after:wN #1 \exp:w \l__regex\_input\_tl \exp_end:
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \_\_regex\_peek\_remove\_end:n #1
\bool_if:NTF \g__regex\_success\_bool
{ \__regex\_peek\_reinsert:N \l__regex\_peek\_true\_tl #1 }
{ \__regex\_peek\_reinsert:N \l__regex\_peek\_false\_tl }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__regex\_peek\_reinsert:N #1
\tl_build_end:N \l__regex\_input\_tl
\cs_set_eq:NN \__regex\_input\_item:n \__regex\_reinsert\_item:n
\exp_after:wN \exp_after:wN \exp_after:wN \exp_after:wN \exp_after:wN
#1 \exp:w \l__regex\_input\_tl \exp_end:
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \_\_peek\_regex\_replace\_once:nnTF \langle cs \rangle \{ \}
\cs_new_protected:Npn \_\_peek\_regex\_replace\_once:NnTF \langle cs \rangle \{ \}
\end{verbatim}

Similar to \texttt{\_\_peek\_regex:nTF} above.
\__regex_peek_replace:nnTF

Same as \__regex_peek:nTF (used for \peek_regex:nTF above), but without disabling
submatches, and with a different end. The replacement text #2 is stored, to be analyzed
later.

\__regex_peek_replace_end:

If the match failed \__regex_peek_reinsert:N reinserts the tokens found. Otherwise,
finish storing the submatch information using \__regex_extract:, and store the input
into \toks. Redefine a few auxiliaries to change slightly their expansion behaviour as
explained below. Analyse the replacement text with \__regex_replacement:n, which
as usual defines \__regex_replacement_do_one_match:n to insert the tokens from the
start of the match attempt to the beginning of the match, followed by the replacement
text. The \use:x expands for instance the trailing \__regex_query_range:nn down to a
sequence of \__regex_reinsert_item:n {⟨tokens⟩} where ⟨tokens⟩ o-expand to a single
token that we want to insert. After x-expansion, \use:x does \use:n, so we have \exp-af
\__regex_reinsert_item:n \__regex_reinsert_true_tl \exp:w ... \exp:end:. This is set up such as to
obtain \__regex_reinsert_true_tl followed by the replaced tokens (possibly unbalanced)
in the input stream.

\__regex_peek_replace:nnTF

(End definition for \peek_regex_replace_once:nnTF and \peek_regex_replace_once:NnTF. These func-
tions are documented on page 195.)
(End definition for \__regex_peek_replace_end:)

\__regex_query_set_from_input_tl:

The input was stored into \__regex_input_tl as successive items \__regex_input_item:n \{⟨tokens⟩\}. Store that in successive \toks. It's not clear whether the empty entries before and after are both useful.

(End definition for \__regex_query_set_from_input_tl: and \__regex_query_set_item:n)

\__regex_peek_replacement_put:n

While building the replacement function \__regex_peek_replacement_do_one_match:n, we often want to put simple material, given as #1, whose \texttt{x}-expansion \texttt{o}-expands to a single token. Normally we can just add the token to \__regex_build_tl, but for \texttt{peek_replacement_once:nnNTF} we eventually want to do some strange expansion that is basically using \texttt{exp_after:wN} to jump through numerous tokens (we cannot use \texttt{x}-expansion like for \texttt{regex_replace_once:nnNTF} because it is ok for the result to be unbalanced since we insert it in the input stream rather than storing it. When within a csname we don’t do any such shenanigan because \texttt{cs:w ... cs_end:} does all the expansion we need.

554
\fi:
\}

(End definition for \_\_\_regex.peek.replacement.put:n)
\_\_\_regex.peek.replacement.token:n
When hit with \exp:w, \_\_\_regex.peek.replacement.token:n \{(token)\} stops \exp:-end: and does \exp_after:wN \{token\} \exp:w to continue expansion after it.
\cs_new_protected:Npn \_\_\_regex.peek.replacement.token:n #1
\{ \exp_after:wN \exp_end: \exp_after:wN #1 \exp:w \}

(End definition for \_\_\_regex.peek.replacement.token:n)
\_\_\_regex.peek.replacement.put.submatch_aux:n
While analyzing the replacement we also have to insert submatches found in the query. Since query items \_\_\_regex.input.item:n \{\{tokens\}\} expand correctly only when surrounded by \exp:w ... \exp_end:, and since these expansion controls are not there within csnames (because \cs:w ... \cs_end: make them unnecessary in most cases), we have to put \exp:w and \exp_end: by hand here.
\cs_new_protected:Npn \_\_\_regex.peek.replacement.put.submatch_aux:n #1
\{ \if_case:w \l__regex.replacement.csnames_int
\tl_build_put_right:Nn \l__regex.build_tl \{ \_\_\_regex.query.submatch:n \{ \int_eval:n \{ #1 + ##1 \} \} \}
\else:
\tl_build_put_right:Nn \l__regex.build_tl \{ \exp:w \_\_\_regex.query.submatch:n \{ \int_eval:n \{ #1 + ##1 \} \} \exp_end: \}
\fi:
\}

(End definition for \_\_\_regex.peek.replacement.put.submatch_aux:n)
\_\_\_regex.peek.replacement.var:N
This is used for \u outside csnames. It makes sure to continue expansion with \exp:w before expanding the variable \#1 and stopping the \exp:w that precedes.
\cs_new_protected:Npn \_\_\_regex.peek.replacement.var:N #1
\{ \exp_after:wN \exp_last_unbraced:NV \exp_after:wN \exp_end: \exp_after:wN #1 \exp:w \}

(End definition for \_\_\_regex.peek.replacement.var:N)

45.8 Messages
Messages for the pre-parsing phase.
\use:x
\{ \msg_new:nnn { regex } { trailing-backslash } \{ Trailing-\iow_char:N\\-in-regex-or-replacement. \} \msg_new:nnn { regex } { x-missing-rbrace } \{ Missing-brace-\iow_char:N\\-in-regex- \...\iow_char:N\x\iow_char:N\\{...\#1’.

555
\msg_new:nnn { regex } { x-overflow }
   { Character-code-##1-too-large-in-
     \iow_char:N\x\iow_char:N\{\#2\iow_char:N\}-regex. }
\}

Invalid quantifier.
\msg_new:nnnn { regex } { invalid-quantifier }
   { Braced-quantifier-’#1’-may-not-be-followed-by-’#2’. }
   { The-character-’#2’-is-invalid-in-the-braced-quantifier-’#1’.-
     The-only-valid-quantifiers-are-’*’,-’?’,-’+’,-’{<int>}’,-’{<min>,}’-and-’{<min>,<max>}’,-optionally-followed-by-’?’.
   }

Messages for missing or extra closing brackets and parentheses, with some fancy
singular/plural handling for the case of parentheses.
\msg_new:nnnn { regex } { missing-rbrack }
   { Missing-right-bracket-inserted-in-regular-expression. }
   { LaTeX-was-given-a-regular-expression-where-a-character-class-
     was-started-with-’[’,-but-the-matching-’]’-is-missing.
   }
\msg_new:nnnn { regex } { missing-rparen }
   { Missing-right-
     \int_compare:nTF { #1 = 1 } { parenthesis } { parentheses } -
     inserted-in-regular-expression. }
   { LaTeX-was-given-a-regular-expression-with-\int_eval:n {#1} -
     more-left-parentheses-than-right-parentheses.
   }
\msg_new:nnnn { regex } { extra-rparen }
   { Extra-right-parenthesis-ignored-in-regular-expression. }
   { LaTeX-came-across-a-closing-parenthesis-when-no-submatch-group-
     was-open.-The-parenthesis-will-be-ignored.
   }

Some escaped alphanumerics are not allowed everywhere.
\msg_new:nnnn { regex } { bad-escape }
\_\regex_if_in_cs:TF { within-a-control-sequence. }
\_\regex_if_in_class:TF
   { in-a-character-class. }
   { following-a-category-test. }
\}
\{ The-escape-sequence-’\iow_char:N\{\#1’-may-not-appear-
\__regex_if_in_cs:TF
{
  within-a-control-sequence-test-introduced-by-
   \iow_char:N\c\iow_char:N\{'
  }
{
  \__regex_if_in_class:TF
    { within-a-character-class- }
    { following-a-category-test-such-as-\iow_char:N\cL- }
    because-it-does-not-match-exactly-one-character.
  }
}

Range errors.
\msg_new:nnnn { regex } { range-missing-end }
{ Invalid-end-point-for-range-`#1-#2'-in-character-class. }
{
  The-end-point-`#2'-of-the-range-`#1-#2'-may-not-serve-as-an-
  end-point-for-a-range: alphanumeric-characters-should-not-be-
  escaped, and-non-alphanumeric-characters-should-be-escaped.
}
\msg_new:nnnn { regex } { range-backwards }
{ Range-`[#1-#2]'-out-of-order-in-character-class. }
{
  In-ranges-of-characters-`[x-y]'-appearing-in-character-classes,-
  the-first-character-code-must-not-be-larger-than-the-second.-
  Here, `#1'-has-character-code \int_eval:n{`#1}, while-
  `#2'-has-character-code \int_eval:n{`#2}.
}

Errors related to \c and \u.
\msg_new:nnnn { regex } { c-bad-mode }
{ Invalid-nested-`\iow_char:N\c'-escape-in-regular-expression. }
{
  The-`\iow_char:N\c'-escape-cannot-be-used-within-
  a-control-sequence-test-`\iow_char:N\c{...}'-
  nor-another-category-test.-
  To-combine-several-category-tests, use-`\iow_char:N\c[...]'.
}
\msg_new:nnnn { regex } { c-C-invalid }
{ `\iow_char:N\cC'-should-be-followed-by-`,'-or-', `',-not-`#1'. }
{
  The-`\iow_char:N\cC'-construction-restricts-the-next-item-to-be-a-
  control-sequence-or-the-next-group-to-be-made-of-control-sequences.-
  It-only-makes-sense-to-follow-it-by-, or-by-a-group.
}
\msg_new:nnnn { regex } { cu-lbrace }
{ Left-braces-must-be-escaped-in-`\iow_char:N\#{...}'.
{
  Constructions-such-as-`\iow_char:N\#{...\iow_char:N\{...}'-are-
  not-allowed-and-should-be-replaced-by-
  `\iow_char:N\#{...\token_to_str:N\{...}'.
}
\msg_new:nnnn { regex } { c-lparen-in-class }
{ Catcode-test-cannot-apply-to-group-in-character-class }
Construction such as `\iow_char:N\cL(abc)` are not allowed inside a class `'[...]' because classes do not match multiple characters at once.

\msg_new:nnnn { regex } { c-missing-rbrace }
\{ Missing-right-brace-inserted-for-`\iow_char:N\c'-escape. \}
\{ \LaTeX\ was\ given\ a\ regular\ expression\ where\ a-
`\iow_char:N\c\iow_char:N\{...' construction was not ended-
with\ a\ closing\ brace-`\iow_char:N\}'. \}

\msg_new:nnnn { regex } { c-missing-brack }
\{ Missing-right-bracket-inserted-for-`\iow_char:N\c'-escape. \}
\{ A-construction-`\iow_char:N\c[...'-appears\ in\ a-
 regular-expression,-but\ the\ closing-']'-is\ not\ present. \}

\msg_new:nnnn { regex } { c-missing-category }
\{ Invalid-character-'#1'-following-`\iow_char:N\c'-escape. \}
\{ In-regular-expressions,-the-`\iow_char:N\c'-escape-sequence-
 may-only-be-followed-by-a-left-brace,-a-left-bracket,-or-a-
capital-letter-representing-a-character-category,-namely-
one-of-`ABCDELMOPSTU'. \}

\msg_new:nnnn { regex } { c-trailing }
\{ Trailing-category-code-escape-`\iow_char:N\c'... \}
\{ A-regular-expression-ends-with-`\iow_char:N\c'-followed-
 by\ a\ letter.-It\ will\ be\ ignored. \}

\msg_new:nnnn { regex } { u-missing-lbrace }
\{ Missing-left-brace-following-`\iow_char:N\u'-escape. \}
\{ The-`\iow_char:N\u'-escape-sequence-must-be-followed-by-
a-brace-group-with-the-name-of-the-variable-to-use. \}

\msg_new:nnnn { regex } { u-missing-rbrace }
\{ Missing-right-brace-inserted-for-`\iow_char:N\u'-escape. \}
\{ \LaTeX-\\\str_if_eq:eeTF { } {#2}
\{ reached\ the\ end\ of\ the\ string\ }\}
\{ encountered\ an\ escaped\ alphanumeric-character
`\iow_char:N\#2'- \}
\{ when-parsing\ the\ argument\ of\ an-
`\iow_char:N\u\iow_char:N\{...\}'-escape. \}

Errors when encountering the POSIX syntax `[:...:]`.
\msg_new:nnnn { regex } { posix-unsupported }
\{ POSIX-collating-element-`[#1 - #1]'-not-supported. \}
\{ The-`[..foo..]'-and-`[..bar..]'-syntases have a special meaning-
in-POSIX-regular-expressions.-This\ is\ not\ supported-by-LaTeX.-
Maybe\ you\ forgot\ to\ escape\ a\ left\ bracket\ in\ a\ character\ class? \}

558
In various cases, the result of a \texttt{l3regex} operation can leave us with an unbalanced token list, which we must re-balance by adding begin-group or end-group character tokens.

Error message for unknown options.

Errors in the replacement text.
\msg_new:nnnn { regex } { replacement-g }
{
  Missing-brace-for-the-\iow_char:N\g\{-construction-in-a-replacement-text.
}
\msg_new:nnnn { regex } { replacement-catcode-end }
{
  Missing-character-for-the-\iow_char:N\c<category><character>-construction-in-a-replacement-text.
}
\msg_new:nnnn { regex } { replacement-catcode-escaped }
{
  Escaped-letter-or-digit-after-category-code-in-replacement-text.
}
\msg_new:nnnn { regex } { replacement-catcode-in-cs }
{
  Category-code\iow_char:N\c\{\#1\#3\}'-ignored-inside-\iow_char:N\c\{\ldots\}'-in-a-replacement-text.
}
\msg_new:nnnn { regex } { replacement-null-space }
{ TeX-cannot-build-a-space-token-with-character-code-0. }
\msg_new:nnnn { regex } { replacement-missing-rbrace }
{
  Missing-right-brace-inserted-in-replacement-text.
}
\msg_new:nnnn { regex } { replacement-base }
{
  There-\int_compare:nTF { #1 = 1 } { was } { were } - #1-
8310  \msg_new:nnnn { regex } { replacement-missing-rparen }
8311  { Missing-right-parenthesis-inserted-in-replacement-text. }
8312  { There- \int_compare:nTF { #1 = 1 } { was } { were } - #1-
8313  \int_compare:nTF { #1 = 1 } { parenthesis } { parentheses } .
8314  }
8315  \msg_new:nnn { regex } { submatch-too-big }
8316  { Submatch-#1-used-but-regex-only-has-#2-group(s) } 
8317  \msg_new:nnn { regex } { backwards-quantifier }
8318  { Quantifier-"{#1,#2}"-is-backwards. }
8319  \msg_new:nnn { regex } { case-odd }
8320  { #1-with-odd-number-of-items }
8321  \msg_new:nnn { regex } { show }
8322  { >-Compiled-regex-
8323  \tl_if_empty:nTF {#1} { variable- #2 } { {#1} } : #3
8324  }
8325  \prop_gput:Nnn \g_msg_module_name_prop { regex } { LaTeX3 }
8326  \prop_gput:Nnn \g_msg_module_type_prop { regex } { }

\_regex_msg_repeated:nnN  This is not technically a message, but seems related enough to go there. The arguments are: #1 is the minimum number of repetitions; #2 is the number of allowed extra repetitions (−1 for infinite number), and #3 tells us about lazyness.

\cs_new:Npn \_regex_msg_repeated:nnN #1#2#3
\str_if_eq:eeF { #1 #2 } { 1 0 }
\int_case:nnF {#2}
\{ -1 } { #1-or-more-times,-\bool_if:NTF #3 { lazy } { greedy } }
\{ 0 } { #1-times }
\}
\between-#1-and-\int_eval:n {#1+#2}-times,-
\bool_if:NTF #3 { lazy } { greedy }
\}
\}

(End definition for \_regex_msg_repeated:nnN.)
45.9 Code for tracing

There is a more extensive implementation of tracing in the l3trial package \texttt{l3trace}. Function names are a bit different but could be merged.

```latex
\begin{verbatim}
\__regex_trace_push:nnN \__regex_trace_pop:nnN \__regex_trace:nnx
Here \#1 is the module name (\texttt{regex}) and \#2 is typically 1. If the module's current tracing level is less than \#2 show nothing, otherwise write \#3 to the terminal.

\cs_new_protected:Npn \__regex_trace_push:nnN #1#2#3
  { \__regex_trace:nnx {#1} {#2} { entering \token_to_str:N #3 } }
\cs_new_protected:Npn \__regex_trace_pop:nnN #1#2#3
  { \__regex_trace:nnx {#1} {#2} { leaving \token_to_str:N #3 } }
\cs_new_protected:Npn \__regex_trace:nnx #1#2#3
  { \int_compare:nNnF { \int_use:c { g__regex_trace_#1_int } } < {#2} }\iow_term:x { Trace:~#3 }
\end{verbatim}

(End definition for \texttt{\__regex_trace_push:nnN}, \texttt{\__regex_trace_pop:nnN}, and \texttt{\__regex_trace:nnx}.)

\begin{verbatim}
\g__regex_trace_regex_int
No tracing when that is zero.
\int_new:N \g__regex_trace_regex_int
(End definition for \texttt{\g__regex_trace_regex_int}.)
\__regex_trace_states:n
This function lists the contents of all states of the NFA, stored in \texttt{\toks} from 0 to \texttt{\l__regex_max_state_int} (excluded).
\begin{verbatim}
\cs_new_protected:Npn \__regex_trace_states:n #1
  { \int_step_inline:nnn { \l__regex_min_state_int } { \l__regex_max_state_int - 1 } { \__regex_trace:nnx { regex } {#1}
    \iow_char:N \toks ##1 = { \__regex_toks_use:w ##1 } }
\end{verbatim}

(End definition for \texttt{\__regex_trace_states:n}.)
```

\hspace*{1cm}

\begin{verbatim}
\end{verbatim}
Chapter 46

\textbf{l3prg implementation}

The following test files are used for this code: \texttt{m3prg001.lvt}, \texttt{m3prg002.lvt}, \texttt{m3prg003.lvt}.

\texttt{\langle package\rangle}

\section{Primitive conditionals}

Those two primitive \TeX conditionals are synonyms. \texttt{\if_bool:N} is defined in \texttt{l3basics}, as it’s needed earlier to define quark test functions.

\texttt{\cs_new_eq:NN \if_predicate:w \tex_ifodd:D}

(End definition for \texttt{\if_bool:N} and \texttt{\if_predicate:w}. These functions are documented on page 70.)

\section{Defining a set of conditional functions}

These are all defined in \texttt{l3basics}, as they are needed “early”. This is just a reminder!

(End definition for \texttt{\prg_set_conditional:Npnn} and others. These functions are documented on page 62.)

\section{The boolean data type}

Boolean variables have to be initiated when they are created. Other than that there is not much to say here.

\texttt{\cs_new_protected:Npp \bool_new:N \bool_new:c}

A merger between \texttt{\tl_const:Nn} and \texttt{\bool_set:Nn}.

\texttt{\cs_new_protected:Np \bool_new:N \bool_wrong:N #1 \c_false_bool}

\texttt{\cs_generate_variant:Nn \bool_new:N { c }}

(End definition for \texttt{\bool_new:N}. This function is documented on page 65.)

563
Setting is already pretty easy. When check-declarations is active, the definitions are patched to make sure the boolean exists. This is needed because booleans are not based on token lists nor on TeX registers.

\begin{Verbatim}
\bool_set_true:N     \cs_new_protected:Npn \bool_set_true:N #1 { \cs_set_eq:NN #1 \c_true_bool }
\bool_set_false:N    \cs_new_protected:Npn \bool_set_false:N #1 { \cs_set_eq:NN #1 \c_false_bool }
\bool_gset_true:N    \cs_new_protected:Npn \bool_gset_true:N #1 { \cs_gset_eq:NN #1 \c_true_bool }
\bool_gset_false:N   \cs_new_protected:Npn \bool_gset_false:N #1 { \cs_gset_eq:NN #1 \c_false_bool }
\bool_set_eq:NN      \cs_new_eq:NN \bool_set_eq:NN \tl_set_eq:NN
\bool_set_eq:cN      \cs_new_eq:NN \bool_set_eq:cN \tl_set_eq:cN
\bool_set_eq:Nc      \cs_new_eq:NN \bool_set_eq:Nc \tl_set_eq:Nc
\bool_set_eq:cc      \cs_new_eq:NN \bool_set_eq:cc \tl_set_eq:cc
\end{Verbatim}

The usual copy code. While it would be cleaner semantically to copy the \cs_set_eq:NN family of functions, we copy \tl_set_eq:NN because that has the correct checking code.

\begin{Verbatim}
\bool_set:Nn \cs_new_protected:Npn \bool_set:Nn #1#2 { \exp_last_unbraced:NNNf \tex_chardef:D #1 = { \bool_if_p:n {#2} } }
\bool_gset:Nn \cs_new_protected:Npn \bool_gset:Nn #1#2 { \exp_last_unbraced:NNNNf \tex_global:D \tex_chardef:D #1 = { \bool_if_p:n {#2} } }
\end{Verbatim}

This function evaluates a boolean expression and assigns the first argument the meaning \c_true_bool or \c_false_bool. Again, we include some checking code. It is important to evaluate the expression before applying the \chardef primitive, because that primitive sets the left-hand side to \scan_stop: before looking for the right-hand side.

\begin{Verbatim}
\bool_set:Nn \cs_new_protected:Npn \bool_set:Nn #1#2 { }
\bool_set:cn \cs_new_protected:Npn \bool_set:cn #1#2 { }
\bool_gset:Nn \cs_new_protected:Npn \bool_gset:Nn #1#2 { }
\bool_gset:cn \cs_new_protected:Npn \bool_gset:cn #1#2 { }
\end{Verbatim}

This function evaluates a boolean expression and assigns the first argument the meaning \c_true_bool or \c_false_bool. Again, we include some checking code. It is important to evaluate the expression before applying the \chardef primitive, because that primitive sets the left-hand side to \scan_stop: before looking for the right-hand side.

\begin{Verbatim}
\bool_set:Nn \cs_new_protected:Npn \bool_set:Nn #1#2 { }
\bool_set:cn \cs_new_protected:Npn \bool_set:cn #1#2 { }
\bool_gset:Nn \cs_new_protected:Npn \bool_gset:Nn #1#2 { }
\bool_gset:cn \cs_new_protected:Npn \bool_gset:cn #1#2 { }
\end{Verbatim}

(End definition for \bool_set:N and others. These functions are documented on page 65.)
46.4 Internal auxiliaries

\q__bool_recursion_tail \q__bool_recursion_stop

Internal recursion quarks.

\quark_new:N \q__bool_recursion_tail
\quark_new:N \q__bool_recursion_stop

(End definition for \q__bool_recursion_tail and \q__bool_recursion_stop.)

\bool_use_i_delimit_by_q_recursion_stop:nw

Functions to gobble up to a quark.

\cs_new:Npn \__bool_use_i_delimit_by_q_recursion_stop:nw
\#1 \#2 \q__bool_recursion_stop \{#1\}

(End definition for \__bool_use_i_delimit_by_q_recursion_stop:nw.)

\__bool_if_recursion_tail_stop_do:nn

Functions to query recursion quarks.

\__kernel_quark_new_test:N \__bool_if_recursion_tail_stop_do:nn

(End definition for \__bool_if_recursion_tail_stop_do:nn.)

\bool_if_p:N \bool_if_p:c
\bool_if:NNTF \bool_if:cNTF

Straight forward here. We could optimize here if we wanted to as the boolean can just be input directly.

\prg_new_conditional:Npnn \bool_if:N #1 { p , T , F , TF }
\{ \if_bool:N #1 \prg_return_true: \else: \prg_return_false: \fi: \}
\prg_generate_conditional_variant:Nnn \bool_if:N { c } { p , T , F , TF }

(End definition for \bool_if:NTF. This function is documented on page 65.)

\bool_to_str:N \bool_to_str:c
\bool_to_str:n

Expands to true or false with category code letter.

\cs_new:Npn \bool_to_str:N \bool_to_str:c \bool_to_str:n
\#1 \{ \bool_if:NNTF \#1 \{ \{ true \} \{ false \} \}
\{ \cs_generate_variant:Nn \bool_to_str:N \bool_to_str:c \bool_to_str:n \#1 \{ \bool_if:NNTF \#1 \{ \{ true \} \{ false \} \}

(End definition for \bool_to_str:n and \bool_to_str:n. These functions are documented on page 65.)

Show the truth value of the boolean.

\cs_new_protected:Npn \bool_show:n \bool_log:n
\{ \msg_show_eval:Nn \bool_to_str:n \}
\cs_new_protected:Npn \bool_show:c \bool_log:c
\{ \msg_log_eval:Nn \bool_to_str:n \}

(End definition for \bool_show:n and \bool_log:n. These functions are documented on page 66.)

Show the truth value of the boolean, as true or false.

\cs_new_protected:Npn \bool_show:N \bool_log:N \bool_show:c \bool_log:c
\{ \_bool_show:NN \tl_show:n \}
\cs_new_protected:Npn \bool_show:N \bool_log:N \bool_show:c \bool_log:c
\{ \_bool_show:NN \tl_log:n \}

565
\bool_new:N \l_tmpa_bool
\bool_new:N \l_tmpb_bool
\bool_new:N \g_tmpa_bool
\bool_new:N \g_tmpb_bool

A few booleans just if you need them.

\bool_if_exist_p:N \bool_if_exist_p:c
\bool_if_exist:N \cs_if_exist:N
\bool_if_exist:c \cs_if_exist:c

Copies of the cs functions defined in l3basics.

\bool_if_p:n \bool_if:n

Evaluating the truth value of a list of predicates is done using an input syntax somewhat similar to the one found in other programming languages with \texttt{(} and \texttt{)} for grouping, \texttt{!} for logical “Not”, \texttt{&&} for logical “And” and \texttt{||} for logical “Or”. However, they perform eager evaluation. We shall use the terms Not, And, Or, Open and Close for these operations.

Any expression is terminated by a Close operation. Evaluation happens from left to right in the following manner using a GetNext function:

- If an Open is seen, start evaluating a new expression using the Eval function and call GetNext again.
- If a Not is seen, remove the \texttt{!} and call a GetNext function with the logic reversed.
- If none of the above, reinsert the token found (this is supposed to be a predicate function) in front of an Eval function, which evaluates it to the boolean value \texttt{⟨true⟩} or \texttt{⟨false⟩}.

The Eval function then contains a post-processing operation which grabs the instruction following the predicate. This is either And, Or or Close. In each case the truth value is used to determine where to go next. The following situations can arise:
And Current truth value is true, logical And seen, continue with GetNext to examine truth value of next boolean (sub-)expression.

And Current truth value is false, logical And seen, stop using the values of predicates within this sub-expression until the next Close. Then return \texttt{false}.

Or Current truth value is true, logical Or seen, stop using the values of predicates within this sub-expression until the nearest Close. Then return \texttt{true}.

Or Current truth value is false, logical Or seen, continue with GetNext to examine truth value of next boolean (sub-)expression.

Close Current truth value is true, Close seen, return \texttt{true}.

Close Current truth value is false, Close seen, return \texttt{false}.

\begin{verbatim}
\bool_if:nTF \#1 \{ T , F , TF \}
\{
  \if_predicate:w \bool_if_p:n \#1
  \prg_return_true:
  \else:
  \prg_return_false:
  \fi:
\}
\end{verbatim}

(End definition for \texttt{\bool_if:nTF}. This function is documented on page 67.)

\begin{verbatim}
\bool_if_p:n \\bool_if_p:n \\_bool_if_p_aux:w
To speed up the case of a single predicate, f-expand and check whether the result is one token (possibly surrounded by spaces), which must be \texttt{c_true_bool} or \texttt{c_false_bool}. We use a version of \texttt{tl_if_single:nTF} optimized for speed since we know that an empty \#1 is an error. The auxiliary \texttt{\_bool_if_p_aux:w} removes the trailing parenthesis and gets rid of any space. For the general case, first issue a \texttt{\group_align_safe_begin}: as we are using \&\& as syntax shorthand for the And operation and we need to hide it for \TeX. This group is closed after \texttt{\_bool_get_next:NN} returns \texttt{c_true_bool} or \texttt{c_false_bool}. That function requires the trailing parenthesis to know where the expression ends.

\begin{verbatim}
\cs_new:Npn \bool_if_p:n \{ \exp_args:Nf \_bool_if_p:n \}
\cs_new:Npn \_bool_if_p:n \#1
\{
  \tl_if_empty:oT { \use_none:nn \#1 . } { \_bool_if_p_aux:w }
  \group_align_safe_begin:
  \exp_after:wN \group_align_safe_end:
  \exp:w \exp_end_continue_f:w \%
  \_bool_get_next:NN \use_i:nnnn \#1
\}
\end{verbatim}

(End definition for \texttt{\bool_if_p:n}, \texttt{\_bool_if_p:n}, and \texttt{\_bool_if_p_aux:w}. This function is documented on page 67.)

\begin{verbatim}
\_bool_get_next:NN
The GetNext operation. Its first argument is \texttt{use_i:nnnn}, \texttt{use_ii:nnnn}, \texttt{use_iii:nnnn}, or \texttt{use_iv:nnnn} (we call these “states”). In the first state, this function eventually expand to the truth value \texttt{c_true_bool} or \texttt{c_false_bool} of the expression
\end{verbatim}

567
which follows until the next unmatched closing parenthesis. For instance “\_bool-_get_next:NN \use_i:nnnn \c_true_bool && \c_true_bool )” (including the closing parenthesis) expands to \c_true_bool. In the second state (after a !) the logic is reversed. We call these two states “normal” and the next two “skipping”. In the third state (after \c_true_bool|) it always returns \c_true_bool. In the fourth state (after \c_false_bool&&) it always returns \c_false_bool and also stops when encountering ||, not only parentheses. This code itself is a switch: if what follows is neither ! nor (, we assume it is a predicate.

\__bool_!:Nw

The Not operation reverses the logic: it discards the ! token and calls the GetNext operation with the appropriate first argument. Namely the first and second states are interchanged, but after \c_true_bool| or \c_false_bool&& the ! is ignored.

\__bool_(:Nw

The Open operation starts a sub-expression after discarding the open parenthesis. This is done by calling GetNext (which eventually discards the corresponding closing parenthesis), with a post-processing step which looks for And, Or or Close after the group.

\__bool_p:Nw

If what follows GetNext is neither ! nor (, evaluate the predicate using the primitive \int_value:w. The canonical true and false values have numerical values 1 and 0 respectively. Look for And, Or or Close afterwards.
The arguments are #1: a function such as \use_i:nnnn, #2: 0 or 1 encoding the current truth value, #3: the next operation, And, Or or Close. We distinguish three cases according to a combination of #1 and #2. Case 2 is when #1 is \use_i:nnnn (state 3), namely after \c_true_bool ||. Case 1 is when #1 is \use_i:nnnn and #2 is true or when #1 is \use_i:nnnn and #2 is false, for instance for !\c_false_bool. Case 0 includes the same with true/false interchanged and the case where #1 is \use_iv:nnnn namely after \c_false_bool &&.

When seeing ) the current subexpression is done, leave the appropriate boolean.

In case 1, namely when the argument is true and we are in a normal state continue in the normal state 1. In case 2, namely when skipping alternatives in an Or, continue in the same state. When seeing | in case 0, continue in a normal state; in particular stop skipping for \c_false_bool && because that binds more tightly than ||. In the other two cases start skipping for \c_true_bool ||.

\bool_lazy_all_p:n \bool_lazy_all:n TF \__bool_lazy_all:n

Go through the list of expressions, stopping whenever an expression is false. If the end is reached without finding any false expression, then the result is true.
\bool_lazy_and_p:nn
\bool_lazy_and:nnTF

Only evaluate the second expression if the first is true. Note that #2 must be removed as an argument, not just by skipping to the \else: branch of the conditional since \#2 may contain unbalanced \TeX\ conditionals.

\bool_lazy_any_p:n
\bool_lazy_any:n
\__bool_lazy_any:n

Go through the list of expressions, stopping whenever an expression is true. If the end is reached without finding any true expression, then the result is false.

\bool_lazy_or_p:nn
\bool_lazy_or:nnTF

Only evaluate the second expression if the first is false.
\bool_not_p:n
The Not variant just reverses the outcome of \bool_if_p:n. Can be optimized but this
is nice and simple and according to the implementation plan. Not even particularly useful
to have it when the infix notation is easier to use.
\bool_xor_p:nn
\bool_xor:nnTF
Exclusive or. If the boolean expressions have same truth value, return false, otherwise
return true.
\bool_while_do:Nn
\bool_until_do:Nn
A while loop where the boolean is tested before executing the statement. The “while”
version executes the code as long as the boolean is true; the “until” version executes the
code as long as the boolean is false.
\bool_do_while:Nn
\bool_do_until:Nn
A do-while loop where the body is performed at least once and the boolean is tested
after executing the body. Otherwise identical to the above functions.
\bool_do:nn
Loop functions with the test either before or after the first body expansion.
\bool_while_do:nn {#1} {#2}
\cs_new:Npn \bool_do_while:nn #1#2
{
  #2
  \bool_if:nT {#1} { \bool_do_while:nn {#1} {#2} }
}
\cs_new:Npn \bool_until_do:nn #1#2
{
  \bool_if:nF {#1}
    {
      #2
      \bool_until_do:nn {#1} {#2}
    }
}
\cs_new:Npn \bool_do_until:nn #1#2
{
  #2
  \bool_if:nF {#1} { \bool_do_until:nn {#1} {#2} }
}

(End definition for \bool_while_do:nn and others. These functions are documented on page 69.)

46.7 Producing multiple copies

\prg_replicate:nn
\__prg_replicate:N
\__prg_replicate_first:N
\__prg_replicate_0:n
\__prg_replicate_1:n
\__prg_replicate_2:n
\__prg_replicate_3:n
\__prg_replicate_4:n
\__prg_replicate_5:n
\__prg_replicate_6:n
\__prg_replicate_7:n
\__prg_replicate_8:n
\__prg_replicate_9:n
\__prg_replicate_first_:n
\__prg_replicate_first_0:n
\__prg_replicate_first_1:n
\__prg_replicate_first_2:n
\__prg_replicate_first_3:n
\__prg_replicate_first_4:n
\__prg_replicate_first_5:n
\__prg_replicate_first_6:n
\__prg_replicate_first_7:n
\__prg_replicate_first_8:n
\__prg_replicate_first_9:n

This function uses a cascading csname technique by David Kastrup (who else :-) The idea is to make the input 25 result in first adding five, and then 20 copies of the code to be replicated. The technique uses cascading csnames which means that we start building several csnames so we end up with a list of functions to be called in reverse order. This is important here (and other places) because it means that we can for instance make the function that inserts five copies of something to also hand down ten to the next function in line. This is exactly what happens here: in the example with 25 then the next function is the one that inserts two copies but it sees the ten copies handed down by the previous function. In order to avoid the last function to insert say, 100 copies of the original argument just to gobble them again we define separate functions to be inserted first. These functions also close the expansion of \exp:w, which ensures that \prg_replicate:nn only requires two steps of expansion.

This function has one flaw though: Since it constantly passes down ten copies of its previous argument it severely affects the main memory once you start demanding hundreds of thousands of copies. Now I don’t think this is a real limitation for any ordinary use, and if necessary, it is possible to write \prg_replicate:nn {1000} \prg_replicate:nn {1000} \langle\text{code}\rangle. An alternative approach is to create a string of m’s with \exp:w which can be done with just four macros but that method has its own problems since it can exhaust the string pool. Also, it is considerably slower than what we use here so the few extra csnames are well spent I would say.
Then comes all the functions that do the hard work of inserting all the copies. The first function takes \texttt{:n} as a parameter.

\begin{verbatim}
\cs_new:Npn \__prg_replicate_:n #1 \{ \cs_end: \}
\cs_new:cpn { __prg_replicate_0:n } #1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \}
\cs_new:cpn { __prg_replicate_1:n } #1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \#1\#1 \}
\cs_new:cpn { __prg_replicate_2:n } #1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \#1\#1\#1 \}
\cs_new:cpn { __prg_replicate_3:n } #1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \#1\#1\#1\#1 \}
\cs_new:cpn { __prg_replicate_4:n } #1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \#1\#1\#1\#1\#1 \}
\cs_new:cpn { __prg_replicate_5:n } #1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \#1\#1\#1\#1\#1\#1 \}
\cs_new:cpn { __prg_replicate_6:n } #1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \#1\#1\#1\#1\#1\#1\#1 \}
\cs_new:cpn { __prg_replicate_7:n } #1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \#1\#1\#1\#1\#1\#1\#1\1 \}
\cs_new:cpn { __prg_replicate_8:n } #1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \#1\#1\#1\#1\#1\#1\#1\1\#1 \}
\cs_new:cpn { __prg_replicate_9:n } #1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \#1\#1\#1\#1\#1\#1\#1\1\#1\#1 \}
\end{verbatim}

Users shouldn’t ask for something to be replicated once or even not at all but...

\begin{verbatim}
\cs_new:cpn \__prg_replicate_first_-:n \#1 \{ \cs_end: \}
\cs_new:cpn \__prg_replicate_first_0:n \#1 \{ \cs_end: \}
\cs_new:cpn \__prg_replicate_first_1:n \#1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \}
\cs_new:cpn \__prg_replicate_first_2:n \#1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \#1\1\1 \}
\cs_new:cpn \__prg_replicate_first_3:n \#1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \#1\1\1\#1\1 \}
\cs_new:cpn \__prg_replicate_first_4:n \#1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \#1\1\1\#1\1\#1 \}
\cs_new:cpn \__prg_replicate_first_5:n \#1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \#1\1\1\#1\1\#1\#1 \}
\cs_new:cpn \__prg_replicate_first_6:n \#1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \#1\1\1\#1\1\#1\#1\#1 \}
\cs_new:cpn \__prg_replicate_first_7:n \#1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \#1\1\1\#1\1\#1\#1\#1\#1 \}
\cs_new:cpn \__prg_replicate_first_8:n \#1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \#1\1\1\#1\1\#1\#1\#1\1\1 \}
\cs_new:cpn \__prg_replicate_first_9:n \#1 \{ \cs_end: \#1\#1\#1\#1\#1\#1\#1\#1\#1\#1 \#1\1\1\#1\1\#1\#1\#1\1\1\1 \}
\end{verbatim}

Users shouldn’t ask for something to be replicated once or even not at all but...
46.8 Detecting \TeX’s mode

\mode_if_vertical_p: \mode_if_vertical:TF

For testing vertical mode. Strikes me here on the bus with David, that as long as we
are just talking about returning true and false states, we can just use the primitive
conditionals for this and gobbling the \texttt{exp_end}: in the input stream. However this
requires knowledge of the implementation so we keep things nice and clean and use the
return statements.

\begin{verbatim}
\prg_new_conditional:Npnn \mode_if_vertical: { p , T , F , TF } \end{verbatim}
\begin{verbatim}
{ \if_mode_vertical: \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}

(End definition for \mode_if_vertical:TF. This function is documented on page 70.)

\mode_if_horizontal_p: \mode_if_horizontal:TF

For testing horizontal mode.

\begin{verbatim}
\prg_new_conditional:Npnn \mode_if_horizontal: { p , T , F , TF } \end{verbatim}
\begin{verbatim}
{ \if_mode_horizontal: \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}

(End definition for \mode_if_horizontal:TF. This function is documented on page 70.)

\mode_if_inner_p: \mode_if_inner:TF

For testing inner mode.

\begin{verbatim}
\prg_new_conditional:Npnn \mode_if_inner: { p , T , F , TF } \end{verbatim}
\begin{verbatim}
{ \if_mode_inner: \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}

(End definition for \mode_if_inner:TF. This function is documented on page 70.)

\mode_if_math_p: \mode_if_math:TF

For testing math mode. At the beginning of an alignment cell, this should be used only
inside a non-expandable function.

\begin{verbatim}
\prg_new_conditional:Npnn \mode_if_math: { p , T , F , TF } \end{verbatim}
\begin{verbatim}
{ \if_mode_math: \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}

(End definition for \mode_if_math:TF. This function is documented on page 70.)

46.9 Internal programming functions

\texttt{\group_align_safe_begin:} \texttt{\group_align_safe_end:}

\TeX’s alignment structures present many problems. As Knuth says himself in \TeX: The
Program: “It’s sort of a miracle whenever \texttt{\halign} or \texttt{\valign} work, […]” One problem
relates to commands that internally issue a \texttt{\cr} but also peek ahead for the next character
for use in, say, an optional argument. If the next token happens to be a \texttt{&} with category
code 4 we get some sort of weird error message because the underlying \texttt{\futurelet}
stores the token at the end of the alignment template. This could be a \texttt{&} giving a message
like! \texttt{Misplaced \cr.} or even worse: it could be the \texttt{\endtemplate} token causing even
more trouble! To solve this we have to open a special group so that \TeX still thinks
it’s on safe ground but at the same time we don’t want to introduce any brace group
that may find its way to the output. The following functions help with this by using
behaviour documented only in Appendix D of \TeXbook...In short evaluating \texttt{"}{ and
\texttt{`} as numbers will not change the counter \TeX uses to keep track of its state in an
alignment, whereas gobbling a brace using \texttt{\if_false:} will affect \TeX’s state without
producing any real group. We place the \texttt{\if_false:} \{ \texttt{\fi:} part at that place so that the
successive expansions of \texttt{\group_align_safe_begin/end:} are always brace balanced.

\begin{verbatim}
\prg_new:Npn \group_align_safe_begin: \end{verbatim}
\begin{verbatim}
\tex_catcode:D '\^\@ = 2 \exp_stop_f:
\cs_new:Npn \group_align_safe_begin:
\begin{verbatim}
\{ \exp:w \if_false: \{ \fi: '^^@ \exp_stop_f: } \end{verbatim}
\end{verbatim}
\begin{verbatim}
\group_begin:
\end{verbatim}
\begin{verbatim}
\cs_new:Npn \group_align_safe_end:
\end{verbatim}
\begin{verbatim}
\end{verbatim}

574
\group_end:
\cs_new:Npn \group_align_safe_end:
  { \if_int_compare:w '{ = \c_zero_int } \fi: }

(End definition for \group_align_safe_begin: and \group_align_safe_end:. These functions are documented on page 71.)

\__kernel_prg_map_int
\int_new:N \__kernel_prg_map_int

(End definition for \__kernel_prg_map_int.)

\prg_break_point:Nn \prg_map_break:Nn
These are defined in \texttt{l3basics}, as they are needed “early”. This is just a reminder that is the case!

(End definition for \prg_break_point:Nn and \prg_map_break:Nn. These functions are documented on page 70.)

\prg_break_point: \prg_break: \prg_break:n
Also done in \texttt{l3basics}.

(End definition for \prg_break_point:, \prg_break:, and \prg_break:n. These functions are documented on page 71.)

{/package}
Chapter 47

\textbf{l3sys implementation}

47.1 Kernel code

\begin{verbatim}
\__sys_const:nn\ Set the T, F, TF, p forms of \texttt{#1} to be constants equal to the result of evaluating the boolean expression \texttt{#2}.
\cs_new_protected:Npn \__sys_const:nn \#1\#2
\begin{verbatim}
{ \bool_if:nTF {#2} { \cs_new_eq:cN { #1 :T } \use:n \cs_new_eq:cN { #1 :F } \use_none:n \cs_new_eq:cN { #1 :TF } \use_i:nn \cs_new_eq:cN { #1 _p: } \c_true_bool } { \cs_new_eq:cN { #1 :T } \use_none:n \cs_new_eq:cN { #1 :F } \use:n \cs_new_eq:cN { #1 :TF } \use_ii:nn \cs_new_eq:cN { #1 _p: } \c_false_bool }
\end{verbatim}
\end{verbatim}
\end{verbatim} (End definition for \texttt{\__sys_const:nn}.)

\sys_if_engine_luatex_p:\ Set up the engine tests on the basis exactly one test should be true. Mainly a case of looking for the appropriate marker primitive.
\begin{verbatim}
\sys_if_engine_luatex:TF\ \str_const:Nx \c_sys_engine_str
\sys_if_engine_pdftex:TF\ \{ \cs_if_exist:NT \tex_luatexversion:D \{ lualatex \}
\sys_if_engine_ptex:TF\ \{ \cs_if_exist:NT \tex_pdftexversion:D \{ pdftex \}
\sys_if_engine_uptex_p:\ \{ \cs_if_exist:NT \tex_kanjiskip:D
\sys_if_engine_uptex:TF\ \{ \cs_if_exist:NT \tex_kanjiskip:D
\sys_if_engine_xetex_p:\ \{ \cs_if_exist:NT \tex_kanjiskip:D
\sys_if_engine_xetex:TF\ \{ \cs_if_exist:NT \tex_kanjiskip:D
\end{verbatim}

576
The functions defined above, and set up the engine and format names. \texttt{\_c\_sys\_engine\_exec\_str} differs from \texttt{\_c\_sys\_engine\_str} as it is the actual engine name, not a "filtered" version. It differs for \TeX{} and \Uptex{}, which have a leading e, and for \LaTeXe{}, because \LaTeX{} uses the LuaHB\TeX{} engine.

\texttt{\_c\_sys\_engine\_format\_str} is quite similar to \texttt{\_c\_sys\_engine\_str}, except that it differentiates pdflatex from latex (which is pdf\TeX{} in DVI mode). This differentiation, however, is reliable only if the user doesn’t change \texttt{\_tex\_pdfoutput:D} before loading this code.

```latex
\begin{verbatim}
\cs_set_eq:NN \lua_now:e \tex_directlua:D
\str_const:Nx \c_sys_engine_exec_str
{ \sys_if_engine_pdftex:T { pdf } \sys_if_engine_xetex:T { xe } \sys_if_engine_ptex:T { ep } \sys_if_engine_uptex:T { eup } \sys_if_engine_luatex:T
  { lua \lua_now:e
    { if (pcall(require, 'luaharfbuzz')) then ~
tex.print("hb") ~
      end
    }
  }
}\group_end:
\str_const:Nx \c_sys_engine_format_str
{ \cs_if_exist:NTF \fmtname
  { \bool_lazy_or:nnTF
    { \str_if_eq_p:Vn \fmtname { plain } }
    { \str_if_eq_p:Vn \fmtname { LaTeX2e } }
    { \sys_if_engine_pdftex:T \int_compare:nNnT { \tex_pdfoutput:D } = { 1 } { pdf } }
  }
}\sys_if_engine_xetex:T { xe }
\end{verbatim}

(End definition for \texttt{\_sys\_if\_engine\_luatex:TF} and others. These functions are documented on page 73.)

\texttt{\_c\_sys\_engine\_exec\_str} and \texttt{\_c\_sys\_engine\_format\_str}

\newpage
\sys_if_rand_exist:TF

Currently, randomness exists under pdf\TeX, Lua\TeX, \p\TeX and \up\TeX.

\sys const:nn { \sys_if_rand_exist }
\{ \cs_if_exist_p:N \tex_uniformdeviate:D \}

(End definition for \sys_if_rand_exist:TF. This function is documented on page 306.)

47.1.3 Platform

Setting these up requires the file module (file lookup), so is actually implemented there.

(End definition for \sys_if_platform_unix:TF, \sys_if_platform_windows:TF, and \c_sys_platform_str. These functions are documented on page 74.)

47.1.4 Configurations

Loading the backend code is pretty simply: check that the backend is valid, then load it up.

\cs new protected:Npn \sys_load_backend:n #1
\{ \sys_finalise:
\str_if_exist:NTF \c sys_backend_str
\{ \str_if_eq:VnF \c sys_backend_str \#
\{ \msg_error:nn { \sys } { backend-set } \}
\}

578
\tl_if_blank:nF {#1}
\tl_gset:Nn \g__sys_backend_tl {#1}
\__sys_load_backend_check:N \g__sys_backend_tl
\str_const:Nx \c_sys_backend_str { \g__sys_backend_tl }
\__kernel_sys_configuration_load:n
{ l3backend- \c_sys_backend_str }
\cs_new_protected:Npn \__sys_load_backend_check:N #1
{ \sys_if_engine_xetex:TF
  \str_case:VnF #1
  { \dvisvgm } { }
  { \xdvipdfmx } \{ \tl_gset:Nn #1 { xetex } \}
  { \xetex } { }
}
{ \msg_error:nnxx { sys } { wrong-backend }
  \{ xetex \}
  \tl_gset:Nn #1 { xetex }
}
\sys_if_output_pdf:TF
{ \str_if_eq:VnF #1 { pdfmode }
  \{ \sys_if_engine_luatex:TF
    \{ \tl_gset:Nn #1 { luatex } \}
    \{ \tl_gset:Nn #1 { pdftex } \}
  \}
  \bool_lazy_or:nnF
  \{ \str_if_eq_p:Vn #1 { luatex } \}
  \{ \str_if_eq_p:Vn #1 { pdfTeX } \}
  { \msg_error:nnxx { sys } { wrong-backend }
    \{ \sys_if_engine_luatex:TF \{ luatex \} \{ pdftex \} \}
    \{ \sys_if_engine_luatex:TF
      \{ \tl_gset:Nn #1 { luatex } \}
      \{ \tl_gset:Nn #1 { pdftex } \}
    \}
  \}
}
\str_case:VnF #1
{ \dvisvgm } { }
{ \xdvipdfmx } { }
{ \dvips } { }
{ \dvisvgm } { }

8850 \msg_error:nxx { sys } { wrong-backend }
8851 #1 { dvips }
8852 \tl_gset:Nn #1 { dvips }
8853 }
8854 }
8855 }
8856 }

(End definition for \sys_load_backend:n, \_sys_load_backend_check:N, and \c_sys_backend_str. These functions are documented on page 75.)

\g__sys_debug_bool
8857 \bool_new:N \g__sys_debug_bool

(End definition for \g__sys_debug_bool.)

\sys_load_debug:
Simple.

\cs_new_protected:Npn \sys_load_debug:
8858 { \bool_if:NF \g__sys_debug_bool
8859 { \__kernel_sys_configuration_load:n { l3debug } }
8860 \bool_gset_true:N \g__sys_debug_bool
8861 }

(End definition for \sys_load_debug:. This function is documented on page 76.)

47.1.5 Access to the shell

\l__sys_internal_tl
\tl_new:N \l__sys_internal_tl

(End definition for \l__sys_internal_tl.)

\c__sys_marker_tl
The same idea as the marker for rescanning token lists.

\tl_const:Nx \c__sys_marker_tl { : \token_to_str:N : }

(End definition for \c__sys_marker_tl.)

\sys_get_shell:nnN\sys_get_shell:nnN\sys_get_shell:nnN\sys_get_shell:nnN\sys_get_shell:nnN\sys_get_shell:nnN
Setting using a shell is at this level just a slightly specialised file operation, with an additional check for quotes, as these are not supported.

\cs_new_protected:Npn \sys_get_shell:nnN #1#2#3
8866 { \sys_get_shell:nnNF {#1} {#2} #3
8867 \tl_set:Nn #3 { \q_no_value } }
8868 }
8869 \prg_new_protected_conditional:Npnn \sys_get_shell:nnN #1#2#3 { T , F , TF }
8870 { \sys_if_shell:TF
8871 { \exp_args:No \__sys_get:nnN { \tl_to_str:n {#1} } {#2} #3 }
8872 \prg_return_false: }
8873 }
8874 \cs_new_protected:Npn \__sys_get:nnN #1#2#3
8875 {
End definition for \sys_get_shell:nnNTF and others. These functions are documented on page 74.

\c__sys_shell_stream_int

This is not needed for Lua\TeX{}: shell escape there isn’t done using a \TeX{} interface.

\sys_if_engine_luatex:F
\{ \int_const:Nn \c__sys_shell_stream_int { 18 } \}

(End definition for \c__sys_shell_stream_int.)

\sys_shell_now:n

Execute commands through shell escape immediately.

For Lua\TeX{}, we use a pseudo-primitive to do the actual work.

\langle/tex\rangle
\langle\ast lua\rangle
\do{
  local os_exec = os.execute
  local function shellescape(cmd)
    local status, msg = os_exec(cmd)
    if status == nil then
      write_nl("log","runsystem(" .. cmd .. ")...(" .. msg .. ")\n")
    elseif status == 0 then
      write_nl("log","runsystem(" .. cmd .. ")...executed\n")
    else
      write_nl("log","runsystem(" .. cmd .. ")...failed " .. (msg or ") .. "\n")
    end
  end
  luacmd("\_\_sys_shell_now:e", function()
\sys_shell_shipout:n Execute commands through shell escape at shipout.

\sys_shell_shipout:e For LuaTeX, we use the same helper as above but delayed using a late_lua whatsit.

Creating a late_lua whatsit works a bit different if we are running under Con\TeXt.
47.2 Dynamic (every job) code

\sys_everyjob:
\__sys_everyjob:n
\g__sys_everyjob_tl

\cs_new_protected:Npn \sys_everyjob:
\tl_use:N \g__sys_everyjob_tl
\tl_gclear:N \g__sys_everyjob_tl
}
\cs_new_protected:Npn \__sys_everyjob:n #1
{ \tl_gput_right:Nn \g__sys_everyjob_tl {#1} }
\tl_new:N \g__sys_everyjob_tl

\__sys_everyjob:n
{ \cs_new_eq:NN \c_sys_jobname_str \tex_jobname:D }

\__sys_everyjob:n
{ \group_begin:
\cs_set:Npn \__sys_tmp:w #1
{ \str_if_eq:eeTF { \cs_meaning:N #1 } { \token_to_str:N #1 } { #1 }
\cs_if_exist:NTF \tex_primitive:D
{ \bool_lazy_and:nnTF
\{ \sys_if_engine_xetex_p: \}
\int_compare_p:nNn{ \exp_after:wN \use_none:n \tex_XeTeXrevision:D }{583}
{ \boolexpr:nnTF
\{ \sys_if_engine_xetex_p: \}
{ \int_compare_p:nNn{ \exp_after:wN \use_none:n \tex_XeTeXrevision:D }{583}
}
}
}

47.2.1 The name of the job

\c_sys_jobname_str
Inherited from the \LaTeX3 name for the primitive. This has to be the primitive as it’s set in \everyjob. If the user does

\pdflatex \input some-file-name

then \everyjob is inserted before \jobname is changed from texput, and thus we would have the wrong result.

\__sys_everyjob:n
{ \cs_new_eq:NN \c_sys_jobname_str \tex_jobname:D }

47.2.2 Time and date

\c_sys_minute_int
\c_sys_hour_int
\c_sys_day_int
\c_sys_month_int
\c_sys_year_int

Copies of the information provided by \TeX. There is a lot of defensive code in package mode: someone may have moved the primitives, and they can only be recovered if we have \primitiv and it is working correctly. For Ini\TeX of course that is all redundant but does no harm.

\__sys_everyjob:n
{ \group_begin:
\cs_set:Npn \__sys_tmp:w #1
{ \str_if_eq:eeTF { \cs_meaning:N #1 } { \token_to_str:N #1 } { #1 }
\cs_if_exist:NTF \tex_primitive:D
{ \bool_lazy_and:nnTF
\{ \sys_if_engine_xetex_p: \}
\int_compare_p:nNn{ \exp_after:wN \use_none:n \tex_XeTeXrevision:D }{583}
{ \boolexpr:nnTF
\{ \sys_if_engine_xetex_p: \}
{ \int_compare_p:nNn{ \exp_after:wN \use_none:n \tex_XeTeXrevision:D }{583}
}
}
}
47.2.3 Random numbers

\sys_rand_seed:  Unpack the primitive. When random numbers are not available, we return zero after an error (and incidentally make sure the number of expansions needed is the same as with random numbers available).

\sys_gset_rand_seed:n  The primitive always assigns the seed globally.

(End definition for \c_sys_minute_int and others. These variables are documented on page 72.)

(End definition for \sys_rand_seed:. This function is documented on page 74.)

(End definition for \sys_gset_rand_seed:. This function is documented on page 74.)
In LuATeX, create a pseudo-primitive, otherwise try to locate the real primitive. The elapsed time will be available if this succeeds.

```plaintext
definition for \sys_gset_rand_seed:n. This function is documented on page 74.)

\sys_timer:
\_\_sys_elapsedtime:
\sys_if_timer_exist_p:
\sys_if_timer_exist:TF

\cs_new:Npn \sys_timer:n{ \_\_sys_elapsedtime: }
\cs_new:Npn \sys_timer:n{ \int_value:w \tex_elapsedtime:D }
\msg_new:nnnn { kernel } { no-elapsed-time }
{ No-clock-detected-for-#1. } { The-current-engine-provides-no-way-to-access-the-system-time. }
\cs_new:Npn \sys_timer:n{ \int_value:w \msg_expandable_error:nnn { kernel } { no-elapsed-time }{ \sys_timer: }{ \c_zero_int }
\__sys_const:nn { sys_if_timer_exist } { \cs_if_exist_p:N \_\_sys_elapsedtime: \cs_if_exist_p:N \_\_sys_elapsedtime: }

(End definition for \sys_gset_rand_seed:n. This function is documented on page 74.)

47.2.4 Access to the shell

\c_sys_shell_escape_int Expose the engine’s shell escape status to the user.
\_=\_sys_everyjob:n
```
\sys_if_shell:TF\sys_if_shell_unrestricted:TF\sys_if_shell_restricted:TF\g_file_curr_name_str

47.2.5 Held over from l3file
See comments about \c_sys_jobname_str: here, as soon as there is file input/output, things get “tided up”.

\sys_finalise:
\__sys_finalise:n\g__sys_finalise_tl

47.3 Last-minute code
A simple hook to finalise the system-dependent layer. This is forced by the backend loader, which is forced by the main loader, so we do not need to include that here.
47.3.1 Detecting the output

This is a simple enough concept: the two views here are complementary.

47.3.2 Configurations

As the backend has to be checked and possibly adjusted, the approach here is to create a variable and use that in a one-shot to set a constant.
\str_case:nnT {#1}
{
  { dvipdfmx }
    { \tl_gset:Nn \g__sys_backend_tl { dvipdfmx } }
  { dvips }
    { \tl_gset:Nn \g__sys_backend_tl { dvips } }
  { dvisvgm }
    { \tl_gset:Nn \g__sys_backend_tl { dvisvgm } }
  { pdftex }
    { \tl_gset:Nn \g__sys_backend_tl { pdfmode } }
  { xetex }
    { \tl_gset:Nn \g__sys_backend_tl { xdvipdfmx } }
  { \clist_remove_all:Nn \@unusedoptionlist {#1} }
}

(End definition for \g__sys_backend_tl.)

{/tex}

{/package}
Chapter 48

l3msg implementation

\l_msg_internal_tl A general scratch for the module.
\tl_new:N \l__msg_internal_tl
(End definition for \l__msg_internal_tl.)
\l__msg_name_str
\l__msg_text_str
\str_new:N \l__msg_name_str
\str_new:N \l__msg_text_str
(End definition for \l__msg_name_str and \l__msg_text_str.)

48.1 Internal auxiliaries
\s_msg_mark Internal scan marks.
\s_msg_stop \scan_new:N \s__msg_mark
\scan_new:N \s__msg_stop
(End definition for \s__msg_mark and \s__msg_stop.)
\msg_use_none_delimit_by_s_stop:w Functions to gobble up to a scan mark.
\cs_new:Npn \__msg_use_none_delimit_by_s_stop:w #1 \s__msg_stop { }
(End definition for \__msg_use_none_delimit_by_s_stop:w.)

48.2 Creating messages
Messages are created and used separately, so there two parts to the code here. First, a mechanism for creating message text. This is pretty simple, as there is not actually a lot to do.
\c_msg_text_prefix_tl \c__msg_more_text_prefix_tl
\tl_const:Nn \c__msg_text_prefix_tl { msg~text~>~ }
\tl_const:Nn \c__msg_more_text_prefix_tl { msg~extra~text~>~ }

589
Test whether the control sequence containing the message text exists or not.

\begin{verbatim}
\prg_new_conditional:Npnn \msg_if_exist:nn #1#2 { p , T , F , TF } 
{ \cs_if_exist:cTF { \c__msg_text_prefix_tl #1 / #2 } 
{ \prg_return_true: } { \prg_return_false: } }
\end{verbatim}

This auxiliary is similar to \_\_kernel_chk_if_free_cs:N, and is used when defining messages with \msg_new:nnnn.

\begin{verbatim}
\cs_new_protected:Npn \__msg_chk_free:nn #1#2
{ \msg_if_exist:nnT {#1} {#2} 
{ \msg_error:nnxx { msg } { already-defined } 
{#1} {#2} }
}
\end{verbatim}

Setting a message simply means saving the appropriate text into two functions. A sanity check first.

\begin{verbatim}
\cs_new_protected:Npn \msg_new:nnnn #1#2
{ \__msg_chk_free:nn {#1} {#2} 
\msg_gset:nnnn {#1} {#2} }
\cs_new_protected:Npn \msg_new:nnn #1#2#3
{ \msg_new:nnnn {#1} {#2} {#3} { } }
\end{verbatim}

(End definition for \_\_msg_chk_if_free:nn.)

(End definition for \c__msg_text_prefix_tl and \c__msg_more_text_prefix_tl.)

(End definition for \msg_if_exist:nnTF. This function is documented on page 78.)

(End definition for \msg_new:nnnn and others. These functions are documented on page 78.)
48.3 Messages: support functions and text

Simple pieces of text for messages.

\c__msg_coding_error_text_tl
\c__msg_continue_text_tl
\c__msg_critical_text_tl
\c__msg_fatal_text_tl
\c__msg_no_info_text_tl
\c__msg_on_line_text_tl
\c__msg_return_text_tl
\c__msg_trouble_text_tl

\tl_const:Nn \c__msg_coding_error_text_tl
\{ This-is-a-coding-error. \}
\\\n
\tl_const:Nn \c__msg_continue_text_tl
\{ Type\langle\text{return}\rangle\text{-to-continue} \}

\tl_const:Nn \c__msg_critical_text_tl
\{ Reading-the-current-file-'\g_file_curr_name_str'-will-stop. \}
\tl_const:Nn \c__msg_fatal_text_tl
\{ This-is-a-fatal-error:-LaTeX-will-abort. \}
\tl_const:Nn \c__msg_no_info_text_tl
\{ \LaTeX\text{-does-not-know-anything-more-about-this-error,-sorry.} \}
\tl_const:Nn \c__msg_return_text_tl
\{ on-line \}
\tl_const:Nn \c__msg_return_text_tl
\{ \text{Try-typing-\langle\text{return}\rangle\text{-to-proceed.}} \}
\tl_const:Nn \c__msg_trouble_text_tl
\{ \text{If-that-doesn't-work,-type-X-\langle\text{return}\rangle\text{-to-quit.}} \}
\tl_const:Nn \c__msg_trouble_text_tl
\{ \text{More-errors-will-almost-certainly-follow:} \}
\text{the-LaTeX-run-should-be-aborted.} \}

(End definition for \c__msg_coding_error_text_tl and others.)

\msg_line_number: For writing the line number nicely. \msg_line_context: was set up earlier, so this is not new.
\cs_new:Npn \msg_line_number: \{ \int_use:N \tex_inputlineno:D \}
\cs_gset:Npn \msg_line_context:  \{}
\c__msg_on_line_text_tl
\c_space_tl
\msg_line_number:
\}

(End definition for \msg_line_number: and \msg_line_context: These functions are documented on page 79.)
48.4 Showing messages: low level mechanism

The low-level interruption macro is rather opaque, unfortunately. Depending on the availability of more information there is a choice of how to set up the further help. We feed the extra help text and the message itself to a wrapping auxiliary, in this order because we must first setup \TeX’s \texttt{\textbackslash errhelp} register before issuing an \texttt{\textbackslash errmessage}. To deal with the various cases of critical or fatal errors with and without help text, there is a bit of argument-passing to do.

\begin{verbatim}
\cs_new_protected:Npn \__msg_interrupt:Nnnn #1#2#3#4#5
\{\str_set:Nx \l__msg_text_str { #1 {#2} } \str_set:Nx \l__msg_name_str { \msg_module_name:n {#2} } \cs_if_eq:cNTF \{ { \c__msg_more_text_prefix_tl #2 / #3 } \__msg_no_more_text:nnnn \} \__msg_no_more_text:nnnn \{ \use:c { \c__msg_text_prefix_tl #2 / #3 } \__msg_interrupt_wrap:nnn \{ \use:c { \c__msg_continue_text_tl } \} \\\c__msg_no_info_text_tl \tl_if_empty:NF \l_iow_line_count_int \{ \\\\#5 \} \} \__msg_interrupt_wrap:nnn \{ \use:c { \c__msg_text_prefix_tl #2 / #3 } \c__msg_help_text_tl \} \\\use:c { \c__msg_more_text_prefix_tl #2 / #3 } \tl_if_empty:NF \l_iow_line_count_int \{ \\\\#5 \} \}
\cs_new:Npn \__msg_no_more_text:nnnn #1#2#3#4 { }
\end{verbatim}

\begin{verbatim}
\__msg_interrupt_wrap:nnn \__msg_interrupt_text:n \__msg_interrupt_more_text:n
\end{verbatim}

First setup \TeX’s \texttt{\textbackslash errhelp} register with the extra help \#1, then build a nice-looking error message with \#2. Everything is done using \texttt{x}-type expansion as the new line markers are different for the two type of text and need to be correctly set up. The auxiliary \texttt{\__msg_interrupt_more_text:n} receives its argument as a line-wrapped string, which is thus unaffected by expansion. We ave to split the main text into two parts as only the “message” itself is wrapped with a leader: the generic help is wrapped at full width. We also have to allow for the two characters used by \texttt{\textbackslash errmessage} itself.

\begin{verbatim}
\cs_new_protected:Npn \__msg_interrupt_wrap:nnn \__msg_interrupt_inc_str:n \__msg_interrupt_more_text:n
\end{verbatim}
The business end of the process starts by producing some visual separation of the message from the main part of the log. The error message needs to be printed with everything made “invisible”: \TeX’s own information involves the macro in which \errmessage is called, and the end of the argument of the \errmessage, including the closing brace. We use an active \! to call the \errmessage primitive, and end its argument with \use-\none:n \{ ⟨spaces⟩ \} which fills the output with spaces. Two trailing closing braces are turned into spaces to hide them as well. The group in which we alter the definition of the active \! is closed before producing the message: this ensures that tokens inserted by typing I in the command-line are inserted after the message is entirely cleaned up.

The \_\_kernel_iow_with:Nnn auxiliary, defined in l3file, expects an \langle integer variable⟩, an integer \langle value⟩, and some \langle code⟩. It runs the \langle code⟩ after ensuring that the \langle integer variable⟩ takes the given \langle value⟩, then restores the former value of the \langle integer variable⟩ if needed. We use it to ensure that the \newlinechar is 10, as needed for \iow_newline: to work, and that \errorcontextlines is −1, to avoid showing irrelevant context. Note that restoring the former value of these integers requires inserting tokens after the \errmessage, which go in the way of tokens which could be inserted by the user. This is unavoidable.
48.5 Displaying messages

\LaTeX{} is handling error messages and so the \TeX{} ones are disabled.

A function for issuing messages: both the text and order could in principle vary. The module name may be empty for kernel messages, hence the slightly contorted code path for space.

(End definition for \texttt{\_msg\_interrupt:n}.)

\begin{verbatim}
  \int_gset:Nn \tex_errorcontextlines:D { -1 }
  \msg_fatal_text:n \msg_critical_text:n \msg_error_text:n \msg_warning_text:n \msg_info_text:n \__msg_text:nn \__msg_text:n

  \cs_new:Npn \msg_fatal_text:n #1
  { Fatal ~ \msg_error_text:n {#1} }
  \cs_new:Npn \msg_critical_text:n #1
  { Critical ~ \msg_error_text:n {#1} }
  \cs_new:Npn \msg_error_text:n #1
  { \__msg_text:nn {#1} { Error } }
  \cs_new:Npn \msg_warning_text:n #1
  { \__msg_text:nn {#1} { Warning } }
  \cs_new:Npn \msg_info_text:n #1
  { \__msg_text:nn {#1} { Info } }
  \cs_new:Npn \__msg_text:nn #1#2
  { \exp_args:Nf \__msg_text:n { \msg_module_type:n {#1} } \exp_args:Nf \__msg_text:n { \msg_module_name:n {#1} } }
\end{verbatim}

594
\__msg_class_new:nn
\group_begin:
\cs_set_protected:Npn \__msg_text:n #1
\{
\tl_if_blank:nF {#1}
\{ #1 ~ 
\}
\}

(End definition for \msg_fatal_text:n and others. These functions are documented on page 79.)
\g_msg_module_name_prop
\g_msg_module_type_prop
For storing public module information: the kernel data is set up in advance.
\prop_new:N \g_msg_module_name_prop
\prop_gput:Nnn \g_msg_module_name_prop { LaTeX } { LaTeX3 }
\prop_new:N \g_msg_module_type_prop
\prop_gput:Nnn \g_msg_module_type_prop { LaTeX } { }

(End definition for \g_msg_module_name_prop and \g_msg_module_type_prop. These variables are documented on page 78.)
\msg_module_type:n
Contextual footer information, with the potential to give modules an alternative name.
\cs_new:Npn \msg_module_type:n #1
\{
\prop_if_in:NnTF \g_msg_module_type_prop {#1}
\{ \prop_item:Nn \g_msg_module_type_prop {#1} 
\}
\}
\cs_new:Npn \msg_module_name:n #1
\{
\prop_if_in:NnTF \g_msg_module_name_prop {#1}
\{ \prop_item:Nn \g_msg_module_name_prop {#1} 
\}
\}
\cs_new:Npn \msg_see_documentation_text:n #1
\{ See~the~ \msg_module_name:n {#1} ~ documentation~for~further~information. 
\}

(End definition for \msg_module_type:n. This function is documented on page 78.)
\msg_module_name:n
\msg_see_documentation_text:n
Contextual footer information, with the potential to give modules an alternative name.
\cs_new:Npn \msg_module_name:n #1
\{
\prop_if_in:NnTF \g_msg_module_name_prop {#1}
\{ \prop_item:Nn \g_msg_module_name_prop {#1} 
\}
\}
\cs_new:Npn \msg_see_documentation_text:n #1
\{ See-the- \msg_module_name:n {#1} - documentation-for-further-information. 
\}

(End definition for \msg_module_name:n and \msg_see_documentation_text:n. These functions are documented on page 78.)
For fatal errors, after the error message \TeX{} bails out. We force a bail out rather than using `\end` as this means it does not matter if we are in a context where normally the run cannot end.

\verb|\__msg_fatal_exit:|
For an error, the interrupt routine is called. We check if there is a “more text” by comparing that control sequence with a permanently empty text. We have to undefine the bootstrap versions here.

\cs_undefine:N \msg_error:nnxx
\cs_undefine:N \msg_error:nxx
\cs_undefine:N \msg_error:nnn
\cs_undefine:N \msg_error:nn

\__msg_class_new:nn { error }
\__msg_interrupt:NnnnN \msg_error_text:n {#1} {#2}
\empty_tl

(End definition for $\textbackslash \text{msg\_critical}:nnnnnn$ and others. These functions are documented on page 81.)

Warnings and information messages have no decoration. Warnings are printed to the terminal while information can either go to the log or both log and terminal.

\cs_new_protected:Npn \__msg_info_aux:NNnnnnnn #1#2#3#4#5#6#7#8
\str_set:Nx \l__msg_text_str { #2 {#3} }
\str_set:Nx \l__msg_name_str { \msg_module_name:n {#3} }
#1 { }
\iow_wrap:nxnN
\{ \l__msg_text_str : ~
\use:c { \c__msg_text_prefix_tl #3 / #4 } {#5} {#6} {#7} {#8} }
\prg_replicate:nn
{ \l__msg_name_str }
( \l__msg_text_str )
\prg_replicate:nn
{ \l__msg_text_prefix_tl #3 / #4 }
\prg_replicate:nn
\{ #1 }
#1 { }
\__msg_class_new:nn { warning }
\__msg_info_aux:NNnnnnnn \iow_term:n \msg_warning_text:n {#1} {#2} {#3} {#4} {#5} {#6}

(End definition for $\textbackslash \text{msg\_error}:nnnnnn$ and others. These functions are documented on page 81.)
\msg_log:nnnnnnn
\msg_log:nnxxxx
\msg_log:nnnnnn
\msg_log:nnxxx
\msg_log:nnnn
\msg_log:nnxx
\msg_log:nnn
\msg_log:nnx
\msg_log:nn

"Log" data is very similar to information, but with no extras added. "Term" is used for communicating with the user through the terminal, like diagnostic messages, and debugging. This is similar to "log" messages, but uses the terminal output.

\msg_term:nnnnnn
\msg_term:nnxxxx
\msg_term:nnnnnn
\msg_term:nnxxx
\msg_term:nnnn
\msg_term:nnxx
\msg_term:nnn
\msg_term:nnx
\msg_term:nn

(End definition for \msg_log:nnnnnnn and others. These functions are documented on page 81.)

The \texttt{none} message type is needed so that input can be gobbled.

\msg_none:nnnnnn
\msg_none:nnxxxx
\msg_none:nnnnnn
\msg_none:nnxxx
\msg_none:nnnn
\msg_none:nnxx
\msg_none:nnn
\msg_none:nnx
\msg_none:nn

(End definition for \msg_none:nnnnnn and others. These functions are documented on page 82.)

The \texttt{show} message type is used for \texttt{useq_show:N} and similar complicated data structures. Wrap the given text with a trailing dot (important later) then pass it to \__msg_show:n. If there is \\>~ (or if the whole thing starts with >) we split there, print the first part and show the second part using \\showtokens (the \exp_after:wN ensure a nice display).

Note that this primitive adds a leading \texttt{>} and trailing dot. That is why we included a trailing dot before wrapping and removed it afterwards. If there is no \texttt{\>~} do the same but with an empty second part which adds a spurious but inevitable \texttt{>}.

\msg_show:nnnnnn
\msg_show:nnxxxx
\msg_show:nnnnnn
\msg_show:nnxxx
\msg_show:nnnn
\msg_show:nnxx
\msg_show:nnn
\msg_show:nnx
\msg_show:nn

598
(End definition for \msg_show:nnnnnn and others. These functions are documented on page 83.)

End the group to eliminate \_msg_class_new:nn.

\__msg_class_chk_exist:nT Checking that a message class exists. We build this from \cs_if_free:cTF rather than \cs_if_exist:cTF because that avoids reading the second argument earlier than necessary.

(End definition for \__msg_class_chk_exist:nT.)

\_msg_class_tl Support variables needed for the redirection system.

(End definition for \_msg_class_tl and \_msg_current_class_tl.)

\_msg_redirect_prop For redirection of individually-named messages

(End definition for \_msg_redirect_prop.)

\_msg_hierarchy_seq During redirection, split the message name into a sequence: {/module/submodule}, {/module}, and {}.

599
Actually using a message is a multi-step process. First, some safety checks on the message and class requested. The code and arguments are then stored to avoid passing them around. The assignment to \texttt{\_msg\_use\_code:} is similar to \texttt{\_tl\_set:Nn}. The message is eventually produced with whatever \texttt{\_msg\_class\_tl} is when \texttt{\_msg\_use\_code:} is called. Here is also a good place to suppress tracing output if the \texttt{trace} package is loaded since all (non-expandable) messages go through this auxiliary.

The first check is for an individual message redirection. If this applies then no further redirection is attempted. Otherwise, split the message name into \langle module \rangle, \langle submodule \rangle and \langle message \rangle (with an arbitrary number of slashes), and store \\
\langle /module/submodule \rangle, \langle /module \rangle and \langle \rangle into \texttt{\_msg\_hierarchy\_seq}. We then map through this sequence, applying the most specific redirection.
At this point, the items of \l__msg_hierarchy_seq are the various levels at which we should look for a redirection. Redirections which are less specific than the argument of \__msg_use_redirect_module:n are not attempted. This argument is empty for a class redirection, /module for a module redirection, etc. Loop through the sequence to find the most specific redirection, with module ##1. The loop is interrupted after testing for a redirection for ##1 equal to the argument #1 (least specific redirection allowed). When a redirection is found, break the mapping, then if the redirection targets the same class, output the code with that class, and otherwise set the target as the new current class, and search for further redirections. Those redirections should be at least as specific as ##1.

\msg_redirect_name:nnn

Named message always use the given class even if that class is redirected further. An empty target class cancels any existing redirection for that message.
If the target class is empty, eliminate the corresponding redirection. Otherwise, add the redirection. We must then check for a loop: as an initialization, we start by storing the initial class in \l__msg_current_class_tl.

Since multiple redirections can only happen with increasing specificity, a loop requires that all steps are of the same specificity. The new redirection can thus only create a loop with other redirections for the exact same module, \#1, and not submodules. After some initialization above, follow redirections with \l__msg_class_tl, and keep track in \l__msg_class_loop_seq of the various classes encountered. A redirection from a class to itself, or the absence of redirection both mean that there is no loop. A redirection to the initial class marks a loop. To break it, we must decide which redirection to cancel. The user most likely wants the newly added redirection to hold with no further redirection. We thus remove the redirection starting from \#2, target of the new redirection. Note that no message is emitted by any of the underlying functions: otherwise we may get an infinite loop because of a message from the message system itself.
48.6 Kernel-specific functions

These are all retained purely for older \texttt{xparse} support.

\begin{verbatim}
\_kernel_msg_new:nnnn \_kernel_msg_new:nnn
\_kernel_msg_info:nnxx \_kernel_msg_warning:nnxx \_kernel_msg_error:nnxx
\_kernel_msg_error:nnxxx
\_kernel_msg_expandable_error:nnn \_kernel_msg_expandable_error:nnf
\_kernel_msg_expandable_error:nnff
\end{verbatim}

(End definition for \texttt{\_kernel_msg_new:nnn} and \texttt{\_kernel_msg_new:nnn}.)

(End definition for \texttt{\_kernel_msg_info:nnxx} and others.)

(End definition for \texttt{\_kernel_msg_expandable_error:nnn} and \texttt{\_kernel_msg_expandable_error:nnff}.)
48.7 Internal messages

Error messages needed to actually implement the message system itself.

\msg_new:nnnn { msg } { already-defined }
\{ Message-'#2'-for-module-'#1'-already-defined. }
\{ \c__msg_coding_error_text_tl
LaTeX-was-asked-to-define-a-new-message-called-'#2'\\\
by-the-module-'#1':-this-message-already-exists.
\c__msg_return_text_tl
\}
\msg_new:nnnn { msg } { unknown }
\{ Unknown-message-'#2'-for-module-'#1'. }
\{ \c__msg_coding_error_text_tl
LaTeX-was-asked-to-display-a-message-called-'#2'\\\
by-the-module-'#1':-this-message-does-not-exist.
\c__msg_return_text_tl
\}
\msg_new:nnnn { msg } { class-unknown }
\{ Unknown-message-class-'#1'. }
\{ \LaTeX-has-been-asked-to-redirect-messages-to-a-class-'#1':\\\
this-was-never-defined.
\c__msg_return_text_tl
\}
\msg_new:nnnn { msg } { redirect-loop }
\{ Message-redirection-loop-caused-by- {#1} \=\=\{#2}
\tl_if_empty:nF {#3} { -for-module- \use_none:n #3 } .
\}
\{ \LaTeX-has-been-asked-to-redirect-messages-to-a-class-'#1':\\\
created-an-infinite-loop\\\\
\iow_indent:n { #4 } \\\
\}

Messages for earlier kernel modules plus a few for l3keys which cover coding errors.

\msg_new:nnnn { kernel } { bad-number-of-arguments }
\{ Function-'#1'-cannot-be-defined-with-#2-arguments. }
\{ \c__msg_coding_error_text_tl
LaTeX-has-been-asked-to-define-a-function-'#1'-with-
#2-arguments.-
\c__msg_return_text_tl
\}
\msg_new:nnnn { kernel } { command-already-defined }
\{ Control-sequence-'#1'-already-defined. }
\{ \c__msg_coding_error_text_tl
LaTeX-has-been-asked-to-create-a-new-control-sequence-'#1'-
but-this-name-has-already-been-used-elsewhere. \\\
\}

The-current-meaning-is:\\
\\ #2

\msg_new:nnnn { kernel } { command-not-defined }
{ Control-sequence '#1' undefined. }
\c__msg_coding_error_text_tl
LaTeX has been asked to use a control sequence '-'#1'':\ 
this has not been defined yet.

\msg_new:nnnn { kernel } { empty-search-pattern }
{ Empty-search-pattern. }
\c__msg_coding_error_text_tl
LaTeX has been asked to replace an empty pattern by '-'#1':-that-
would-lead-to-an-infinite-loop!

\msg_new:nnnn { kernel } { non-base-function }
{ Function '-'#1' is not a base function }
\c__msg_coding_error_text_tl
Functions defined through \iow_char:N\cs_new:Nn must have-
a signature consisting of only normal arguments '-'N' and '-'n'.-
The signature '-'#2' of '-'#1' contains other arguments '-'#3'.-
To define variants use \iow_char:N\cs_generate_variant:Nn
and to define other functions use \iow_char:N\cs_new:Npn.

\msg_new:nnnn { kernel } { missing-colon }
{ Function '-'#1' contains no '-'.'. }
\c__msg_coding_error_text_tl
Code-level functions must contain '-'':-' to separate the-
argument specification from the function name. This is-
needed when defining conditionals or variants, or when building a-
parameter text from the number of arguments of the function.

\msg_new:nnnn { kernel } { overflow }
{ Integers larger than 2^{30} - 1 cannot be stored in arrays. }
{ An attempt was made to store '-'#3-
\tl_if_empty:nF {#2} { at-position '-'#2' } in-the-array '-'#1'.-
The largest allowed value '-'#4' will be used instead.

\msg_new:nnnn { kernel } { out-of-bounds }
{ Access to an entry beyond an array's bounds. }
{ An attempt was made to access or store data at-position '-'#2' of-the-
array '-'#1', but this array has entries at positions from 1 to '-'#3'.

\msg_new:nnnn { kernel } { protected-predicate }
{ Predicate '-'#1' must be expandable. }
\c__msg_coding_error_text_tl
LaTeX has been asked to define '-'#1' as a protected predicate.-
Only expandable tests can have a predicate version.
\msg_new:nnnn { keys } { no-property } { No-property-given-in-definition-of-key-"#1". }

{ property-boolean-values-only }
\msg_new:nnnn { keys } { property-boolean-values-only } { The-property-"#1"-accepts-boolean-values-only. }

{ property-requires-value }
\msg_new:nnnn { keys } { property-requires-value } { The-property-"#1"-requires-a-value. }

{ property-unknown }
\msg_new:nnnn { keys } { property-unknown } { The-key-property-"#1"-is-unknown. }

\msg_new:nnnn { quark } { invalid-function } { Quark-test-function-"#1"-is-invalid. }

{ invalid-quark-variable-"#1". }
\msg_new:nnnn { scanmark } { already-defined } { Scan-mark-"#1"-already-defined. }

{ shuffle-too-large }
\msg_new:nnnn { seq } { shuffle-too-large } { The-sequence-"#1"-is-too-long-to-be-shuffled-by-TeX. }

TeX-has- \int_eval:n { \c_max_register_int + 1 } -
Some errors are only needed in package mode if debugging is enabled by one of the options `enable-debug`, `check-declarations`, `log-functions`, or on the contrary if debugging is turned off. In format mode the error is somewhat different.

Some errors only appear in expandable settings, hence don’t need a “more-text” argument.
\{ Erroneous-variable-#1 used! \}
\msg_new:nnn { seq } { misused }
{ A-sequence-was-misused. }
\msg_new:nnn { prop } { misused }
{ A-property-list-was-misused. }
\msg_new:nnn { prg } { negative-replication }
{ Negative-argument-for-\texttt{iow_char:N\prg_replicate:nn}. }
\msg_new:nnn { prop } { prop-keyval }
{ Missing-’=’-in-’#1’-(in-’..._keyval:Nn’) }
\msg_new:nnn { kernel } { unknown-comparison }
{ Relation-’#1’-not-among=,<,>,==,!=,<,>. }
\msg_new:nnn { kernel } { zero-step }
{ Zero-step-size-for-function-#1. }
\cs_if_exist:NF \tex_expanded:D
\{ E-type \}
\msg_new:nnn { kernel } { e-type }
{ #1 - in-e-type-argument }
\msg_new:nnn { clist } { show }
{ The-comma-list- \texttt{\tl_if_empty:nF \{#1\} \{ #1 - \}}
\tl_if_empty:nF \{#2\}
{ is-empty \<\> .. }
{ contains-the-items-(without-outer-braces): #2 . }
\msg_new:nnn { intarray } { show }
{ The-integer-array-#1-contains-#2-items: \#3 . }
\msg_new:nnn { prop } { show }
{ The-property-list-#1-
\tl_if_empty:nTF \{#2\}
{ is-empty \<\> .. }
{ contains-the-pairs-(without-outer-braces): #2 . }
\msg_new:nnn { seq } { show }
{ The-sequence-#1-
\tl_if_empty:nTF \{#2\}
{ is-empty \<\> .. }
{ contains-the-items-(without-outer-braces): #2 . }
\msg_new:nnn { kernel } { show-streams }
{ No- } { The-following- }
\str_case:nn \{#1\}
{ ior } { input - }
{ iow } { output - }
streams-are-
\tl_if_empty:nTF \{#2\} { open } { in-use: #2 . }
}\endinput
System layer messages

10017 \msg_new:nnnn \{ sys \} \{ backend-set \}
10018 \{ Backend-configuration-already-set. \}
10019 \}
10020 \msg_new:nnnn \{ sys \} \{ backend-set \}
10021 \{ Run-time-backend-selection-may-only-be-carried-out-once-during-a-run.-
10022 \This-second-attempt-to-set-them-will-be-ignored. \}
10023 \}
10024 \msg_new:nnnn \{ sys \} \{ wrong-backend \}
10025 \{ Backend-request-inconsistent-with-engine:-using-'#2'-backend. \}
10026 \{ You-have-requested-backend-'#1',-but-this-is-not-suitable-for-use-with-the-
10027 \active-engine.-LaTeX3-will-use-the-'#2'-backend-instead. \}

48.8 Expandable errors

In expansion only context, we cannot use the normal means of reporting errors. Instead, we rely on a low-level \TeX error caused by expanding a macro \texttt{??} with parameter text “?” (this could be any token) which we used followed by something else (here, a space). This shows the context, which thanks to the odd-looking \use:n is

<argument> \???

! mypkg Error: The error message.

In other words, \TeX is processing the argument of \use:n, which is \texttt{??} \texttt{space} \texttt{error type} \texttt{error message}.

The command built from the csname \texttt{c__msg_text_prefix_tl #1 / #2} takes four arguments and builds the error text, which is fed to \texttt{__msg_expandable_error:n} with appropriate expansion: just as for usual messages the arguments are first turned to strings, then the message is fully expanded. The module name also has to be determined.
48.9 Message formatting
Chapter 49

\texttt{l3file implementation}

The following test files are used for this code: \texttt{m3file001}.

\section{Input operations}

\subsection{Variables and constants}

\__ior_internal_tl

\begin{verbatim}
\tl_new:N \l__ior_internal_tl
\end{verbatim}

Used as a short-term scratch variable.

\c__ior_term_ior

\begin{verbatim}
\int_const:Nn \c__ior_term_ior { 16 }
\end{verbatim}

Reading from the terminal (with a prompt) is done using a positive but non-existent stream number. Unlike writing, there is no concept of reading from the log.

\g__ior_streams_seq

\begin{verbatim}
\seq_new:N \g__ior_streams_seq
\end{verbatim}

A list of the currently-available input streams to be used as a stack.

\l__ior_stream_tl

\begin{verbatim}
\tl_new:N \l__ior_stream_tl
\end{verbatim}

Used to recover the raw stream number from the stack.

\g__ior_streams_prop

\begin{verbatim}
\prop_new:N \g__ior_streams_prop
\end{verbatim}

The name of the file attached to each stream is tracked in a property list. To get the correct number of reserved streams in package mode the underlying mechanism needs to be queried. For \LaTeX{} and plain \TeX{} this data is stored in \texttt{\count16}: with the \texttt{etex} package loaded we need to subtract 1 as the register holds the number of the next stream to use. In Con\TeXt{}, we need to look at \texttt{\count38} but there is no subtraction: like the original plain \TeX{}/\LaTeX{} mechanism it holds the value of the last stream allocated.

\prop_new:N \g__ior_streams_prop
49.1.2 Stream management

Reserving a new stream is done by defining the name as equal to using the terminal.

\ior_new:N
\ior_new:c

Reserving a new stream is done by defining the name as equal to using the terminal.

\ior_open:Nn
\ior_open:cn

Use the conditional version, with an error if the file is not found.

\l__ior_file_name_tl

Data storage.

\ior_open:NnTF
\ior_open:cnTF

An auxiliary searches for the file in the \TeX, \LaTeXe and \LaTeX3 paths. Then pass the file found to the lower-level function which deals with streams. The full_name is empty when the file is not found.
Streams are reserved using `\newread` before they can be managed by `ior`. To prevent `ior` from being affected by redefinitions of `\newread` (such as done by the third-party package `morewrites`), this macro is saved here under a private name. The complicated code ensures that `\__ior_new:N` is not `outer` despite plain TeX’s `\newread` being `\outer`. For ConTeXt, we have to deal with the fact that `\newread` works like our own: it actually checks before altering definition.

\begin{verbatim}
\cs_new_protected:Npn \__ior_new:N #1
{} \ior_close:N #1 \seq_gpop:NNTF \g__ior_streams_seq \l__ior_stream_tl
{ \__ior_open_stream:Nn #1 {#2} }
{ \cs_undefine:N #1 \__ior_new_aux:N #1 }
\cs_set_protected:Npn \__ior_new:N #1
{} \cs_new_eq:NN \__ior_new_aux:N \__ior_new:N
\cs_set_protected:Npn \__ior_new:N #1
{ \cs_undefine:N #1 \__ior_new_aux:N #1 }
\end{verbatim}

The stream allocation itself uses the fact that there is a list of all of those available. Life gets more complex as it’s important to keep things in sync. That is done using a two-part approach: any streams that have already been taken up by `ior` but are now free are tracked, so we first try those. If that fails, ask plain TeX or \LaTeX\ for a new stream and use that number (after a bit of conversion).

\begin{verbatim}
\cs_new_protected:Npx \__ior_open_stream:Nn #1#2
{ \tex_global:D \tex_chardef:D #1 = \exp_not:N \l__ior_stream_tl \scan_stop:
\prop_gput:NVn \exp_not:N \g__ior_streams_prop #1 {#2}
\tex_openin:D #1 \sys_if_engine_luatex:TF
{ \exp_not:N \__kernel_file_name_quote:n {#2} \scan_stop: }
{ \exp_not:N \__kernel_file_name_quote:n {#2} \scan_stop: }
\end{verbatim}

Here, we act defensively in case Lua\TeX\ is in use with an extensionless file name.

\begin{verbatim}
\cs_new_protected:Npx \__ior_open_stream:Nn \__kernel_iorsream:Nn { No }
\end{verbatim}

(End definition for `\ior_open:NNTF`. This function is documented on page \pageref{ior_open:NNTF}.)
Closing a stream means getting rid of it at the TeX level and removing from the various data structures. Unless the name passed is an invalid stream number (outside the range [0, 15]), it can be closed. On the other hand, it only gets added to the stack if it was not already there, to avoid duplicates building up.

```
\ior_close:N
\ior_close:c
```

Seek the stream in the `\g__ior_streams_prop` list, then show the stream as open or closed accordingly.

```
\ior_show:N
\ior_log:N
\__ior_show:NN
```

Show the property lists, but with some “pretty printing”. See the l3msg module. The first argument of the message is `ior` (as opposed to `iow`) and the second is empty if no read stream is open and non-empty (the list of streams formatted using `msg_show-item_unbraced:nn`) otherwise. The code of the message `show-streams` takes care of translating `ior/iow` to English.

```
\ior_show_list:
\ior_log_list:
\__ior_list:N
```
49.1.3 Reading input

The primitive conditional \texttt{\if_eof:w} provides a conditional to test if some particular input stream is exhausted. The primitive test can only deal with numbers in the range $[0, 15]$ so we catch outliers (they are exhausted).

\begin{verbatim}
\ior_if_eof_p:N  \ior_if_eof:N  TF
\ior_if_eof:NTF
\end{verbatim}

To test if some particular input stream is exhausted the following conditional is provided. The primitive test can only deal with numbers in the range $[0, 15]$ so we catch outliers (they are exhausted).

\begin{verbatim}
\ior_get:NN  \ior_get:NN  TF
\ior_get:NN
\ior_get:NN
\end{verbatim}

And here we read from files.
Reading as strings is a more complicated wrapper, as we wish to remove the endline character and restore it afterwards.

\begin{verbatim}
\cs_new_protected:Npn \ior_str_get:NN #1#2 
  { \ior_str_get:NNF #1 #2 { \tl_set:Nn #2 { \q_no_value } } }
\cs_new_protected:Npn \__ior_str_get:NN #1#2 
  { \exp_args:Nno \use:n
    { \int_set:Nn \tex_endlinechar:D { -1 }
      \tex_readline:D #1 to #2
      \int_set:Nn \tex_endlinechar:D
    } { \int_use:N \tex_endlinechar:D } }
\prg_new_protected_conditional:Npnn \ior_str_get:NN #1#2 { T , F , TF }
  { \ior_if_eof:NTF #1 
    { \prg_return_false: }
    { \__ior_str_get:NN #1 #2 
      \prg_return_true: }
  }
\end{verbatim}

(End definition for \ior_str_get:NN, \__ior_str_get:NN, and \ior_str_get:NNTF. These functions are documented on page \pageref{page1}.)

\begin{verbatim}
\c__ior_term_noprompt_ior
\end{verbatim}

For reading without a prompt. \c__ior_term_noprompt_ior

(End definition for \c__ior_term_noprompt_ior.)

\begin{verbatim}
\ior_get_term:nN
\ior_str_get_term:nN
\__ior_get_term:NnN
\end{verbatim}

Getting from the terminal is better with pretty-printing.

\begin{verbatim}
\cs_new_protected:Npn \ior_get_term:nN #1#2 
  { \__ior_get_term:NnN \__ior_get:NN {#1} #2 }
\cs_new_protected:Npn \ior_str_get_term:nN #1#2 
  { \__ior_get_term:NnN \__ior_str_get:NN {#1} #2 }
\cs_new_protected:Npn \__ior_get_term:NnN #1#2#3 
  { \group_begin: 
      \tex_escapechar:D = -1 \scan_stop: 
      \tl_if_blank:nTF {#2} 
      { \exp_args:NNc #1 \c__ior_term_noprompt_ior }
      { \exp_args:NNc #1 \c__ior_term_ior }
      {#2}
      \exp_args:NNv \group_end: 
      \tl_set:Nn #3 {#2} }
\end{verbatim}

(End definition for \ior_get_term:nN, \ior_str_get_term:nN, and \__ior_get_term:NnN. These functions are documented on page \pageref{page2}.)

\begin{verbatim}
\ior_map_break:
\ior_map_break:n
\end{verbatim}

Usual map breaking functions.

\begin{verbatim}
\cs_new:Npn \ior_map_break:
  { \prg_map_break:Nn \ior_map_break: { } }
\cs_new:Npn \ior_map_break:n
  { \prg_map_break:Nn \ior_map_break: }
\end{verbatim}

617
Mapping over an input stream can be done on either a token or a string basis, hence the set up. Within that, there is a check to avoid reading past the end of a file, hence the two applications of \ior_if_eof:N and its lower-level analogue \if_eof:w. This mapping cannot be nested with twice the same stream, as the stream has only one “current line”.

Since the \TeX primitive (\read or \readline) assigns the tokens read in the same way as a token list assignment, we simply call the appropriate primitive. The end-of-loop is checked using the primitive conditional for speed.
49.2 Output operations

There is a lot of similarity here to the input operations, at least for many of the basics. Thus quite a bit is copied from the earlier material with minor alterations.

49.2.1 Variables and constants

\l__iow_internal_tl Used as a short-term scratch variable.
\tl_new:N \l__iow_internal_tl

\c_log_iow \c_term_iow Here we allocate two output streams for writing to the transcript file only (\c_log_iow) and to both the terminal and transcript file (\c_term_iow). Recent LuaTeX provide 128 write streams; we also use \c_term_iow as the first non-allowed write stream so its value depends on the engine.
\int_const:Nn \c_log_iow \{-1\}
\int_const:Nn \c_term_iow
\bool_lazy_and:nnTF\{\sys_if_engine_luatex_p:\}\{\int_compare_p:nNn\tex_luatexversion:D>\{80\}\}\{128\}\{16\}\}

\g__iow_streams_seq A list of the currently-available output streams to be used as a stack.
\seq_new:N \g__iow_streams_seq

\l__iow_stream_tl Used to recover the raw stream number from the stack.
\tl_new:N \l__iow_stream_tl

\g__iow_streams_prop As for reads with the appropriate adjustment of the register numbers to check on.
\cs_if_exist:NT \loccount { - 1 }
\}
\{
\prop_gput:Nnn \g__iow_streams_prop {#1} { Reserved-by-format }
\}

(End definition for \g__iow_streams_prop.)

49.2.2 Internal auxiliaries

\s__iow_mark
\s__iow_stop

Internal scan marks.
\scan_new:N \s__iow_mark
\scan_new:N \s__iow_stop

(End definition for \s__iow_mark and \s__iow_stop.)

\_iow_use_i_delimit_by_s_stop:nw

Functions to gobble up to a scan mark.
\cs_new:Npn \_iow_use_i_delimit_by_s_stop:nw #1 #2 \s__iow_stop {#1}

(End definition for \_iow_use_i_delimit_by_s_stop:nw.)

\q__iow_nil

Internal quarks.
\quark_new:N \q__iow_nil

(End definition for \q__iow_nil.)

49.3 Stream management

\iow_new:N
\iow_new:c

Reserving a new stream is done by defining the name as equal to writing to the terminal:
odd but at least consistent.
\cs_new_protected:Npn \iow_new:N #1 { \cs_new_eq:NN #1 \c_term_iow }
\cs_generate_variant:Nn \iow_new:N { c }

(End definition for \iow_new:N. This function is documented on page 86.)

\g_tmpa_iow
\g_tmpb_iow

The usual scratch space.
\iow_new:N \g_tmpa_iow
\iow_new:N \g_tmpb_iow

(End definition for \g_tmpa_iow and \g_tmpb_iow. These variables are documented on page 93.)

\_iow_new:N

As for read streams, copy \newwrite, making sure that it is not \outer.
\exp_args:NNf \cs_new_protected:Npn \_iow_new:N #1 { \cs_new_eq:NN #1 \c_term_iow }
\cs_generate_variant:Nn \iow_new:N { c }
\{ \exp_args:NNc \exp_after:wN \exp_stop_f: { newwrite } \}

(End definition for \_iow_new:N.)

\l__iow_file_name_tl

Data storage.
\tl_new:N \l__iow_file_name_tl

(End definition for \l__iow_file_name_tl.)
The same idea as for reading, but without the path and without the need to allow for a conditional version.

10344 \cs_new_protected:Npn \iow_open:Nn \#1 \#2
10345 { \__kernel_tl_set:Nx \l__iow_file_name_tl { \__kernel_file_name_sanitize:n \#2 } \iow_close:N \#1
10349 \seq_gpop:NNTF \g__iow_streams_seq \l__iow_stream_tl
10350 { \__iow_open_stream:NV \#1 \l__iow_file_name_tl }
10352 { \__iow_new:N \#1 \__kernel_tl_set:Nx \l__iow_stream_tl { \int_eval:n \#1 } \__iow_open_stream:NV \#1 \l__iow_file_name_tl
10356 }
10357 \cs_generate_variant:Nn \iow_open:Nn { c }
10358 \cs_new_protected:Npn \__iow_open_stream:Nn \#1 \#2
10359 { \tex_global:D \tex_chardef:D \#1 = \l__iow_stream_tl \scan_stop: \prop_gput:NVn \g__iow_streams_prop \#1 \#2 \tex_immediate:D \tex_openout:D \#1 \__kernel_file_name_quote:n \#2 \scan_stop: }
10365 \cs_generate_variant:Nn \__iow_open_stream:Nn { NV }
10366 (End definition for \iow_open:Nn and \__iow_open_stream:Nn. This function is documented on page 86.)

Closing a stream is not quite the reverse of opening one. First, the close operation is easier than the open one, and second as the stream is actually a number we can use it directly to show that the slot has been freed up.

10366 \cs_new_protected:Npn \iow_close:N \#1
10367 { \int_compare:nT { - \c_log_iow < \#1 < \c_term_iow } \int_compare:nF { \c_log_iow < \#1 < \c_term_iow } \int_compare:nF { \c_log_iow < \#1 < \c_term_iow } \tex_immediate:D \tex_closeout:D \#1 \prop_gremove:NV \g__iow_streams_prop \#1 \seq_if_in:NVF \g__iow_streams_seq \#1 { \seq_gpush:NV \g__iow_streams_seq \#1 } \cs_gset_eq:NN \#1 \c_term_iow
10375 }
10376 \cs_generate_variant:Nn \iow_close:N { c }
10377 (End definition for \iow_close:N. This function is documented on page 86.)

Seek the stream in the \g__iow_streams_prop list, then show the stream as open or closed accordingly.

10378 \cs_new_protected:Npn \iow_show:N \__iow_show:NN \tl_show:n
10379 \cs_generate_variant:Nn \iow_show:N { c }
10380 \cs_new_protected:Npn \iow_log:N \__iow_show:NN \tl_log:n
10381 \cs_generate_variant:Nn \iow_log:N { c }
10382 \cs_new_protected:Npn \__iow_show:NN \#1 \#2
10383 { \cs_generate_variant:Nn \iow_show:N \__iow_show:NN \#1 \#2
621
\begin{verbatim}
\_kernel_chk_defined:NT #2
{
 \prop_get:NVNTF \g__iow_streams_prop #2 \l__iow_internal_tl
 { \exp_args:Nx \l__iow_internal_tl #1
   { \token_to_str:N \l__iow_internal_tl open: - \l__iow_internal_tl }
   \exp_args:Nx \l__iow_internal_tl #2 { \token_to_str:N \l__iow_internal_tl closed }
 }
}

(End definition for \iow_show:N, \iow_log:N, and \__iow_show:NN. These functions are documented on page 86.)

\iow_list:
\iow_log:
\__iow_list:N
\cs_new_protected:Npn \iow_list:N #1
{ #1 { kernel } { show-streams }
  { iow }
  { \prop_map_function:NN \g__iow_streams_prop \msg_show_item_unbraced:nn }
}

(End definition for \iow_list:, \iow_log:, and \__iow_list:N. These functions are documented on page 87.)

49.3.1 Deferred writing

\iow_shipout_x:Nn
\iow_shipout_x:Nx
\iow_shipout_x:cn
\iow_shipout_x:cx
\cs_new_protected:Npm \iow_shipout_x:Nn #1 #2
\cs_generate_variant:Nn \iow_shipout_x:Nn { c, Nx, cx }

(End definition for \iow_shipout_x:Nn. This function is documented on page 90.)

\iow_shipout:Nn
\iow_shipout:Nx
\iow_shipout:cn
\iow_shipout:cx
\cs_new_protected:Npm \iow_shipout:Nn #1 #2
\cs_generate_variant:Nn \iow_shipout:Nn { c, Nx, cx }

(End definition for \iow_shipout:Nn. This function is documented on page 90.)

49.3.2 Immediate writing

\_kernel_iow_with:Nnn
\__iow_with:nNnn
\cs_new_protected:Npm \__kernel_iow_with:Nnn #1 #2
\end{verbatim}

49.3.1 Deferred writing

First the easy part, this is the primitive, which expects its argument to be braced.

With \verb+\text-TEX+ available deferred writing without expansion is easy.

If the integer \#1 is equal to \#2, just leave \#3 in the input stream. Otherwise, pass the old value to an auxiliary, which sets the integer to the new value, runs the code, and restores the integer.
This routine writes the second argument onto the output stream without expansion. If this stream isn’t open, the output goes to the terminal instead. If the first argument is no output stream at all, we get an internal error. We don’t use the expansion done by \write to get the \texttt{Nx} variant, because it differs in subtle ways from \texttt{x}-expansion, namely, macro parameter characters would not need to be doubled. We set the \texttt{\newlinechar} to 10 using \_\texttt{kernel_iow_with:Nnn} to support formats such as plain \TeX: otherwise, \iow_newline: would not work. We do not do this for \iow_shipout:Nn or \iow_shipout_x:Nn, as \TeX looks at the value of the \texttt{\newlinechar} at shipout time in those cases.

Writing to the log and the terminal directly are relatively easy.

\texttt{\newline}:

Global variable holding the character that forces a new line when something is written to an output stream.

Function to write any escaped char to an output stream.

(End definition for \_\texttt{kernel_iow_with:Nnn} and \_\texttt{iow_with:nNNn}.)
49.3.4 Hard-wrapping lines to a character count

The code here implements a generic hard-wrapping function. This is used by the messaging system, but is designed such that it is available for other uses.

\l_iow_line_count_int

This is the “raw” number of characters in a line which can be written to the terminal. The standard value is the line length typically used by \TeX{} Live and MiK\TeX{}.

\VerbatimInput{\int_new:N \l_iow_line_count_int
\int_set:Nn \l_iow_line_count_int { 78 }}

(End definition for \l_iow_line_count_int. This variable is documented on page 92.)

\l_iow_newline_tl

The token list inserted to produce a new line, with the (run-on text).

\VerbatimInput{\tl_new:N \l_iow_newline_tl}

(End definition for \l_iow_newline_tl.)

\l_iow_line_target_int

This stores the target line count: the full number of characters in a line, minus any part for a leader at the start of each line.

\VerbatimInput{\int_new:N \l_iow_line_target_int}

(End definition for \l_iow_line_target_int.)

\__iow_set_indent:n\__iow_unindent:w
\l_iow_one_indent_tl\l_iow_one_indent_int

The one_indent variables hold one indentation marker and its length. The \__iow_unindent:w auxiliary removes one indentation. The function \__iow_set_indent:n (that could possibly be public) sets the indentation in a consistent way. We set it to four spaces by default.

\VerbatimInput{\tl_new:N \l_iow_one_indent_tl\int_new:N \l_iow_one_indent_int\cs_new:Npn \__iow_unindent:w \l_iow_one_indent_tl \cs_new_protected:Npn \__iow_set_indent:n #1 \{\exp_args:No \__kernel_str_to_other_fast:n \tl_to_str:n {#1}\} \int_set:Nn \l_iow_one_indent_int \{\str_count:N \l_iow_one_indent_tl \exp_last_unbraced:NNo \cs_set:Npn \__iow_unindent:w \l_iow_one_indent_tl \{\} \exp_args:Nx \__iow_set_indent:n \l_iow_one_indent_int\}}

(End definition for \__iow_set_indent:n and others.)

\l_iow_indent_tl\l_iow_indent_int

The current indentation (some copies of \l_iow_one_indent_tl) and its number of characters.

\VerbatimInput{\tl_new:N \l_iow_indent_tl\int_new:N \l_iow_indent_int}

(End definition for \l_iow_indent_tl and \l_iow_indent_int.)

\l_iow_line_tl\l_iow_line_part_tl

These hold the current line of text and a partial line to be added to it, respectively.

\VerbatimInput{\tl_new:N \l_iow_line_tl\tl_new:N \l_iow_line_part_tl}

(End definition for \l_iow_line_tl and \l_iow_line_part_tl.)
\_\_\_iow\_line\_break\_bool Indicates whether the line was broken precisely at a chunk boundary.

10457 \bool_new:N \l__iow_line_break_bool

(End definition for \_\_\_iow\_line\_break\_bool.)

\_\_\_iow\_wrap\_tl Used for the expansion step before detokenizing, and for the output from wrapping text: fully expanded and with lines which are not overly long.

10458 \tl_new:N \l__iow_wrap_tl

(End definition for \_\_\_iow\_wrap\_tl.)

\c__iow\_wrap\_marker\_tl \c__iow\_wrap\_end\_marker\_tl \c__iow\_wrap\_newline\_marker\_tl \c__iow\_wrap\_allow\_break\_marker\_tl \c__iow\_wrap\_indent\_marker\_tl \c__iow\_wrap\_unindent\_marker\_tl

Every special action of the wrapping code starts with the same recognizable string, \c__iow\_wrap\_marker\_tl. Upon seeing that “word”, the wrapping code reads one space-delimited argument to know what operation to perform. The setting of \escapechar here is not very important, but makes \c__iow\_wrap\_marker\_tl look marginally nicer.

10459 \group_begin:
10460 \int_set:Nn \tex_escapechar:D { -1 }
10461 \tl_const:Nx \c__iow\_wrap\_marker\_tl { \tl_to_str:n { \^^I \^^O \^^W \^^_ \^^W \^^R \^^A \^^P } }
10462 \group_end:
10463 \tl_map_inline:nn { { end } { newline } { allow\_break } { indent } { unindent } }
10464 { \tl_const:cx { c__iow\_wrap\_ #1 \_marker\_tl }
10465 \c__iow\_wrap\_marker\_tl #1 \c_catcode\_other\_space\_tl
10466 }

(End definition for \c__iow\_wrap\_marker\_tl and others.)

\iow\_allow\_break: \_\_\_iow\_allow\_break: \_\_\_iow\_allow\_break\_error:

We set \iow\_allow\_break:n to produce an error when outside messages. Within wrapped message, it is set to \_\_\_iow\_allow\_break: when valid and otherwise to \_\_\_iow\_allow\_break\_error:. The second produces an error expandably.

10473 \cs_new_protected:Npn \iow\_allow\_break:
10474 { \msg_error:nnnn { kernel } { iow\_indent } { \iow\_wrap:nnnN } { \iow\_allow\_break: } }
10475 \cs_new:Npx \__iow\_allow\_break: { \c__iow\_wrap\_allow\_break\_marker\_tl }
10476 \cs_new:Npn \__iow\_allow\_break\_error:
10477 { \msg\_expandable\_error:nnnn { kernel } { iow\_indent } { \iow\_wrap:nnnN } { \iow\_allow\_break: } }

(End definition for \iow\_allow\_break:, \_\_\_iow\_allow\_break:, and \_\_\_iow\_allow\_break\_error:. This function is documented on page 302.)
\texttt{\iow_indent:n} \texttt{\__iow_indent:n} \texttt{\__iow_indent_error:n} We set \texttt{\iow_indent:n} to produce an error when outside messages. Within wrapped message, it is set to \texttt{\__iow_indent:n} when valid and otherwise to \texttt{\__iow_indent_error:n}. The first places the instruction for increasing the indentation before its argument, and the instruction for unindenting afterwards. The second produces an error expandably. Note that there are no forced line-break, so the indentation only changes when the next line is started.

\begin{verbatim}
\cs_new_protected:Npn \iow_indent:n #1
\msg_error:nnnnn { kernel } { iow-indent }
\{ \iow_wrap:nnnN \} \{ \iow_indent:n \} \{ #1 \}
\}
\cs_new:Npx \__iow_indent:n #1
\c__iow_wrap_indent_marker_tl #1
\c__iow_wrap_unindent_marker_tl
\}
\cs_new:Npn \__iow_indent_error:n #1
\msg_expandable_error:nnnnn { kernel } { iow-indent }
\{ \iow_wrap:nnnN \} \{ \iow_indent:n \} \{ #1 \}
\}
\end{verbatim}

\textit{(End definition for \texttt{\iow_indent:n}, \texttt{\__iow_indent:n}, and \texttt{\__iow_indent_error:n}. This function is documented on page 92.)}

\texttt{\iow_wrap:nnnN} \texttt{\iow_wrap:nxnN} The main wrapping function works as follows. First give \\, \underline{\text{\_}} and other formatting commands the correct definition for messages and perform the given setup \#3. The definition of \underline{\text{\_}} uses an “other” space rather than a normal space, because the latter might be absorbed by \TeX{} to end a number or other \verb|$|$-type expansions. Use \texttt{\conditionally@traceoff} if defined; it is introduced by the \texttt{trace} package and suppresses uninteresting tracing of the wrapping code.

\begin{verbatim}
\cs_new_protected:Npn \iow_wrap:nnnN #1#2#3#4
\group_begin:
\cs_if_exist_use:N \conditionally@traceoff
\int_set:Nn \tex_escapechar:D { -1 }
\cs_set:Npx { \token_to_str:N \{ }
\cs_set:Npx \# { \token_to_str:N \# }
\cs_set:Npx { \token_to_str:N \} }
\int_set:Nn \tex_escapechar:D { 92 }
\cs_set_eq:NN \\iow_newline: \iow_newline:
\cs_set_eq:NN \c_catcode_other_space_tl \c_catcode_other_space_tl
\cs_set_eq:NN \iow_allow_break: \__iow_allow_break:
\cs_set_eq:NN \iowIndent:n \__iow_indent:n \__iow_indent:n #3
\end{verbatim}

Then fully-expand the input: in package mode, the expansion uses \LaTeX{} \texttt{\protect} mechanism in the same way as \texttt{\typeout}. In generic mode this setting is useless but
harmless. As soon as the expansion is done, reset \texttt{iow\_indent:n} to its error definition: it only works in the first argument of \texttt{iow\_wrap:nnnN}.

\begin{verbatim}
\cs_set_eq:NN \protect \token_to_str:N \__kernel_tl_set:Nx \l__iow_wrap_tl {#1}
\cs_set_eq:NN \iow_allow_break: \__iow_allow_break_error:
\cs_set_eq:NN \iow_indent:n \__iow_indent_error:n
\end{verbatim}

Afterwards, set the newline marker (two assignments to fully expand, then convert to a string) and initialize the target count for lines (the first line has target count \texttt{l\_iow\_line\_count\_int} instead).

\begin{verbatim}
\__kernel_tl_set:Nx \l__iow_newline_tl { \iow_newline: #2 }
\__kernel_tl_set:Nx \l__iow_newline_tl { \tl_to_str:N \l__iow_newline_tl }
\int_set:Nn \l__iow_line_target_int { \l_iow_line_count_int - \str_count:N \l__iow_newline_tl + 1 }
\end{verbatim}

Sanity check.

\begin{verbatim}
\int_compare:nNnT { \l__iow_line_target_int } < 0 {
\tl_set:Nn \l__iow_newline_tl { \iow_newline: }
\int_set:Nn \l__iow_line_target_int { \l_iow_line_count_int + 1 }
}
\end{verbatim}

There is then a loop over the input, which stores the wrapped result in \texttt{l\_iow\_wrap\_tl}. After the loop, the resulting text is passed on to the function which has been given as a post-processor. The \texttt{tl\_to\_str:N} step converts the “other” spaces back to normal spaces. The \texttt{f}\-expansion removes a leading space from \texttt{l\_iow\_wrap\_tl}.

\begin{verbatim}
\__iow_wrap_do:
\exp_args:NNf \group_end: #4 { \tl_to_str:N \l__iow_wrap_tl }
\cs_generate_variant:Nn \iow_wrap:nnnN { nx }
\end{verbatim}

\texttt{(End definition for \texttt{iow\_wrap:nnnN}. This function is documented on page 92.)}

\begin{verbatim}
\__iow_wrap_do:
\__iow_wrapFixNewline:w
\__iow_wrapStart:w
\end{verbatim}

Escape spaces and change newlines to \texttt{c\_iow\_wrap\_newline\_marker\_tl}. Set up a few variables, in particular the initial value of \texttt{l\_iow\_wrap\_tl}: the space stops the \texttt{f}\-expansion of the main wrapping function and \texttt{\use\_none:n} removes a newline marker inserted by later code. The main loop consists of repeatedly calling the \texttt{chunk} auxiliary to wrap chunks delimited by (newline or indentation) markers.

\begin{verbatim}
\__iow_wrap_do:
\_\_iow_wrapFixNewline:w
\_\_iow_wrapStart:w
\end{verbatim}

\texttt{(End definition for \texttt{iow\_wrap:nnnN}. This function is documented on page 92.)}
The chunk and next auxiliaries are defined indirectly to obtain the expansions of \texttt{\_\_iow_wrap_newline:tl} and \texttt{\_\_iow_wrap_marker:tl} in their definition. The next auxiliary calls a function corresponding to the type of marker (its \#2), which can be \texttt{newline} or \texttt{indent} or \texttt{unindent} or \texttt{end}. The first argument of the chunk auxiliary is a target number of characters and the second is some string to wrap. If the chunk is empty simply call next. Otherwise, set up a call to \texttt{\_\_iow_wrap_line:nw}, including the indentation if the current line is empty, and including a trailing space (\#1) before the \texttt{\_\_iow_wrap_end_chunk:nw} auxiliary.

\texttt{\cs_set_protected:Npn \__iow_tmp:w \#1\#2}
\texttt{\cs_new_protected:Npn \__iow_wrap_chunk:nw \#2 #1 \__iow_wrap_next:nw}
\texttt{\cs_new_protected:Npn \__iow_wrap_start:w}
\texttt{\cs_new_protected:Npn \__iow_wrap_end_chunk:nw}

\texttt{\__iow_wrap_chunk:nw}
\texttt{\__iow_wrap_start:w}
\texttt{\__iow_wrap_next:nw}
\texttt{\__iow_wrap_end_chunk:nw}
This is followed by \{\langle\text{string}\rangle\} \langle\text{intexpr}\rangle ;. It stores the \langle\text{string}\rangle and up to \langle\text{intexpr}\rangle characters from the current chunk into \l__iow_line_part_tl. Characters are grabbed 8 at a time and left in \l__iow_line_part_tl by the line_loop auxiliary. When \(k < 8\) remain to be found, the line_aux auxiliary calls the line_end auxiliary followed by (the single digit) \(k\), then \(7 - k\) empty brace groups, then the chunk’s remaining characters. The line_end auxiliary leaves \(k\) characters from the chunk in the line part, then ends the assignment. Ignore the \use_none:nnnn line for now. If the next character is a space the line can be broken there: store what we found into the result and get the next line. Otherwise some work is needed to find a break-point. So far we have ignored what happens if the chunk is shorter than the requested number of characters: this is dealt with by the end_chunk auxiliary, which gets treated like a character by the rest of the code. It ends up being called either as one of the arguments \#2–\#9 of the line_loop auxiliary or as one of the arguments \#2–\#8 of the line_end auxiliary. In both cases stop the assignment and work out how many characters are still needed. Notice that when we have exactly seven arguments to clean up, a \exp_stop_f: has to be inserted to stop the \exp:w. The weird \use_none:nnnn ensures that the required data is in the right place.
Functions here are defined indirectly: \_\_iow\_tmp:w is eventually called with an “other” space as its argument. The goal is to remove from \l__iow_line_part_tl the part after the last space. In most cases this is done by repeatedly calling the break_loop auxiliary, which leaves “words” (delimited by spaces) until it hits the trailing space: then its argument ##3 is ? \_\_iow\_break\_end:w instead of a single token, and that break_end auxiliary leaves in the assignment the line until the last space, then calls \_\_iow\_wrap\_line\_end:nw to finish up the line and move on to the next. If there is no space in \l__iow_line_part_tl then the break_first auxiliary calls the break_none auxiliary. In that case, if the current line is empty, the complete word (including ##4, characters beyond what we had grabbed) is added to the line, making it over-long. Otherwise, the word is used for the following line (and the last space of the line so far is removed because it was inserted due to the presence of a marker).
\use_none:nn \_\_iow\_wrap\_break\_none:w
\_\_iow\_wrap\_break\_loop:w \#1 \#2
}
\cs_new:Npn \_\_iow\_wrap\_break\_none:w \#1\#2 \#3 \s\_iow\_mark \#4 \#1
{\tl_if_empty:NTF \l\_iow\_line_tl
{\#2 \#4 \_\_iow\_wrap\_end:nw \{ \}
{\_\_iow\_wrap\_end:nw \{ \_\_iow\_trim:N \} \#2 \#4 \#1}
}
\cs_new:Npn \_\_iow\_wrap\_break\_loop:w \#1 \#1 \#2 \#1 \#3
{\use_none:n \#1 \#1\_\_iow\_wrap\_break\_loop:w \#2 \#1 \#3}
\cs_new:Npn \_\_iow\_wrap\_break\_end:w \#1 \#1 \#2 \#3 \#1 \#4 \s\_iow\_mark
{\#1 \_\_iow\_wrap\_line\_end:nw \{ \} \#3}
\exp_args:NV \_\_iow\_tmp:w \c\_catcode\_other\_space_tl
(End definition for \_\_iow\_wrap\_break:w and others.)
\_\_iow\_wrap\_next\_line:w
The special case where the end of a line coincides with the end of a chunk is detected here, to avoid a spurious empty line. Otherwise, call \_\_iow\_wrap\_line:nw to find characters for the next line (remembering to account for the indentation).
\cs_new_protected:Npn \_\_iow\_wrap\_next\_line:w \#1\#2 \s\_iow\_stop
{\tl_clear:N \l\_iow\_line_tl\_\_iow\_wrap\_end\_chunk:w
{\tl_clear:N \l\_iow\_line\_part_tl\bool_set\_true:N \l\_iow\_line\_break\_bool
\_\_iow\_wrap\_next:nw \{ \l\_iow\_line\_target\_int \}
}
{\_\_iow\_wrap\_line:nw\{ \_\_iow\_indent\_tl \}
\l\_iow\_line\_target\_int - \_\_iow\_indent\_int ;\#1 \#2 \s\_iow\_stop
}
(End definition for \_\_iow\_wrap\_next\_line:w)
\_\_iow\_wrap\_allow\_break:n
This is called after a chunk has been wrapped. The \_\_iow\_line\_part\_tl typically ends with a space (except at the beginning of a line?), which we remove since the allow\_break marker should not insert a space. Then move on with the next chunk, making sure to adjust the target number of characters for the line in case we did remove a space.
\cs_new_protected:Npn \_\_iow\_wrap\_allow\_break:n \#1
{\_\_kernel\_tl_set:Nx \l\_iow\_line_tl\_\_kernel\_tl_set:Nx \l\_iow\_line_tl
\_\_iow\_trim:N \l\_iow\_line\_part\_tl
\bool_set\_false:N \l\_iow\_line\_break\_bool
\tl_if_empty:NTF \_\_iow\_line\_part\_tl
631
These functions are called after a chunk has been wrapped, when encountering indent/unindent markers. Add the line part (last line part of the previous chunk) to the line so far and reset a boolean denoting the presence of a line-break. Most importantly, add or remove one indent from the current indent (both the integer and the token list). Finally, continue wrapping.

\__iow_wrap_newline:n These functions are called after a chunk has been line-wrapped, when encountering a newline/end marker. Unless we just took a line-break, store the line part and the line so far into the whole \l__iow_wrap_tl, trimming a trailing space. In the newline case look for a new line (of length \l__iow_line_target_int) in a new chunk.

\__iow_wrap_store_do:n First add the last line part to the line, then append it to \l__iow_wrap_tl with the appropriate new line (with “run-on” text), possibly with its last space removed (#1 is empty or \__iow_wrap_trim:N).
\begin{verbatim}
\cs_new_protected:Npn \__iow_wrap_store_do:n #1
\{
  \__kernel_tl_set:Nx \l__iow_line_tl
  { \l__iow_line_tl \l__iow_line_part_tl }
  \__kernel_tl_set:Nx \l__iow_wrap_tl
  { \l__iow_wrap_tl \l__iow_newline_tl
    \l__iow_line_tl
  }
  \tl_clear:N \l__iow_line_tl
\}
(End definition for \__iow_wrap_store_do:n.)
\__iow_wrap_trim:N
\__iow_wrap_trim:w
\__iow_wrap_trim_aux:w
\cs_set_protected:Npn \__iow_tmp:w #1
{ \exp_after:wN \__iow_wrap_trim:w #1 \s__iow_mark #1 \s__iow_mark \s__iow_stop }
\cs_new:Npn \__iow_wrap_trim:w ##1 \s__iow_mark
{ \__iow_wrap_trim_aux:w ##1 \s__iow_mark }
\cs_new:Npn \__iow_wrap_trim_aux:w ##1 \s__iow_mark ##2 \s__iow_stop
{##1}
\exp_args:NV \__iow_tmp:w \c_catcode_other_space_tl
(End definition for \__iow_wrap_trim:N, \__iow_wrap_trim:w, and \__iow_wrap_trim_aux:w.)
\end{verbatim}

\section*{49.4 \ File operations}

\_\_file_internal_tl

Used as a short-term scratch variable.

\_\_file_internal_tl

(End definition for \_\_file_internal_tl.)

\_\_file_internal_tl

The name of the current file should be available at all times: the name itself is set
dynamically.

\_\_file_internal_tl

(End definition for \_\_file_internal_tl, \_\_file_internal_tl, and \_\_file_internal_tl. These
variables are documented on page 93.)

\_\_file_stack_seq

The input list of files is stored as a sequence stack. In package mode we can recover the
information from the details held by \LaTeX{}2\epsilon{} (we must be in the preamble and loaded
using \usepackage or \RequirePackage). As \LaTeX{}2\epsilon{} doesn't store directory and name
separately, we stick to the same convention here. In pre-loading, \@currnamestack is
empty so is skipped.

\_\_file_stack_seq

(End definition for \_\_file_stack_seq, \_\_file_stack_seq, and \_\_file_stack_seq. These
variables are documented on page 93.)
The total list of files used is recorded separately from the current file stack, as nothing is ever popped from this list. The current file name should be included in the file list! We will eventually copy the contents of \@filelist.

The current search path.

Scratch space for comma list conversion.

\cs_set:Npn \__file_tmp:w ##1 " ##2 " ##3 \s__file_stop
\seq_gput_right:Nx \g__file_stack_seq
\exp_after:wN \__file_tmp:w \tex_jobname:D " \tex_jobname:D " \s__file_stop
\seq_gput_right:Nn \g__file_stack_seq { { } {#1} {#2} }

\cs_if_exist:NT \@currnamestack
\tl_if_empty:NF \@currnamestack
\exp_after:wN \__file_tmp:w \@currnamestack
\seq_new:N \g__file_record_seq
\seq_new:N \l__file_base_name_tl
\seq_new:N \l__file_full_name_tl
\seq_new:N \l__file_dir_str
\seq_new:N \l__file_ext_str
\seq_new:N \l__file_name_str
\seq_new:N \l_file_search_path_seq
\seq_new:N \l__file_tmp_seq
The total list of files used is recorded separately from the current file stack, as nothing is ever popped from this list. The current file name should be included in the file list! We will eventually copy the contents of \@filelist.

\l__file_base_name_tl
\l__file_full_name_tl
\l__file_dir_str
\l__file_ext_str
\l__file_name_str
\l_file_search_path_seq
\l__file_tmp_seq

For storing the basename and full path whilst passing data internally.

In contrast to the above, these are never used outside the current module.

This variable is documented on page 94.
49.4.1 Internal auxiliaries

\s__file_stop

Internal scan marks.

\scan_new:N \s__file_stop

(End definition for \s__file_stop.)

\q__file_nil

Internal quarks.

\quark_new:N \q__file_nil

(End definition for \q__file_nil.)

\__file_quark_if_nil_p:n \__file_quark_if_nil:n

Branching quark conditional.

\__kernel_quark_new_conditional:Nn \__file_quark_if_nil:n { TF }

(End definition for \__file_quark_if_nil:nTF.)

\q__file_recursion_tail \q__file_recursion_stop

Internal recursion quarks.

\quark_new:N \q__file_recursion_tail
\quark_new:N \q__file_recursion_stop

(End definition for \q__file_recursion_tail and \q__file_recursion_stop.)

\__file_if_recursion_tail_break:NN \__file_if_recursion_tail_stop_do:Nn

Functions to query recursion quarks.

\__kernel_quark_new_test:N \__file_if_recursion_tail_stop:N
\__kernel_quark_new_test:N \__file_if_recursion_tail_stop_do:nn

(End definition for \__file_if_recursion_tail_break:NN and \__file_if_recursion_tail_stop_do:NN.)

\__kernel_file_name_sanitize:n \__file_name_expand:n \__file_name_expand_cleanup:Nw \__file_name_expand_cleanup:w
\__file_name_expand_end:
\__file_name_expand_error:Nw \__file_name_expand_error_aux:Nw
\__file_name_strip_quotes:n \__file_name_strip_quotes nosw
\__file_name_strip_quotes:nom
\__file_name_trim_spaces:n \__file_name_trim_spaces:nw
\__file_name_trim_spaces:aux:n \__file_name_trim_spaces:aux:w

Expanding the file name uses a \csname-based approach, and relies on active characters (for example from UTF-8 characters) being properly set up to expand to an expansion-safe version using \ifcsname. This is less conservative than the token-by-token approach used before, but it is much faster.

\cs_new:Npn \__kernel_file_name_sanitize:n #1
\exp_args:Ne \__file_name_trim_spaces:n
\exp_args:Ne \__file_name_strip_quotes:n
\__file_name_expand:n {#1}

We’ll use \cs:w to start expanding the file name, and to avoid creating csnames equal to \relax with “common” names, there’s a prefix __file_name= to the csname. There’s also a guard token at the end so we can check if there was an error during the process and (try to) clean up gracefully.

\cs_new:Npn \__kernel_file_name_sanitize:n #1
\exp_after:wN \__file_name_expand_cleanup:Nw
\cs:w __file_name = #1 \cs_end:
\__file_name_expand_end:
With the csname built, we grab it, and grab the remaining tokens delimited by \_\_file_name_expand_end:. If there are any remaining tokens, something bad happened, so we'll call the error procedure \_\_file_name expand error:Nw. If everything went according to plan, then use \token_to_str:N on the csname built, and call \_\_file_name expand cleanup:w to remove the prefix we added a while back. \_\file_name_ expand cleanup:w takes a leading argument so we don’t have to bother about the value of \text escapechar:D.

\cs_new:Npn \__file_name_expand_cleanup:Nw #1 #2 \__file_name_expand_end:
\tl_if_empty:nF {#2}
\__file_name_expand_error:Nw #2 \__file_name_expand_end:
\exp_after:wN \__file_name_expand_cleanup:w \token_to_str:N #1
\exp_last_unbraced:NNNNo
\cs_new:Npn \__file_name_expand_cleanup:w #1 \tl_to_str:n { __file_name = } { }

In non-error cases \_\file_name_expand_end: should not expand. It will only do so in case there is a \csname too much in the file name, so it will throw an error (while expanding), then insert the missing \cs_end: and yet another \_\file_name-expand-end: that will be used as a delimiter by \_\file_name expand cleanup:Nw (or that will expand again if yet another \endcsname is missing).

\cs_new:Npn \__file_name_expand_end:
\msg_expandable_error:nn
\cs_end: \_\file_name_expand_end:

Now to the error case. \_\file_name expand error:Nw adds an extra \cs_end: so that in case there was an extra \csname in the file name, then \_\file_name expand error aux:Nw throws the error.

\cs_new:Npn \__file_name expand error:Nw #1 #2 \__file_name expand end:
\__file_name expand error aux:Nw #1 #2 \cs_end: \_\file_name expand end:
\cs_new:Npn \__file_name expand error aux:Nw #1 #2 \cs_end: #3
\__file_name expand end:
\msg_expandable_error:nnff
\cs_end: \_\file_name expand end:

Quoting file name uses basically the same approach as for \luaquotejobname: count the " tokens and remove them.

\cs_new:Npn \__file_name strip quotes:nw #1
\__file_name strip quotes:nw { 0 }
#1 " \q_file_recursion_tail " \q_file_recursion_stop {#1}
\cs_new:Npn \__file_name strip quotes:nw #1#2 "
\if_meaning:w \q_file_recursion_tail #2
\__file_name strip quotes_end:w
\fi:
#2

636
Spaces need to be trimmed from the start of the name and from the end of any extension. However, the name we are passed might not have an extension: that means we have to look for one. If there is no extension, we still use the standard trimming function but deliberately prevent any spaces being removed at the end.

(End definition for \_file_name_sanitize:n and others.)

The same idea as the marker for rescanning token lists: this pair of tokens cannot appear in a file that is being input.

(End definition for \c__file_marker_tl.)

The approach here is similar to that for tl_rescan:N. The file contents are grabbed as an argument delimited by \c__file_marker_tl. A few subtleties: braces in \iffalse: ... \fi: to deal with possible alignment tabs, \tracingnesting to avoid
a warning about a group being closed inside the \scantokens, and \prg_return_true: is placed after the end-of-file marker.

\cs_new_protected:Npn \file_get:nnN \#1\#2\#3
\{ \file_get:nnNF \{\#1\} \{\#2\} \#3
\{ \tl_set:Nn \#3 \{ \q_no_value \} \}
\}
\prg_new_protected_conditional:Npn \file_get:nnN \#1\#2\#3 \{ T , F , TF \}
\{ \file_get_full_name:nNTF \{\#1\} \l__file_full_name_tl
\{ \exp_args:NV \__file_get_aux:nnN \l__file_full_name_tl
\{\#2\} \#3
\prg_return_true:
\}
\{ \prg_return_false: \}
\}
\cs_new_protected:Npx \__file_get_aux:nnN \#1\#2\#3
\{ \exp_not:N \if_false: \exp_not:N \fi:
\group_begin:
\int_set_eq:NN \tex_tracingnesting:D \c_zero_int
\exp_not:N \exp_args:No \tex_everyeof:D
\{ \exp_not:N \c__file_marker_tl \}
\#2 \scan_stop:
\exp_not:N \exp_after:wN \exp_not:N \__file_get_do:Nw \l__file_full_name_tl
\exp_not:N \exp_after:wN \exp_not:N \prg_do_nothing:
\sys_if_engine_luatex:TF
\{ \#1\} \exp_not:N \exp_after:wN \exp_not:N \prg_do_nothing:
\exp_not:N \exp_after:wN \exp_not:N \sysoptions:TF
\{ \#1\} \exp_not:N \exp_after:wN \exp_not:N \prg_return_true:
\}
\exp_not:N \exp_after:wN \exp_not:N \__kernel_file_name_quote:n \{\#1\} \scan_stop:
\exp_not:N \if_false: \exp_not:N \fi:
\}
\exp_args:Nno \use:nn \{ \cs_new_protected:Npn \__file_get_do:Nw \#1\#2 \}
\{ \c__file_marker_tl \}
\group_end:
\tl_set:No \#1 \{\#2\}
\}
\exp_args:Nno \use:nn \{ \cs_new_protected:Npn \__file_get_do:Nw \#1\#2 \}
\{ \c__file_marker_tl \}
\group_end:
\tl_set:No \#1 \{\#2\}
\}
\exp_args:Nno \use:nn \{ \cs_new_protected:Npn \__file_get_do:Nw \#1\#2 \}
\{ \c__file_marker_tl \}
\group_end:
\tl_set:No \#1 \{\#2\}
\}
\exp_args:Nno \use:nn \{ \cs_new_protected:Npn \__file_get_do:Nw \#1\#2 \}
\{ \c__file_marker_tl \}
\group_end:
\tl_set:No \#1 \{\#2\}
\}
(End definition for \file_get:nnNTF and others. These functions are documented on page 94.)

\__file_size:n A copy of the primitive where it’s available.
\cs_new_eq:NN \__file_size:n \tex_filesize:D
(End definition for \__file_size:n.)

\file_full_name:n File searching can be carried out if the \pdffilesize primitive or an equivalent is available. That of course means we need to arrange for everything else to here to be done by expansion too. We start off by sanitizing the name and quoting if required: we may need to remove those quotes, so the raw name is passed too.
First, we check if the file is just here: no mapping so we do not need the break part of the broader auxiliary. We are using the fact that the primitive here returns nothing if the file is entirely absent. To avoid unnecessary filesystem lookups, the result of `\pdffilesize` is kept available as an argument. For package mode, `\input@path` is a token list not a sequence.

Two pars to the auxiliary here so we can avoid doing quoting twice in the event we find the right file.

As \TeX automatically adds `.tex` if there is no extension, there is a little cleanup to do here. First, make sure we are not in the directory part, saving that. Then check for an extension.
Deal with the fact that the primitive might not be available.

\cs_if_exist:NF \tex_filesize:D
\cs_gset:Npn \file_full_name:n #1
\msg_expandable_error:nnn { kernel } { primitive-not-available }
\{ \(pdf\)filesize \}
\}{ kernel } { primitive-not-available }
\{ \(pdf\)filesize \}
\) (End definition for \file_full_name:n and others. This function is documented on page 94.)

\file_get_full_name:nN
\file_get_full_name:VN
\file_get_full_name:nTF
\file_get_full_name:VN
\file_get_full_name:nTF
These functions pre-date using \tex_filesize:D for file searching, so are get functions with protection. To avoid having different search set ups, they are simply wrappers
around the code above.

```latex
\cs_new_protected:Npn \file_get_full_name:nN #1#2
{ \file_get_full_name:nNF {#1} #2
   { \tl_set:Nn #2 { \q_no_value } }
}
\cs_generate_variant:Nn \file_get_full_name:nN { V }
\prg_new_protected_conditional:Nppnn \file_get_full_name:nN #1#2 { T , F , TF }
{ \__kernel_tl_set:Nx #2
   { \file_full_name:n {#1} }
   \tl_if_empty:NTF #2
   { \prg_return_false: }
   { \prg_return_true: }
}
\cs_generate_variant:Nn \file_get_full_name:nNT { V }
\cs_generate_variant:Nn \file_get_full_name:nNF { V }
\cs_generate_variant:Nn \file_get_full_name:nNTF { V }
```

If \texttt{\textbackslash tex\_filesize:D} is not available, the way to test if a file exists is to try to open it: if it does not exist then \TeX{} reports end-of-file. A search is made looking at each potential path in turn (starting from the current directory). The first location is of course treated as the correct one: this is done by jumping to \texttt{\prg\_break\_point:}. If nothing is found, \texttt{#2} is returned empty. A special case when there is no extension is that once the first location is found we test the existence of the file with \texttt{.tex} extension in that directory, and if it exists we include the \texttt{.tex} extension in the result.

```latex
\cs_if_exist:NF \tex_filesize:D
{ \prg_set_protected_conditional:Nppnn \file_get_full_name:nN #1#2 { T , F , TF }
{ \__kernel_tl_set:Nx \l__file_base_name_tl
{ \__kernel_file_name_sanitize:n {#1} }
\__file_get_full_name_search:nN { } \use:n
\seq_map_inline:Nn \l_file_search_path_seq
{ \__file_get_full_name_search:n { ##1 / } \seq_map_break:n }
\cs_if_exist:NT \input@path
{ \tl_map_inline:Nn \input@path
{ \__file_get_full_name_search:n { ##1 } \tl_map_break:n }
}
\tl_set:Nn \l__file_full_name_tl { \q_no_value }
\prg_break_point:
\quark_if_no_value:NTF \l__file_full_name_tl
{ \ior_close:N \g__file\_internal\_ior
\prg\_return\_false: }
{ \file_parse_full_name:VNNNN \l__file_full_name_tl
\l__file_dir_str \l__file_name_str \l__file_ext_str
\str_if_empty:NT \l__file_ext_str
{ \__kernel_ior_open:No \g__file\_internal\_ior
{ \l__file_full_name_tl .tex } }
```

641
\ior_{if\_eof}:NF \g__file\_internal\_ior
\{ \tl\_put\_right:Nn \l__file\_full\_name\_tl { .tex } \}
\ior\_close:N \g__file\_internal\_ior
\tl\_set\_eq:NN \#2 \l__file\_full\_name\_tl
\prg\_return\_true:
\}
\}
\}
\cs\_new\_protected:Npn \__file\_get\_full\_name\_search:nN \#1#2
\{ \__kernel\_tl\_set:Nx \l__file\_full\_name\_tl
\{ tl\_to\_str:n \{#1\} \l__file\_base\_name\_tl \}
\__kernel\_ior\_open:N No \g__file\_internal\_ior \l__file\_full\_name\_tl
\ior\_if\_eof:NF \g__file\_internal\_ior \{ \prg\_break: \}
\}
(End definition for \file\_get\_full\_name:nN, \file\_get\_full\_name:nNTF, and \__file\_get\_full\_name\_search:nN. These functions are documented on page 94.)
\g__file\_internal\_ior
A reserved stream to test for file existence (if required), and for opening a shell.
\ior\_new:N \g__file\_internal\_ior
(End definition for \g__file\_internal\_ior.)
\file\_mdfive\_hash:n
\file\_size:n
\file\_timestamp:n
\__file\_details:nn
\__file\_details\_aux:nn
\__file\_mdfive\_hash:n
Getting file details by expansion is relatively easy if a bit repetitive. As the MD5 function
has a slightly different syntax from the other commands, there is a little cleaning up to
do.
\cs\_new:Npn \file\_size:n \#1
\{ \__file\_details:nn \{#1\} \{ size \} \}
\cs\_new:Npn \file\_timestamp:n \#1
\{ \__file\_details:nn \{#1\} \{ moddate \} \}
\cs\_new:Npn \__file\_details:nn \#1#2
\{ \exp\_args:Ne \__file\_details\_aux:nn
\{ \file\_full\_name:n \{#1\} \} \{#2\} \}
\}
\cs\_new:Npn \__file\_details\_aux:nn \#1#2
\{ \tl\_if\_blank:NF \{#1\}
\{ \use:c \{ tex\_file \#2 :D \} \{#1\} \}
\}
\cs\_new:Npn \file\_mdfive\_hash:n \#1
\{ \exp\_args:Ne \__file\_mdfive\_hash:n \{ \file\_full\_name:n \{#1\} \} \}
\cs\_new:Npn \__file\_mdfive\_hash:n \#1
\{ \tex\_mdfivesum:D file \{#1\} \}
(End definition for \file\_mdfive\_hash:n and others. These functions are documented on page 96.)
\file\_hex\_dump:nnn
\__file\_hex\_dump\_auxi:nnn
\__file\_hex\_dump\_auxii:nnnn
\__file\_hex\_dump\_auxiii:nnnn
\__file\_hex\_dump\_auxiv:nnnn
\file\_hex\_dump:n
\__file\_hex\_dump:n
These are separate as they need multiple arguments or the file size. For Lua\TeX, the
emulation does not need the file size so we save a little on expansion.
\cs\_new:Npn \file\_hex\_dump:nnn \#1#2#3
\{ \exp\_args:Neee \__file\_hex\_dump\_auxi:nnn
\{ \file\_full\_name:n \{#1\} \}
\}
642
\cs_new:Npn \__file_hex_dump_auxi:nnn #1#2#3
{\bool_lazy_any:nF
{\tl_if_blank_p:n {#1}}
{\int_compare_p:nNn {#2} = 0}
{\int_compare_p:nNn {#3} = 0}
}

{\exp_args:Ne \__file_hex_dump_auxii:nnnn
{\__file_details_aux:nn {#1} { size }}
{#1} {#2} {#3}}

\cs_new:Npn \__file_hex_dump_auxii:nnnn #1#2#3#4
{\int_compare:nNnTF {#3} > 0
{\__file_hex_dump_auxiii:nnnn {#3}}
{\exp_args:Ne \__file_hex_dump_auxiii:nnnn
{\int_eval:n { #1 + #3 }}
{#1} {#2} {#4}}}

\cs_new:Npn \__file_hex_dump_auxiii:nnnn #1#2#3#4
{\int_compare:nNnTF {#4} > 0
{\__file_hex_dump_auxiv:nnn {#4}}
{\exp_args:Ne \__file_hex_dump_auxiv:nnn
{\int_eval:n { #2 + #4 }}
{#1} {#3}}}

\cs_new:Npn \__file_hex_dump_auxiv:nnn #1#2#3
{\tex_filedump:D
offset - \int_eval:n { #2 - 1 } -
length - \int_eval:n { #1 - #2 + 1 }
(#3)}

\cs_new:Npn \hex_dump:n #1
{\exp_args:NNe \hex_dump:nnn
{\file_full_name:n {#1}}}

\sys_if_engine_luatex:TF
{\cs_new:Npn \hex_dump:n #1
{\tl_if_blank:nF {#1}
{\tl_if_blank:nF {#1}
{\tex_filedump:D whole {#1} {#1}}}
\begin{verbatim}
\cs_new:Npn \__file_hex_dump:n #1
\{
\tl_if_blank:nF {#1}
\{ \textfiledump:D \textfilesize:D {#1} \}
\}
\end{verbatim}

(End definition for \file_hex_dump:nnn and others. These functions are documented on page 95.)

Non-expandable wrappers around the above in the case where appropriate primitive support exists.

\begin{verbatim}
\cs_new_protected:Npn \file_get_hex_dump:nN #1#2
{ \file_get_hex_dump:nNF {#1} #2 { \tl_set:Nn #2 { \q_no_value } } }
\cs_new_protected:Npn \file_get_mdfive_hash:nN #1#2
{ \file_get_mdfive_hash:nNF {#1} #2 { \tl_set:Nn #2 { \q_no_value } } }
\cs_new_protected:Npn \file_get_size:nN #1#2
{ \file_get_size:nNF {#1} #2 { \tl_set:Nn #2 { \q_no_value } } }
\cs_new_protected:Npn \file_get_timestamp:nN #1#2
{ \file_get_timestamp:nNF {#1} #2 { \tl_set:Nn #2 { \q_no_value } } }
\prg_new_protected_conditional:Npnn \file_get_hex_dump:nN #1#2 { T , F , TF }
{ \__file_get_details:nnN {#1} { hex_dump } #2 }
\prg_new_protected_conditional:Npnn \file_get_mdfive_hash:nN #1#2 { T , F , TF }
{ \__file_get_details:nnN {#1} { mdfive_hash } #2 }
\prg_new_protected_conditional:Npnn \file_get_size:nN #1#2 { T , F , TF }
{ \__file_get_details:nnN {#1} { size } #2 }
\prg_new_protected_conditional:Npnn \file_get_timestamp:nN #1#2 { T , F , TF }
{ \__file_get_details:nnN {#1} { timestamp } #2 }
\cs_new_protected:Npn \__file_get_details:nnN #1#2#3
{ \__kernel_tl_set:Nx #3 \textuse:c { file_ #2 :n } {#1} }
\tl_if_empty:NTF #3
{ \prg_return_false: }
{ \prg_return_true: }
\end{verbatim}

Where the primitive is not available, issue an error: this is a little more conservative than absolutely needed, but does work.

\begin{verbatim}
\cs_if_exist:NF \tex_filesize:D
{ \cs_set_protected:Npn \__file_get_details:nnN #1#2#3
{ \tl_clear:N #3 \msg_error:nnx { kernel } { primitive-not-available }
{ \token_to_str:N \textpdf)file \str_case:nn {#2}
{ hex_dump } { dump }
{ mdfive_hash } { mdfivesum }
{ timestamp } { moddate }
{ size } { size }
} } 644
\end{verbatim}
\file_get_hex_dump:nnnN
\file_get_hex_dump:nnnNTF

Custom code due to the additional arguments.
\cs_new_protected:Npn \file_get_hex_dump:nnnN \#1\#2\#3\#4
{ \file_get_hex_dump:nnnNF \{\#1\} \{\#2\} \{\#3\} \#4
  \tl_set:Nn \#4 { \q_no_value } }
\prg_new_protected_conditional:Npnn \file_get_hex_dump:nnnN \#1\#2\#3\#4
{ T , F , TF }
{ \__kernel_tl_set:Nx \#4
  \file_hex_dump:nnn \{\#1\} \{\#2\} \{\#3\}
  \tl_if_empty:NTF \#4
  \prg_return_false: }
{ \prg_return_true: }
}

(End definition for \file_get_hex_dump:nnnNTF and others. These functions are documented on page 95.)

\__file_str_cmp:nn

As we are doing a fixed-length “big” integer comparison, it is easiest to use the low-level behavior of string comparisons.
\cs_new_eq:NN \__file_str_cmp:nn \tex_strcmp:D

(End definition for \__file_str_cmp:nn.)

\file_compare_timestamp_p:nNn
\file_compare_timestamp:nNn
\__file_compare_timestamp:nnN
\__file_timestamp:n

Comparison of file date can be done by using the low-level nature of the string comparison functions.
\prg_new_conditional:Npnn \file_compare_timestamp:nNn \#1\#2\#3
{ p , T , F , TF }
{ \exp_args:Nee \__file_compare_timestamp:nnN
  \file_full_name:n \{\#1\}
  \file_full_name:n \{\#3\}
  \#2
 }
\cs_new:Npn \__file_compare_timestamp:nnN \#1\#2\#3
{ \tl_if_blank:nTF \{\#1\}
  \if_charcode:w \#3 <
    \prg_return_true:
  \else:
    \prg_return_false:
  \fi:
  \tl_if_blank:nTF \{\#2\}
  \if_charcode:w \#3 >
  \prg_return_false:
  \else:
    \prg_return_true:
  \fi:
  \tl_if_blank:nTF \{\#3\}
  \if_charcode:w \#3 <
    \prg_return_true:
  \else:
    \prg_return_false:
  \fi:
  \tl_if_blank:nTF \{\#1\}
  \if_charcode:w \#3 >
  \prg_return_false:
  \else:
    \prg_return_true:
  \fi:
  \tl_if_blank:nTF \{\#2\}
  \if_charcode:w \#3 >
  \prg_return_true:
  \else:
    \prg_return_false:
  \fi:
  \tl_if_blank:nTF \{\#3\}
  \if_charcode:w \#3 <
    \prg_return_true:
  \else:
    \prg_return_false:
  \fi:
  \tl_if_blank:nTF \{\#1\}
  \if_charcode:w \#3 <
    \prg_return_true:
  \else:
    \prg_return_false:
  \fi:
  \tl_if_blank:nTF \{\#2\}
  \if_charcode:w \#3 >
  \prg_return_false:
  \else:
    \prg_return_true:
  \fi:
  \tl_if_blank:nTF \{\#3\}
  \if_charcode:w \#3 <
    \prg_return_true:
  \else:
    \prg_return_false:
  \fi:
  \tl_if_blank:nTF \{\#1\}
  \if_charcode:w \#3 >
  \prg_return_true:
  \else:
    \prg_return_false:
  \fi:
  \tl_if_blank:nTF \{\#2\}
  \if_charcode:w \#3 >
  \prg_return_true:
  \else:
    \prg_return_false:
  \fi:
  \tl_if_blank:nTF \{\#3\}
  \if_charcode:w \#3 <
    \prg_return_true:
  \else:
    \prg_return_false:
  \fi:
  \tl_if_blank:nTF \{\#1\}
  \if_charcode:w \#3 <
    \prg_return_true:
  \else:
    \prg_return_false:
  \fi:
  \tl_if_blank:nTF \{\#2\}
  \if_charcode:w \#3 >
  \prg_return_false:
  \else:
    \prg_return_true:
  \fi:
  \tl_if_blank:nTF \{\#3\}
  \if_charcode:w \#3 <
    \prg_return_true:
  \else:
    \prg_return_false:
  \fi:

645
\prg_return_true:
\else:
\prg_return_false:
\fi:
{
\if_int_compare:w
\__file_str_cmp:nn
{ \__file_timestamp:n {#1} }
{ \__file_timestamp:n {#2} }
\c_zero_int
\prg_return_true:
\else:
\prg_return_false:
\fi:
}
\cs_new_eq:NN \__file_timestamp:n \tex_filemoddate:D
\cs_if_exist:NF \__file_timestamp:n
{ \msg_expandable_error:nnn
{ kernel } { primitive-not-available }
\prg_return_false: }
(End definition for \file_compare_timestamp:nNn \__file_compare_timestamp:nNn \__file_compare_timestamp:nNn. This function is documented on page 97.)

\file_if_exist:nTF
The test for the existence of a file is a wrapper around the function to add a path to a file. If the file was found, the path contains something, whereas if the file was not located then the return value is empty.

\cs_new_protected:Nn \file_if_exist:n #1 \{ T , F , TF \}
\{ \__file_input:V \l__file_full_name_tl \}
(End definition for \file_if_exist:nTF. This function is documented on page 94.)

\file_if_exist_input:n
\file_if_exist_input:nF
Input of a file with a test for existence. We do not define the T or TF variants because the most useful place to place the true code would be inconsistent with other conditionals.
A simple rename.

\cs_new_protected:Npn \file_input_stop: { \tex_endinput:D }

(End definition for \file_input_stop:. This function is documented on page 97.)

An error message for a missing file, also used in \ior_open:Nn.

\cs_new_protected:Npn \__kernel_file_missing:n #1
  { \msg_error:nnx { kernel } { file-not-found } \__kernel_file_name_sanitize:n {#1} }

(End definition for \__kernel_file_missing:n.)

Loading a file is done in a safe way, checking first that the file exists and loading only if it does. Push the file name on the \g__file_stack_seq, and add it to the file list, either \g__file_record_seq, or \@filelist in package mode.

\cs_new_protected:Npn \__file_input:n #1
  { \file_get_full_name:nNTF {#1} \l__file_full_name_tl
    \__file_input:V \l__file_full_name_tl
    \__kernel_file_missing:n {#1} }

\cs_new_protected:Npx \__file_input:n #1
  { \exp_not:N \clist_if_exist:NTF \exp_not:N \@filelist
    { \exp_not:N \@addtofilelist {#1} }
    { \seq_gput_right:Nn \exp_not:N \g__file_record_seq {#1} \__file_input_push:n {#1}
      \exp_not:N \__kernel_file_name_quote:n {#1} \scan_stop: \__file_input_pop:nnn }

(End definition for \__file_input:n.)

Keeping a track of the file data is easy enough: we store the separated parts so we do not need to parse them twice.

\cs_new_protected:Npn \__file_input_push:n #1
  { \exp_not:N \clist_if_exist:NTF \exp_not:N \@filelist
    \exp_not:N \@addtofilelist {#1} }

\cs_new_protected:Npx \__file_input_push:n #1
  { \seq_gput_right:Nn \exp_not:N \g__file_stack_seq
    \exp_not:N \__kernel_file_name_quote:n {#1} \scan_stop: }

647
The main parsing macro \file_parse_full_name_apply:n passes the file name #1 through \kernel_file_name_sanitize:n so that we have a single normalised way to treat files internally. \file_parse_full_name:n uses the former, with \prg_do_nothing: to leave each part of the name within a pair of braces.

\file_parse_full_name:n splits the file name into chunks separated by /, until the last one is reached. The last chunk is the file name plus the extension, and everything before that is the path. When \file_parse_full_name_area:nw is done, it leaves the path within braces after the scan mark \s_file_stop and proceeds parsing the actual file name.

(End definition for \file_input:n and others. This function is documented on page 97.)
\_file_parse_full_name_base:nw does roughly the same as above, but it separates the chunks at each period. However here there’s some extra complications: In case #1 is empty, it is assumed that the extension is actually empty, and the file name is #2. Besides, an extra . has to be added to #2 because it is later removed in \_file_parse_full_name_tidy:nnnN. In any case, if there’s an extension, it is returned with a leading ..

\_file_parse_full_name_tidy:nnnN

Now we just need to tidy some bits left loose before. The loop used in the two macros above start with a leading / and . in the file path an name, so here we need to remove them, except in the path, if it is a single /, in which case it’s left as is. After all’s done, pass to #4.

\file_parse_full_name:nNNN
\file_parse_full_name:VNNN

(End definition for \file_parse_full_name:n and others. These functions are documented on page 95.)

\file_parse_full_name:nnNN

(End definition for \file_parse_full_name:nNNN. This function is documented on page 95.)
A function to list all files used to the log, without duplicates. In package mode, if \@filelist is still defined, we need to take this list of file names into account (we capture it \AtBeginDocument into \g__file_record_seq), turning it to a string (this does not affect the commas of this comma list).

\CsNewProtectedNpn \file_show_list: { \__file_list:N \msg_show:nnxxx }
\CsNewProtectedNpn \file_log_list: { \__file_list:N \msg_log:nnxxx }
\CsNewProtectedNpn \__file_list:N \__file_list_aux:n

(End definition for \file_show_list: and others. These functions are documented on page 97.)

When used as a package, there is a need to hold onto the standard file list as well as the new one here. File names recorded in \@filelist must be turned to strings before being added to \g__file_record_seq.

\CsIfExistNT \@filelist
\AtBeginDocument
\{ \exp_args:NNx \seq_set_from_clist:Nn \__file_tmp_seq
\{ \tl_to_str:N \@filelist }
\}
\seq_gconcat:NNN \g__file_record_seq \g__file_record_seq \__file_tmp_seq
\seq_remove_duplicates:N \__file_tmp_seq
\seq_map_function:NN \__file_tmp_seq \__file_list_aux:n
\}
\CsNewNpn \__file_list_aux:n \iow_newline: #1 \{ \iow_newline: #1 \}

49.5 GetIdInfo

As documented in expl3.dtx this function extracts file name etc from an svn Id line. This used to be how we got version number and so on in all modules, so it had to be defined in b3bootstrap. Now it’s more convenient to define it after we have set up quite a lot of tools, and l3file seems the least unreasonable place for it.

The idea here is to extract out the information needed from a standard svn Id line, but to avoid a line that would get changed when the file is checked in. Hence the fact that none of the lines here include both a dollar sign and the Id keyword!

\CsNewProtectedNpn \GetIdInfo
\{ \tl_clear_new:N \ExplFileDescription
\tl_clear_new:N \ExplFileDate

650
A first check for a completely empty svn field. If that is not the case, there is a second case when a file created using svn cp but has not been checked in. That leaves a special marker -1 version, which has no further data. Dealing correctly with that is the reason for the space in the line to use __file_id_info_auxii:w.

Here, #1 is Id, #2 is the file name, #3 is the extension, #4 is the version, #5 is the check in date and #6 is the check in time and user, plus some trailing spaces. If #4 is the marker -1 value then #5 and #6 are empty.

Convert an SVN-style date into a \LaTeX-style one.

(End definition for \GetIdInfo and others. This function is documented on page 10.)

49.6 Checking the version of kernel dependencies

This function is responsible for checking if dependencies of the \LaTeXX kernel match the version preloaded in the \LaTeX\TeX 2\epsilon kernel. If versions don't match, the function attempts to tell why by searching for a possible stray format file.

The function starts by checking that the kernel date is defined, and if not zero is used to force the error route. The kernel date is then compared with the argument requested
date (usually the packaging date of the dependency). If the kernel date is less than the
required date, it’s an error and the loading should abort.

\cs_new_protected:Npn \_kernel_dependency_version_check:Nn #1
\exp_args:NV \_kernel_dependency_version_check:nn #1
\cs_new_protected:Npn \_kernel_dependency_version_check:nn #1
\cs_if_exist:NTF \c__kernel_expl_date_tl
\exp_args:NV \_file_kernel_dependency_compare:nnn
\c__kernel_expl_date_tl {#1}
\{ \_file_kernel_dependency_compare:nnn { 0000-00-00 } {#1} \}
\cs_new_protected:Npn \_file_kernel_dependency_compare:nnn #1 #2 #3
\int_compare:nNnT \__file_parse_version:w #1 \s__file_stop < \__file_parse_version:w #2 \s__file_stop
\__file_mismatched_dependency_error:nn {#2} {#3}
\cs_new:Npn \__file_parse_version:w #1 - #2 - #3 \s__file_stop {#1#2#3}
\__file_mismatched_dependency_error:nn

If the versions differ, then we try to give the user some guidance. This function starts by
taking the engine name \c_sys_engine_str and replacing \textit{tex} by \textit{latex}, then building
a command of the form: \texttt{kpsewhich --all --engine=⟨engine⟩ ⟨format⟩[-dev].fmt} to query
the format files available. A shell is opened and each line is read into a sequence.

\cs_new_protected:Npn \_file_mismatched_dependency_error:nn #1 #2
\exp_args:NNx \ior_shell_open:Nn \g__file_internal_ior
kpsewhich --all --engine=\c_sys_engine_exec_str \c_space_tl \c_sys_engine_format_str
\bool_lazy_and:nnT \tl_if_exist_p:N \development@branch@name \tl_if_empty_p:N \development@branch@name { -dev } .fmt
\seq_clear:N \l__file_tmp_seq \ior_map_inline:Nn \g__file_internal_ior \seq_put_right:Nn \l__file_tmp_seq {##1}
\ior_close:N \g__file_internal_ior \msg_error:nnnn { kernel } { mismatched-support-file } {#1} {#2}
\tex_endinput:D
% \begin{macrocode}
% Now define the actual error message:
\msg_new:nnnn { kernel } { mismatched-support-file } { kernel } { mismatched-support-file }
% \end{macrocode}

And finish by ending the current file.

\tex_endinput:D
%
Mismatched LaTeX support files detected. \ 

Loading-’#2’-aborted!

\c__kernel_expl_date_tl may not exist, due to an older format, so only print the dates when the sentinel token list exists:

\tl_if_exist:NT \c__kernel_expl_date_tl
{ \ \
  The-L3-programming-layer-in-the-LaTeX-format \ 
  is-dated-\c__kernel_expl_date_tl,-but-in-your-TeX-
  tree-the-files-require \ at-least-#1. 
}

{ }

The sequence containing the format files should have exactly one item: the format file currently being run. If that’s the case, the cause of the error is not that, so print a generic help with some possible causes. If more than one format file was found, then print the list to the user, with appropriate indications of what’s in the system and what’s in the user tree.

\int_compare:nNnTF { \seq_count:N \l__file_tmp_seq } > 1
{ \ 
  The-cause-seems-to-be-an-old-format-file-in-the-user-tree. \ 
  LaTeX-found-these-files:
  \seq_map_tokens:Nn \l__file_tmp_seq { \-~\use:n } \ 
  Try-deleting-the-file-in-the-user-tree-then-run-LaTeX-again.
}

{ }

{ 
  The-most-likely-causes-are:
  \-A-recent-format-generation-failed;
  \-A-stray-format-file-in-the-user-tree-which-needs-
  to-be-removed-or-rebuilt;
  \-You-are-running-a-manually-installed-version-of-#2 \ 
  which-is-incompatible-with-the-version-in-LaTeX. \ 
}

LaTeX-will-abort-loading-the-incompatible-support-files-
but-this-may-lead-to \ later-errors.-Please-ensure-that-
your-LaTeX-format-is-correctly-regenerated.

(End definition for \__kernel_dependency_version_check:Nn and others.)

49.7 Messages

\msg_new:nmm { kernel } { file-not-found }
{ File-’#1’-not-found. }

The-requested-file-could-not-be-found-in-the-current-directory,-
in-theTeX-search-path-or-in-theLaTeX-search-path.

\msg_new:nmm { kernel } { file-list }
{ }
\msg_new:nnnn { kernel } { filename-chars-lost }
{ \#1~invalid~in~file~name.~Lost:~\#2. }
{ There was an invalid token in the file name that caused the characters following it to be lost. }

\msg_new:nnnn { kernel } { filename-missing-endcsname }
{ Missing \texttt{\iow_char:N\endcsname} inserted in filename. }
{ The file name had more \texttt{\iow_char:N\csname} commands than \texttt{\iow_char:N\endcsname} ones. \LaTeX{} will add the missing \texttt{\iow_char:N\endcsname} and try to continue as best as it can. }

\msg_new:nnnn { kernel } { filename-missing-endcsname }
{ \unbalanced-quote-in-filename }
{ Unbalanced quotes in file name ‘\#1’. }
{ File names must contain balanced numbers of quotes (‘) . }

\msg_new:nnnn { kernel } { iow-indent }
{ Only \#1 allows \#2 }
{ The command \#2 can only be used in messages which will be wrapped using \#1. }
{ \tl_if_empty:nF {\#3} { - It was called with argument ‘\#3’. } }

\msg_new:nnnn { kernel } { iow-indent }
{ Only \#1 allows \#2 }
{ The command \#2 can only be used in messages which will be wrapped using \#1. }
{ \tl_if_empty:nF {\#3} { - It was called with argument ‘\#3’. } }

49.8 Functions delayed from earlier modules

\c_sys_platform_str Detecting the platform on \LaTeX{} is easy: for other engines, we use the fact that the two common cases have special null files. It is possible to probe further (see package \texttt{platform}), but that requires shell escape and seems unlikely to be useful. This is set up here as it requires file searching.

\sys_if_engine_luatex:TF
{ \str_const:Nx \c_sys_platform_str
{ \tex_directlua:D { tex.print(os.type) } }
}
{ \file_if_exist:nTF { nul: }
{ \file_if_exist:nF { /dev/null }
{ \str_const:Nn \c_sys_platform_str { windows } }
}
{ \file_if_exist:nF { /dev/null }
{ \str_const:Nn \c_sys_platform_str { unix } }
}
}
{ \cs_if_exist:NF \c_sys_platform_str

654
We can now set up the tests.

\begin{verbatim}
\clist_map_inline:nn { unix , windows }
\{
  \__file_const:nn { sys_if_platform_ #1 }
  \__if_eq_p:Vn \c_sys_platform_str { #1 }
\}
\end{verbatim}

(End definition for \sys_if_platform_unix:TF and \sys_if_platform_windows:TF. These functions are documented on page 74.)

\end{verbatim}

(End definition for \c_sys_platform_str. This variable is documented on page 74.)

\end{verbatim}
Chapter 50

\texttt{l3luatex} implementation

50.1 Breaking out to Lua

Copies of primitives.

\begin{verbatim}
\cs_new_eq:NN \__lua_escape:n \tex_luaescapestring:D
\cs_new_eq:NN \__lua_now:n \tex_directlua:D
\cs_new_eq:NN \__lua_shipout:n \tex_latelua:D
\end{verbatim}

(End definition for \texttt{\_\_lua\_escape:n, \_\_lua\_now:n, and \_\_lua\_shipout:n}.)

These functions are set up in \texttt{l3str} for bootstrapping: we want to replace them with a “proper” version at this stage, so clean up.

\begin{verbatim}
\cs_undefine:N \lua_escape:e
\cs_undefine:N \lua_now:e
\cs_undefine:N \lua_shipout_e:n
\lua_now:n \lua_now:e
\lua_shipout_e:n \lua_shipout:n
\lua_escape:n \lua_escape:e
\end{verbatim}

Wrappers around the primitives. As with engines other than Lua\TeX{} these have to be macros, we give them the same status in all cases. When Lua\TeX{} is not in use, simply give an error message/

\begin{verbatim}
\sys_if_engine_luatex:F {
\clist_map_inline:nn {
\lua_escape:n , \lua_escape:e ,
\lua_now:n , \lua_now:e
}{ \cs_set:Npn #1 ##1 }
\sys_if_engine_luatex:F {
\end{verbatim}

656
50.2 Messages

\msg_new:nnnn { luatex } { luatex-required } { LuaTeX~engine~not~in~use!~Ignoring~#1. } { The~feature~you~are~using~is~only~available~with~the~LuaTeX~engine.~\LaTeX{}3~ignored~'##1'. }

50.3 Lua functions for internal use

Most of the emulation of pdfTeX here is based heavily on Heiko Oberdiek’s \texttt{pdftex-cmds} package.

\texttt{ltx.utils} Create a table for the kernel’s own use.

\texttt{ltx} = \texttt{ltx} or \{\texttt{utils=}{}\}
\texttt{ltx.utils} = \texttt{ltx.utils} or \{\}
local \texttt{ltxutils} = \texttt{ltx.utils}

(End definition for \texttt{ltx.utils}. This function is documented on page 99.)

Local copies of global tables.

local io = io
local kpse = kpse
local lfs = lfs
local math = math
local md5 = md5
local os = os
local string = string
local tex = tex
local texio = texio
local tonumber = tonumber
Local copies of standard functions.

```lua
local abs = math.abs
local byte = string.byte
local floor = math.floor
local format = string.format
local gsub = string.gsub
local lfs_attr = lfs.attributes
local open = io.open
local os_date = os.date
local setcatcode = tex.setcatcode
local sprint = tex.sprint
local cprint = tex.cprint
local write = tex.write
local write_nl = texio.write_nl
local utf8_char = utf8.char
local scan_int = token.scan_int or token.scan_integer
local scan_string = token.scan_string
local scan_keyword = token.scan_keyword
local put_next = token.put_next
local token_create = token.create
local token_new = token.new
```

Since `token.create` only returns useful values after the tokens has been added to TeX's hash table, we define a variant which defines it first if necessary.

```lua
local true_tok = token_create_safe'prg_return_true:'
local false_tok = token_create_safe'prg_return_false:'
```

In ConTeXt lmtx `token.command_id` does not exist, but it can easily be emulated with ConTeXt's `tokens.commands`.

```lua
local command_id = token.command_id
if not command_id and tokens and tokens.commands then
  local id_map = tokens.commands
```

658
function command_id(name)
  return id_map[name]
end

local kpse_find = (resolvers and resolvers.findfile) or kpse.find_file

escapehex An internal auxiliary to convert a string to the matching hex escape. This works on a byte basis: extension to handled UTF-8 input is covered in pdftexcmds but is not currently required here.

function escapehex(str)
  return (gsub(str, ".", function (ch) return format("%02X", byte(ch)) end))
end

(lEnd definition for escapehex.)

ltx.utils.filedump Similar comments here to the next function: read the file in binary mode to avoid any line-end weirdness.

local function filedump(name,offset,length)
  local file = kpse_find(name,"tex",true)
  if not file then return end
  local f = open(file,"rb")
  if not f then return end
  if offset and offset > 0 then
    f:seek("set", offset)
  end
  local data = f:read(length or 'a')
  f:close()
  return escapehex(data)
end
ltxutils.filedump = filedump

(lEnd definition for ltx.utils.filedump. This function is documented on page 99.)

md5.HEX Hash a string and return the hash in uppercase hexadecimal format. In some engines, this is build-in. For traditional LuaTeX, the conversion to hexadecimal has to be done by us.

local md5_HEX = md5.HEX
if not md5_HEX then
  local md5_sum = md5.sum
  function md5_HEX(data)
    return escapehex(md5_sum(data))
  end
  md5.HEX = md5_HEX
end

(lEnd definition for md5.HEX. This function is documented on page ??.)

ltx.utils.filemd5sum Read an entire file and hash it: the hash function itself is a built-in. As Lua is byte-based there is no work needed here in terms of UTF-8 (see pdftexcmds and how it handles strings that have passed through LuaTeX). The file is read in binary mode so that no line ending normalisation occurs.
local function filemd5sum(name)
    local file = kpse_find(name, "tex", true) if not file then return end
    local f = open(file, "rb") if not f then return end
    local data = f:read("*a")
    f:close()
    return md5_HEX(data)
end
ltxutils.filemd5sum = filemd5sum

(End definition for ltx.utils.filemd5sum. This function is documented on page 99.)

ltx.utils.filemoddate

There are two cases: If the C standard library is C99 compliant, we can use %z to get
the timezone in almost the right format. We only have to add primes and replace a zero
or missing offset with Z.

Of course this would be boring, so Windows does things differently. There we have
to manually calculate the offset. See procedure makepdftime in utils.c of pdfTeX.

local filemoddate
if os_date('%z':match'~~[+-]%d%d%d$' then
    local pattern = lpeg.Cs(16 *
        (lpeg.Cg(lpeg.S'+-' * '0000' * lpeg.Cc'Z')
        + 3 * lpeg.Cc"" * 2 * lpeg.Cc'Z')
        * -1)
    function filemoddate(name)
        local file = kpse_find(name, "tex", true)
        if not file then return end
        local date = lfs_attr(file, "modification")
        if not date then return end
        return pattern:match(os_date("D:%Y%m%d%H%M%S%z", date))
    end
else
    local function filemoddate(name)
        local file = kpse_find(name, "tex", true)
        if not file then return end
        local date = lfs_attr(file, "modification")
        if not date then return end
        local d = os_date("*t", date)
        local u = os_date("!*t", date)
        local off = 60 * (d.hour - u.hour) + d.min - u.min
        if d.year ~= u.year then
            if d.year > u.year then
                off = off + 1440
            else
                off = off - 1440
            end
        elseif d.yday ~= u.yday then
            if d.yday > u.yday then
                off = off + 1440
            else
                off = off - 1440
            end
        end
        local timezone

660
if off == 0 then
    timezone = "Z"
else
    if off < 0 then
        timezone = "-"
        off = -off
    else
        timezone = "+
    end
    timezone = format("%s%02d'\%02d:\%02d\%02d%\%02d%a", d.year, d.month, d.day, d.hour, d.min, d.sec, timezone)
end

ltxutils.filemoddate = filemoddate

ltx.utils.filesize

A simple disk lookup.

local function filesize(name)
    local file = kpse_find(name, "tex", true)
    if file then
        local size = lfs_attr(file, "size")
        if size then
            return size
        end
    end
end

ltxutils.filesize = filesize

luadef

An internal function for defining control sequences form Lua which behave like primitives. This acts as a wrapper around token.set_lua which accepts a function instead of an index into the functions table.

local luacmd do
    local set_lua = token.set_lua
    local undefined_cs = command_id'undefined_cs'
    if not context and not luatexbase then require'ltluatex' end
    if luatexbase then
        local new_luafunction = luatexbase.new_luafunction
        local functions = lua.get_functions_table()
        function luacmd(name, func, ...)
            local id
            local tok = token_create(name)
            if tok.command == undefined_cs then
                id = new_luafunction(name)
                set_lua(name, id, ...)
            else
                id = tok.index or tok.mode
            end
            functions[id] = func
        end
    end

End definition for ltx.utils.filemoddate. This function is documented on page 99.)

End definition for ltx.utils.filesize. This function is documented on page 99.)

End definition for luadef. This function is documented on page 99.)
end
elseif context then
  local register = context.functions.register
  local functions = context.functions.known
  function luacmd(name, func, ...)
    local tok = token_create(name)
    if tok.command == undefined_cs then
      token.set_lua(name, register(func), ...)
    else
      functions[tok.index or tok.mode] = func
    end
  end
end
end

(End definition for luadef.)

50.4 Preserving iniTeX Lua data for runs

The Lua state is not dumped when a forat is written, therefore any Lua variables
filled doing format building need to be restored in order to be accessible during normal
runs.

We provide some kernel-internal helpers for this. They will only be available if
luatexbase is available. This is not a big restriction though, because ConTeXt (which
does not use luatexbase) does not load expl3 in the format.

if luatexbase then
  local register = token_create'@expl@luadata@bytecode'.index
  if status.ini_version then
    register_luadata
    register_luadata

register_luadata

register_luadata is only available during format generation. It accept a string which
uniquely identifies the data object and has to be provided to retrieve it later. Additionally
it accepts a function which is called in the pre_dump callback and which has to return a
string that evaluates to a valid Lua object to be preserved.

function register_luadata(name, func)
  if luadata[name] then
    error(format("LaTeX error: data name %q already in use", name))
  end
  luadata[name] = func
  luadata_order[#luadata_order + 1] = func and name
end

(End definition for register_luadata. This function is documented on page ??.)

The actual work is done in pre_dump. The luadata_order is used to ensure that
the order is consistent over multiple runs.

luatexbase.add_to_callback("pre_dump", function()
  if next(lua) then
    local str = "return \{
"
11864
11865
11866
11867
11868
11869
11870
11871

get_luadata

for i=1, #luadata_order do
local name = luadata_order[i]
str = format(’%s[%q]=%s,’, str, name, luadata[name]())
end
lua.bytecode[register] = assert(load(str .. "}"))
end
end, "ltx.luadata")
else

get_luadata is only available if data should be restored. It accept the identifier which
was used when the data object was registered and returns the associated object. Every
object can only be retrieved once.
11872
11873
11874
11875
11876
11877
11878
11879
11880
11881
11882
11883
11884

local luadata = lua.bytecode[register]
if luadata then
lua.bytecode[register] = nil
luadata = luadata()
end
function get_luadata(name)
if not luadata then return end
local data = luadata[name]
luadata[name] = nil
return data
end
end
end

(End definition for get_luadata. This function is documented on page ??.)
11885

⟨/lua⟩

11886

⟨/package⟩

663


Chapter 51

\textbf{l3legacy Implementation}

\begin{verbatim}
\legacy_if_p:n \legacy_if:nTF
\legacy_if_set_true:n \legacy_if_set_false:n \legacy_if_gset_true:n \legacy_if_gset_false:n
\end{verbatim}

\begin{verbatim}
\legacy_if_set:nn \legacy_if_gset:nn
\end{verbatim}

A friendly wrapper. We need to use the \texttt{if:w} approach here, rather than testing against \texttt{iftrue/iffalse} as the latter approach fails for primitive conditionals such as \texttt{ifmmode}. The \texttt{reverse_if:N} here means that we get a slightly more useful error if the name is undefined.

\begin{verbatim}
\prg_new_conditional:Npnn \legacy_if:n #1 { p , T , F , TF }
{ \exp_after:wN \reverse_if:N \cs:w if#1 \cs_end: \prg_return_false: \else: \prg_return_true: \fi: }
\end{verbatim}

(End definition for \texttt{\legacy_if:nTF}. This function is documented on page 100.)

A friendly wrapper.

\begin{verbatim}
\cs_new_protected:Npn \legacy_if_set_true:n #1 { \cs_set_eq:cN { if#1 } \if_true: }
\cs_new_protected:Npn \legacy_if_set_false:n #1 { \cs_set_eq:cN { if#1 } \if_false: }
\cs_new_protected:Npn \legacy_if_gset_true:n #1 { \cs_gset_eq:cN { if#1 } \if_true: }
\cs_new_protected:Npn \legacy_if_gset_false:n #1 { \cs_gset_eq:cN { if#1 } \if_false: }
\end{verbatim}

(End definition for \texttt{\legacy_if_set_true:n} and others. These functions are documented on page 100.)

A more elaborate wrapper.

\begin{verbatim}
\cs_new_protected:Npn \legacy_if_set:nn #1\#2
{ \bool_if:nTF {\#2} \legacy_if_set_true:n \legacy_if_set_false:n {\#1} }
\cs_new_protected:Npn \legacy_if_gset:nn #1\#2
\end{verbatim}
{ \bool_if:nTF {#2} \legacy_if_gset_true:n \legacy_if_gset_false:n
  (#1)
  }

(End definition for \legacy_if_set:n and \legacy_if_gset:n. These functions are documented on page 100.)

{/package}
A token list variable is a \TeX macro that holds tokens. By using the ε-\TeX primitive \texttt{\unexpanded} inside a \TeX \texttt{\edef} it is possible to store any tokens, including \texttt{#}, in this way.

### 52.1 Functions

\texttt{\__kernel_tl_set:Nx} and \texttt{\__kernel_tl_gset:Nx} These two are supplied to get better performance for macros which would otherwise use \texttt{\tl_set:Nx} or \texttt{\tl_gset:Nx} internally.

\texttt{\cs_new_eq:NN \__kernel_tl_set:Nx \cs_set_nopar:Npx}
\texttt{\cs_new_eq:NN \__kernel_tl_gset:Nx \cs_gset_nopar:Npx}

\texttt{\tl_new:N} Creating new token list variables is a case of checking for an existing definition and doing the definition.

\texttt{\cs_new_protected:Npn \tl_new:N \#1}
\texttt{\__kernel_chk_if_free_cs:N \#1}
\texttt{\cs_gset_eq:NN \#1 \c_empty_tl}

\texttt{\cs_generate_variant:Nn \tl_new:N \{ c \}}

\texttt{\tl_new:cn} \texttt{\tl_new:cx} Constants are also easy to generate. They use \texttt{\cs_gset_nopar:Npx} instead of \texttt{\__kernel_tl_gset:Nx} so that the correct scope checking is applied if \texttt{l3debug} is used.

\texttt{\cs_new_protected:Npn \tl_const:N \#1\#2}
\texttt{\__kernel_chk_if_free_cs:N \#1}
\texttt{\cs_gset_nopar:Npx \#1 \{ \__kernel_exp_not:w \#2 \} \#2}

\texttt{\cs_new_protected:Npn \tl_const:Nx \#1\#2}
\texttt{\__kernel_chk_if_free_cs:N \#1}
\texttt{\cs_gset_nopar:Npx \#1 \{ \__kernel_exp_not:w \#2 \} \#2}
Clearing a token list variable means setting it to an empty value. Error checking is sorted out by the parent function.

```
\tl_clear:N \tl_clear:c
\tl_gclear:N \tl_gclear:c
```

(End definition for \tl_clear:N and \tl_gclear:N. These functions are documented on page 103.)

```
\tl_clear_new:N \tl_clear_new:c
\tl_gclear_new:N \tl_gclear_new:c
```

Clearing a token list variable means setting it to an empty value. Error checking is sorted out by the parent function.

```
\cs_new_protected:Npm \tl_clear:N #1
\cs_new_protected:Npm \tl_gclear:N #1
\cs_generate_variant:Nn \tl_clear:N { c }
\cs_generate_variant:Nn \tl_gclear:N { c }
```

(End definition for \tl_clear:N and \tl_gclear:N. These functions are documented on page 103.)

```
\tl_set_eq:NN \tl_set_eq:Nc \tl_set_eq:cN
\tl_gset_eq:NN \tl_gset_eq:Nc \tl_gset_eq:cN
```

For setting token list variables equal to each other. To allow for patching, the arguments have to be explicit. In addition this ensures that a braced second argument will not cause problems.

```
\cs_new_protected:Npm \tl_set_eq:NN #1#2
\cs_new_protected:Npm \tl_gset_eq:NN #1#2
\cs_generate_variant:Nn \tl_set_eq:NN { cN, Nc, cc }
\cs_generate_variant:Nn \tl_gset_eq:NN { cN, Nc, cc }
```

(End definition for \tl_set_eq:NN and \tl_gset_eq:NN. These functions are documented on page 103.)

```
\tl_concat:NNN \tl_concat:ccc
\tl_gconcat:NNN \tl_gconcat:ccc
```

Concatenating token lists is easy. When checking is turned on, all three arguments must be checked: a token list \#2 or \#3 equal to \scan_stop: would lead to problems later on.

```
\cs_new_protected:Npm \tl_concat:NNN #1#2#3
\cs_new_protected:Npm \tl_gconcat:NNN #1#2#3
\__kernel_tl_set:Nx #1
\__kernel_exp_not:w \exp_after:wN \#2
\__kernel_exp_not:w \exp_after:wN \#3
\__kernel_tl_gset:Nx #1
```

667
52.2 Constant token lists

\c_empty_tl

Never full. We need to define that constant before using \tl_new:N.

(End definition for \c_empty_tl. This variable is documented on page 118.)

\c_novalue_tl

A special marker: as we don’t have \char_generate:nn yet, has to be created the old-fashioned way.

(End definition for \c_novalue_tl. This variable is documented on page 118.)

\c_space_tl

A space as a token list (as opposed to as a character).

(End definition for \c_space_tl. This variable is documented on page 118.)

52.3 Adding to token list variables

\tl_set:Nn
\tl_set:NV
\tl_set:Nv
\tl_set:No
\tl_set:Nf
\tl_set:Nx
\tl_gset:Nn
\tl_gset:NV
\tl_gset:Nv
\tl_gset:No
\tl_gset:Nf
\tl_gset:Nx
\tl_gset:cn
\tl_gset:cV
\tl_gset:cv
\tl_gset:co
\tl_gset:cf
\tl_gset:cx

By using \exp_not:n token list variables can contain # tokens, which makes the token list registers provided by \TeX{} more or less redundant. The \tl_set:No version is done by hand as it is used quite a lot.
{ \_kernel_tl_gset:Nx #1 \{#2\} }
\cs_generate_variant:Nn \tl_set:Nn \{ NV , Nv , Nf \}
\cs_generate_variant:Nn \tl_set:Nn \{ c , cV , cv , cf \}
\cs_generate_variant:Nn \tl_set:Nx \{ c \}
\cs_generate_variant:Nn \tl_gset:Nn \{ NV , Nv , Nf \}
\cs_generate_variant:Nn \tl_gset:Nn \{ c , cV , cv , cf \}
\cs_generate_variant:Nn \tl_gset:Nx \{ c \}
\cs_generate_variant:Nn \tl_gset:No \{ c \}

(End definition for \tl_set:Nn and \tl_gset:Nn. These functions are documented on page 103.)

\tl_put_left:Nn
\tl_put_left:NV
\tl_put_left:No
\tl_put_left:cn
\tl_put_left:cV
\tl_put_left:co
\tl_put_left:cx
\tl_gput_left:Nn
\tl_gput_left:NV
\tl_gput_left:No
\tl_gput_left:cn
\tl_gput_left:cV
\tl_gput_left:co
\tl_gput_left:cx

Adding to the left is done directly to gain a little performance.

\cs_new_protected:Npn \tl_put_left:Nn \tl_put_left:No \1 \2
\{ \_kernel_tl_set:Nx \1 \{ \_kernel_exp_not:w \{#2\} \_kernel_exp_not:w \exp_after:wN \1 \} \}
\cs_new_protected:Npn \tl_put_left:NV \tl_put_left:No \1 \2
\{ \exp_not:V \1 \_kernel_exp_not:w \exp_after:wN \1 \}
\cs_new_protected:Npn \tl_put_left:No \tl_put_left:No \1 \2
\{ \_kernel_tl_set:Nx \1 \{ \_kernel_exp_not:w \exp_after:wN \1 \} \}
\cs_new_protected:Npn \tl_put_left:cn \tl_put_left:cn \1 \2
\{ \exp_after:wN \1 \_kernel_exp_not:w \exp_after:wN \1 \}
\cs_new_protected:Npn \tl_put_left:cV \tl_put_left:cV \1 \2
\{ \_kernel_tl_set:Nx \1 \{ \_kernel_exp_not:w \exp_after:wN \1 \} \}
\cs_new_protected:Npn \tl_put_left:co \tl_put_left:co \1 \2
\{ \_kernel_exp_not:w \exp_after:wN \1 \}
\cs_new_protected:Npn \tl_gput_left:Nn \tl_gput_left:NV \1 \2
\{ \_kernel_tl_gset:Nx \1 \{ \_kernel_exp_not:w \{#2\} \_kernel_exp_not:w \exp_after:wN \1 \} \}
\cs_new_protected:Npn \tl_gput_left:NV \tl_gput_left:No \1 \2
\{ \exp_not:V \1 \_kernel_exp_not:w \exp_after:wN \1 \}
\cs_new_protected:Npn \tl_gput_left:No \tl_gput_left:No \1 \2
\{ \_kernel_tl_gset:Nx \1 \{ \_kernel_exp_not:w \exp_after:wN \1 \} \}
\cs_new_protected:Npn \tl_gput_left:cn \tl_gput_left:cn \1 \2
\{ \exp_after:wN \1 \_kernel_exp_not:w \exp_after:wN \1 \}
\cs_new_protected:Npn \tl_gput_left:cV \tl_gput_left:cV \1 \2
\{ \_kernel_tl_gset:Nx \1 \{ \_kernel_exp_not:w \exp_after:wN \1 \} \}
\cs_new_protected:Npn \tl_gput_left:co \tl_gput_left:co \1 \2
\{ \_kernel_exp_not:w \exp_after:wN \1 \}
\cs_new_protected:Npn \tl_gput_left:cx \tl_gput_left:cx \1 \2
\{ \exp_after:wN \1 \_kernel_exp_not:w \exp_after:wN \1 \}
\cs_generate_variant:Nn \tl_put_left:Nn \{ c \}
\cs_generate_variant:Nn \tl_gput_left:Nn \{ c \}
(End definition for \texttt{\tl_put_left:Nn} and \texttt{\tl_gput_left:Nn}. These functions are documented on page 103.)

\texttt{\tl_put_right:Nn}
\texttt{\tl_put_right:NV}
\texttt{\tl_put_right:No}
\texttt{\tl_put_right:Nx}
\texttt{\tl_put_right:cn}
\texttt{\tl_put_right:cV}
\texttt{\tl_put_right:co}
\texttt{\tl_put_right:cx}
\texttt{\tl_gput_right:Nn}
\texttt{\tl_gput_right:NV}
\texttt{\tl_gput_right:No}
\texttt{\tl_gput_right:Nx}
\texttt{\tl_gput_right:cn}
\texttt{\tl_gput_right:cV}
\texttt{\tl_gput_right:co}
\texttt{\tl_gput_right:cx}

The same on the right.

\cs_new_protected:Npn \tl_put_right:Nn #1#2
\cs_new_protected:Npn \tl_put_right:NV #1#2
\cs_new_protected:Npn \tl_put_right:No #1#2
\cs_new_protected:Npn \tl_put_right:Nx #1#2
\cs_generate_variant:Nn \tl_put_right:Nn { c }
\cs_generate_variant:Nn \tl_put_right:NV { c }
\cs_generate_variant:Nn \tl_put_right:No { c }
\cs_generate_variant:Nn \tl_put_right:Nx { c }
\cs_generate_variant:Nn \tl_gput_right:Nn { c }
\cs_generate_variant:Nn \tl_gput_right:NV { c }
\cs_generate_variant:Nn \tl_gput_right:No { c }
\cs_generate_variant:Nn \tl_gput_right:Nx { c }

\cs_new_protected:Npn \tl_gput_right:Nn #1#2
\cs_new_protected:Npn \tl_gput_right:NV #1#2
\cs_new_protected:Npn \tl_gput_right:No #1#2
\cs_new_protected:Npn \tl_gput_right:Nx #1#2
\cs_generate_variant:Nn \tl_gput_right:Nn { c }
\cs_generate_variant:Nn \tl_gput_right:NV { c }
\cs_generate_variant:Nn \tl_gput_right:No { c }
\cs_generate_variant:Nn \tl_gput_right:Nx { c }

\cs_generate_variant:Nn \tl_put_left:Nn { c }
\cs_generate_variant:Nn \tl_put_left:NV { c }
\cs_generate_variant:Nn \tl_put_left:No { c }
\cs_generate_variant:Nn \tl_put_left:Nx { c }
\cs_generate_variant:Nn \tl_gput_left:Nn { c }
\cs_generate_variant:Nn \tl_gput_left:NV { c }
\cs_generate_variant:Nn \tl_gput_left:No { c }
\cs_generate_variant:Nn \tl_gput_left:Nx { c }

\cs_new_protected:Npn \tl_put_left:Nn #1#2
\cs_new_protected:Npn \tl_put_left:NV #1#2
\cs_new_protected:Npn \tl_put_left:No #1#2
\cs_new_protected:Npn \tl_put_left:Nx #1#2
\cs_generate_variant:Nn \tl_put_left:Nn { c }
\cs_generate_variant:Nn \tl_put_left:NV { c }
\cs_generate_variant:Nn \tl_put_left:No { c }
\cs_generate_variant:Nn \tl_put_left:Nx { c }
\cs_generate_variant:Nn \tl_gput_left:Nn { c }
\cs_generate_variant:Nn \tl_gput_left:NV { c }
\cs_generate_variant:Nn \tl_gput_left:No { c }
\cs_generate_variant:Nn \tl_gput_left:Nx { c }

670
52.4 Internal quarks and quark-query functions

Internal quarks.
\q__tl_nil
\q__tl_mark
\q__tl_stop

(End definition for \q__tl_nil, \q__tl_mark, and \q__tl_stop.)

Internal recursion quarks.
\q__tl_recursion_tail
\q__tl_recursion_stop

(End definition for \q__tl_recursion_tail and \q__tl_recursion_stop.)

Functions to query recursion quarks.
\__tl_if_recursion_tail_break:nN
\__tl_if_recursion_tail_stop_p:n
\__tl_if_recursion_tail_stop:n

(End definition for \__tl_if_recursion_tail_break:nN and \__tl_if_recursion_tail_stop:pT.)

52.5 Reassigning token list category codes

The rescanning code needs a special token list containing the same character (chosen here to be a colon) with two different category codes: it cannot appear in the tokens being rescanned since all colons have the same category code.
\c__tl_rescan_marker_tl
\tl_const:Nx \c__tl_rescan_marker_tl { : \token_to_str:N : }

(End definition for \c__tl_rescan_marker_tl.)

In a group, after some initial setup explained below and the user setup #3 (followed by \scan_stop: to be safe), there is a call to \__tl_set_rescan:nNN. This shared auxiliary defined later distinguishes single-line and multi-line "files". In the simplest case of multi-line files, it calls (with the same arguments) \__tl_set_rescan_multi:nNN, whose code is included here to help understand the approach. This function rescans its argument #1, closes the group, and performs the assignment.

One difficulty when rescanning is that \scantokens treats the argument as a file, and without the correct settings a \TeX error occurs:

File ended while scanning definition of ... 

A related minor issue is a warning due to opening a group before the \scantokens and closing it inside that temporary file; we avoid that by setting \tracingnesting. The standard solution to the "File ended" error is to grab the rescanned tokens as a delimited argument of an auxiliary, here \__tl_rescan:NNw, that performs the assignment, then let \TeX "execute" the end of file marker. As usual in delimited arguments we use \prg_do_nothing: to avoid stripping an outer set braces: this is removed by using o-expanding assignments. The delimiter cannot appear within the rescanned token list because it contains twice the same character, with different catcodes.
For \texttt{\tl_rescan:nn} we cannot simply call \texttt{\_\_tl_set_rescan:NNnn \prg_do_nothing: \use:n} because that would leave the end-of-file marker after the result of rescanning. If that rescanned result is code that looks further in the input stream for arguments, it would break.

For multi-line files the only subtlety is that \texttt{\newlinechar} should be equal to \texttt{\endlinechar} because \texttt{\newlinechar} characters become new lines and then become \texttt{\endlinechar} characters when writing to an abstract file and reading back. This equality is ensured by setting \texttt{\newlinechar} equal to \texttt{\endlinechar}. Prior to this, \texttt{\endlinechar} is set to \texttt{-1} if it was \texttt{32} (in particular true after \texttt{\ExplSyntaxOn}) to avoid unreasonable line-breaks at every space for instance in error messages triggered by the user setup. Another side effect of reading back from the file is that spaces (catcode \texttt{10}) are ignored at the beginning of lines, and spaces and tabs (character code \texttt{32} and \texttt{9}) are ignored at the end of lines.

The two \texttt{\if_false: ... \fi:} are there to prevent alignment tabs to cause a change of tabular cell while rescanning. We put the “opening” one after \texttt{\group_begin:} so that if one accidentally \texttt{f}-expands \texttt{\tl_set_rescan:nn} braces remain balanced. This is essential in e-type arguments when \texttt{\expanded} is not available.
\begin{verbatim}
{ \group_end:
#1 #2 {#3}
}
\cs_generate_variant:Nn \tl_set_rescan:Nnn { Nno , Nnx }
\cs_generate_variant:Nn \tl_set_rescan:Nnn { c , cno , cnx }
\cs_generate_variant:Nn \tl_gset_rescan:Nnn { Nno , Nnx }
\cs_generate_variant:Nn \tl_gset_rescan:Nnn { c , cno }
\end{verbatim}

(End definition for \tl_set_rescan:Nnn and others. These functions are documented on page 117.)

The function \_\_\_tl_set_rescan:nNN calls \_\_\_tl_set_rescan_multi:nNN or \_\_\_tl_set_rescan_single:nNN \{ ’ \} depending on whether its argument is a single-line fragment of code/data or is made of multiple lines by testing for the presence of a newlinechar character. If \newlinechar is out of range, the argument is assumed to be a single line.

For a single line, no \endlinechar should be added, so it is set to −1, and spaces should not be removed. Trailing spaces and tabs are a difficult matter, as \TeX removes these at a very low level. The only way to preserve them is to rescan not the argument but the argument followed by a character with a reasonable category code. Here, 11 (letter) and 12 (other) are accepted, as these are convenient, suitable for delimiting an argument, and it is very unlikely that none of the ASCII characters are in one of these categories. To avoid selecting one particular character to put at the end, whose category code may have been modified, there is a loop through characters from ’ (ASCII 39) to ~ (ASCII 127). The choice of starting point was made because this is the start of a very long range of characters whose standard category is letter or other, thus minimizing the number of steps needed by the loop (most often just a single one). If no valid character is found (very rare), fall-back on \_\_\_tl_set_rescan_multi:nNN.

Otherwise, once a valid character is found (let us use ’ in this explanation) run some code very similar to \_\_\_tl_set_rescan_multi:nNN but with ’ added at both ends of the input. Of course, we need to define the auxiliary \_\_\_tl_set_rescan_single:nnNN \{ ’ \} on the fly to remove the additional ’ that is just before ’: (by which we mean \c__tl_rescan_marker_tl). Note that the argument must be delimited by ’ with the current catcode; this is done thanks to \char_generate:nn. Yet another issue is that the rescanned token list may contain a comment character, in which case the ’ we expected is not there. We fix this as follows: rather than just :: we set \everyeof to ::\{code1\}\':\{code2\}\}\_\_\_tl_stop. The auxiliary \_\_\_tl_set_rescan_single:nnNN runs the o-expanding assignment, expanding either \{code1\} or \{code2\} before its the main argument #3. In the typical case without comment character, \{code1\} is expanded, removing the leading ’. In the rarer case with comment character, \{code2\} is expanded, calling \_\_\_tl_set_rescan_single_aux:w, which removes the trailing ::\{code1\}\} and the leading ’.

\begin{verbatim}
\cs_new_protected:Npn \_\_\_tl_set_rescan:nNN #1
\int_compare:nNnTF \tex_newlinechar:D < 0
{ \use_ii:nn }
\exp_args:Nnf \tl_if_in:nnTF {#1}
{ \char_generate:nn \{ \tex_newlinechar:D \} \{ 12 \} }
\}
{ \_\_\_tl_set_rescan_multi:nNN }
\end{verbatim}

673
52.6 Modifying token list variables

All of the replace functions call \_\_tl_replace:NnnNNnn with appropriate arguments. The first two arguments are explained later. The next controls whether the replacement function calls itself (\_\_tl_replace_next:w) or stops (\_\_tl_replace_wrap:w) after

\tl_replace_all:Nnn
\tl_replace_all:cnn
\tl_greplace_all:Nnn
\tl_greplace_all:cnn
\tl_replace_once:Nnn
\tl_replace_once:cnn
\tl_greplace_once:Nnn
\tl_greplace_once:cnn
To implement the actual replacement auxiliary \texttt{\tl_replace_all:Nnn} we need a \{\texttt{delimiter}\} with the following properties:

- all occurrences of the \{\texttt{pattern}\} \#6 in \{\texttt{token list}\} \{\texttt{delimiter}\} belong to the \{\texttt{token list}\} and have no overlap with the \{\texttt{delimiter}\},
- the first occurrence of the \{\texttt{delimiter}\} in \{\texttt{token list}\} \{\texttt{delimiter}\} is the trailing \{\texttt{delimiter}\}.

We first find the building blocks for the \{\texttt{delimiter}\}, namely two tokens \(\langle A\rangle\) and \(\langle B\rangle\) such that \(\langle A\rangle\) does not appear in \#6 and \#6 is not \(\langle B\rangle\) (this condition is trivial if \#6 has more than one token). Then we consider the delimiters “\(\langle A\rangle\)” and “\(\langle A\rangle^n \langle B\rangle \langle A\rangle^n \langle B\rangle\)”, for \(n \geq 1\), where \(\langle A\rangle^n\) denotes \(n\) copies of \(\langle A\rangle\), and we choose as our \{\texttt{delimiter}\} the first one which is not in the \{\texttt{token list}\}.

Every delimiter in the set obeys the first condition: \#6 does not contain \(\langle A\rangle\) hence cannot be overlapping with the \{\texttt{token list}\} and the \{\texttt{delimiter}\}, and it cannot be within the \{\texttt{delimiter}\} since it would have to be in one of the two \(\langle B\rangle\) hence be equal to this single token (or empty, but this is an error case filtered separately). Given the particular form of these delimiters, for which no prefix is also a suffix, the second condition is actually a consequence of the weaker condition that the \{\texttt{delimiter}\} we choose does not appear in the \{\texttt{token list}\}. Additionally, the set of delimiters is such that a \{\texttt{token list}\} of \(n\) tokens can contain at most \(O(n^{3/2})\) of them, hence we find a \{\texttt{delimiter}\} with at most \(O(n^{3/2})\) tokens in a time at most \(O(n^{3/2})\). Bear in mind that these upper bounds are reached only in very contrived scenarios: we include the case “\(\langle A\rangle\)” in the list of delimiters to try, so that the \{\texttt{delimiter}\} is simply \texttt{\q__tl_mark} in the most common situation where neither the \{\texttt{token list}\} nor the \{\texttt{pattern}\} contains \texttt{\q__tl_mark}.

Let us now ahead, optimizing for this most common case. First, two special cases: an empty \{\texttt{pattern}\} \#6 is an error, and if \#1 is absent from both the \{\texttt{token list}\} \#5 and the \{\texttt{pattern}\} \#6 then we can use it as the \{\texttt{delimiter}\} through \texttt{\tl_replace_auxii:nNNNnn \#1}. Otherwise, we end up calling \texttt{\tl_replace:NnnNNNnn} repeatedly with the first two arguments \texttt{\q__tl_mark \#1}, \texttt{\q__tl_mark \#1}, \texttt{\q__tl_mark \#1}, and so on, until \#6 does not contain the control sequence \#1, which we take as our \(\langle A\rangle\). The argument \#2 only serves to collect \(n\) characters for \#1. Note that the order of the tests means that the
first two are done every time, which is wasteful (for instance, we repeatedly test for the emptiness of \#6). However, this is rare enough not to matter. Finally, choose \((B)\) to be \(\texttt{\textbackslash q\_tl\_nil}\) or \(\texttt{\textbackslash q\_tl\_stop}\) such that it is not equal to \#6.

The \_\_tl_replace_auxi:NnnNNNnn auxiliary receives \{\(A\)\} and \{\((A)^n(B)\)\} as its arguments, initially with \(n = 1\). If “\((A)\ (A)^n(B)\ (A)^n(B)\)” is in the \{token list\} then increase \(n\) and try again. Once it is no more in the \{token list\} we take it as our \{delimiter\} and pass this to the auxii auxiliary.

\begin{verbatim}
\cs_new_protected:Npn \_\_tl_replace:NnNNNnn #1#2#3#4#5#6#7 #3#4#5#6#7
  \tl_if_empty:nTF {#6}
    { \msg_error:nnx { kernel } { empty-search-pattern } #7 }
    { \tl_if_in:onTF { #5 #6 } #1
      \tl_if_in:nnTF {#6} #1
        \tl_if_in:nTF {#6}
          \_\_tl_replace_auxi:NnnNNNnn \#5 \#1 \#1 \texttt{\textbackslash q\_tl\_stop}
        \_\_tl_replace_auxi:NnnNNNnn \#5 \#1 \texttt{\textbackslash q\_tl\_nil}
      \__tl_quark_if_nil:nTF {#6}
        \_\_tl_replace_auxi:NnnNNNnn #5 #1 \#1 \texttt{\textbackslash q\_tl\_nil}
      \__tl_replace_auxii:nNNNnn #1
    }

\__tl_replace_auxii:nNNNnn #1 #1 #1 #1 #1
  \__tl_replace_next:w \{ \__tl_replace_wrap:w \}
  \__tl_replace_next:w \{ \__tl_replace_wrap:w \}
\end{verbatim}

The auxiliary \_\_tl_replace_auxi:NnnNNNnn receives the following arguments:

\begin{verbatim}
\((\text{delimiter})\) \{\(\text{function}\) \{\(\text{assignment}\) \{\(\text{replacement}\)\}}
\end{verbatim}

All of its work is done between \texttt{\textbackslash group\_align\_safe\_begin:} and \texttt{\textbackslash group\_align\_safe\_end:} to avoid issues in alignments. It does the actual replacement within \#3 \#4 \ldots, an x-expanding \{assignment\} \#3 to the \{tl var\} \#4. The auxiliary \_\_tl_replace_next:w is called, followed by the \{token list\}, some tokens including the \{delimiter\} \#1, followed by the \{pattern\} \#5. This auxiliary finds an argument delimited by \#5 (the presence of a trailing \#5 avoids runaway arguments) and calls \_\_tl_replace_wrap:w to test whether this \#5 is found within the \{token list\} or is the trailing one.

If on the one hand it is found within the \{token list\}, then \#1 cannot contain the \{delimiter\} \#1 that we worked so hard to obtain, thus \_\_tl_replace_wrap:w gets \#1 as its own argument \#1, and protects it against the x-expanding assignment. It also finds \texttt{\textbackslash exp\_not:n} as \#2 and does nothing to it, thus letting through \texttt{\textbackslash exp\_not:n \{\(\text{replacement}\)\}} into the assignment. Note that \_\_tl_replace_next:w and \_\_tl_replace_wrap:w are always called followed by two empty brace groups. These are safe
because no delimiter can match them. They prevent losing braces when grabbing delimited arguments, but require the use of \texttt{\exp_not:o} and \texttt{\use_none:nn}, rather than simply \texttt{\exp_not:n}. Afterwards, \texttt{\_tl_replace_next:w} is called to repeat the replacement, or \texttt{\_tl_replace_wrap:w} if we only want a single replacement. In this second case, \#1 is the (remaining tokens) in the (token list) and \#2 is some (ending code) which ends the assignment and removes the trailing tokens \#5 using some \texttt{\if_false: { \fi: }} trickery because \#5 may contain any delimiter.

If on the other hand the argument \#1 of \texttt{\_tl_replace_next:w} is delimited by the trailing \texttt{(pattern)} \#5, then \#1 is \texttt{\{ \} \langle \text{token list} \rangle \langle \text{delimiter} \rangle \{ \langle \text{ending code} \rangle \}}", hence \texttt{\_tl_replace_wrap:w} finds \texttt{\{ \} \langle \text{token list} \rangle \langle \text{delimiter} \rangle \}} as \#1 and the \texttt{(ending code)} as \#2. It leaves the \texttt{(token list)} into the assignment and unbraces the \texttt{(ending code)} which removes what remains (essentially the \texttt{(delimiter)} and \texttt{(replacement)}).

\begin{verbatim}
\cs_new_protected:Npn \_tl_replace_auxii:nNNNnn #1#2#3#4#5#6
\group_align_safe_begin:
\cs_set:Npn \_tl_replace_wrap:w \#1 \#1 \#2
\group_after:wN \_tl_replace_wrap:ww \#1 #1
\exp_not:n \{ \#1 \}
\exp_not:n { \exp_not:n \{#6 \} }
\exp_not:n { \#2 \{ } \} \}
\}
\group_align_safe_end:
\cs_set:Npx \_tl_replace_next:w \#1 \#5
\exp_not:N \_tl_replace_wrap:w \#1
\exp_not:n \{ \#1 \}
\exp_not:n { \exp_not:n \{#6 \} }
\exp_not:n { \#2 \{ } \} \}
\}
\cs_new:Npn \_tl_replace_next_aux:w \{} \}
\cs_new_eq:NN \_tl_replace_wrap:w ?
\cs_new_eq:NN \_tl_replace_next:w ?
\end{verbatim}

(End definition for \texttt{\_tl_replace:NnNNNnn} and others.)

\begin{verbatim}
\cs_new_protected:Npn \tl_remove_once:Nn #1#2
\cs_set:Nn \_tl_replace_next:ww #1 \{ \}
\cs_set:Npx \_tl_replace_next:ww \#1 #1
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \tl_gremove_once:Nn #1#2
\cs_set:Nn \_tl_replace_next:ww #1 \{ \}
\cs_set:Npx \_tl_replace_next:ww \#1 #1
\end{verbatim}

\begin{verbatim}
\cs_generate_variant:Nn \_tl_remove_once:ww { c }
\cs_generate_variant:Nn \_tl_gremove_once:ww { c }
\end{verbatim}

(End definition for \texttt{\_tl_remove_once:Nn} and \texttt{\_tl_gremove_once:Nn}. These functions are documented on page 116.)
Removal is just a special case of replacement.

\begin{verbatim}
12273 \cs_new_protected:Npn \tl_remove_all:Nn #1#2
12274 { \tl_replace_all:Nnn #1 {#2} { } }
12275 \cs_new_protected:Npn \tl_gremove_all:Nn #1#2
12276 { \tl_greplace_all:Nnn #1 {#2} { } }
12277 \cs_generate_variant:Nn \tl_remove_all:Nn { c }
12278 \cs_generate_variant:Nn \tl_gremove_all:Nn { c }
\end{verbatim}

(End definition for \tl_remove_all:Nn and \tl_gremove_all:Nn. These functions are documented on page 117.)

\section{Token list conditionals}

These functions check whether the token list in the argument is empty and execute the proper code from their argument(s).

\begin{verbatim}
12279 \prg_new_conditional:Npnn \tl_if_empty:N #1 { p , T , F , TF }
12280 { \if_meaning:w #1 \c_empty_tl \prg_return_true: \else: \prg_return_false: \fi: }
12281 \prg_generate_conditional_variant:Nnn \tl_if_empty:N { c } { p , T , F , TF }
\end{verbatim}

(End definition for \tl_if_empty:NTF. This function is documented on page 104.)

The \texttt{if:w} triggers the expansion of \texttt{tl_to_str:n} which converts the argument to a string: this is empty if and only if the argument is. Then \texttt{if:w} \texttt{\scan_stop: ... \scan_stop:} is true if and only if the string \ldots{} is empty. It could be tempting to use \texttt{if:w} \texttt{\scan_stop: \#1 \scan_stop:} directly. But this fails on a token list expanding to anything starting with \texttt{\scan_stop:} leaving everything that follows in the input stream.

\begin{verbatim}
12289 \prg_new_conditional:Npnn \tl_if_empty:n #1 { p , TF , T , F }
12290 { \if:w \scan_stop: \tl_to_str:n {#1} \scan_stop: \prg_return_true: \else: \prg_return_false: \fi: }
12291 \prg_generate_conditional_variant:Nnn \tl_if_empty:n { V } { p , TF , T , F }
\end{verbatim}

(End definition for \tl_if_empty:nTF. This function is documented on page 104.)

The auxiliary function \texttt{\_tl_if_empty_if:o} is for use in various token list conditionals which reduce to testing if a given token list is empty after applying a simple function to it. The test for emptiness is based on \texttt{tl_if_empty:nTF}, but the expansion is hard-coded for efficiency, as this auxiliary function is used in several places. We don’t put \texttt{\prg_return_true:} and so on in the definition of the auxiliary, because that would prevent an optimization applied to conditionals that end with this code. Also the
\_\_if\_empty\_if:o is expanded once in \_\_if\_empty:oTF for efficiency as well (and to reduce code doubling).

\begin{verbatim}
\cs_new:Npn \_\_if\_empty\_if:o #1
{ \if:w \scan_stop: \_\_kernel_tl_to_str:w \exp_after:wN \{#1\} \scan_stop: }
\exp_args:Nno \use:n
{ \prg_new_conditional:Npnn \_\_if\_empty\_if:o #1 { p , TF , T , F } }
{ \__tl_if_empty_if:o {#1} \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}

(TeX skips spaces when reading a non-delimited arguments. Thus, a (token list) is blank if and only if \use_none:n \langle token list \rangle \textcolor{red}{?} is empty after one expansion. The auxiliary \_\_if\_empty\_if:o is a fast emptiness test, converting its argument to a string (after one expansion) and using the test \if:w \scan_stop: ... \scan_stop:.

\begin{verbatim}
\exp_args:Nno \use:n
{ \prg_new_conditional:Npnn \_\_if\_empty\_if:o #1 { p , TF , T , F } }
{ \__tl_if_empty_if:o { \use_none:n #1 ? } \prg_return_true: \else: \prg_return_false: \fi: }
\prg_generate_conditional_variant:Nnn \_\_if\_empty\_if:o { e , V , o } { p , TF , T , F }
\end{verbatim}

(End definition for \_\_if\_empty:oTF and \_\_if\_empty_if:o. This function is documented on page 104.)

\begin{verbatim}
\_\_if\_blank\_p:n
\_\_if\_blank\_p:V
\_\_if\_blank\_p:o
\_\_if\_blank:nTF
\_\_if\_blank:VTF
\_\_if\_blank:OTF
\_\_if\_blank:p:NNw
\end{verbatim}

\_\_if\_blank\_p:NNw
\_\_if\_blank\_p:V
\_\_if\_blank\_p:o
\_\_if\_blank\_p:V
\_\_if\_blank\_p:OTF
\_\_if\_blank\_p:NNw
\_\_if\_blank\_p:NNw

\_\_if\_eq\_p:NN
\_\_if\_eq\_p:Nc
\_\_if\_eq\_p:cN
\_\_if\_eq\_p:cc
\_\_if\_eq\_p:NN
\_\_if\_eq\_p:NN
\_\_if\_eq\_p:Nc
\_\_if\_eq\_p:cN
\_\_if\_eq\_p:cc
\_\_if\_eq\_p:cc

\begin{verbatim}
\tl_new:N \l__tl_internal_a_tl
\tl_new:N \l__tl_internal_b_tl
\end{verbatim}

Temporary storage.
(End definition for \l__tl_internal_a_tl and \l__tl_internal_b_tl.)

\tl_if_eq:NnTF
A simple store and compare routine.
\begin{verbatim}
\prg_new_protected_conditional:Nppn \tl_if_eq:Nn #1#2 { T , F , TF }
\{
\group_begin:
\tl_set:Nn \l__tl_internal_b_tl {#2}
\exp_after:wN
\group_end:
\if_meaning:w #1 \l__tl_internal_b_tl
\prg_return_true:
\else:
\prg_return_false:
\fi:
\}
\prg_generate_conditional_variant:Nnn \tl_if_eq:Nn { c } { TF , T , F }
\end{verbatim}

(End definition for \tl_if_eq:nnTF. This function is documented on page 104.)

\tl_if_eq:nnTF
A simple store and compare routine.
\begin{verbatim}
\prg_new_protected_conditional:Nppn \tl_if_eq:nn #1#2 { T , F , TF }
\{
\group_begin:
\tl_set:Nn \l__tl_internal_a_tl {#1}
\tl_set:Nn \l__tl_internal_b_tl {#2}
\exp_after:wN
\group_end:
\if_meaning:w \l__tl_internal_a_tl \l__tl_internal_b_tl
\prg_return_true:
\else:
\prg_return_false:
\fi:
\}
\prg_generate_conditional_variant:Nnn \tl_if_eq:nn { c } { TF , T , F }
\end{verbatim}

(End definition for \tl_if_eq:nnTF. This function is documented on page 105.)

\tl_if_in:NnTF \tl_if_in:cnTF
\tl_if_in:nnTF \tl_if_in:onTF
\tl_if_in:noTF
See \tl_if_in:nnTF for further comments. Here we simply expand the token list variable and pass it to \tl_if_in:nnTF.
\begin{verbatim}
\cs_new_protected:Npn \tl_if_in:NnT { \exp_args:No \tl_if_in:nnT }
\cs_new_protected:Npn \tl_if_in:NnF { \exp_args:No \tl_if_in:nnF }
\cs_new_protected:Npn \tl_if_in:NnTF { \exp_args:No \tl_if_in:nnTF }
\prg_generate_conditional_variant:Nnn \tl_if_in:Nn { c } { T , F , TF }
\end{verbatim}

(End definition for \tl_if_in:nnTF. This function is documented on page 105.)

\tl_if_in:nnTF \tl_if_in:VnTF \tl_if_in:onTF \tl_if_in:noTF
Once more, the test relies on the emptiness test for robustness. The function \__tl__tmp:w removes tokens until the first occurrence of \#2. If this does not appear in \#1, then the final \#2 is removed, leaving an empty token list. Otherwise some tokens remain, and the test is false. See \tl_if_empty:nTF for details on the emptiness test.

Treating correctly cases like \tl_if_in:nnTF {a state}{states}, where \#1\#2 contains \#2 before the end, requires special care. To cater for this case, we insert {} before the two token lists. This marker may not appear in \#2 because of \TeXX limitations on what can delimit a parameter, hence we are safe. Using two brace groups makes the
test work also for empty arguments. The \if_false:\ constructions are a faster way to do
\group_align_safe_begin: and \group_align_safe_end:. The \scan_stop: ensures
that f-expanding \tl_if_in:nTF does not lead to unbalanced braces.
\begin{verbatim}
\prg_new_protected_conditional:Npnn \tl_if_in:nn \tl_if_in:nn #1#2 \{ T , F , TF \}
 \{ \scan_stop:
 \if_false: { \fi:
 \cs_set:Npn \__tl_tmp:w ##1 #2 { }
 \tl_if_empty:oTF { \__tl_tmp:w #1 {} {} #2 }
 \{ \prg_return_false: \} \{ \prg_return_true: 
 \if_false: \} \fi:
 \}
 \prg_generate_conditional_variant:Nnn \tl_if_in:nn \tl_if_in:nn \tl_if_in:nn \tl_if_in:nn \tl_if_in:nn \tl_if_in:nn \tl_if_in:nn \tl_if_in:nn \tl_if_in:nn (\End definition for \tl_if_in:nTF. This function is documented on page 105.)
\end{verbatim}

\tl_if_novalue_p:n\tl_if_novalue:nTF\tl_if_novalue:w Tests whether ##1 matches \-NoValue- exactly (with suitable catcodes): this is similar
to \quark_if_nil:nTF. The first argument of \__tl_if_novalue:w is empty if and
only if ##1 starts with \-NoValue-, while the second argument is empty if ##1 is exactly
\-NoValue- or if it has a question mark just following \-NoValue-. In this second case,
however, the material after the first ?! remains and makes the emptiness test return
false.
\begin{verbatim}
\cs_set_protected:Npn \__tl_if_novalue:w ##1 #1 ##2 ? ##3 ? ! { ##1 ##2 }
\end{verbatim}
(End definition for \tl_if_novalue:nTF and \__tl_if_novalue:w. This function is documented on page 105.)

\tl_if_single_p:N\tl_if_single:NTFExpand the token list and feed it to \tl_if_single:nTF.
\begin{verbatim}
\cs_new:Npn \tl_if_single_p:N { \exp_args:No \tl_if_single_p:n }
\cs_new:Npn \tl_if_single:NT { \exp_args:No \tl_if_single:nT }
\cs_new:Npn \tl_if_single:NF { \exp_args:No \tl_if_single:nF }
\cs_new:Npn \tl_if_single:NTF { \exp_args:No \tl_if_single:nTF }
\end{verbatim}
(End definition for \tl_if_single:NTF. This function is documented on page 105.)

\tl_if_single_p:n\tl_if_single:nTF\tl_if_single:nw This test is similar to \tl_if_empty:nTF. Expanding \use_none:nn \#1 ?? once yields
an empty result if \#1 is blank, a single ? if \#1 has a single item, and otherwise yields
some tokens ending with ???. Then, \__kernel_tl_to_str:w makes sure there are no
odd category codes. An earlier version would compare the result to a single ? using string
comparison, but the Lua call is slow in Lua\TeX. Instead, \__tl_if_single:nw picks
the second token in front of it. If \#1 is empty, this token is the trailing \texttt{?} and the \texttt{if:w} test yields \texttt{false}. If \#1 has a single item, the token is \texttt{\scan_stop:} and the \texttt{if:w} test yields \texttt{true}. Otherwise, it is one of the characters resulting from \texttt{\tl_to_str:}, and the \texttt{if:w} test yields \texttt{false}. Note that \texttt{if:w} and \texttt{\__kernel_tl_to_str:w} are primitives that take care of expansion.

\begin{verbatim}
\prg_new_conditional:Npn \tl_if_single:n #1 { p , T , F , TF }
    { \if:w \scan_stop: \exp_after:wN \__tl_if_single:nnw
        \__kernel_tl_to_str:w
        \exp_after:wN { \use_none:nn #1 ?? } \scan_stop: ? \s__tl_stop
        \prg_return_true:
    \else:
        \prg_return_false:
    \fi: }
\cs_new:Npn \__tl_if_single:nnw #1#2#3 \s__tl_stop {#2}
\end{verbatim}

(End definition for \texttt{\tl_if_single:nTF} and \texttt{\__tl_if_single:nnw}. This function is documented on page 105.)

\tl_if_single_token_p:n\tl_if_single_token:nTF

There are four cases: empty token list, token list starting with a normal token, with a brace group, or with a space token. If the token list starts with a normal token, remove it and check for emptiness. For the next case, an empty token list is not a single token. Finally, we have a non-empty token list starting with a space or a brace group. Applying \texttt{f}\texttt{-expansion} yields an empty result if and only if the token list is a single space.

\begin{verbatim}
\prg_new_conditional:Npn \tl_if_single:n #1 { p , T , F , TF }
    { \tl_if_head_is_N_type:nTF {#1}
        { \__tl_if_empty_if:o { \use_none:n #1 } }
        { \tl_if_empty:nTF {#1}
            { \__tl_if_empty_if:o { \exp:w \exp_end_continue_f:w #1 } }
            \prg_return_true:
        }
        \else:
            \prg_return_false:
        \fi: }
\end{verbatim}

(End definition for \texttt{\tl_if_single_token:nTF}. This function is documented on page 105.)

\tl_case:Nn\tl_case:cn\tl_case:NnTF\tl_case:cnTF\tl_case:NnTF\tl_case:cnTF\tl_case:Nn\tl_case:cn
\tl_case_end:nw

The aim here is to allow the case statement to be evaluated using a known number of expansion steps (two), and without needing to use an explicit “end of recursion” marker. That is achieved by using the test input as the final case, as this is always true. The trick is then to tidy up the output such that the appropriate case code plus either the \texttt{true} or \texttt{false} branch code is inserted.

\begin{verbatim}
\cs_new:Npn \tl_case:Nn \tl_case:Nn #1#2
    { \exp:w \__tl_case:NnTF #1 {#2} { } { } }
\cs_new:Npn \tl_case:NnT #1#2#3
\end{verbatim}

(End definition for \texttt{\tl_case:NnTF}. This function is documented on page 105.)
To tidy up the recursion, there are two outcomes. If there was a hit to one of the cases searched for, then \#1 is the code to insert, \#2 is the next case to check on and \#3 is all of the rest of the cases code. That means that \#4 is the true branch code, and \#5 tidies up the spare \s__tl_mark and the false branch. On the other hand, if none of the cases matched then we arrive here using the “termination” case of comparing the search with itself. That means that \#1 is empty, \#2 is the first \s__tl_mark and so \#4 is the false code (the true code is mopped up by \#3).

End definition for \tl_case:NnTF and others. This function is documented on page 106.

52.8 Mapping over token lists

Expandable loop macro for token lists. We use the internal scan mark \s__tl_stop (defined later), which is not allowed to show up in the token list \#1 since it is internal to l3tl. This allows us a very fast test of whether some \item is the end-marker \s__tl_stop, namely call \_tl_use_none_delimit_by_s_stop:w \item (function) \s__tl_stop, which calls (function) if the \item is the end-marker. To speed up the loop even more, only test one out of eight items, and once we hit one of the eight end-markers, go more slowly through the last few items of the list using \_tl_map_function_end:w.
The inline functions are straightforward by now. We use a little trick with the counter \g__kernel_prg_map_int to make them nestable. We can also make use of \__tl_map\_function:Nnnnnnnnnn from before.

\begin{verbatim}
\cs_new_protected:Npn \tl_map_inline:nn #1#2
  { \int_gincr:N \g__kernel_prg_map_int
    \cs_gset_protected:cpn
      { __tl_map_ \int_use:N \g__kernel_prg_map_int :w } ##1 {#2}
    \exp_args:Nc \__tl_map_function:Nnnnnnnnnn
      { __tl_map_ \int_use:N \g__kernel_prg_map_int :w }
      #1
    \s__tl_stop \s__tl_stop \s__tl_stop \s__tl_stop
    \prg_break_point:Nn \tl_map_break: { }\int_gdecr:N \g__kernel_prg_map_int }
\end{verbatim}

(End definition for \tl_map_inline:nn and others. These functions are documented on page 110.)

The inline functions are straight forward by now. We use a little trick with the counter \g__kernel_prg_map_int to make them nestable. We can also make use of \__tl_map\_function:Nnnnnnnnnn from before.

\begin{verbatim}
\cs_new_protected:Npn \tl_map_tokens:nn #1#2
  { \__tl_map_tokens:nnnnnnnnn {#2} #1
    \s__tl_stop \s__tl_stop \s__tl_stop \s__tl_stop
    \prg_break_point:Nn \tl_map_break: { }\int_gdecr:N \g__kernel_prg_map_int }
\end{verbatim}

(End definition for \tl_map_inline:nn and \tl_map_inline:Nn. These functions are documented on page 110.)

Much like the function mapping.

\begin{verbatim}
\cs_new:Npn \tl_map_tokens:Nn { \exp_args:No \tl_map_tokens:nn }
\cs_generate_variant:Nn \tl_map_tokens:Nn { c }
\end{verbatim}

(End definition for \tl_map_inline:nn and \tl_map_inline:Nn. These functions are documented on page 110.)
\tl_map_variable:nNn  \tl_map_variable:nNn \{ \langle token list \rangle \} \langle \langle tl var \rangle \{ \langle action \rangle \} \} assigns \langle \langle tl var \rangle \} to each element and executes \langle \langle action \rangle \}. The assignment to \langle \langle tl var \rangle \} is done after the quark test so that this variable does not get set to a quark.

\tl_map_variable:NNn  \tl_map_variable:NNn \tl_map_variable:cNn  \tl_map_variable:Nnn
(End definition for \tl_map_variable:nNn and others. These functions are documented on page 111.)

\tl_map_break:  \tl_map_break:n
The break statements use the general \prg_map_break:Nn.

\tl_to_str:n  \tl_to_str:V  \tl_to_str:c
Another name for a primitive: defined in l3basics.  These functions return the replacement text of a token list as a string.

52.9 Using token lists

(End definition for \tl_to_str:n. This function is documented on page 107.)
Token lists which are simply not defined give a clear \TeX error here. No such luck for ones equal to \texttt{\textbackslash scan\_stop}: so instead a test is made and if there is an issue an error is forced.

\begin{verbatim}
cs_new:Npn \tl_use:N #1
{ \tl_if_exist:NTF #1 {#1}
{ \msg_expandable_error:nnn { kernel } { bad-variable } {#1} }
}
cs_generate_variant:Nn \tl_use:N { c }
\end{verbatim}

(End definition for \texttt{\tl_use:N}. This function is documented on page 108.)

\section{Working with the contents of token lists}

\begin{verbatim}
cs_new:Npn \tl_count:n #1
{ \int_eval:n { 0 \tl_map_function:nN {#1} \__tl_count:n } }
cs_new:Npn \tl_count:N #1
{ \int_eval:n { 0 \tl_map_function:NN #1 \__tl_count:n } }
cs_new:Npn \__tl_count:n #1 { + 1 }
cs_generate_variant:Nn \tl_count:n { V , o }
cs_generate_variant:Nn \tl_count:N { c }
\end{verbatim}

(End definition for \texttt{\tl_count:n}, \texttt{\tl_count:N}, and \texttt{\__tl_count:n}. These functions are documented on page 108.)

\begin{verbatim}
cs_new:Npn \tl_count_tokens:n #1
{ \int_eval:n
{ 0 \tl_map_function:nN \__tl_count_normal:N \__tl_count_group:n \__tl_count_space:n {#1} }
}
cs_new:Npn \__tl_count_tokens:n #1
{ \int_eval:n
{ 0 \tl_map_function:nN \__tl_count_normal:N \__tl_count_group:n \__tl_count_space:n {#1} }
}
cs_generate_variant:Nn \tl_count_tokens:n { V , o }
cs_generate_variant:Nn \tl_count_tokens:n { c }
\end{verbatim}

(The token count is computed through an \texttt{\int_eval:n} construction. Each \texttt{1+} is output to the left, into the integer expression, and the sum is ended by the \texttt{\exp_end:} inserted by \texttt{\__tl_act_end:wn} (which is technically implemented as \texttt{\c_zero_int}). Somewhat a hack!)

\begin{verbatim}
cs_new:Npn \__tl_count_tokens:n #1
{ \int_eval:n
{ \__tl_act:NNNn \__tl_act_count_normal:n \__tl_act_count_group:n \__tl_act_count_space:n {#1} }
}
cs_new:Npn \__tl_count_tokens:n #1
{ \int_eval:n
{ \__tl_act:NNNn \__tl_act_count_normal:n \__tl_act_count_group:n \__tl_act_count_space:n {#1} }
}
cs_generate_variant:Nn \tl_count_tokens:n { V , o }
cs_generate_variant:Nn \tl_count_tokens:n { c }
\end{verbatim}

(End definition for \texttt{\tl_count_tokens:n}, \texttt{\tl_count_tokens:n}, and \texttt{\__tl_count_tokens:n}.)

686
Reversal of a token list is done by taking one item at a time and putting it after \s__-tl_stop.

Trimming spaces from around the input is deferred to an internal function whose first argument is the token list to trim, augmented by an initial \__tl_trim_mark:␣, and whose second argument is a ⟨continuation⟩, which receives as a braced argument \__tl_trim-_mark:␣⟨trimmed token list⟩. The control sequence \__tl_trim_mark: expands to nothing in a single expansion. In the case at hand, we take \__kernel_exp_not:w \exp_after:wN \\use_none:nn #2 \} \}

Trimming spaces from around the input is done using delimited arguments and quarks, and to get spaces at odd places in the definitions, we nest those in \__tl_tmp:w, which then receives a single space as its argument: \\ #1 is \ . Removing leading spaces is done with \__tl_trim_spaces_auxi:w, which loops until \__tl_trim_mark:\ matches the
end of the token list: then #1 is the token list and #3 is \__tl_trim_spaces_auxii:w. This hands the relevant tokens to the loop \__tl_trim_spaces_auxiii:w, responsible for trimming trailing spaces. The end is reached when \s__tl_nil matches the one present in the definition of \tl_trim_spaces:n. Then \__tl_trim_spaces_auxiv:w puts the token list into a group, with a lingering \__tl_trim_mark: at the start (which will expand to nothing in one step of expansion), and feeds this to the (continuation).

\begin{verbatim}
c_set_protected:Npn \__tl_tmp:w #1
c_new:Npn \__tl_trim_spaces:nn ##1
c_new:Npn \__tl_trim_spaces_auxi:w ##1 \__tl_trim_mark: #1 ##2 \__tl_trim_mark: ##3
c_new:Npn \__tl_trim_spaces_auxii:w \__tl_trim_spaces_auxi:w \__tl_trim_mark: \s__tl_nil \__tl_trim_spaces_auxiii:w
c_new:Npn \__tl_trim_spaces_auxiii:w \__tl_trim_spaces_auxii:w \__tl_trim_spaces_auxi:w \__tl_trim_mark: \s__tl_nil \s__tl_nil
c_new:Npn \__tl_trim_spaces_auxiv:w \__tl_trim_spaces_auxi:w \__tl_trim_spaces_auxii:w \__tl_trim_spaces_auxiii:w \s__tl_stop
c_new:Npn \__tl_trim_mark: {}
\end{verbatim}

(End definition for \tl_trim_spaces:n and others. These functions are documented on page 109.)

\begin{verbatim}
\tl_sort:Nn
\tl_sort:cn
\tl_gsort:Nn
\tl_gsort:cn
\tl_sort:nN
\end{verbatim}

Implemented in \texttt{l3sort}.

(End definition for \tl_sort:Nn, \tl_gsort:Nn, and \tl_sort:nN. These functions are documented on page 116.)
52.11 The first token from a token list

Finding the head of a token list expandably always strips braces, which is fine as this is consistent with for example mapping over a list. The empty brace groups in `\tl_head:n` ensure that an empty argument gives an empty result. The result is returned within the `unexpanded` primitive. The approach here is to use `\if_false:` to allow us to use `{}` as the closing delimiter: this is the only safe choice, as any other token would not be able to parse its own code. If the `expanded` primitive is available it is used to get a fast and safe code variant in which we don’t have to ensure that the left-most token is an internal to not break in an f-type expansion. If `expanded` isn’t available, using a marker, we can see if what we are grabbing is exactly the marker, or there is anything else to deal with. If there is, there is a loop. If not, tidy up and leave the item in the output stream. More detail in http://tex.stackexchange.com/a/70168.

To correctly leave the tail of a token list, it’s important not to absorb any of the tail part
as an argument. For example, the simple definition

\cs_new:Npn \tl_tail:n #1 \{ \tl_tail:w #1 \q_stop \}
\cs_new:Npn \tl_tail:w #1#2 \q_stop

would give the wrong result for \tl_tail:n \{ a \{ bc \} \} (the braces would be stripped). Thus the only safe way to proceed is to first check that there is an item to grab (i.e. that the argument is not blank) and assuming there is to dispose of the first item. As with \tl_head:n, the result is protected from further expansion by \unexpanded. While we could optimise the test here, this would leave some tokens “banned” in the input, which we do not have with this definition.

\exp_args:Nno \use:n \{ \cs_new:Npn \tl_tail:n #1 \}
\exp_after:wN \__kernel_exp_not:w
\tl_if_blank:nTF {#1}
{ { } }
{ \exp_after:wN { \use_none:n #1 } }
\cs_generate_variant:Nn \tl_tail:n { V , v , f }
\cs_new:Npn \tl_tail:N { \exp_args:No \tl_tail:n }

(End definition for \tl_head:N and others. These functions are documented on page 112.)

Accessing the first token of a token list is tricky in three cases: when it has category code 1 (begin-group token), when it is an explicit space, with category code 10 and character code 32, or when the token list is empty (obviously).

Forgetting temporarily about this issue we would use the following test in \tl_if_head_eq_charcode:nN. Here, \tl_head:w yields the first token of the token list, then passed to \exp_not:N.\if_charcode:w
\exp_after:wN \exp_not:N \tl_head:w #1 \q_nil \q_stop \exp_not:N #2
\prg_return_true:
\else:
\prg_return_false:
\fi:
\prg_generate_conditional_variant:Nnn \tl_if_head_eq_charcode:nN
{ f } { p , T , F , TF }

\tl_if_head_eq_p:nN \tl_if_head_eq_p:Nn \tl_if_head_eq_p:nTF \tl_if_head_eq_p:NN \tl_if_head_eq_p:NFF \tl_if_head_eq_p:NNF \tl_if_head_eq_p:NNNF \tl_if_head_eq_p:NNN \tl_if_head_eq_p:NFFN \tl_if_head_eq_p:NFFF \tl_if_head_eq_p:NFFF
\tl_if_head_eq_end:w \tl_if_head_eq_end:pp \tl_if_head_eq_end:PN \tl_if_head_eq_end:P \tl_if_head_eq_end:PNF \tl_if_head_eq_end:PF \tl_if_head_eq_end:PNFF \tl_if_head_eq_end:PNFFF
\tl_if_head_eq_end:NN \tl_if_head_eq_end:NNF \tl_if_head_eq_end:NNNF \tl_if_head_eq_end:NFF \tl_if_head_eq_end:NFFF \tl_if_head_eq_end:NFFF
\tl_if_head_eq_end:N \tl_if_head_eq_end:NF \tl_if_head_eq_end:NNF \tl_if_head_eq_end:NNNF
\tl_if_head_eq_end:NNNF \tl_if_head_eq_end:NNFF \tl_if_head_eq_end:NNFFF
\tl_if_head_eq_end:NNN \tl_if_head_eq_end:NNFFN \tl_if_head_eq_end:NNFFFN
\tl_if_head_eq_end:NNNN \tl_if_head_eq_end:NNNFF \tl_if_head_eq_end:NNFFF \tl_if_head_eq_end:NNNNF
\tl_if_head_eq_end:NNNNF \tl_if_head_eq_end:NNNNNF \tl_if_head_eq_end:NNNNFF 
\tl_if_head_eq_end:NNNNFFN \tl_if_head_eq_end:NNNNFFF
\tl_if_head_eq_end:NNNNN \tl_if_head_eq_end:NNNNNF \tl_if_head_eq_end:NNNNFF 
\tl_if_head_eq_end:NNNNFFN \tl_if_head_eq_end:NNNNFFF
\tl_if_head_eq_end:NNNNNN \tl_if_head_eq_end:NNNNFFN \tl_if_head_eq_end:NNNNFFF 
\tl_if_head_eq_end:NNNNFFNN \tl_if_head_eq_end:NNNNFFFN
\tl_if_head_eq_end:NNNNNNF \tl_if_head_eq_end:NNNNFFFN \tl_if_head_eq_end:NNNNNNNF
\tl_if_head_eq_end:NNNNNNFF \tl_if_head_eq_end:NNNNNNFFF

The two first special cases are detected by testing if the token list starts with an N-type token (the extra ? sends empty token lists to the true branch of this test). In those cases, the first token is a character, and since we only care about its character code, we can use \str_head:n to access it (this works even if it is a space character). An empty argument results in \tl_head:w leaving two token: \^ and \tl_if_head_eq_empty_arg:w which will result in the \if_charcode:w test being false and remove \exp_not:N and #2.
For \texttt{\tl_if_head_eq_catcode:nN}, again we detect special cases with a \texttt{\tl_if_head_is_N_type:n}. Then we need to test if the first token is a begin-group token or an explicit space token, and produce the relevant token, either \texttt{\c_group_begin_token} or \texttt{\c_space_token}. Again, for an empty argument, a hack is used, removing the token given by the user and leaving two tokens in the input stream which will make the \texttt{\if_catcode:w} test return false.

\begin{verbatim}
12701 \prg_new_conditional:Npn \tl_if_head_eq_catcode:nN #1 #2 { p , T , F , TF }
12702 { \if_catcode:w
12703 \tl_if_head_is_N_type:nTF { #1 ? }
12704 { \tl_head:w #1 { ^ \tl_if_head_eq_empty_arg:w } \tl_stop }
12705 { \tl_if_head_is_group:nTF {#1}
12706 \c_group_begin_token
12707 \c_space_token
12708 }
12709 \exp_not:N #2
12710 \prg_return_true:
12711 \else:
12712 \prg_return_false:
12713 \fi:
12714 }
12715 \prg_generate_conditional_variant:Nnn \tl_if_head_eq_catcode:nN
12716 \o { p , TF , T , F }
\end{verbatim}

For \texttt{\tl_if_head_eq_meaning:nN}, again, detect special cases. In the normal case, use \texttt{\tl_head:w}, with no \texttt{\exp_not:N} this time, since \texttt{\if_meaning:w} causes no expansion. With an empty argument, the test is true, and \texttt{\use_none:nnn} removes \#2 and \texttt{\prg_return_true:} and \texttt{\else:} (it is safe this way here as in this case \texttt{\prg_new_conditional:Npn} didn’t optimize these two away). In the special cases, we know that the first token is a character, hence \texttt{\if_charcode:w} and \texttt{\if_catcode:w} together are enough. We combine them in some order, hopefully faster than the reverse. Tests are not nested because the arguments may contain unmatched primitive conditionals.

\begin{verbatim}
12719 \prg_new_conditional:Npn \tl_if_head_eq_meaning:nN #1#2 { p , T , F , TF }
12720 { \tl_if_head_is_N_type:nTF { #1 ? }
12721 \__tl_if_head_eq_meaning_normal:nN
12722 \__tl_if_head_eq_meaning_special:nN
12723 {#1} #2
12724 }
12725 \cs_new:Npn \__tl_if_head_eq_meaning_normal:nN #1 #2
12726 { \exp_after:wN \if_meaning:w
12727 \__tl_tl_head:w #1 { ?? \use_none:nnn } \tl_stop #2
12728 \prg_return_true:
12729 \else:
12730 \prg_return_false:
12731 \fi:
12732 }
12733 \cs_new:Npn \__tl_if_head_eq_meaning_special:nN #1 #2
12734 { \if_charcode:w \str_head:n {#1} \exp_not:N #2
12735 \exp_after:wN \use_ii:nn
12736 \fi:
12737 }
12738 \end{verbatim}

691
Both \tl_if_head_eq_charcode:nN and \tl_if_head_eq_catcode:nN will need to get the first token of their argument and apply \exp_not:N to it. \__tl_head_exp_not:w does exactly that.

If the argument of \tl_if_head_eq_charcode:nN and \tl_if_head_eq_catcode:nN was empty \__tl_if_head_eq_empty_arg:w will be left in the input stream. This macro has to remove \exp_not:N and the following token from the input stream to make sure no unbalanced if-construct is created and leave tokens there which make the two tests return false.

A token list can be empty, can start with an explicit space character (catcode 10 and charcode 32), can start with a begin-group token (catcode 1), or start with an N-type argument. In the first two cases, the line involving \__tl_if_head_is_N_type_auxi:w produces $f$ (and otherwise nothing). In the third case (begin-group token), the lines involving \token_to_str:N produce a single closing brace. The category code test is thus true exactly in the fourth case, which is what we want. One cannot optimize by moving one of the \scan_stop: to the beginning: if #1 contains primitive conditionals, all of its occurrences must be dealt with before the \if:w tries to skip the true branch of the conditional.
\tl_if_head_is_N_type:nTF \tl_if_head_is_N_type_auxii:n

(End definition for \tl_if_head_is_N_type:nTF and others. This function is documented on page 107.)

\tl_if_head_is_group_p:n \tl_if_head_is_group:nTF \__tl_if_head_is_group_fi_false:w

Pass the first token of #1 through \token_to_str:N, then check for the brace balance. The extra ? caters for an empty argument. This could be made faster, but we need all brace tricks to happen in one step of expansion, keeping the token list brace balanced at all times.

\tl_if_head_is_space_p:n \tl_if_head_is_space:nTF \__tl_if_head_is_space:w

The auxiliary’s argument is all that is before the first explicit space in \prg_do_nothing:#1?. If that is a single \prg_do_nothing: the test yields true. Otherwise, that is more than one token, and the test yields false. The work is done within braces (with an \if: \if: \__tl_if_head_is_space:w \prg_do_nothing: #1 ?) construction) both to hide potential alignment tab characters from \TeX{} in a table, and to allow for removing what remains of the token list after its first space. The use of \if:w ensures that the result of a single step of expansion directly yields a balanced token list (no trailing closing brace).

693
52.12 Token by token changes

The \_\_tl\_act\_... functions may be applied to any token list. Hence, we use a private quark, to allow any token, even quarks, in the token list. Only \s\_\_tl\_act\_stop may not appear in the token lists manipulated by \_\_tl\_act\_:NNN functions.

To help control the expansion, \_\_tl\_act\_:NNN should always be preceded by \exp:w and ends by producing \exp_end: once the result has been obtained. This way no internal token of it can be accidentally end up in the input stream. Because \s\_\_tl\_act\_stop can’t appear without braces around it in the argument #1 of \_\_tl\_act\_loop:w, we can use this marker to set up a fast test for leading spaces.

\exp_args:Nnx \use:n { \cs_new:Npn \_\_tl\_act\_loop:w #1 \s\_\_tl\_act\_stop }

\exp_not:o { \__tl_act_if_head_is_space:nTF {#1} }
\exp_not:N \__tl_act_space:wwNNN
\exp_not:N \__tl_act_normal:NwNNN
\exp_not:N \__tl_act_group:nwNNN
\exp_not:N \__tl_act_end:wn
\exp_not:o { \tl_if_head_is_space:nTF \__tl_act_stop \__tl_act_if_head_is_space_true:w \__tl_act_stop #1 }
\__tl_act:NNNn loops over tokens, groups, and spaces in #4. \exp_after:wN serves as the end of token list marker, the ? after it avoids losing outer braces. The result is stored as an argument for the dummy function \__tl_act_result:n.

Typically, the output is done to the right of what was already output, using \__tl_act_output:n, but for the \__tl_act_reverse functions, it should be done to the left.

\tl_reverse:n
\tl_reverse:o
\tl_reverse:V
\__tl_reverse_normal:n
\__tl_reverse_group_preserve:n
\__tl_reverse_space:n

The goal here is to reverse without losing spaces nor braces. This is done using the general internal function \__tl_act:NNNn. Spaces and “normal” tokens are output on the left of the current output. Grouped tokens are output to the left but without any reversal within the group.

(End definition for \__tl_act:NNNn and others.)
\exp:w
\__tl_act:NNNn
\__tl_reverse_normal:N
\__tl_reverse_group_preserve:n
\__tl_reverse_space:
{#1}
\}
\)
\cs_generate VARIANT:NN \tl_reverse:n \{ o , V \}
\cs_new:Npn \__tl_reverse_normal:N
{ \__tl_act_reverse_output:n }
\cs_new:Npn \__tl_reverse_group_preserve:n #1
{ \__tl_act_reverse_output:n \{ (#1) \} }
\cs_new:Npn \__tl_reverse_space:
{ \__tl_act_reverse_output:n \{ ~ \} }

\end definition for \tl_reverse:n and others. This function is documented on page 108.
\tl_reverse:N
\tl_reverse:c
\tl_greverse:N
\tl_greverse:c
\end definition for \tl_reverse:N and \tl_greverse:N. These functions are documented on page 108.

52.13 Using a single item
\tl_item:nn
\tl_item:Nn
\tl_item:cn
\__tl_item_aux:nn
\__tl_item:nn
\__tl_item:nn
The idea here is to find the offset of the item from the left, then use a loop to grab the correct item. If the resulting offset is too large, then \__tl_if_recursion_tail_break:nN terminates the loop, and returns nothing at all.
\cs_new:Npn \tl_item:nn #1#2
{ \exp_args:Nf \__tl_item:nn
{ \exp_args:Nf \__tl_item_aux:nn { \int_eval:n {#2} } {#1} } #1
\int_compare:nNnTF {#1} < 0
{ \int_eval:n { \tl_count:n {#2} + 1 + #1 } } #1
\int_compare:nNnTF {#1} = 1
{ \prg_break:n { \exp_not:n {#2} } } #1
\prg_break_point:
\cs_new:Npn \__tl_item:nn #1#2
{ \exp_args:Nf \__tl_item:nn
\exp_args:Nf \__tl_item_aux:nn { \int_eval:n {#2} } {#1} } #1
\int_compare:nNnTF {#1} < 0
{ \int_eval:n { \tl_count:n {#2} + 1 + #1 } } #1
\int_compare:nNnTF {#1} = 1
{ \prg_break:n { \exp_not:n {#2} } } #1
\prg_break_point:
\begin{verbatim}
{ \exp_args:Nf \__tl_item:nn \{ \int_eval:n \{ #1 - 1 \} \} }
\}
\cs_new:Npn \tl_item:nn \{ \exp_args:No \tl_item:nn \}
\cs_new:generate_variant:Nn \tl_item:nn \{ c \}
(End definition for \tl_item:nn and others. These functions are documented on page 113.)

\tl_rand_item:n
\tl_rand_item:N
\tl_rand_item:c

Importantly \tl_item:nn only evaluates its argument once.
\cs_new:Npn \tl_rand_item:n \#1
{ \tl_if_blank:nF \{#1\}
{ \tl_item:nn \{#1\} \{ \int_rand:nn \{ 1 \} \{ \tl_count:n \{#1\} \} \} }
}
\cs_new:Npn \tl_rand_item:N \{ \exp_args:No \tl_rand_item:n \}
\cs_new:generate_variant:Nn \tl_rand_item:N \{ c \}
(End definition for \tl_rand_item:n and \tl_rand_item:N. These functions are documented on page 114.)

\tl_range:Nnn
\tl_range:cnn
\__tl_range:Nnnn
\__tl_range:nnnNn
\__tl_range:nnNn
\__tl_range_items:nnNn
\__tl_range_skip:w
\__tl_range_collect:nn
\__tl_range_collect:ff
\__tl_range_collect_space:nw
\__tl_range_collect_group:nN
\__tl_range:nnnNn
\_\_tl_range:Nnnn
\_\_tl_range:nnnNn
\_\_tl_range:nnNn
\_\_tl_range_skip:w
\_\_tl_range_skip_spaces:n
\_\_tl_range_collect:nn
\_\_tl_range_collect:ff
\_\_tl_range_collect_space:nw
\_\_tl_range_collect_group:nN

To avoid checking for the end of the token list at every step, start by counting the number \l of items and “normalizing” the bounds, namely clamping them to the interval \[0, l\] and dealing with negative indices. More precisely, \_\_tl_range_items:nnnn receives the number of items to skip at the beginning of the token list, the index of the last item to keep, a function which is either \_\_tl_range:w or the token list itself. If nothing should be kept, leave \{\}: this stops the f-expansion of \tl_head:f and that function produces an empty result. Otherwise, repeatedly call \_\_tl_range_skip:w to delete \#1 items from the input stream (the extra brace group avoids an off-by-one shift). For the braced version \_\_tl_range_braced:w sets up \_\_tl_range_collect_braced:w which stores items one by one in an argument after the semicolon. Depending on the first token of the tail, either just move it (if it is a space) or also decrement the number of items left to find. Eventually, the result is a brace group followed by the rest of the token list, and \tl_head:f cleans up and gives the result in \exp_not:n.
\cs_new:Npn \tl_range:Nnn \{ \exp_args:No \tl_range:nnn \}
\cs_new:generate_variant:Nn \tl_range:Nnn \{ c \}
\cs_new:Npn \tl_range:nnn \{ \__tl_range:Nnnn \__tl_range:w \}
\cs_new:Npn \__tl_range:Nnnn \#1\#2\#3\#4
{ \tl_head:f
{ \exp_args:Nf \__tl_range:nnnNn
\{ \tl_count:n \{#2\} \{#3\} \{#4\} \#1 \{#2\} \}
}
}
\cs_new:Npn \__tl_range:nnnNn \#1\#2\#3
{ \exp_args:Nff \__tl_range:nnNn
\{ \exp_args:Nf \__tl_range_normalize:nn
\{ \int_eval:n \{ #2 - 1 \} \} \{#1\} \}
}
\cs_new:Npn \__tl_range:nnNn \#1\#2\#3
{ \exp_args:Nff \__tl_range:nnNn
\{ \exp_args:Nf \__tl_range_normalize:nn
\{ \int_eval:n \{ #3 \} \} \{#1\} \}
}
\end{verbatim}
\cs_new:Npn \__tl_range:nnNn #1#2#3#4
{
    \if_int_compare:w #2 > #1 \exp_stop_f: \else:
\exp_after:wN { \exp_after:wN }
\fi:
\exp_after:wN #3
\int_value:w \int_eval:n { #2 - #1 } \exp_after:wN ;
\exp_after:wN { \exp:w \__tl_range_skip:w #1 ; { } #4 }
}
\cs_new:Npn \__tl_range_skip:w #1 ; #2
{
    \if_int_compare:w #1 > \c_zero_int
\exp_after:wN \__tl_range_skip:w
\int_value:w \int_eval:n { #1 - 1 } \exp_after:wN ;
\else:
\exp_after:wN \exp_end:
\fi:
}
\cs_new:Npn \__tl_range:w #1 ; #2
{
    \exp_args:Nf \__tl_range_collect:nn
    \__tl_range_skip_spaces:n {#2} {#1}
}
\cs_new:Npn \__tl_range_skip_spaces:n #1
{
    \tl_if_head_is_space:nTF {#1}
    { \exp_args:Nf \__tl_range_skip_spaces:n {#1} }
    { { } #1 }
}
\cs_new:Npn \__tl_range_collect:nn #1#2
{
    \int_compare:nNnTF {#2} = 0
    {#1}
    {\exp_args:No \tl_if_head_is_space:nTF { \use_none:n #1 }
    { \exp_args:Nf \__tl_range_collect:nn
    \__tl_range_skip_spaces:n {#2} }
    { \__tl_range_collect:ff
    { \exp_args:No \tl_if_head_is_N_type:nTF { \use_none:n #1 }
    \__tl_range_collect_N:nN }
    \__tl_range_collect_group:nn
    #1
    { \int_eval:n { #2 - 1 } }
    }
    { \__tl_range_collect_space:nw #1 - { { #1 - } }
    }
}
\_tl\_range\_normalize:nn

This function converts an \langle index \rangle argument into an explicit position in the token list (a result of 0 denoting “out of bounds”). Expects two explicit integer arguments: the \langle index \rangle #1 and the string count #2. If #1 is negative, replace it by #1 + #2 + 1, then limit to the range [0, #2].

\cs\_new\_Npn \_\_tl\_range\_collect:NN #1#2 \{ \{ #1 \#2 \} \}
\cs\_new\_Npn \_\_tl\_range\_collect\_group:nn #1#2 \{ \{ #1 \{#2\} \} \}
\cs\_generate\_variant:NN \_\_\_tl\_range\_collect:nn { ff }

(End definition for \_tl\_range:NN and others. These functions are documented on page 115.)

52.14 Viewing token lists

Showing token list variables is done after checking that the variable is defined (see \_\_kernel\_register\_show:N).

\cs\_new\_protected:Npn \tl\_show:N \{ \_\_tl\_show:NN \tl\_show:n \}
\cs\_generate\_variant:NN \_\_\_tl\_show:NN \_\_\tl\_show:n
\cs\_new\_protected:Npn \tl\_log:N \{ \_\_\_tl\_show:NN \_\_\tl\_log:n \}
\cs\_generate\_variant:NN \_\_\_\_tl\_log:n
\cs\_new\_protected:Npn \_\_\_\_tl\_show:NN #1#2
\{ \_\_\_kernel\_chk\_defined:NT #2
\{ \exp\_args:Nf \tl\_if\_empty:nTF
\{ \cs\_prefix\_spec:N #2 \cs\_argument\_spec:N #2 \}
\{ \exp\_args:Ne #1
\{ \token\_to\_str:N #2 = \_\_\_kernel\_exp\_not:u \exp\_after:wN \{#2\} \}
\}
\{ \msg\_error:nxxx { kernel } \{ bad:\_type \}
\{ \token\_to\_str:N #2 \{ \token\_to\_meaning:N #2 \} \{ \tl \} }
Many `show` functions are based on `\tl_show:n`. The argument of `\tl_show:n` is linewrapped using `\iow_wrap:nnnN` but with a leading `~`- and trailing period, both removed before passing the wrapped text to the `\showtokens` primitive. This primitive shows the result with a leading `~`- and trailing period.

The token list `\tl_internal_a_tl` containing the result of all these manipulations is displayed to the terminal using `\text_showtokens:D` and an odd `\exp_after:wN` which expand the closing brace to improve the output slightly. The calls to `\kernel_iow_with:Nn` ensure that the `newline` is set to 10 so that the `\newline` is correctly recognized by TeX, and that `\errorcontextlines` is -1 to avoid printing irrelevant context.

Claim.

```
\tl_show:n #1
\__tl_show:n
\__tl_show:w
```

(End definition for `\tl_show:n`, `\tl_log:n`, and `\__tl_show:NN`. This function is documented on page 109.)

### `\tl_log:n`

Logging is much easier, simply line-wrap. The `~`- and trailing period is there to match the output of `\tl_show:n`. Make sure that it is defined and that it is a token list, namely a macro with no \long or \protected prefix. Then compare \#1 to an attempt at reconstructing a valid structure of the given type using \#2 (see implementation of `\seq_show:N` for instance). If that is successful run the requested code \#4.

Claim.

```
\tl_log:n #1
```

(End definition for `\tl_log:n`. This function is documented on page 110.)

### `\__kernel_chk_tl_type:NnnT`

Helper for checking that \#1 has the correct internal structure to be of a certain type. Make sure that it is defined and that it is a token list, namely a macro with no \long nor \protected prefix. Then compare \#1 to an attempt at reconstructing a valid structure of the given type using \#2 (see implementation of `\seq_show:N` for instance). If that is successful run the requested code \#4.

Claim.

```
\__kernel_chk_tl_type:NnnT #1#2#3#4
```

(End definition for `\__kernel_chk_tl_type:NnnT`. This function is documented on page 110.)
52.15 Internal scan marks

\s_tl_nil
\s_tl_mark
\s_tl_stop

Internal scan marks. These are defined here at the end because the code for \scan_new:N depends on some l3tl functions.

52.16 Scratch token lists

\g_tmpa_tl
\g_tmpb_tl

Global temporary token list variables. They are supposed to be set and used immediately, with no delay between the definition and the use because you can’t count on other macros not to redefine them from under you.

\g_tmpa_tl
\g_tmpb_tl

These are local temporary token list variables. Be sure not to assume that the value you put into them will survive for long—see discussion above.

\l_tmpa_tl
\l_tmpb_tl

We finally clean up a temporary control sequence that we have used at various points to set up some definitions.
Chapter 53

\textbf{l3str implementation}

\begin{verbatim}
\texttt{\textbackslash s\_str\_mark} \texttt{\textbackslash s\_str\_stop} \par
\texttt{\_\_str\_use\_none\_delimit\_by\_s\_stop:w} \texttt{\_\_str\_use\_i\_delimit\_by\_s\_stop:nw} \par
\texttt{\_\_str\_if\_recursion\_tail\_break:NN} \texttt{\_\_str\_if\_recursion\_tail\_stop\_do:NN} \par
\end{verbatim}

53.1 Internal auxiliaries

\texttt{\_\_str\_use\_none\_delimit\_by\_s\_stop:w} \texttt{\_\_str\_use\_i\_delimit\_by\_s\_stop:nw} \par

(End definition for \texttt{\_\_str\_use\_none\_delimit\_by\_s\_stop:w} and \texttt{\_\_str\_use\_i\_delimit\_by\_s\_stop:nw}.)

\texttt{\_\_str\_recursion\_tail} \texttt{\_\_str\_recursion\_stop} \par

(End definition for \texttt{\_\_str\_recursion\_tail} and \texttt{\_\_str\_recursion\_stop}.)

(End definition for \texttt{\_\_kernel\_quark\_new\_test:N} \texttt{\_\_str\_if\_recursion\_tail\_break:NN} and \texttt{\_\_str\_if\_recursion\_tail\_stop\_do:NN}.)
53.2 Creating and setting string variables

A string is simply a token list. The full mapping system isn’t set up yet so do things by hand.

\(\text{\str_new:N}\)
\(\text{\str_new:c}\)
\(\text{\str_use:N}\)
\(\text{\str_use:c}\)
\(\text{\str_clear:N}\)
\(\text{\str_clear:c}\)
\(\text{\str_gclear:N}\)
\(\text{\str_gclear:c}\)
\(\text{\str_clear_new:N}\)
\(\text{\str_clear_new:c}\)
\(\text{\str_gclear_new:N}\)
\(\text{\str_gclear_new:c}\)
\(\text{\str_set_eq:NN}\)
\(\text{\str_set_eq:cN}\)
\(\text{\str_set_eq:Nc}\)
\(\text{\str_set_eq:cc}\)
\(\text{\str_gset_eq:NN}\)
\(\text{\str_gset_eq:cN}\)
\(\text{\str_gset_eq:Nc}\)
\(\text{\str_gset_eq:cc}\)
\(\text{\str_concat:NNN}\)
\(\text{\str_concat:ccc}\)
\(\text{\str_gconcat:NNN}\)
\(\text{\str_gconcat:ccc}\)

A string is simply a token list. The full mapping system isn’t set up yet so do things by hand.

\(\text{\cs_set_protected:Npn \__str_tmp:n #1}\)
\(\{\)
\(\text{\tl_if_blank:nF \{#1\}}\)
\(\}\)

\(\text{\cs_new_eq:cc \{ str_ #1 :N \}} \{ \text{tl_ #1 :N} \}\)
\(\text{\exp_args:Nc \text{cs_generate_variant:NN} \{ str_ #1 :N \} c}\)
\(\text{\str_set_eq:NN}\)
\(\text{\str_set_eq:cN}\)
\(\text{\str_set_eq:Nc}\)
\(\text{\str_set_eq:cc}\)
\(\text{\str_gset_eq:NN}\)
\(\text{\str_gset_eq:cN}\)
\(\text{\str_gset_eq:Nc}\)
\(\text{\str_gset_eq:cc}\)
\(\text{\str_concat:NNN}\)
\(\text{\str_concat:ccc}\)
\(\text{\str_gconcat:NNN}\)
\(\text{\str_gconcat:ccc}\)

Simply convert the token list inputs to \(\langle\text{strings}\rangle\).

\(\text{\cs_set_protected:Npn \__str_tmp:n #1}\)
\(\{\)
\(\text{\tl_if_blank:nF \{#1\}}\)
\(\{\)
\(\text{\cs_new_protected:cpx \{ str_ #1 :Nn \} ##1##2}\)
\(\{\)
\(\text{\exp_not:c \{ tl_ #1 :Nx \} ##1}\)
\(\{\)
\(\text{\exp_not:N \text{tl_to_str:n} \{##2\}}\)
\(\}\)
\(\}\)
\(\cs_generate_variant:cn \{ str_ #1 :Nn \} \{ cN, cc \}\)
\(\cs_generate_variant:cc \{ str_ #1 :Nn \} \{ cN, cc \}\)
\(\cs_generate_variant:NN \{ str_ #1 :Nn \} \{ cN, cc \}\)
\(\cs_generate_variant:NNN \{ str_ #1 :Nn \} \{ cN, cc \}\)
\(\text{\__str_tmp:n}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
\(\}\)
53.3 Modifying string variables

Start by applying \texttt{\tl_to_str:n} to convert the old and new token lists to strings, and also apply \texttt{\tl_to_str:N} to avoid any issues if we are fed a token list variable. Then
the code is a much simplified version of the token list code because neither the delimiter
nor the replacement can contain macro parameters or braces. The delimiter \texttt{\s__str_mark}
cannot appear in the string to edit so it is used in all cases. Some \texttt{x}\texttt{-expansion is}
unnecessary. There is no need to avoid losing braces nor to protect against expansion.
The ending code is much simplified and does not need to hide in braces.
Removal is just a special case of replacement.

\[ \text{str_remove_once:Nn} \]
\[ \text{str_remove_once:cn} \]
\[ \text{str_gremove_once:Nn} \]
\[ \text{str_gremove_once:cn} \]

Removal is just a special case of replacement.

\[ \text{str_remove_all:Nn} \]
\[ \text{str_remove_all:cn} \]
\[ \text{str_gremove_all:Nn} \]
\[ \text{str_gremove_all:cn} \]

53.4 String comparisons

\[ \text{str_if_empty_p:N} \]
\[ \text{str_if_empty_p:c} \]
\[ \text{str_if_empty:NTF} \]
\[ \text{str_if_empty:cFF} \]
\[ \text{str_if_empty:nFF} \]
\[ \text{str_if_empty:p:N} \]
\[ \text{str_if_empty:p:c} \]
\[ \text{str_if_empty:nTF} \]
\[ \text{str_if_empty:NTF} \]
\[ \text{str_if_empty:cTF} \]

String comparisons rely on the primitive \( \text{pdf} \text{strcmp} \), so we define a new name for it.

\[ \text{__str_if_eq:nn} \]
Simply rely on \_\_str_if_eq:nn, which expands to -1, 0 or 1. The ee version is created directly because it is more efficient.

Modern engines provide a direct way of comparing two token lists, but returning a number. This set of conditionals therefore makes life a bit clearer. The nn and ee versions are created directly as this is most efficient. Since \_\_str_if_eq:nn will expand to 0 as an explicit character with category 12 if the two lists match (and either -1 or 1 if they don’t) we can use \if:w here which is faster than using \if_int_compare:w.

Note that \str_if_eq:NNTF is different from \tl_if_eq:NNTF because it needs to ignore category codes.

Everything here needs to be detokenized but beyond that it is a simple token list test. It would be faster to fine-tune the T, F, TF variants by calling the appropriate variant of \tl_if_in:nnTF directly but that takes more code.

(End definition for \str_compare:nnTF. This function is documented on page 124.)

(End definition for \str_if_eq:nnTF. This function is documented on page 123.)

(End definition for \str_if_eq:NNTF. This function is documented on page 122.)
\use:x
\use:x
{ \tl_if_in:nnTF { \tl_to_str:N #1 } { \tl_to_str:n {#2} } }
{ \prg_return_true: } { \prg_return_false: }
\prg_generate_conditional_variant:Nnn \str_if_in:Nn
{ c } { T , F , TF }
\prg_new_protected_conditional:Npnn \str_if_in:nn #1#2 { T , F , TF }
{ \use:x
{ \tl_if_in:nnTF { \tl_to_str:n {#1} } { \tl_to_str:n {#2} } }
{ \prg_return_true: } { \prg_return_false: }
}

(End definition for \str_if_in:NnTF and \str_if_in:nnTF. These functions are documented on page 123.)

\str_case:nn
\str_case:Vn
\str_case:on
\str_case:nV
\str_case:nv
\str_case:nn TF
\str_case:Vn TF
\str_case:on TF
\str_case:nV TF
\str_case:nv TF
\str_case:Nn
\str_case:Nn TF
\str_case:NnF
\str_case:e:nn
\__str_case:nnTF
\__str_case:nnTF
\__str_case:nnTF
\__str_case:nnTF
\__str_case:nnTF
\__str_case:nnTF
\__str_case:nnTF
\__str_case:nnTF
\__str_case:nnTF
\__str_case:nnTF
\__str_case:nnTF
\__str_case:nnTF
\__str_case:nnTF
\__str_case:nnTF
\__str_case:nnTF
\__str_case:nnTF

Much the same as \tl_case:nnTF here: just a change in the internal comparison.
The inline and variable mappings are similar to the usual token list mappings but start out by turning the argument to an “other string”. Doing the same for the expandable function mapping would require \q__kernel_str_to_other:n, quadratic in the string length. To deal with spaces in that case, \q__str_map_function:w replaces the following space by a braced space and a further call to itself. These are received by \q__str_map_function:nn, which passes the space to #1 and calls \q__str_map_function:w to deal with the next space. The space before the braced space allows to optimize the \q__str_recursion_tail test. Of course we need to include a trailing space (the question mark is needed to avoid losing the space when \TeX
tokenizes the line). At the cost of about three more auxiliaries this code could get a 9 times speed up by testing only every 9-th character for whether it is \q__str_recursion_tail (also by converting 9 spaces at a time in the \q__str_map_function:nN case).

For the \q_map_variable functions we use a string assignment to store each character because spaces are made catcode 12 before the loop.
\_q\_str_recursion_tail ? -
\prg_break_point:Nn \str_map_break: { }
}
cs_new:Npn \str_map_function:NN
{ \exp_args:No \str_map_function:nN }
cs_new:Npn \__str_map_function:w #1 ~
{ \#1 { - { - } \__str_map_function:w } }
cs_new:Npn \__str_map_function:nn #1#2
{ }
\if_meaning:w \q\_str_recursion_tail #2
\exp_after:wN \str_map_break:
\f1:
\#1 \#2 \__str_map_function:nn {#1} }
cs_generate_variant:Nn \str_map_function:NN \cs_new_protected:Npn \str_map_inline:nn #1#2
{ \int_gincr:N \g__kernel_prg_map_int
cs_gset_protected:cpn
{ \__str_map___\int_use:N \g__kernel_prg_map_int :w } \#1 \#2
\use:x
\begin{verbatim}
\exp_not:N \__str_map_inline:NN \exp_not:c { \__str_map_\int_use:N \g__kernel_prg_map_int :w }
\__kernel_str_to_other_fast:n {#1}
\end{verbatim}
\q\_str_recursion_tail
\prg_break_point:Nn \str_map_break:
\begin{verbatim}
{ \int_gdecr:N \g__kernel_prg_map_int }
\end{verbatim}
cs_new_protected:Npn \str_map_inline:NN
{ \exp_args:No \str_map_inline:nn }
cs_generate_variant:Nn \str_map_inline:NN \cs_new_protected:Npn \str_map_variable:nNn #1#2#3
{ \exp_not:n { \__str_map_variable:NnN \#2 \#3 }
\begin{verbatim}
\__kernel_str_to_other_fast:n {#1}
\end{verbatim}
\q\_str_recursion_tail
\prg_break_point:Nn \str_map_break: { }
cs_new_protected:Npn \str_map_variable:NNn
{ \exp_args:No \str_map_variable:nNn }
cs_new_protected:Npn \__str_map_variable:NNn #1#2#3
{ \exp_args:No \__str_map_variable:NNn }
cs_new_protected:Npn \__str_if_recursion_tail_break:NNn #2 \str_map_break:
\begin{verbatim}
\exp_args:No #1 { \token_to_str:N \#2 }
\begin{verbatim}
\__str_map_inline:NN \#1
\end{verbatim}
\cs_new_protected:Npn \str_map_variable:nNn #1#2#3
{ { \exp_args:No \str_map_variable:nNn }
cs_new_protected:Npn \__str_map_variable:NNn #1#2#3
{ \exp_args:No \__str_map_variable:NNn }
\begin{verbatim}
\__str_if_recursion_tail_break:NNn \#3 \str_map_break:
\end{verbatim}
709
53.6 Accessing specific characters in a string

First apply \texttt{tl_to_str:n}, then replace all spaces by “other” spaces, 8 at a time, storing the converted part of the string between the \texttt{s__str_mark} and \texttt{s__str_stop} markers. The end is detected when \texttt{__str_to_other_loop:w} finds one of the trailing A, distinguished from any contents of the initial token list by their category. Then \texttt{__str_to_other_end:w} is called, and finds the result between \texttt{s__str_mark} and the first A (well, there is also the need to remove a space).
The difference with \texttt{\_\_kernel\_str\_to\_other\_fast:n} is that the converted part is left in the input stream, making these commands only restricted-expandable.

\begin{verbatim}
\cs_new:Npn \__kernel_str_to_other_fast:n \#1 { \exp_after:wN \__str_to_other_fast_loop:w \tl_to_str:n {#1} A ~ A ~ A ~ A ~ A ~ A ~ A ~ A ~ A ~ \s__str_stop }
\group_begin:
\tex_lccode:D '*= ' \ %
\tex_lccode:D 'A = A \ %
\tex_lowercase:D
\group_end:
\cs_new:Npn \__str_to_other_fast_loop:w #1 ~ #2 ~ #3 ~ #4 ~ #5 ~ #6 ~ #7 ~ #8 ~ #9 ~ { \if_meaning:w A #9 \__str_to_other_fast_end:w \fi:\ #1 * #2 * #3 * #4 * #5 * #6 * #7 * #8 * #9 \__str_to_other_fast_loop:w * }
\cs_new:Npn \__str_to_other_fast_end:w #1 * A #2 \s__str_stop {#1}
\end{verbatim}

\end{document}
\exp_args:Nf \__str_item:nn
\{ \__kernel_str_to_other:n {#1} \} \{#2\}
\}
\cs_new:Npn \str_item_ignore_spaces:nn #1 { \exp_args:No \__str_item:nn { \tl_to_str:n {#1} } }
\cs_new:Npn \__str_item:nn #1#2
{ \exp_after:wN \__str_item:w 
\int_value:w \int_eval:n {#2} \exp_after:wN ; 
\int_value:w \__str_count:n {#1} ; 
#1 \s__str_stop 
}
\cs_new:Npn \__str_item:w #1; #2;
{ \int_compare:nNnTF {#1} < 0 
{ \int_compare:nNnTF {#1} < {-#2} 
{ \__str_use_none_delimit_by_s_stop:w } 
{ \__str_use_i_delimit_by_s_stop:nw \exp:w \exp_after:wN \__str_skip_exp_end:w \int_value:w \int_eval:n { #1 + #2 } ; 
} 
} 
{ \__str_use_none_delimit_by_s_stop:w } 
{ \__str_use_i_delimit_by_s_stop:nw \exp:w \__str_skip_exp_end:w #1 ; { } 
} 
}
(End definition for \str_item:Nn and others. These functions are documented on page 127.)

\__str_skip_end:w
\__str_skip_loop:wNNNNNNNN
\__str_skip_end:w
\__str_skip_end:NNNNNNNN
Removes \texttt{max(#1,0)} characters from the input stream, and then leaves \texttt{\exp_end:}. This
should be expanded using \texttt{\exp:w}. We remove characters 8 at a time until there are
at most 8 to remove. Then we do a dirty trick: the \texttt{\if_case:w} construction leaves between 0 and 8 times the \texttt{\or:} control sequence, and those \texttt{\or:} become arguments of
\texttt{\__str_skip_end:NNNNNNNN}. If the number of characters to remove is 6, say, then there
are two \texttt{\or:} left, and the 8 arguments of \texttt{\__str_skip_end:NNNNNNNN} are the two \texttt{\or:},
and 6 characters from the input stream, exactly what we wanted to remove. Then close the \texttt{\if_case:w} conditional with \texttt{\fi:}, and stop the initial expansion with \texttt{\exp_end:}
(see places where \texttt{\__str_skip_exp_end:w} is called).
\cs_new:Npn \__str_skip_exp_end:w #1; 
{ 
\if_int_compare:w #1 > 8 \exp_stop_f: 
\exp_after:wN \__str_skip_loop:wNNNNNNNN
\else: 
\exp_after:wN \__str_skip_end:w 
\int_value:w \int_eval:w \fi: 
} 
712
Sanitize the string. Then evaluate the arguments. At this stage we also decrement the \textit{start index}, since our goal is to know how many characters should be removed. Then limit the range to be non-negative and at most the length of the string (this avoids needing to check for the end of the string when grabbing characters), shifting negative numbers by the appropriate amount. Afterwards, skip characters, then keep some more, and finally drop the end of the string.
This function converts an \textit{index} argument into an explicit position in the string (a result of 0 denoting “out of bounds”). Expects two explicit integer arguments: the \textit{index} \#1 and the string count \#2. If \#1 is negative, replace it by \#1 + \#2 + 1, then limit to the range [0, \#2].

\begin{verbatim}
\cs_new:Npn \str_range_normalize:nn #1#2 {
  \int_eval:n {
    \if_int_compare:w #1 < \c_zero_int
      \if_int_compare:w #1 < -#2 \exp_stop_f:
        0
      \else:
        \#1 + \#2 + 1
      \fi:
    \else:
      \if_int_compare:w #1 < \#2 \exp_stop_f:
        \#1
      \else:
        \#2
      \fi:
    \fi:
  }
}
\end{verbatim}

Collects max(\#1,0) characters, and removes everything else until \s__str_stop. This is somewhat similar to \_\_\_str\_skip\_exp\_end:w, but accepts integer expression arguments. This time we can only grab 7 characters at a time. At the end, we use an \texttt{\_\_\_str\_collect\_end:nnnnnnnw} trick again, so that the 8 first arguments of \_\_\_str\_collect\_end:nnnnnnnw are some \texttt{\or:}, followed by an \texttt{\fi:}, followed by \#1 characters from the input stream. Simply leaving this in the input stream closes the conditional properly and the \texttt{\or:} disappear.

\begin{verbatim}
\cs_new:Npn \str_collect_delimit_by_q_stop:w #1; {
  \str_collect_loop:wn #1 ; { }
}
\cs_new:Npn \str_collect_loop:wn #1 ; {
  \if_int_compare:w #1 > 7 \exp_stop_f:
    \exp_after:wN \str_collect_loop:wnNNNNNNN
  \else:
    \exp_after:wN \str_collect_end:wn
    \int_value:w \int_eval:n { #1 - 7 } ;
  \fi:
}
\cs_new:Npn \str_collect_loop:wnNNNNNNN #1; #2 #3#4#5#6#7#8#9 {
  \exp_after:wN \str_collect_loop:wnNNNNNNN
  \int_value:w \int_eval:n { \#1 - 7 } ;
  \#2 #3#4#5#6#7#8#9
}
\cs_new:Npn \str_collect_end:wn #1 ; {
  \}
\end{verbatim}
53.7 Counting characters

To speed up this function, we grab and discard 9 space-delimited arguments in each iteration of the loop. The loop stops when the last argument is one of the trailing \( X\langle\text{number}\rangle \), and that \langle\text{number}\rangle is added to the sum of 9 that precedes, to adjust the result.

To count characters in a string we could first escape all spaces using \texttt{\_kernel_str-\_to_other:n}, then pass the result to \texttt{\tl\_count:n}. However, the escaping step would be quadratic in the number of characters in the string, and we can do better. Namely, sum the number of spaces (\texttt{\_str\_count_spaces:n}) and the result of \texttt{\tl\_count:n}, which ignores spaces. Since strings tend to be longer than token lists, we use specialized functions to count characters ignoring spaces. Namely, loop, grabbing 9 non-space characters at each step, and end as soon as we reach one of the 9 trailing items. The internal function \texttt{\_str\_count:n}, used in \texttt{\str\_item:nn} and \texttt{\str\_range:nnn}, is similar to \texttt{\_str\_count\_ignore\_spaces:n} but expects its argument to already be a string or a string with spaces escaped.
53.8 The first character in a string

The \_ignore\_spaces variant applies \tl\_to\_str:n then grabs the first item, thus skipping spaces. As usual, \str\_head:N expands its argument and hands it to \str\_head:n.

To circumvent the fact that \TeX{} skips spaces when grabbing undelimited macro parameters, \__str\_head:w takes an argument delimited by a space. If \#1 starts with a non-space character, \__str\_use\_i\_delimit\_by\_s\_stop:w leaves that in the input stream. On the other hand, if \#1 starts with a space, the \__str\_head:w takes an empty argument, and the single (initially braced) space in the definition of \__str\_head:w makes its way to the output. Finally, for an empty argument, the (braced) empty brace group in the definition of \str\_head:n gives an empty result after passing through \__str\_use\_i\_delimit\_by\_s\_stop:nw.

\cs\_new:Npn \str\_count:n \#1
\{ \__str\_count\_aux:n
\{ \str\_count\_spaces:n \#1
+ \exp\_after:wN \__str\_count\_loop:NNNNNNNN \tl\_to\_str:n \#1
\}
\}
\cs\_new:Npn \__str\_count:n \#1
\{ \__str\_count\_aux:n
\{ \__str\_count\_loop:NNNNNNNN \#1 \}
\}
\cs\_new:Npn \str\_count\_ignore\_spaces:n \#1
\{ \__str\_count\_aux:n
\{ \exp\_after:wN \__str\_count\_loop:NNNNNNNN \tl\_to\_str:n \#1 \}
\}
\cs\_new:Npn \__str\_count\_aux:n \#1
\{ \int\_eval:n
\{ \#1
\{ X 8 \} \{ X 7 \} \{ X 6 \}
\{ X 5 \} \{ X 4 \} \{ X 3 \}
\{ X 2 \} \{ X 1 \} \{ X 0 \}
\s\_str\_stop
\}
\}
\cs\_new:Npn \__str\_count\_loop:NNNNNNNN \#1\#2\#3\#4\#5\#6\#7\#8\#9
\{ \if\meaning:w X \#9
\exp\_after:wN \__str\_use\_none\_delimit\_by\_s\_stop:w
\fi:
9 + \__str\_count\_loop:NNNNNNNN
\}

(End definition for \str\_count:N and others. These functions are documented on page 126.)
Getting the tail is a little bit more convoluted than the head of a string. We hit the front of the string with \reverse_if:N \if_charcode:w \scan_stop:. This removes the first character, and necessarily makes the test true, since the character cannot match \scan_stop:. The auxiliary function then inserts the required \fi: to close the conditional, and leaves the tail of the string in the input stream. The details are such that an empty string has an empty tail (this requires in particular that the end-marker X be unexpandable and not a control sequence). The \ignore_spaces is rather simpler: after converting the input to a string, \__str_tail_auxii:w removes one undelimited argument and leaves everything else until an end-marker \s__str_mark. One can check that an empty (or blank) string yields an empty tail.

Case changing for programmatic reasons is done by first detokenizing input then doing a simple loop that only has to worry about spaces and everything else. The output is detokenized to allow data sharing with text-based case changing. Similarly, for 8-bit engines the multi-byte information is shared.
\cs_new:Npn \str_lowercase:n { \__str_change_case:nn {#1} { lower } }
\cs_new:Npn \str_uppercase:n { \__str_change_case:nn {#1} { upper } }
\cs_generate_variant:Nn \str_foldcase:n { V }
\cs_generate_variant:Nn \str_lowercase:n { f }
\cs_generate_variant:Nn \str_uppercase:n { f }
\cs_new:Npn \__str_change_case:nn #1
\exp_after:wN \__str_change_case_aux:nn \exp_after:wN
\tl_to_str:n {#1} }
\cs_new:Npn \__str_change_case_aux:nn #1#2
\exp_last_unbraced:NNNNo
\cs_new:Npn \__str_change_case_output:nw #1#2 \__str_change_case_result:n #3
{ #2 \__str_change_case_result:n { #3 #1 } }
\cs_generate_variant:Nn \__str_change_case_output:nw { f }
\cs_new:Npn \__str_change_case_end:wn #1 \__str_change_case_result:n #2
\tl_to_str:n {#2} }
\cs_new:Npn \__str_change_case_loop:nw #1#2 \q__str_recursion_stop
\exp_last_unbraced:NNNNo
\cs_new:Npn \__str_change_case_space:n #1 \c_space_tl
\__str_change_case_output:nw { ~ }
\__str_change_case_loop:nw {#1}
\cs_new:Npn \__str_change_case_char:nN #1#2
\__str_if_recursion_tail_stop_do:Nn #2
{ \__str_change_case_end:wn }
\use:c { char_str_ #1 case:N } #2
\__str_change_case_loop:nw {#1}
\if_int_compare:w 0 \cs_if_exist:NT \tex_XeTeXversion:D { 1 }
\cs_if_exist:NT \tex_luatexversion:D { 1 }
> 0 \exp_stop_f:
\else:
\cs_gset:Npn \__str_change_case_char:nN #1#2
\__str_if_recursion_tail_stop_do:Nn #2
{ \__str_change_case_end:wn }
\__str_change_case_output:fw
\use:c { char_str_ #1 case:N } #2 }
\__str_change_case_loop:nw {#1}
\if_int_compare:w 0
Skip high chars for the Japanese engines.

\cs_if_exist:NF \tex_pdftexversion:D {
  \cs_gset:Npn \__str_change_case_char_UTFviii:nNNN #1#2#3#4
  { \__str_change_case_output:nw {#2#3#4} \__str_change_case_loop:nw {#1} }
  \cs_gset:Npn \__str_change_case_char_UTFviii:nNNNN #1#2#3#4#5
  { \__str_change_case_output:nw {#2#3#4#5} \__str_change_case_loop:nw {#1} }
\}
\cs_new:Npn \__str_change_case_char_UTFviii:nn #1#2
  { \use:c { \__str_change_case_char_UTFviii_#1:nn } {#1} {#2} }
\cs_new:Npn \__str_change_case_char_UTFviii_upper:nn #1#2
  { \__str_change_case_output:fw { \cs_if_exist:cTF { c__kernel_#1 case_#2_tl } { \__kernel_tl_to_str:w \exp_after:wN \exp_after:wN \exp_after:wN { \cs:w c__kernel_ #1 case_ #2_tl \cs_end: } } #2 } \__str_change_case_loop:nw {#1} }
\cs_new_eq:NN \__str_change_case_char_UTFviii_lower:nn \__str_change_case_char_UTFviii_upper:nn
\cs_new:Npn \__str_change_case_char_UTFviii_fold:nn #1#2
  { #2 } \__str_change_case_loop:nw {#1} }
\__str_change_case_output:fw
{\cs_if_exist:cTF {c__kernel_#1 case_#2_tl}
{\__kernel_tl_to_str:w\exp_after:wN\exp_after:wN\exp_after:wN\cs:w c__kernel_#1 case_#2_tl \cs_end:}
}{\cs_if_exist:cTF {c__kernel_lowercase_#2_tl}
{\__kernel_tl_to_str:w\exp_after:wN\exp_after:wN\exp_after:wN\cs:w c__kernel_lowercase_#2_tl \cs_end:}
{#2}}
\__str_change_case_loop:nw {#1}
} \fi:

(End definition for \str_foldcase:n and others. These functions are documented on page 131.)
\c_ampersand_str \c_atsign_str \c_backslash_str \c_left_brace_str \c_right_brace_str \c_circumflex_str \c_colon_str \c_dollar_str \c_hash_str \c_percent_str \c_tilde_str \c_underscore_str \c_zero_str

For all of those strings, use \cs_to_str:N to get characters with the correct category code without worries
\str_new:N \l_tmpa_str \str_new:N \l_tmpb_str \str_new:N \g_tmpa_str \str_new:N \g_tmpb_str

(End definition for \c_ampersand_str and others. These variables are documented on page 132.)
\l_tmpa_str \l_tmpb_str \g_tmpa_str \g_tmpb_str

(End definition for \l_tmpa_str and others. These variables are documented on page 132.)

53.10 Viewing strings

\str_show:n Displays a string on the terminal.
\str_show:N \cs_new_eq:NN \str_show:n \tl_show:n
\str_show:c \str_log:n \str_log:N \str_log:c
\cs_new_protected:Npn \str_show:N #1
\begin{verbatim}
\__kernel_chk_tl_type:NnnT #1 { str } { \tl_to_str:N #1 }
\__kernel_chk_tl_type:NnnT #1 { str } { \tl_show:N #1 }
\end{verbatim}
\cs_generate_variant:Nn \str_show:N { c }
\cs_new_eq:NN \str_log:n \tl_log:n
\cs_new_protected:Npn \str_log:N #1
\begin{verbatim}
\__kernel_chk_tl_type:NnnT #1 { str } { \tl_to_str:N #1 }
\__kernel_chk_tl_type:NnnT #1 { str } { \tl_log:N #1 }
\end{verbatim}
\cs_generate_variant:Nn \str_log:N { c }

(End definition for \str_show:n and others. These functions are documented on page 131.)

\end{verbatim}
Chapter 54

l3str-convert implementation

54.1 Helpers

54.1.1 Variables and constants

\_\_\_str\_tmp:w
Internal scratch space for some functions.

\l\_str\_internal\_tl
\cs\_new\_protected:Npn \_\_\_str\_tmp:w { }
\tl\_new:N \l\_str\_internal\_tl

(End definition for \_\_\_str\_tmp:w and \l\_str\_internal\_tl.)

\g\_str\_result\_tl
The \g\_str\_result\_tl variable is used to hold the result of various internal string operations (mostly conversions) which are typically performed in a group. The variable is global so that it remains defined outside the group, to be assigned to a user-provided variable.

\tl\_new:N \g\_str\_result\_tl

(End definition for \g\_str\_result\_tl.)

\c\_str\_replacement\_char\_int
When converting, invalid bytes are replaced by the Unicode replacement character "FFFD.

\int\_const:Nn \c\_str\_replacement\_char\_int \{ "FFFD \}

(End definition for \c\_str\_replacement\_char\_int.)

\c\_str\_max\_byte\_int
The maximal byte number.

\int\_const:Nn \c\_str\_max\_byte\_int \{ 255 \}

(End definition for \c\_str\_max\_byte\_int.)

\s\_str
Internal scan marks.

\scan\_new:N \s\_str

(End definition for \s\_str.)
\_\_str\_nil Internal quarks.
\(\text{\textbackslash{quark\_new:N}}\ \_\_str\_nil\)
\(\text{(End definition for } _\_str\_nil.)\)

\_\_str\_alias\_prop To avoid needing one file per encoding/escaping alias, we keep track of those in a property list.
\(\text{\textbackslash{prop\_new:N}}\ \_\_str\_alias\_prop\)
\(\text{\textbackslash{prop\_gput:Nnn}}\ \_\_str\_alias\_prop\ { latin1 } { iso88591 }\)
\(\text{\textbackslash{prop\_gput:Nnn}}\ \_\_str\_alias\_prop\ { latin2 } { iso88592 }\)
\(\text{\textbackslash{prop\_gput:Nnn}}\ \_\_str\_alias\_prop\ { latin3 } { iso88593 }\)
\(\text{\textbackslash{prop\_gput:Nnn}}\ \_\_str\_alias\_prop\ { latin4 } { iso88594 }\)
\(\text{\textbackslash{prop\_gput:Nnn}}\ \_\_str\_alias\_prop\ { latin5 } { iso88599 }\)
\(\text{\textbackslash{prop\_gput:Nnn}}\ \_\_str\_alias\_prop\ { latin6 } { iso885910 }\)
\(\text{\textbackslash{prop\_gput:Nnn}}\ \_\_str\_alias\_prop\ { latin7 } { iso885913 }\)
\(\text{\textbackslash{prop\_gput:Nnn}}\ \_\_str\_alias\_prop\ { latin8 } { iso885914 }\)
\(\text{\textbackslash{prop\_gput:Nnn}}\ \_\_str\_alias\_prop\ { latin9 } { iso885915 }\)
\(\text{\textbackslash{prop\_gput:Nnn}}\ \_\_str\_alias\_prop\ { latin10 } { iso885916 }\)
\(\text{\textbackslash{prop\_gput:Nnn}}\ \_\_str\_alias\_prop\ { utf16le } { utf16 }\)
\(\text{\textbackslash{prop\_gput:Nnn}}\ \_\_str\_alias\_prop\ { utf16be } { utf16 }\)
\(\text{\textbackslash{prop\_gput:Nnn}}\ \_\_str\_alias\_prop\ { utf32le } { utf32 }\)
\(\text{\textbackslash{prop\_gput:Nnn}}\ \_\_str\_alias\_prop\ { utf32be } { utf32 }\)
\(\text{\textbackslash{prop\_gput:Nnn}}\ \_\_str\_alias\_prop\ { hexadecimal } { hex }\)
\(\text{\textbackslash{bool\_lazy\_any:nTF}}\)
\{ \\
\text{\textbackslash{sys\_if\_engine\_luatex:p}:} \\
\text{\textbackslash{sys\_if\_engine\_xetex:p}:} \\
\} \\
\{ \\
\text{\textbackslash{prop\_gput:Nnn}}\ \_\_str\_alias\_prop\ { default } { } \\
\} \\
\{ \\
\text{\textbackslash{prop\_gput:Nnn}}\ \_\_str\_alias\_prop\ { default } { utf8 } \\
\}
\(\text{(End definition for } _\_str\_alias\_prop.)\)

\_\_str\_error\_bool In conversion functions with a built-in conditional, errors are not reported directly to the user, but the information is collected in this boolean, used at the end to decide on which branch of the conditional to take.
\(\text{\textbackslash{bool\_new:N}}\ \_\_str\_error\_bool\)
\(\text{(End definition for } _\_str\_error\_bool.)\)

str\_byte Conversions from one \langle encoding \rangle / \langle escaping \rangle pair to another are done within x-expanding assignments. Errors are signalled by raising the relevant flag.
\(\text{\textbackslash{flag\_new:n}}\ { \text{str\_byte} }\)
\(\text{\textbackslash{flag\_new:n}}\ { \text{str\_error} }\)
\(\text{(End definition for } \text{str\_byte} \text{ and } \text{str\_error}. \text{These variables are documented on page ??.)}\)
### 54.2 String conditionals

\texttt{\_str_if\_contains\_char:NnT} *(token list)* *(char)*

Expects the *(token list)* to be an *(other string)*: the caller is responsible for ensuring that no (too-)special catcodes remain. Loop over the characters of the string, comparing character codes. The loop is broken if character codes match. Otherwise we return “false”.

\begin{verbatim}
\prg_new_conditional:Npnn \__str_if_contains_char:Nn #1#2 { T , TF }
\prg_new_conditional:Npnn \__str_if_contains_char:nnTF #1 {tf,TN}
\cs_new:Npn \__str_if_contains_char:nnTF #1 #2 { \__str_if_contains_char_aux:nn #1 #2 \prg_return_false: }
\cs_new:Npn \__str_if_contains_char_aux:nn #1 #2 { \__str_if_contains_char_auxi:nN {#2} #1 }
\prg_new_conditional:Npnn \__str_if_contains_char_true:nn #1 {TN}
\prg_break_point:
\prg_return_true: \use_none:n
\end{verbatim}

(End definition for \_str_if\_contains\_char:NnT and others.)

\texttt{\_str_octal_use:NTF} *(token)* *(true code)* *(false code)*

If the *(token)* is an octal digit, it is left in the input stream, followed by the *(true code)*. Otherwise, the *(false code)* is left in the input stream.

\textbf{TeXhackers note:} This function will fail if the escape character is an octal digit. We are thus careful to set the escape character to a known value before using it. \TeX{} dutifully detects octal digits for us: if \#1 is an octal digit, then the right-hand side of the comparison is ’\#1’, greater than 1. Otherwise, the right-hand side stops as ’1’, and the conditional takes the false branch.

\begin{verbatim}
\prg_new_conditional:Npnn \__str_octal_use:NTF \string \{ \true \} \{ \false \}
\prg_new_conditional:Npnn \__str_octal_use:NTF \string \{ \true \} \{ \false \}
{ \if_int_compare:w \numexpr \token_to_int:N \string \base \#1 \exp_stop_f:
\begin{verbatim}
\if_int_compare:w \numexpr \token_to_int:N \string \base \#1 \exp_stop_f:
\else:
\fi:
\end{verbatim}

(End definition for \_str_octal_use:NTF.)

\end{verbatim}
\_\_str\_hexadecimal\_use:NTF

\TeX{} detects uppercase hexadecimal digits for us (see \_\_str\_octal\_use:NTF), but not the lowercase letters, which we need to detect and replace by their uppercase counterpart.

\begin{verbatim}
\prg_new_conditional:Npn \_\_str\_hexadecimal\_use:N #1 { TF }
{ \if_int_compare:w 1 < "1 \token_to_str:N #1 \exp_stop_f:
#1 \prg_return_true:
\else: \if_case:w \int_eval:n { \exp_after:wN ' \token_to_str:N #1 - 'a }
A \or: B \or: C \or: D \or: E \or: F \else: \prg_return_false:
\exp_after:wN \use_none:n \fi:
\prg_return_true:
\fi: }
\end{verbatim}

(End definition for \_\_str\_hexadecimal\_use:NTF.)

54.3 Conversions

54.3.1 Producing one byte or character

\c__str\_byte\_0_tl \c__str\_byte\_1_tl \c__str\_byte\_255_tl \c__str\_byte\_\-1_tl

For each integer \(N\) in the range \([0, 255]\), we create a constant token list which holds three character tokens with category code other: the character with character code \(N\), followed by the representation of \(N\) as two hexadecimal digits. The value \(-1\) is given a default token list which ensures that later functions give an empty result for the input \(-1\).

\begin{verbatim}
\group_begin:
\__kernel_tl_set:Nx \l__str\_internal_tl { \tl_to_str:n { 0123456789ABCDEF } }
\tl_map_inline:Nn \l__str\_internal_tl { \tl_map_inline:Nn \l__str\_internal_tl { \tl_const:cx { c__str\_byte_\int_eval:n { "##1##1\} _tl } { \char\_generate:nn { "##1##1\} { 12 } \} \#1 \##1\} } }
\group_end:
\tl_const:cn { c__str\_byte\_\-1_tl } { { } \use\_none:n { } }
\end{verbatim}

(End definition for \c__str\_byte\_0_tl and others.)

\c__str\_output\_byte:n \c__str\_output\_byte:w \c__str\_output\_hexadecimal:n \c__str\_output\_end:

Those functions must be used carefully: feeding them a value outside the range \([-1, 255]\) will attempt to use the undefined token list variable \c__str\_byte\_\langle\number\rangle_tl. Assuming that the argument is in the right range, we expand the corresponding token list, and pick either the byte (first token) or the hexadecimal representations (second and third tokens). The value \(-1\) produces an empty result in both cases.
13937 \cs_new:Npn \__str_output_byte:n #1 
13938 \{ \__str_output_byte:w #1 \__str_output_end: \}
13939 \cs_new:Npn \__str_output_byte:w
13940 \{
13941 \exp_after:wN \exp_after:wN
13942 \exp_after:wN \use_i:nnn
13943 \cs:w c__str_byte_ \int_eval:w
13944 \}
13945 \cs_new:Npn \__str_output_hexadecimal:n #1
13946 \{
13947 \exp_after:wN \exp_after:wN
13948 \exp_after:wN \use_none:n
13949 \cs:w c__str_byte_ \int_eval:n \{#1\} \_tl \cs_end:
13950 \}
13951 \cs_new:Npn \__str_output_end:
13952 \{ \scan_stop: \_tl \cs_end: \}

(End definition for \__str_output_byte:n and others.)

\__str_output_byte_pair_be:n 
\__str_output_byte_pair_le:n 
\__str_output_byte_pair:nnN

Convert a number in the range $[0, 65535]$ to a pair of bytes, either big-endian or little-endian.

13953 \cs_new:Npn \__str_output_byte_pair_be:n #1
13954 \{
13955 \exp_args:Nf \__str_output_byte_pair:nnN
13956 \{ \int_div_truncate:nn \{ #1 \} \{ "100\} \{#1\} \use:nn
13957 \}
13958 \cs_new:Npn \__str_output_byte_pair_le:n #1
13959 \{
13960 \exp_args:Nf \__str_output_byte_pair:nnN
13961 \{ \int_div_truncate:nn \{ #1 \} \{ "100\} \{#1\} \use_ii_i:nn
13962 \}
13963 \cs_new:Npn \__str_output_byte_pair:nnN #1#2#3
13964 \{
13965 \#3
13966 \{ \__str_output_byte:n \{ #1 \}
13967 \{ \__str_output_byte:n \{ #2 - #1 * "100\} \}
13968 \}

(End definition for \__str_output_byte_pair_be:n, \__str_output_byte_pair_le:n, and \__str_output_byte_pair:nnN)

54.3.2 Mapping functions for conversions

This maps the function \#1 over all characters in \g__str_result_tl, which should be a byte string in most cases, sometimes a native string.

13969 \cs_new_protected:Npn \__str_convert_gmap:N #1
13970 \{
13971 \__kernel_tl_gset:Nx \g__str_result_tl
13972 \{ \exp_after:wN \__str_convert_gmap_loop:NN
13973 \exp_after:wN #1
13974 \g__str_result_tl \{ ? \prg_break: \}
13975 \prg_break_point:
13976 \}

54.3.2 Mapping functions for conversions

This maps the function \#1 over all characters in \g__str_result_tl, which should be a byte string in most cases, sometimes a native string.
This maps the function #1 over all character codes in \g__str_result_tl, which must be in the internal representation.

\cs_new_protected:Npn \__str_convert_gmap_internal:N #1
\{ \__kernel_tl_gset:Nx \g__str_result_tl
\exp_after:wN \__str_convert_gmap_internal_loop:Nww
\exp_after:wN #1
\g__str_result_tl \s__str \s__str_stop \prg_break: \s__str
\prg_break_point:
\}

\cs_new:Npn \__str_convert_gmap_internal_loop:Nww #1 #2 \s__str #3 \s__str
\{ \__str_use_none_delimit_by_s_stop:w #3 \s__str_stop
#1 {#3}
\__str_convert_gmap_internal_loop:Nww #1 \}

(End definition for \__str_convert_gmap_internal:N and \__str_convert_gmap_internal_loop:Nw.)

54.3.3 Error-reporting during conversion

When converting using the function \str_set_convert:Nnnn, errors should be reported to the user after each step in the conversion. Errors are signalled by raising some flag (typically \@@_error), so here we test that flag: if it is raised, give the user an error, otherwise remove the arguments. On the other hand, in the conditional functions \str_set_convert:NnnnTF, errors should be suppressed. This is done by changing \__str_if_flag_error:nnx into \__str_if_flag_no_error:nnx locally.

\cs_new_protected:Npn \__str_if_flag_error:nnx #1
\{ \flag_if_raised:nTF {#1}
{ \msg_error:nnx { str } }
{ \use_none:nn }
\}

\cs_new_protected:Npn \__str_if_flag_no_error:nnx #1#2#3
\{ \flag_if_raised:nT {#1} { \bool_gset_true:N \g__str_error Bool } \}

(End definition for \__str_if_flag_error:nnx and \__str_if_flag_no_error:nnx.)

\__str_if_flag_times:nT

At the end of each conversion step, we raise all relevant errors as one error message, built on the fly. The height of each flag indicates how many times a given error was encountered. This function prints #2 followed by the number of occurrences of an error if it occurred, nothing otherwise.
54.3.4 Framework for conversions

Most functions in this module expect to be working with “native” strings. Strings can also be stored as bytes, in one of many encodings, for instance UTF8. The bytes themselves can be expressed in various ways in terms of \TeX{} tokens, for instance as pairs of hexadecimal digits. The questions of going from arbitrary Unicode code points to bytes, and from bytes to tokens are mostly independent.

Conversions are done in four steps:

- “unescape” produces a string of bytes;
- “decode” takes in a string of bytes, and converts it to a list of Unicode characters in an internal representation, with items of the form 
  \(\text{(bytes)} \, \text{s__str (Unicode code point)} \, \text{s__str}\)
  where we have collected the \(\text{(bytes)}\) which combined to form this particular Unicode character, and the \(\text{(Unicode code point)}\) is in the range \([0, \text{10FFFF}]\).
- “encode” encodes the internal list of code points as a byte string in the new encoding;
- “escape” escapes bytes as requested.

The process is modified in case one of the encoding is empty (or the conversion function has been set equal to the empty encoding because it was not found): then the unescape or escape step is ignored, and the decode or encode steps work on tokens instead of bytes. Otherwise, each step must ensure that it passes a correct byte string or internal string to the next step.

The input string is stored in \texttt{\g__str_result_tl}, then we: unescape and decode; encode and escape; exit the group and store the result in the user’s variable. The various conversion functions all act on \texttt{\g__str_result_tl}. Errors are silenced for the conditional functions by redefining \texttt{\__str_if_flag_error:nnx} locally.

\begin{verbatim}
\cs_new:Npn \__str_if_flag_times:nT #1#2 { \flag_if_raised:nT {#1} { #2~(x \flag_height:n {#1} ) } } (End definition for \__str_if_flag_times:nT.)
\end{verbatim}
\begin{verbatim}
{ \bool_gset_false:N \g__str_error_bool \__str_convert:nNNnnn
  { \cs_set_eq:NN \__str_if_flag_error:nnx \__str_if_flag_no_error:nnx }
  \tl_gset_eq:NN #1 {#2} {#3} {#4}
  \bool_if:NTF \g__str_error_bool \prg_return_false: \prg_return_true:
}
\cs_new_protected:Npn \__str_convert:nNNnnn #1#2#3#4#5#6
  {
  \group_begin:
  #1 \__kernel_tl_gset:Nx \g__str_result_tl { \__kernel_str_to_other_fast:n {#4} }
  \exp_after:wN \__str_convert_decode_: \__kernel_tl_gset:Nx \g__str_result_tl { \__kernel_str_to_other_fast:n {#5} }
  \tl_to_str:n {#6} \s__str_stop
  \prg_do_nothing: \use_ii_i:nn \__str_convert_encode_: \__kernel_tl_gset:Nx \g__str_result_tl { \__kernel_str_to_other_fast:n {#6} }
  \group_end:
  #2 #3 \g__str_result_tl
\end{verbatim}

(End definition for \emph{\str_set_convert:Nnnn} and others. These functions are documented on page 135.)

The task of \emph{\__str_convert:wwwnn} is to split \emph{(encoding)/(escaping)} pairs into their components, \#1 and \#2. Calls to \emph{\__str_convert:nnn} ensure that the corresponding conversion functions are defined. The third auxiliary does the main work.

- \#1 is the encoding conversion function;
- \#2 is the escaping function;
- \#3 is the escaping name for use in an error message;
- \#4 is \emph{\prg_do_nothing:} for unescaping/decoding, and \emph{\use_ii_i:nn} for encoding/escaping;
- \#5 is the default encoding function (either “decode” or “encode”), for which there should be no escaping.

Let us ignore the native encoding for a second. In the unescaping/decoding phase, we want to do \#2\#1 in this order, and in the encoding/escaping phase, the order should be reversed: \#4\#2\#1 does exactly that. If one of the encodings is the default (native), then the escaping should be ignored, with an error if any was given, and only the encoding, \#1, should be performed.

\begin{verbatim}
\__str_convert:wwwnn \__str_convert:NNnNN
\end{verbatim}
The arguments of \_\_str_convert:nnn are: enc or esc, used to build filenames, the type of the conversion (unescape, decode, encode, escape), and the encoding or escaping name. If the function is already defined, no need to do anything. Otherwise, filter out all non-alphanumerics in the name, and lowercase it. Feed that, and the same three arguments, to \_\_str_convert:nnnn. The task is then to make sure that the conversion function #3_#1 corresponding to the type #3 and filtered name #1 is defined, then set our initial conversion function #3_#4 equal to that.

How do we get the #3_#1 conversion to be defined if it isn’t? Two main cases.

First, if #1 is a key in \g__str_alias_prop, then the value \l__str_internal_tl tells us what file to load. Loading is skipped if the file was already read, i.e., if the conversion command based on \l__str_internal_tl already exists. Otherwise, try to load the file; if that fails, there is an error, use the default empty name instead.

Second, #1 may be absent from the property list. The \cs_if_exist:cF test is automatically false, and we search for a file defining the encoding or escaping #1 (this should allow third-party .def files). If the file is not found, there is an error, use the default empty name instead.

In all cases, the conversion based on \l__str_internal_tl is defined, so we can set the #3_#1 function equal to that. In some cases (e.g., utf16be), the #3_#1 function is actually defined within the file we just loaded, and it is different from the \l__str_internal_tl-based function: we mustn’t clobber that different definition.
This function keeps only letters and digits, with upper case letters converted to lower case.

\cs_new:Npn \__str_convert_lowercase_alphanum:n #1
{ \exp_after:wN \__str_convert_lowercase_alphanum_loop:N \tl_to_str:n {#1} { ? \prg_break: } \prg_break_point: }
\cs_new:Npn \__str_convert_lowercase_alphanum_loop:N #1
{ \use_none:n #1 \if_int_compare:w '#1 > 'Z \exp_stop_f: \if_int_compare:w '#1 > 'z \exp_stop_f: \else: \__str_output_byte:n { '#1 + 'a - 'A } \fi: \fi: \else: \if_int_compare:w '#1 < 'a \exp_stop_f: \else: \__str_output_byte:n { '#1 + 'a - 'A } \fi: \fi: }
(End definition for \__str_convert_lowercase_alphanum:n and \__str_convert_lowercase_alphanum_loop:N.)
54.3.5 Byte unescape and escape

Strings of bytes may need to be stored in auxiliary files in safe “escaping” formats. Each such escaping is only loaded as needed. By default, on input any non-byte is filtered out, while the output simply consists in letting bytes through.

In the case of 8-bit engines, every character is a byte. For Unicode-aware engines, test the character code; non-bytes cause us to raise the flag `str_byte`. Spaces have already been given the correct category code when this function is called.

```latex
\bool_lazy_any:nTF
\{ \sys_if_engine_luatex_p:
\sys_if_engine_xetex_p:
\}
\{ \cs_new:Npn \__str_filter_bytes:n #1
\{ \__str_filter_bytes_aux:N #1
\{ ? \prg_break: \}
\prg_break_point:
\}
\cs_new:Npn \__str_filter_bytes_aux:N #1
\{ \use_none:n #1
\if_int_compare:w '#1 < 256 \exp_stop_f:
#1
\else:
\flag_raise:n { str_byte }
\fi:
\__str_filter_bytes_aux:N
\}
\cs_new_eq:NN \__str_filter_bytes:n \use:n }
```

(The definition for \__str_filter_bytes:n and \__str_filter_bytes_aux:N.)

The simplest unescaping method removes non-bytes from \g__str_result_tl.

```latex
\bool_lazy_any:nTF
\{ \sys_if_engine_luatex_p:
\sys_if_engine_xetex_p:
\}
\{ \cs_new_protected:Npn \__str_convert_unescape_: \use:n
\{ \flag_clear:n { str_byte }
\__kernel_tl_gset:Nx \g__str_result_tl
\{ \exp_args:No \__str_filter_bytes:n \g__str_result_tl }
\__str_if_flag_error:nnx { str_byte } { non-byte } { bytes }
\}
\}
\{ \cs_new_protected:Npn \__str_convert_unescape_: \}
\cs_new_eq:NN \__str_convert_unescape_bytes: \__str_convert_unescape_: }
```

(End definition for \__str_convert_unescape_: and \__str_convert_unescape_bytes:.)

732
The simplest form of escape leaves the bytes from the previous step of the conversion unchanged.

\[\text{cs_new_protected:Npn \_str_convert_escape_: } \{ \} \]
\[\text{cs_new_eq:NN \_str_convert_escape_bytes: \_str_convert_escape_:} \]

(End definition for \_str_convert_escape_ and \_str_convert_escape_bytes.)

54.3.6 Native strings

Convert each character to its character code, one at a time.

\[\text{cs_new_protected:Npn \_str_convert_decode_:} \]
\[\text{\_str_decode_native_char:N} \]

The conversion from an internal string to native character tokens basically maps \texttt{char_generate:nn} through the code-points, but in non-Unicode-aware engines we use a fall-back character ? rather than nothing when given a character code outside \([0, 255]\). We detect the presence of bad characters using a flag and only produce a single error after the x-expanding assignment.

\[\text{bool_lazy_any:nTF} \]
\[\{ \]
\[\text{\sys_if_engine_luatex_p:} \]
\[\text{\sys_if_engine_xetex_p:} \]
\[\{ \]
\[\text{\cs_new_protected:Npn \_str_convert_encode_:} \]
\[\text{\_str_convert_gmap:n \_str_encode_native_char:n} \]
\[\text{\cs_new:Npn \_str_encode_native_char:n #1} \]
\[\text{\char_generate:nn \{#1\} \{12\}} \]
\[\} \]
\[\} \]
\[\text{\cs_new_protected:Npn \_str_convert_encode_:} \]
\[\{ \]
\[\text{\flag_clear:n \{\str_error\} \_str_convert_gmap:n \_str_encode_native_char:n} \]
\[\text{\_str_if_flag_error:nnx \{\str_error\} \{ native-overflow \} \{ \}} \]
\[\} \]
\[\text{\cs_new:Npn \_str_encode_native_char:n #1} \]
\[\{ \]
\[\text{\if_int_compare:w #1 > \c__str_max_byte_int} \]
\[\text{\flag_raise:n \{\str_error\}} \]
\[\} \]
\[\text{\else:} \]
\[\text{\char_generate:nn \{#1\} \{12\}} \]
\[\text{\fi:} \]
\[\text{\msg_new:nnnn \{ str \} \{ native-overflow \}} \]
\[\{ \text{Character-code-too-large-for-this-engine.}\} \]
This engine only supports 8-bit characters:
valid character codes are in the range \([0,255]\).
To manipulate arbitrary Unicode, use LuaTeX or XeTeX.

\(\text{End definition for } \_\text{str_convert_encode_} \text{ and } \_\text{str_encode_native_char:n}\)

### 54.3.7 clist

\_str_convert_decode_clist: \_str_decode_clist_char:n

Convert each integer to the internal form. We first turn \(\text{g\_str_result_tl}\) into a clist variable, as this avoids problems with leading or trailing commas.

\[
\text{cs\_new\_protected:Npm } \_\text{str_convert_decode_clist:}
\]

\[
\{ \text{clist\_gset:No } \text{g\_str_result_tl } \text{g\_str_result_tl} \\
\text{ } \text{\_kernel\_tl\_gset:Nx } \text{g\_str_result_tl} \\
\{ \text{exp\_args:No } \text{clist\_map\_function:nN} \\
\text{g\_str_result_tl } \_\text{str_decode_clist_char:n} \\
\}
\]

\[
\text{cs\_new:Npm } \_\text{str_decode_clist_char:n } \#1 \\
\{ \#1 \text{/s\_str int\_eval:n } \#1 \} \text{/s\_str}
\]

\(\text{End definition for } \_\text{str_convert_decode_clist: and } \_\text{str_decode_clist_char:n}\)

\_str_convert_encode_clist: \_str_encode_clist_char:n

Convert the internal list of character codes to a comma-list of character codes. The first line produces a comma-list with a leading comma, removed in the next step (this also works in the empty case, since \(\text{tl\_tail:N}\) does not trigger an error in this case).

\[
\text{cs\_new\_protected:Npm } \_\text{str_convert_encode_clist:}
\]

\[
\{ \_\text{str_convert_gmap\_internal:N } \_\text{str_encode_clist_char:n} \\
\text{ \_kernel\_tl\_gset:Nx } \text{g\_str_result_tl } \{ \text{tl\_tail:N } \text{g\_str_result_tl } \}
\]

\[
\text{cs\_new:Npm } \_\text{str_encode_clist_char:n } \#1 \{ , \#1 \}
\]

\(\text{End definition for } \_\text{str_convert_encode_clist: and } \_\text{str_encode_clist_char:n}\)

### 54.3.8 8-bit encodings

It is not clear in what situations 8-bit encodings are used, hence it is not clear what should be optimized. The current approach is reasonably efficient to convert long strings, and it scales well when using many different encodings.

The data needed to support a given 8-bit encoding is stored in a file that consists of a single function call

\[
\text{\_str\_declare\_eight\_bit\_encoding:nnnn } \langle \text{name} \rangle \} \{ \langle \text{modulo} \rangle \} \{ \langle \text{mapping} \rangle \} \{ \langle \text{missing} \rangle \}
\]

This declares the encoding \(\langle \text{name} \rangle\) to map bytes to Unicode characters according to the \(\langle \text{mapping} \rangle\), and map those bytes which are not mentioned in the \(\langle \text{mapping} \rangle\) either to the replacement character (if they appear in \(\langle \text{missing} \rangle\)), or to themselves. The \(\langle \text{mapping} \rangle\) argument is a token list of pairs \(\{ \langle \text{byte} \rangle \} \{ \langle \text{Unicode} \rangle \}\) expressed in uppercase hexadecimal notation. The \(\langle \text{missing} \rangle\) argument is a token list of \(\langle \text{byte} \rangle\). Every \langle byte \rangle which does
not appear in the \langle mapping \rangle nor the \langle missing \rangle lists maps to itself in Unicode, so for instance the \texttt{latin1} encoding has empty \langle mapping \rangle and \langle missing \rangle lists. The \langle modulo \rangle is a (decimal) integer between 256 and 558 inclusive, modulo which all Unicode code points supported by the encodings must be different.

We use two integer arrays per encoding. When decoding we only use the \texttt{decode} integer array, with entry \( n + 1 \) (offset needed because integer array indices start at 1) equal to the Unicode code point that corresponds to the \( n \)-th byte in the encoding under consideration, or \(-1\) if the given byte is invalid in this encoding. When encoding we use both arrays: upon seeing a code point \( n \), we look up the entry \((1 \text{ plus})\ n\) modulo some number \( M \) in the \texttt{encode} array, which tells us the byte that might encode the given Unicode code point, then we check in the \texttt{decode} array that indeed this byte encodes the Unicode code point we want. Here, \( M \) is an encoding-dependent integer between 256 and 558 (it turns out), chosen so that among the Unicode code points that can be validly represented in the given encoding, no pair of code points have the same value modulo \( M \).

Loop through both lists of bytes to fill in the \texttt{decode} integer array, then fill the \texttt{encode} array accordingly. For bytes that are invalid in the given encoding, store \(-1\) in the \texttt{decode} array.
\cs_new_protected:Npn \__str_declare_eight_bit_loop:Nnn #1#2#3 
{ \__str_use_none_delimit_by_s_stop:w #2 \s__str_stop 
\intarray_gset:Nnn #1 \{ 1 + "#2" \} \{ "#3" \} 
\__str_declare_eight_bit_loop:Nnn #1 
} 
\cs_new_protected:Npn \__str_declare_eight_bit_loop:Nn #1#2 
{ \__str_use_none_delimit_by_s_stop:w #2 \s__str_stop 
\intarray_gset:Nnn #1 \{ 1 + "#2" \} \{ -1 \} 
\__str_declare_eight_bit_loop:Nn #1 
} 

\__str_convert_decode_eight_bit:n 
\__str_decode_eight_bit_aux:n 
\__str_decode_eight_bit_aux:Nn 

The map from bytes to Unicode code points is in the \texttt{decode} array corresponding to the given encoding. Define \texttt{\__str\_tmp:w} and pass it successively all bytes in the string. It produces an internal representation with suitable \texttt{\s__str} inserted, and the corresponding code point is obtained by looking it up in the integer array. If the entry is \texttt{-1} then issue a replacement character and raise the flag indicating that there was an error.

\cs_new_protected:Npn \__str_convert_decode_eight_bit:n #1 
{ \cs_set:Npx \__str_tmp:w 
{ \exp_not:N \__str_decode_eight_bit_aux:Nn 
\exp_not:c { g__str_decode_#1_intarray } } 
\flag_clear:n \{ str\_error \} \__str_convert_gmap:N \__str_tmp:w 
\__str_if_flag_error:nnx \{ str\_error \} \{ decode\_8\_bit \} \{#1\} 
} 
\cs_new:Npn \__str_decode_eight_bit_aux:Nn #1#2 
{ #2 \s__str 
\exp_args:Nf \__str_decode_eight_bit_aux:n 
{ \intarray_item:Nn #1 \{ 1 + "#2" \} } \s__str 
} 
\cs_new:Npn \__str_decode_eight_bit_aux:n #1 
{ \if_int_compare:w #1 < \c_zero_int 
\flag_raise:n \{ str\_error \} \int_value:w \c__str_replacement_char_int 
\else: \c__str_replacement_char_int \fi: 
} 

\__str_convert_encode_eight_bit:n 
\__str_encode_eight_bit_aux:nnN 
\__str_encode_eight_bit_aux:NNn 

It is not practical to make an integer array with indices in the full Unicode range, so we work modulo some number, which is simply the size of the \texttt{encode} integer array for the
given encoding. This gives us a candidate byte for representing a given Unicode code point. Of course taking the modulo leads to collisions so we check in the decode array that the byte we got is indeed correct. Otherwise the Unicode code point we started from is simply not representable in the given encoding.

\int_new:N \l__str_modulo_int
\cs_new_protected:Npn \__str_convert_encode_eight_bit:n #1
{\cs_set:Npx \__str_tmp:w
{\exp_not:N \__str_decode_#1_intarray \exp_not:c { g__str_decode_#1_intarray }
\flag_clear:n { str_error }
\__str_convert_gmap_internal:N \__str_tmp:w
\__str_if_flag_error:nnx { str_error } { encode-8-bit } { #1 }
} \cs_new:Npn \__str_encode_eight_bit_aux:NNn #1#2#3
{\exp_args:Nf \__str_encode_eight_bit_aux:nnN
{\intarray_item:Nn #1 \int_mod:nn { #3 } \intarray_count:N #1 }
\int_compare:nNnTF { \intarray_item:Nn #3 { 1 + #1 } } = { #2 }
\__str_output_byte:n { #1 } \flag_raise:n { str_error }
}
\cs_new:Npn \__str_encode_eight_bit_aux:nnN #1#2#3
{\int_compare:nNnTF { \intarray_item:Nn #3 { 1 + #1 } } = { #2 }
\__str_output_byte:n { #1 } \flag_raise:n { str_error }
}

(End definition for \__str_convert_encode_eight_bit:n, \__str_convert_encode_eight_bit:aux:nnN, and \__str_encode_eight_bit:aux:NNn.)

54.4 Messages

General messages, and messages for the encodings and escapings loaded by default (“native”, and “bytes”).

\msg_new:nnn { str } { unknown-esc }{ Escaping-scheme-‘#1’-(filtered:-‘#2’)-unknown. }
\msg_new:nnn { str } { unknown-enc }{ Encoding-scheme-‘#1’-(filtered:-‘#2’)-unknown. }
\msg_new:nnnn { str } { native-escaping }{ The-‘native’-encoding-scheme-does-not-support-any-escaping. }
\msg_new:nnn { str } { file-not-found }{ Since-native-strings-do-not-consist-in-bytes,- none-of-the-escaping-methods-make-sense.- The-specified-escaping,-‘#1’,-will-be-ignored. }

737
Message used when the “bytes” unescaping fails because the string given to `\str_set_convert:Nnnn` contains a non-byte. This cannot happen for the -8-bit engines. Messages used for other escapings and encodings are defined in each definition file.

Those messages are used when converting to and from 8-bit encodings.

Several of those encodings are defined by the pdf file format. The following byte storage methods are defined:

- **bytes** (default), non-bytes are filtered out, and bytes are left untouched (this is defined by default);
- **hex** or **hexadecimal**, as per the pdfTeX primitive `\pdfescapehex`
- **name**, as per the pdfTeX primitive `\pdfescapename`
- **string**, as per the pdfTeX primitive `\pdfescapestring`
- **url**, as per the percent encoding of urls.
54.5.1 Unescape methods

Take chars two by two, and interpret each pair as the hexadecimal code for a byte. Anything else than hexadecimal digits is ignored, raising the flag. A string which contains an odd number of hexadecimal digits gets 0 appended to it: this is equivalent to appending a 0 in all cases, and dropping it if it is alone.

```latex
\cs_new_protected:Npn \__str_convert_unescape_hex: & \cs_new_protected:Npn \__str_unescape_hex_auxi:N & \cs_new_protected:Npn \__str_unescape_hex_auxii:N & \msg_new:nnnn { str } { unescape-hex } { String~invalid~in~escaping~'hex':~only~hexadecimal~digits~allowed. } & \msg_new:nnnn { str } { unescape-hex } { Some~characters~in~the~string~you~asked~to~convert~are~not~hexadecimal~digits~(0-9,~A-F,~a-f)~nor~spaces. } & \end{definition}
```
The \_str_convert_unescape_name: function replaces each occurrence of # followed by two hexadecimal digits in \_\_str_result_tl by the corresponding byte. The \_str_url: function is identical, with escape character % instead of #. Thus we define the two together. The arguments of \_str_tmp:w are the character code of # or % in hexadecimal, the name of the main function to define, and the name of the auxiliary which performs the loop.

The looping auxiliary #3 finds the next escape character, reads the following two characters, and tests them. The test \_str_hexadecimal_use:NTF leaves the uppercase digit in the input stream, hence we surround the test with \_str_output_byte:w " and \_str_output_end:. If both characters are hexadecimal digits, they should be removed before looping; this is done by \use_i:nn. If one of the characters is not a hexadecimal digit, then feed "#1 to \_str_output_byte:w to produce the escape character, raise the flag, and call the looping function followed by the two characters (remove \use_i:nn).

\begin{verbatim}
cs_set_protected:Npn \_str_tmp:w \#1\#2\#3
  \cs_new_protected:cpn { __str_convert_unescape_#2: }
  { 
    \group_begin:
      \flag_clear:n { str_byte }
      \flag_clear:n { str_error }
      \int_set:Nn \tex_escapechar:D { 92 }
      \__kernel_tl_gset:Nx \g__str_result_tl
        { \exp_after:wN #3 \g__str_result_tl
          #1 ? { ? \prg_break: }
          \prg_break_point:
        }
      \__str_if_flag_error:nnx { str_byte } { non-byte } { #2 }
      \__str_if_flag_error:nnx { str_error } { unescape-#2 } { }
    \group_end:
  }
cs_new:Npn #3 ##1#1##2##3
  { \__str_filter_bytes:n {##1} \use_none:n ##3 \__str_hexadecimal_use:NTF \#2
    \__str_output_byte:w "
    \__str_hexadecimal_use:NTF ##2
      { \__str_hexadecimal_use:NTF ##3
        { }
        \flag_raise:n { str_error }
        \use_i:nn
      }
      \flag_raise:n { str_error }
      0 + '#1 \use_i:nn
    }
    \__str_output_end:
    \use_i:nn \#3 \#2\#3
\end{verbatim}

740
The string escaping is somewhat similar to the name and url escapings, with escape character \. The first step is to convert all three line endings, "\^J", "\^M", and "\^M\^J" to the common "\^J", as per the PDF specification. This step cannot raise the flag.

Then the following escape sequences are decoded.

- \n Line feed (10)
- \r Carriage return (13)
- \t Horizontal tab (9)
- \b Backspace (8)
- \f Form feed (12)
- \( Left parenthesis
- \) Right parenthesis
- \ Backslash

\ddd (backslash followed by 1 to 3 octal digits) Byte ddd (octal), subtracting 256 in case of overflow.

If followed by an end-of-line character, the backslash and the end-of-line are ignored. If followed by anything else, the backslash is ignored, raising the error flag.
\__kernel_tl_gset:Nx \__str_result_tl
{
\exp_after:wN \__str_unescape_string_loop:wNNN
\__str_result_tl #1 ?? { ? \prg_break: }
\prg_break_point:
}
\__str_if_flag_error:nnx { str_byte } { non-byte } { string }
\__str_if_flag_error:nnx { str_error } { unescape-string } { }
\group_end:
\exp_args:No \__str_tmp:w { \c_backslash_str }
\exp_last_unbraced:NNNNo
\cs_new:Npn \__str_unescape_string_loop:wNNN #1 \c_backslash_str #2#3#4
{
\__str_filter_bytes:n {#1}
\use_none:n #4
\__str_output_byte:w '
\__str_octal_use:NTF #2
{ \__str_octal_use:NTF #3
{ \__str_octal_use:NTF #4
{ \if_int_compare:w #2 > 3 \exp_stop_f:
- 256 
\fi:
\__str_unescape_string_repeat:NNNNNN
\__str_unescape_string_repeat:NNNNNN ? }
\__str_unescape_string_repeat:NNNNNN ?? }
\__str_unescape_string_repeat:NNNNNN ? }

\str_case_e:nnF {#2}
{
{ \c_backslash_str } { 134 }
{ ( } { 50 }
{ ) } { 51 }
{ r } { 15 }
{ f } { 14 }
{ n } { 12 }
{ t } { 11 }
{ b } { 10 }
{ "\^J\} { 0 - 1 }
\flag_raise:n { str_error }
0 - 1 \use_i:nn
}
\__str_output_end:
\use_i:nn \__str_unescape_string_loop:wNNN #2#3#4

742
\cs_new:Npn \__str_unescape_string_repeat:NNNNNN #1#2#3#4#5#6
{ \__str_output_end: \__str_unescape_string_loop:wNNN }
\cs_new:Npn \__str_unescape_string_newlines:wN #1 \^{}w #2
{ #1 \if_charcode:w \^{}w \else: \^{}w \fi:
 \__str_unescape_string_newlines:wN #2 }
\msg_new:nnnn { str } { unescape-string }
{ String invalid in escaping `string'. }
\LaTeX~came across an escape character '\c_backslash_str' not followed by any of: `n', `r', `t', `b', `f', `(', `)'), `-\c_backslash_str', one-to-three-octal-digits, or the end of a line.

54.5.2 Escape methods

Currently, none of the escape methods can lead to errors, assuming that their input is made out of bytes.

\__str_convert_escape_hex: \__str_escape_hex_char:N

Loop and convert each byte to hexadecimal.

\cs_new_protected:Npn \__str_convert_escape_hex:
{ \__str_convert_gmap:N \__str_escape_hex_char:N }
\cs_new:Npn \__str_escape_hex_char:N #1
{ \__str_output_hexadecimal:n { '\#1' } }

\__str_convert_escape_name: \__str_escape_name_char:n \__str_if_escape_name:nTF \c__str_escape_name_str \c__str_escape_name_not_str

For each byte, test whether it should be output as is, or be “hash-encoded”. Roughly, bytes outside the range ["2A", "7E"] are hash-encoded. We keep two lists of exceptions: characters in \c__str_escape_name_not_str are not hash-encoded, and characters in the \c__str_escape_name_str are encoded.

\str_const:Nn \c__str_escape_name_not_str { ! " $ & ' } %$
\str_const:Nn \c__str_escape_name_str { {}/<[]> }
\cs_new_protected:Npn \__str_convert_escape_name:
{ \__str_convert_gmap:N \__str_escape_name_char:n }
\cs_new:Npn \__str_escape_name_char:n #1
{ \__str_if_escape_name:nTF {#1} {#1} { \c_hash_str \__str_output_hexadecimal:n {'#1} } }
\prg_new_conditional:Npnn \__str_if_escape_name:n #1 { TF }
{ \if_int_compare:w '#1 < "2A \exp_stop_f:
 \__str_if_contains_char:NnTF \c__str_escape_name_not_str {#1}
 { \prg_return_true: \prg_return_false: }
 \else:
 \if_int_compare:w '#1 > "7E \exp_stop_f:

Any character below (and including) space, and any character above (and including) del, are converted to octal. One backslash is added before each parenthesis and backslash.

This function is similar to \_str_convert escape name; escaping different characters.
54.6 Encoding definitions

The native encoding is automatically defined. Other encodings are loaded as needed. The following encodings are supported:

- utf-8;
- utf-16, big-, little-endian, or with byte order mark;
- utf-32, big-, little-endian, or with byte order mark;
- the iso 8859 code pages, numbered from 1 to 16, skipping the inexistient iso 8859-12.

54.6.1 utf-8 support

Loop through the internal string, and convert each character to its UTF-8 representation. The representation is built from the right-most (least significant) byte to the left-most (most significant) byte. Continuation bytes are in the range [128, 191], taking 64 different values, hence we roughly want to express the character code in base 64, shifting the first digit in the representation by some number depending on how many continuation bytes there are. In the range [0, 127], output the corresponding byte directly. In the range [128, 2047], output the remainder modulo 64, plus 128 as a continuation byte, then output the quotient (which is in the range [0, 31]), shifted by 192. In the next range, [2048, 65535], split the character code into residue and quotient modulo 64, output the residue as a first continuation byte, then repeat; this leaves us with a quotient in the range [0, 15], which we output shifted by 224. The last range, [65536, 1114111], follows the same pattern: once we realize that dividing twice by 64 leaves us with a number larger than 15, we repeat, producing a last continuation byte, and offset the quotient by 240 for the leading byte.

How is that implemented? \_\_str_encode_utf_vii_loop:wwnnw takes successive quotients as its first argument, the quotient from the previous step as its second argument (except in step 1), the bound for quotients that trigger one more step or not, and finally the offset used if this step should produce the leading byte. Leading bytes can be in
the ranges $[0, 127]$, $[192, 223]$, $[224, 239]$, and $[240, 247]$ (really, that last limit should be 244 because Unicode stops at the code point 1114111). At each step, if the quotient $#1$ is less than the limit $#3$ for that range, output the leading byte ($#1$ shifted by $#4$) and stop. Otherwise, we need one more step: use the quotient of $#1$ by 64, and $#1$ as arguments for the looping auxiliary, and output the continuation byte corresponding to the remainder $#2 - 64#1 + 128$. The bizarre construction $- 1 + 0 *$ removes the spurious initial continuation byte (better methods welcome).

When decoding a string that is purportedly in the utf-8 encoding, four different errors can occur, signalled by a specific flag for each (we define those flags using \flag_clear_new:n rather than \flag_new:n, because they are shared with other encoding definition files).

- “Missing continuation byte”: a leading byte is not followed by the right number of continuation bytes.
- “Extra continuation byte”: a continuation byte appears where it was not expected, i.e., not after an appropriate leading byte.
- “Overlong”: a Unicode character is expressed using more bytes than necessary, for instance, “C0“80 for the code point 0, instead of a single null byte.
- “Overflow”: this occurs when decoding produces Unicode code points greater than 1114111.

We only raise one \TeX3 error message, combining all the errors which occurred. In the short message, the leading comma must be removed to get a grammatically correct sentence. In the long text, first remind the user what a correct utf-8 string should look like, then add error-specific information.

\l__str_missing_flag
\l__str_extra_flag
\l__str_overlong_flag
\l__str_overflow_flag
Invalid UTF-8 string:
\exp_last_unbraced:Nf \use_none:n
\__str_if_flag_times:nT { str_missing } { ,missing-continuation-byte }
\__str_if_flag_times:nT { str_extra } { ,extra-continuation-byte }
\__str_if_flag_times:nT { str_overlong } { ,overlong-form }
\__str_if_flag_times:nT { str_overflow } { ,code-point-too-large }

In the UTF-8 encoding, each Unicode character consists in 1 to 4 bytes, with the following bit pattern:

\iou_indent:n
\__str_if_flag_times:nT \use_none:n
\__str_if_flag_times:nT \use_none:n
\__str_if_flag_times:nT \use_none:n
\__str_if_flag_times:nT \use_none:n

Code-point-\textbackslash{} \textbackslash{} \textbackslash{} < 128: 0xxxxxxx
Code-point-\textbackslash{} \textbackslash{} \textbackslash{} 128-2047: 110xxxxx 10xxxxxx
Code-point-\textbackslash{} \textbackslash{} 2048-65535: 1110xxxx 10xxxxxx 10xxxxxx
Code-point-\textbackslash{} \textbackslash{} 65536-1114111: 11110xxx 10xxxxxx 10xxxxxx 10xxxxxx

Bytes of the form 10xxxxxx are called continuation-bytes.

\flag_if_raised:nT { str_missing }
{ \\\n  A leading byte (in the range [192, 255]) was not followed by the appropriate number of continuation-bytes.
}

\flag_if_raised:nT { str_extra }
{ \\\n  \LaTeX{} came across a continuation byte when it was not expected.
}

\flag_if_raised:nT { str_overlong }
{ \\\n  Every Unicode code point must be expressed in the shortest possible form. For instance, \texttt{\textasciitilde{}0xCO\textasciitilde{}0x83\textasciitilde{}} is not a valid representation for the code point 3.
}

\flag_if_raised:nT { str_overflow }
{ \\\n  Unicode limits code points to the range [0, 1114111].
}

(End definition for \texttt{\_str_missing_flag} and others.)
Decoding is significantly harder than encoding. As before, lower some flags, which are tested at the end (in bulk, to trigger at most one \texttt{IPTX3} error, as explained above). We expect successive multi-byte sequences of the form \texttt{(start byte) (continuation bytes)}. The \texttt{\_start} auxiliary tests the first byte:

- \texttt{[0, "7F"]}: the byte stands alone, and is converted to its own character code;
- \texttt{["80", "BF"]}: unexpected continuation byte, raise the appropriate flag, and convert that byte to the replacement character \texttt{"FFFD};
- \texttt{["C0", "FF"]}: this byte should be followed by some continuation byte(s).

In the first two cases, \texttt{\use{none}{delimit_by_q}{stop}} removes data that only the third case requires, namely the limits of ranges of Unicode characters which can be expressed with 1, 2, 3, or 4 bytes.

We can now concentrate on the multi-byte case and the \texttt{\_continuation} auxiliary. We expect \texttt{\_3} to be in the range \texttt{["80", "BF"]}. The test for this goes as follows: if the character code is less than \texttt{"80"}, we compare it to \texttt{"C0"}, yielding \texttt{false}; otherwise to \texttt{"C0"}, yielding \texttt{true} in the range \texttt{["80", "BF"]} and \texttt{false} otherwise. If we find that the byte is not a continuation range, stop the current slew of bytes, output the replacement character, and continue parsing with the \texttt{\_start} auxiliary, starting at the byte we just tested. Once we know that the byte is a continuation byte, leave it behind us in the input stream, compute what code point the bytes read so far would produce, and feed that number to the \texttt{\_aux} function.

The \texttt{\_aux} function tests whether we should look for more continuation bytes or not. If the number it receives as \texttt{\_1} is less than the maximum \texttt{\_4} for the current range, then we are done: check for an overlong representation by comparing \texttt{\_1} with the maximum \texttt{\_3} for the previous range. Otherwise, we call the \texttt{\_continuation} auxiliary again, after shifting the “current code point” by \texttt{\_4} (maximum from the range we just checked).

Two additional tests are needed: if we reach the end of the list of range maxima and we are still not done, then we are faced with an overflow. Clean up, and again insert the code point \texttt{"FFFD} for the replacement character. Also, every time we read a byte, we need to check whether we reached the end of the string. In a correct \texttt{UTF-8} string, this happens automatically when the \texttt{\_start} auxiliary leaves its first argument in the input stream: the end-marker begins with \texttt{\prg{break}}: which ends the loop. On the other hand, if the end is reached when looking for a continuation byte, the \texttt{\use{none}{n}{\_3}} construction removes the first token from the end-marker, and leaves the \texttt{\_end} auxiliary, which raises the appropriate error flag before ending the mapping.

```latex
\cs_new_protected:cpn { \_str_convert_decode_utf8: }
{
  \flag_clear:n { str_error }
  \flag_clear:n { str_missing }
  \flag_clear:n { str_extra }
  \flag_clear:n { str_overlong }
  \flag_clear:n { str_overflow }
  \_kernel_tl_gset:Nx \g__str_result_tl \exp_after:wN \_str_decode_utf_viii_start:N \g__str_result_tl
  \prg_break: \_str_decode_utf_viii_end: }
  \prg_break_point:
  \_str_if_flag_error:nnx { str_error } { utf8-decode } { }
```

748
\cs_new:Npn \__str_decode_utf_viii_start:N #1
{\if_int_compare:w #1 < "C0 \exp_stop_f:
  \s__str
  \if_int_compare:w #1 < "80 \exp_stop_f:
    \int_value:w #1
  \else:
    \flag_raise:n { str_extra }
    \flag_raise:n { str_error }
    \int_use:N \c__str_replacement_char_int
    \fi:
  \else:
    \exp_after:wN \__str_decode_utf_viii_continuation:wwN
    \int_value:w \int_eval:n { #1 - "C0 } \exp_after:wN
  \fi:
\s__str
\__str_use_none_delimit_by_s_stop:w {"80} {"800} {"10000} {"110000} \s__str_stop
\__str_decode_utf_viii_start:N}
\cs_new:Npn \__str_decode_utf_viii_continuation:wwN #1 \s__str #2 \__str_decode_utf_viii_start:N #3
{\use_none:n #3 \if_int_compare:w #3 <
  \if_int_compare:w #3 < "80 \exp_stop_f: - \fi:
  "C0 \exp_stop_f:
#3 \exp_after:wN \__str_decode_utf_viii_aux:wNnnwN
\int_value:w \int_eval:n { #1 * "40 + #3 - "80 } \exp_after:wN
\else:
  \s__str
  \flag_raise:n { str_missing }
  \flag_raise:n { str_error }
  \int_use:N \c__str_replacement_char_int
  \fi:
\s__str
#2
\__str_decode_utf_viii_start:N #3
\cs_new:Npn \__str_decode_utf_viii_aux:wNnnwN #1 \s__str #2#3#4 #5 \__str_decode_utf_viii_start:N #6
{\if_int_compare:w #1 < #4 \exp_stop_f:
  \s__str
  \if_int_compare:w #1 < #3 \exp_stop_f:
    \flag_raise:n { str_overlong }
    \flag_raise:n { str_error }
    \int_use:N \c__str_replacement_char_int
  \else:
    \int_eval:n { #1 - #3 + #4 }
  \fi:
#1
\else:
}}
54.6.2 utf-16 support

The definitions are done in a category code regime where the bytes 254 and 255 used by the byte order mark have catcode 12.

When the endianness is not specified, it is big-endian by default, and we add a byte-order mark. Convert characters one by one in a loop, with different behaviours depending on the character code.

- [0, "D7FF]: converted to two bytes;
- ["D800", "DFFF"] are used as surrogates: they cannot be converted and are replaced by the replacement character;
- ["E000", "FFFF"]: converted to two bytes;
- ["10000", "10FFFF"]: converted to a pair of surrogates, each two bytes. The magic "D7C0 is "D800 - "10000"/400.

For the duration of this operation, \_\_str_tmp:w is defined as a function to convert a number in the range [0, "FFFF"] to a pair of bytes (either big endian or little endian), by feeding the quotient of the division of #1 by "100, followed by #1 to \_\_str_encode_utf_xvi_be:nn or its le analog: those compute the remainder, and output two bytes for the quotient and remainder.
When encoding a Unicode string to utf-16, only one error can occur: code points in the range [U+D800, U+DFFF), corresponding to surrogates, cannot be encoded. We use the all-purpose flag @@@error to signal that error.

When decoding a Unicode string which is purportedly in UTF-16, three errors can occur: a missing trail surrogate, an unexpected trail surrogate, and a string containing an odd number of bytes.
\____str_if_flag_times:nT { str_missing } { ,missing-trail-surrogate }
\____str_if_flag_times:nT { str_extra } { ,extra-trail-surrogate }
\____str_if_flag_times:nT { str_end } { ,odd-number-of-bytes }
\)
\)
\}
\}
{ In-the-UTF-16-encoding,-each-Unicode-character-is-encoded-as-
2-or-4-bytes: \ \}
\ioW_indent:n
\}
{ Code-point-in-[U+0000,-U+D7FF]:-two-bytes \ \}
\{ Code-point-in-[U+D800,-U+DFFF]:-illegal \ \}
\{ Code-point-in-[U+E000,-U+F000]:-two-bytes \ \}
\{ Code-point-in-[U+10000,-U+10FFFF]:-
\a-lead-surrogate-and-a-trail-surrogate \ \}
\}
{ Lead-surrogates-are-pairs-of-bytes-in-the-range-[0xD800,-0xDBFF],-
\ and-trail-surrogates-are-in-the-range-[0xDC00,-0xDFFF].
\}
\flag_if_raised:nT { str_missing }
\}
\\
\{ \LaTeX\-came-across-a-trail-surrogate-when-it-was-not-expected.
\}
\flag_if_raised:nT { str_extra }
\}
\\
\flag_if_raised:nT { str_end }
\}
\\
The-string-contained-an-odd-number-of-bytes.-This-is-invalid:-
\the-basic-code-unit-for-UTF-16-is-16-bits-(2-bytes).\)
\)
\)
(End definition for \l__str_missing_flag, \l__str_extra_flag, and \l__str_end_flag.)
\_\_str_convert_decode_utf16:
\_\_str_convert_decode_utf16be:
\_\_str_convert_decode_utf16le:
\_\_str_decode_utf_xvi_bom:NN
\_\_str_decode_utf_xvi:Nw

As for UTF-8, decoding UTF-16 is harder than encoding it. If the endianness is unknown, check the first two bytes: if those are *FE and *FF in either order, remove them and use the corresponding endianness, otherwise assume big-endianness. The three endianness cases are based on a common auxiliary whose first argument is 1 for big-endian and 2 for little-endian, and whose second argument, delimited by the scan mark \s__str_stop, is expanded once (the string may be long; passing \g__str_result_tl as an argument before expansion is cheaper).

The \_\_str_decode_utf_xvi:Nw function defines \_\_str_tmp:w to take two arguments and return the character code of the first one if the string is big-endian, and the second one if the string is little-endian, then loops over the string using \_\_str_decode_utf_xvi_pair:NN described below.
Bytes are read two at a time. At this stage, `\@_tmp:w #1#2` expands to the character code of the most significant byte, and we distinguish cases depending on which range it lies in:

- `["D8", "DB"]` signals a lead surrogate, and the integer expression yields 1 (ε-TEX rounds ties away from zero);
- `["DC", "DF"]` signals a trail surrogate, unexpected here, and the integer expression yields 2;
- any other value signals a code point in the Basic Multilingual Plane, which stands for itself, and the `\if_case:w` construction expands to nothing (cases other than 1 or 2), leaving the relevant material in the input stream, followed by another call to the `_pair` auxiliary.

The case of a lead surrogate is treated by the `_quad` auxiliary, whose arguments #1, #2, #4 and #5 are the four bytes. We expect the most significant byte of #4#5 to be in the range `["DC", "DF"]` (trail surrogate). The test is similar to the test used for continuation bytes.
in the UTF-8 decoding functions. In the case where \#4\#5 is indeed a trail surrogate, leave 
\#1#2\#4\#5 \texttt{\texttt{\_s\_str}} (code point) \texttt{\_s\_str}, and remove the pair \#4\#5 before looping with \\texttt{\_str\_decode\_utf\_xvi\_pair:NN}. Otherwise, of course, complain about the missing surrogate.

The magic number "D7F7 is such that "D7F7∗"400 = "D800∗"400+"DC00−"10000.

Every time we read a pair of bytes, we test for the end-marker \texttt{\_q\_str\_nil}. When reaching the end, we additionally check that the string had an even length. Also, if the end is reached when expecting a trail surrogate, we treat that as a missing surrogate.

\begin{verbatim}
\cs_new:Npn \_str_decode_utf_xvi_pair:NN \#1\#2
{  
  \if_meaning:w \_q\_str\_nil \#2
    \_str_decode_utf_xvi_pair_end:Nw \#1
  \fi:
  \if_case:w
    \int_eval:n { ( \_str_tmp:w \#1\#2 - "D6 ) / 4 } \scan_stop:
    \or: \exp_after:wN \_str_decode_utf_xvi_quad:NNwNN \#1\#2
    \or: \exp_after:wN \_str_decode_utf_xvi_extra:NNw \#1\#2
  \fi:
  \#1\#2 \s\_str
  \int_eval:n { "100 * \_str_tmp:w \#1\#2 + \_str_tmp:w \#2\#1 } \s\_str
  \_str_decode_utf_xvi_pair:NN
}
\cs_new:Npn \_str_decode_utf_xvi_quad:NNwNN \#1\#2 \#3 \_str_decode_utf_xvi_pair:NN \#4\#5
{  
  \if_meaning:w \_q\_str\_nil \#5
    \_str_decode_utf_xvi_error:nNN { missing } \#1\#2
  \_str_decode_utf_xvi_pair_end:Nw \#4
  \fi:
  \if_int_compare:w
    \if_int_compare:w \_str_tmp:w \#4\#5 < "DC \exp_stop_f:
      0 = 1
    \else:
      \_str_tmp:w \#4\#5 < "E0
    \fi:
    \exp_stop_f:
    \#1 \#2 \#4 \#5 \s\_str
  \int_eval:n
  { ( "100 * \_str_tmp:w \#1\#2 + \_str_tmp:w \#2\#1 - "D7F7 ) * "400
    + "100 * \_str_tmp:w \#4\#5 + \_str_tmp:w \#5\#4
  )}
  \s\_str
  \exp_after:wN \use_i:nnn
  \else:
    \_str_decode_utf_xvi_error:nNN { missing } \#1\#2
  \_str_decode_utf_xvi_pair:NN \#4\#5
  \fi:
}
\cs_new:Npn \_str_decode_utf_xvi_pair_end:Nw \#1 \fi:
{  
  \if_meaning:w \_q\_str\_nil \#1
}\end{verbatim}

754
Else:
  \__str_decode_utf_xvi_error:nNN \{ end } \#1 \prg_do_nothing:
  \if:
  \prg_break:
  \}
\cs_new:Npn \__str_decode_utf_xvi_extra:NNw \#1#2 \s__str \#3 \s__str
{ \__str_decode_utf_xvi_error:nNN \{ extra \} \#1#2 }
\cs_new:Npn \__str_decode_utf_xvi_error:nNN #1#2#3
{ \flag_raise:n \{ str_error \}
  \flag_raise:n \{ str_#1 \}
  \#2 \#3 \s__str
  \int_use:N \c__str_replacement_char_int \s__str
}

(End definition for \__str_decode_utf_xvi_pair:NN and others.)

Restore the original catcodes of bytes 254 and 255.
\group_end:

\__str_convert_encode_utf32:
\__str_convert_encode_utf32be:
\__str_convert_encode_utf32le:
\__str_encode_utf_xxxii_be:n
\__str_encode_utf_xxxii_be_aux:nn
\__str_encode_utf_xxxii_le:n
\__str_encode_utf_xxxii_le_aux:nn

Convert each integer in the comma-list \g__str_result_tl to a sequence of four bytes. The functions for big-endian and little-endian encodings are very similar, but the \__str_output_byte:n instructions are reversed.

\cs_new_protected:cpn { \__str_convert_encode_utf32: }
{ \__str_convert_gmap_internal:N \__str_encode_utf_xxxii_be:n
  \tl_gput_left:Nx \g__str_result_tl \{ ^^00 ^^00 ^^fe ^^ff \}
}
\cs_new_protected:cpn { \__str_convert_encode_utf32be: }
{ \__str_convert_gmap_internal:N \__str_encode_utf_xxxii_be:n
  \tl_gput_left:Nx \g__str_result_tl \{ ^^00 ^^00 ^^fe ^^ff \}
}
\cs_new_protected:cpn { \__str_convert_encode_utf32le: }
{ \__str_convert_gmap_internal:N \__str_encode_utf_xxxii_le:n
  \tl_gput_left:Nx \g__str_result_tl \{ ^^00 ^^00 ^^fe ^^ff \}
}

54.6.3 utf-32 support

The definitions are done in a category code regime where the bytes 0, 254 and 255 used by the byte order mark have catcode “other”.

\group_begin:
\char_set_catcode_other:N \^^00
\char_set_catcode_other:N \^^fe
\char_set_catcode_other:N \^^ff
\__str_convert_gmap_internal:N \__str_encode_utf_xxxii_be:n
\tl_gput_left:Nx \g__str_result_tl \{ ^^00 ^^00 ^^fe ^^ff \}
\group_end:

\__str_convert_gmap_internal:N \__str_encode_utf_xxxii_le:n
\tl_gput_left:Nx \g__str_result_tl \{ ^^00 ^^00 ^^fe ^^ff \}

\__str_convert_gmap_internal:N \__str_encode_utf_xxxii_be:n
\tl_gput_left:Nx \g__str_result_tl \{ ^^00 ^^00 ^^fe ^^ff \}

\__str_convert_gmap_internal:N \__str_encode_utf_xxxii_le:n
\tl_gput_left:Nx \g__str_result_tl \{ ^^00 ^^00 ^^fe ^^ff \}

\__str_convert_gmap_internal:N \__str_encode_utf_xxxii_be:n
\tl_gput_left:Nx \g__str_result_tl \{ ^^00 ^^00 ^^fe ^^ff \}

\__str_convert_gmap_internal:N \__str_encode_utf_xxxii_le:n
\tl_gput_left:Nx \g__str_result_tl \{ ^^00 ^^00 ^^fe ^^ff \}

\exp_args:Nf \__str_encode_utf_xxxii_be_aux:nn
  \{ \int_div_truncate:nn \{#1\} \{ "100 \} \{#1\} \}
\exp_args:Nf \__str_encode_utf_xxxii_be_aux:nn
  \{ \int_div_truncate:nn \{#1\} \{ "100 \} \{#1\} \}

\cs_new:Npn \__str_encode_utf_xxxii_be:n #1
{ \exp_args:Nf \__str_encode_utf_xxxii_be_aux:nn
  \{ \int_div_truncate:nn \{#1\} \{ "100 \} \{#1\} \}
\cs_new:Npn \__str_encode_utf_xxxii_be:n #1
{ \exp_args:Nf \__str_encode_utf_xxxii_be_aux:nn
  \{ \int_div_truncate:nn \{#1\} \{ "100 \} \{#1\} \}
\cs_new:Npn \__str_encode_utf_xxxii_le:n #1
{ \exp_args:Nf \__str_encode_utf_xxxii_le_aux:nn
  \{ \int_div_truncate:nn \{#1\} \{ "100 \} \{#1\} \}
\cs_new:Npn \__str_encode_utf_xxxii_le:n #1
{ \exp_args:Nf \__str_encode_utf_xxxii_le_aux:nn
  \{ \int_div_truncate:nn \{#1\} \{ "100 \} \{#1\} \}
\exp_args:Nf \__str_encode_utf_xxxii_le_aux:nn
\{ \int_div_truncate:nn \{#1\} \{ "100 \}\{#1\}\}
\cs_new:Npn \__str_encode_utf_xxxii_le_aux:nn #1#2
\{ \__str_output_byte:n { #2 - #1 * "100 } \__str_output_byte_pair_le:n \{#1\}\}
\^^00
(End definition for \__str_convert_encode_utf32: and others.)

\exp_last_unbraced:Nf \use_none:n
\{ \__str_if_flag_times:nT { str_overflow } { ,~code~point~too~large } \__str_if_flag_times:nT { str_end } { ,~truncated~string } \}
\flag_clear_new:n { str_overflow }
\flag_clear_new:n { str_end }
\msg_new:nnnn { str } { utf32-decode }
\{ \Invalid~UTF-32~string: \}
\\LaTeX~came~across~a~code~point~larger~than~1114111,~the~maximum~code~point~defined~by~Unicode.~Perhaps~the~string~was~not~encoded~in~the~UTF-32~encoding?
\flag_if_raised:nT { str_end }
\{ \The~length~of~the~string~is~not~a~multiple~of~4.~Perhaps~the~string~was~truncated? \}
(End definition for str_overflow and str_end. These variables are documented on page 77.)

\__str_convert_decode_utf32:
\__str_convert_decode_utf32be:
\__str_convert_decode_utf32le:
\__str_decode_utf_xxxii_bom:NNNN
\__str_decode_utf_xxxii:Nw
\__str_decode_utf_xxxii_loop:NNNN
\__str_decode_utf_xxxii_end:w

There can be no error when encoding in UTF-32. When decoding, the string may not have length \textbf{4}n, or it may contain code points larger than "10FFFF. The latter case often happens if the encoding was in fact not UTF-32, because most arbitrary strings are not valid in UTF-32.

The structure is similar to UTF-16 decoding functions. If the endianness is not given, test the first 4 bytes of the string (possibly \_\_\str_stop if the string is too short) for the presence of a byte-order mark. If there is a byte-order mark, use that endianness, and remove the 4 bytes, otherwise default to big-endian, and leave the 4 bytes in place. The \_\_\str_decode_utf_xxxii:Nw auxiliary receives 1 or 2 as its first argument indicating endianness, and the string to convert as its second argument (expanded or not). It sets
\_\_str\_tmp:w to expand to the character code of either of its two arguments depending on endianness, then triggers the _\_loop auxiliary inside an x-expanding assignment to \_g\_\_str\_result tl.

The _\_loop auxiliary first checks for the end-of-string marker \_s\_\_str\_stop, calling the _\_end auxiliary if appropriate. Otherwise, leave the (4 bytes) \_s\_\_str behind, then check that the code point is not overflowing: the leading byte must be 0, and the following byte at most 16.

In the ending code, we check that there remains no byte: there should be nothing left until the first \_s\_\_str\_stop. Break the map.

\cs_new_protected:cpn { __str_convert_decode_utf32be: } \cs_new_protected:cpn { __str_decode_utf_xxxii:Nw 1 \g__str_result_tl \s__str_stop } \cs_new_protected:cpn { __str_convert_decode_utf32le: } \cs_new_protected:cpn { __str_decode_utf_xxxii:Nw 2 \g__str_result_tl \s__str_stop } \cs_new_protected:cpn { __str_convert_decode_utf32: } \exp_after:wN \__str_decode_utf_xxxii_bom:NNNN \g__str_result_tl \s__str_stop \s__str_stop \s__str_stop \s__str_stop \s__str_stop \cs_new_protected:Npn \__str_decode_utf_xxxii_bom:NNNN #1#2#3#4 \cs_new_protected:Npn \__str_decode_utf_xxxii:Nw #1#2 \s__str_stop \flag_clear:n { str_overflow } \flag_clear:n { str_end } \flag_clear:n { str_error } \cs_set:Npn \__str_tmp:w ##1 ##2 { ' ## #1 } \__kernel_tl_gset:Nx \g__str_result_tl \exp_after:wN \__str_decode_utf_xxxii_loop:NNNN \#2 \s__str_stop \s__str_stop \s__str_stop \s__str_stop \s__str_stop \cs_new_protected:Npn \__str_decode_utf_xxxii_loop:NNNN \cs_new_protected:Npn \__str_decode_utf_xxxii_loop:NNNN #1\#2\#3\#4 \if_meaning:w \s\_\_str\_stop \#4 \exp_after:wN \__str_decode_utf_xxxii_end:w \fi: \#1\#2\#3\#4 \s\_\_str \if_int_compare:w \__str_tmp:w \#1#4 > \c_zero_int \flag_raise:n { str_overflow } \flag_raise:n { str_error } \int_use:N \c__str_replacement_char_int \else: \cs_new_protected:Npn \__str_decode_utf_xxxii_loop:NNNN \cs_new_protected:Npn \__str_decode_utf_xxxii_loop:NNNN #1\#2\#3\#4 \exp_after:wN \__str_decode_utf_xxxii_loop:NNNN \#2 \s\_\_str\_stop \s\_\_str\_stop \s\_\_str\_stop \s\_\_str\_stop \s\_\_str\_stop \prg_break_point: \_\_str_if_flag_error:nnx { str_error } { utf32-decode } { } \cs_new:Npn \__str_decode_utf_xxxii_loop:NNNN \cs_new:Npn \__str_decode_utf_xxxii_loop:NNNN #1\#2\#3\#4 \if_meaning:w \s\_\_str\_stop \#4 \exp_after:wN \__str_decode_utf_xxxii_end:w \fi: \#1\#2\#3\#4 \s\_\_str \if_int_compare:w \__str_tmp:w \#1#4 > \c_zero_int \flag_raise:n { str_overflow } \flag_raise:n { str_error } \int_use:N \c__str_replacement_char_int \else:
54.7 PDF names and strings by expansion

To convert to PDF names by expansion, we work purely on UTF-8 input. The first step is to make a string with “other” spaces, after which we use a simple token-by-token approach. In Unicode engines, we break down everything before one-byte codepoints, but for 8-bit engines there is no need to worry. Actual escaping is covered by the same code as used in the non-expandable route.
\exp_args:Nn \__str_convert_pdfname_bytes_aux:n
\{ \char_to_utfviii_bytes:n {\#1} \}
\cs_new:Npn \__str_convert_pdfname_bytes_aux:n #1
\{ \__str_convert_pdfname_bytes_aux:nnnn #1 \}
\cs_new:Npx \__str_convert_pdfname_bytes_aux:nnnn #1#2#3#4
\{ \c_hash_str \exp_not:N \__str_output_hexadecimal:n {#1}
\c_hash_str \exp_not:N \__str_output_hexadecimal:n {#2}
\exp_not:N \tl_if_blank:nF {#3}
\{ \c_hash_str \exp_not:N \__str_output_hexadecimal:n {#3}
\exp_not:N \tl_if_blank:nF {#4}
\{ \c_hash_str \exp_not:N \__str_output_hexadecimal:n {#4}
\}
\}
\}
\cs_new_eq:NN \__str_convert_pdfname:n \__str_escape_name_char:n
\end{definition}
for \str_convert_pdfname:n and others. This function is documented on page 135.

\section*{54.7.1 iso 8859 support}

The iso-8859-1 encoding exactly matches with the 256 first Unicode characters. For other 8-bit encodings of the iso-8859 family, we keep track only of differences, and of unassigned bytes.

\begin{verbatim}
\__str_declare_eight_bit_encoding:nnnn { iso88591 } { 256 }
\{ { A1 } { 0104 } { A2 } { 02D8 } { A3 } { 0141 } { A5 } { 013D } { A6 } { 015A } { A9 } { 0160 } { AA } { 015E } { AB } { 0164 } { AC } { 0179 } { AE } { 017D } { AF } { 017B } { B1 } { 0105 } { B2 } { 02D8 } { B3 } { 0142 }
\}
\end{verbatim}

759
\_\_str\_declare\_eight\_bit\_encoding:nnnn { iso88593 } { 384 }

{ A1 } { 0126 }
{ A2 } { 02D8 }
{ A6 } { 0124 }
{ A9 } { 0130 }
\__str_declare_eight_bit_encoding:nnnn { iso88594 } { 383 }

{ iso88594 }
\langle iso88593 \rangle
{ B9 } { 0161 }
{ BA } { 0113 }
{ BB } { 0123 }
{ BC } { 0167 }
{ BD } { 014A }
{ BE } { 017E }
{ BF } { 014B }
{ CO } { 0100 }
{ C7 } { 012E }
{ C8 } { 010C }
{ CA } { 0118 }
{ CC } { 0116 }
{ CF } { 012A }
{ D0 } { 0110 }
{ D1 } { 0145 }
{ D2 } { 014C }
{ D3 } { 0136 }
{ D9 } { 0172 }
{ DD } { 0168 }
{ DE } { 016A }
{ E0 } { 0101 }
{ E7 } { 012F }
{ E8 } { 010D }
{ EA } { 0119 }
{ EC } { 0117 }
{ EF } { 012B }
{ F0 } { 0111 }
{ F1 } { 0146 }
{ F2 } { 014D }
{ F3 } { 0137 }
{ F9 } { 0173 }
{ FD } { 0169 }
{ FE } { 016B }
{ FF } { 02D9 }

{ A1 } { 0401 }
{ A2 } { 0402 }
{ A3 } { 0403 }
{ A4 } { 0404 }
{ A5 } { 0405 }
{ A6 } { 0406 }
{ A7 } { 0407 }
{ A8 } { 0408 }
{ A9 } { 0409 }
{ AA } { 040A }
{ AB } { 040B }
{ AC } { 040C }
{ AE } { 040E }

762
763
\__str_declare_eight_bit_encoding:nnnn \{ iso88596 \} \{ 344 \} 

\{ AC \} \{ 060C \}
\{ BB \} \{ 061B \}
\{ BF \} \{ 061F \}
\{ C1 \} \{ 0621 \}
\{ C2 \} \{ 0622 \}
\{ C3 \} \{ 0623 \}
\{ C4 \} \{ 0624 \}
\{ C5 \} \{ 0625 \}
\{ C6 \} \{ 0626 \}
\{ C7 \} \{ 0627 \}
\{ C8 \} \{ 0628 \}
\{ C9 \} \{ 0629 \}
\{ CA \} \{ 062A \}
\{ CB \} \{ 062B \}
\{ CC \} \{ 062C \}
\{ CD \} \{ 062D \}
\{ CE \} \{ 062E \}
\{ CF \} \{ 062F \}
\{ D0 \} \{ 0630 \}
\{ D1 \} \{ 0631 \}
\__str__declared_eight_bit_encoding:nnnn { iso88597 } { 498 }

{ 498 }

{ iso88597 }

{ iso88596 }

{ A1 } { 2018 }
{ A2 } { 2019 }
{ A4 } { 20AC }
{ A5 } { 20AF }
{ AA } { 037A }
{ AF } { 2015 }
{ B4 } { 0384 }
{ B5 } { 0385 }
{ B6 } { 0386 }
{ B8 } { 0388 }
{ B9 } { 0389 }
{ BA } { 038A }
{ BC } { 038C }
{ BE } { 038E }
{ BF } { 038F }
{ C0 } { 0390 }
{ C1 } { 0391 }
{ C2 } { 0392 }
{ C3 } { 0393 }
{ C4 } { 0394 }
{ C5 } { 0395 }
{ C6 } { 0396 }
{ C7 } { 0397 }
{ C8 } { 0398 }
{ C9 } { 0399 }
{ CA } { 039A }
{ CB } { 039B }
{ CC } { 039C }
{ CD } { 039D }
{ CE } { 039E }
{ CF } { 039F }
{ D0 } { 03A0 }
{ D1 } { 03A1 }
{ D3 } { 03A3 }
{ D4 } { 03A4 }
{ D5 } { 03A5 }
{ D6 } { 03A6 }
{ D7 } { 03A7 }
{ D8 } { 03A8 }
{ D9 } { 03A9 }
{ DA } { 03AA }
\__str_declare_eight_bit_encoding:nnnn { iso88598 } { 308 }

{ AE } { D2 }

{ iso88597 }

\__str_declare_eight_bit_encoding:nnnn { iso88598 } { 308 }

{ AA } { 00D7 }

{ BA } { 00F7 }

{ DF } { 2017 }

{ E0 } { 05D0 }

{ E1 } { 05D1 }

{ E2 } { 05D2 }

{ E3 } { 05D3 }

{ E4 } { 05D4 }

{ E5 } { 05D5 }
{ E6 } { 05D6 }
{ E7 } { 05D7 }
{ E8 } { 05D8 }
{ E9 } { 05D9 }
{ EA } { 05DA }
{ EB } { 05DB }
{ EC } { 05DC }
{ ED } { 05DD }
{ EE } { 05DE }
{ EF } { 05DF }
{ F0 } { 05E0 }
{ F1 } { 05E1 }
{ F2 } { 05E2 }
{ F3 } { 05E3 }
{ F4 } { 05E4 }
{ F5 } { 05E5 }
{ F6 } { 05E6 }
{ F7 } { 05E7 }
{ F8 } { 05E8 }
{ F9 } { 05E9 }
{ FA } { 05EA }
{ FD } { 200E }
{ FE } { 200F }

{ A1 }
{ BF }
{ C0 }
{ C1 }
{ C2 }
{ C3 }
{ C4 }
{ C5 }
{ C6 }
{ C7 }
{ C8 }
{ C9 }
{ CA }
{ CB }
{ CC }
{ CD }
{ CE }
{ CF }
{ D0 }
{ D1 }
{ D2 }
{ D3 }
{ D4 }
{ D5 }
{ D6 }
{ D7 }
{ D8 }
{ D9 }
{ DA }

768
\__str_declare_eight_bit_encoding:nnnn { iso88599 } { 352 }
{ DO } { 011E }
{ DD } { 0130 }
{ DE } { 015E }
{ F0 } { 011F }
{ FD } { 0131 }
{ FE } { 015F }

\__str_declare_eight_bit_encoding:nnnn { iso885910 } { 383 }
{ A1 } { 0104 }
{ A2 } { 0112 }
{ A3 } { 0122 }
{ A4 } { 012A }
{ A5 } { 0128 }
{ A6 } { 0136 }
{ A8 } { 013B }
{ A9 } { 0110 }
{ AA } { 0160 }
{ AB } { 0166 }
{ AC } { 017D }
{ AE } { 016A }
{ AF } { 014A }
{ B1 } { 0105 }
{ B2 } { 0113 }
{ B3 } { 0123 }
{ B4 } { 012B }
{ B5 } { 0129 }
{ B6 } { 0137 }
{ B8 } { 013C }
{ B9 } { 0111 }
{ BA } { 0161 }
{ BB } { 0167 }
{ BC } { 017E }
{ BD } { 2015 }
{ BE } { 016B }
{ BF } { 014B }
{ CO } { 0100 }
{ C7 } { 012E }

769
\"__str_declare_eight_bit_encoding:\nnnn { iso885911 } \{ 369 \}
{
    { A1 } { OE01 }
    { A2 } { OE02 }
    { A3 } { OE03 }
    { A4 } { OE04 }
    { A5 } { OE05 }
    { A6 } { OE06 }
    { A7 } { OE07 }
    { A8 } { OE08 }
    { A9 } { OE09 }
    { AA } { OE0A }
    { AB } { OE0B }
    { AC } { OE0C }
    { AD } { OE0D }
    { AE } { OE0E }
    { AF } { OE0F }
    { B0 } { OE10 }
    { B1 } { OE11 }
    { B2 } { OE12 }
    { B3 } { OE13 }
    { B4 } { OE14 }
    { B5 } { OE15 }
    { B6 } { OE16 }
    { B7 } { OE17 }
    { B8 } { OE18 }
    { B9 } { OE19 }
    { BA } { OE1A }
    { BB } { OE1B }
    { BC } { OE1C }
    { BD } { OE1D }
    { BE } { OE1E }
}
| \( \text{BF} \) | \( \text{OE1F} \) |
| \( \text{C0} \) | \( \text{OE20} \) |
| \( \text{C1} \) | \( \text{OE21} \) |
| \( \text{C2} \) | \( \text{OE22} \) |
| \( \text{C3} \) | \( \text{OE23} \) |
| \( \text{C4} \) | \( \text{OE24} \) |
| \( \text{C5} \) | \( \text{OE25} \) |
| \( \text{C6} \) | \( \text{OE26} \) |
| \( \text{C7} \) | \( \text{OE27} \) |
| \( \text{C8} \) | \( \text{OE28} \) |
| \( \text{C9} \) | \( \text{OE29} \) |
| \( \text{CA} \) | \( \text{OE2A} \) |
| \( \text{CB} \) | \( \text{OE2B} \) |
| \( \text{CC} \) | \( \text{OE2C} \) |
| \( \text{CD} \) | \( \text{OE2D} \) |
| \( \text{CE} \) | \( \text{OE2E} \) |
| \( \text{CF} \) | \( \text{OE2F} \) |
| \( \text{D0} \) | \( \text{OE30} \) |
| \( \text{D1} \) | \( \text{OE31} \) |
| \( \text{D2} \) | \( \text{OE32} \) |
| \( \text{D3} \) | \( \text{OE33} \) |
| \( \text{D4} \) | \( \text{OE34} \) |
| \( \text{D5} \) | \( \text{OE35} \) |
| \( \text{D6} \) | \( \text{OE36} \) |
| \( \text{D7} \) | \( \text{OE37} \) |
| \( \text{D8} \) | \( \text{OE38} \) |
| \( \text{D9} \) | \( \text{OE39} \) |
| \( \text{DA} \) | \( \text{OE3A} \) |
| \( \text{DF} \) | \( \text{OE3F} \) |
| \( \text{E0} \) | \( \text{OE40} \) |
| \( \text{E1} \) | \( \text{OE41} \) |
| \( \text{E2} \) | \( \text{OE42} \) |
| \( \text{E3} \) | \( \text{OE43} \) |
| \( \text{E4} \) | \( \text{OE44} \) |
| \( \text{E5} \) | \( \text{OE45} \) |
| \( \text{E6} \) | \( \text{OE46} \) |
| \( \text{E7} \) | \( \text{OE47} \) |
| \( \text{E8} \) | \( \text{OE48} \) |
| \( \text{E9} \) | \( \text{OE49} \) |
| \( \text{EA} \) | \( \text{OE4A} \) |
| \( \text{EB} \) | \( \text{OE4B} \) |
| \( \text{EC} \) | \( \text{OE4C} \) |
| \( \text{ED} \) | \( \text{OE4D} \) |
| \( \text{EE} \) | \( \text{OE4E} \) |
| \( \text{EF} \) | \( \text{OE4F} \) |
| \( \text{F0} \) | \( \text{OE50} \) |
| \( \text{F1} \) | \( \text{OE51} \) |
| \( \text{F2} \) | \( \text{OE52} \) |
| \( \text{F3} \) | \( \text{OE53} \) |
| \( \text{F4} \) | \( \text{OE54} \) |
| \( \text{F5} \) | \( \text{OE55} \) |
| \( \text{F6} \) | \( \text{OE56} \) |
| \( \text{F7} \) | \( \text{OE57} \) |
| \( \text{F8} \) | \( \text{OE58} \) |
1536  { F9 } { 0E59 }
1537  { FA } { 0E5A }
1538  { FB } { 0E5B }
1539  }
1540  {
1541  { DB }
1542  { DC }
1543  { DD }
1544  { DE }
1545  }
1546  ⟨/iso885911⟩
1547  ⟨∗iso885913⟩
1548  \_str_declare_eight_bit_encoding:nmnn { iso885913 } { 399 } {
1549  {
1550  { A1 } { 201D }
1551  { A5 } { 201E }
1552  { A8 } { 00D8 }
1553  { AA } { 0156 }
1554  { AF } { 00C6 }
1555  { B4 } { 201C }
1556  { B8 } { 00F8 }
1557  { BA } { 0157 }
1558  { BF } { 00E6 }
1559  { C0 } { 0104 }
1560  { C1 } { 012E }
1561  { C2 } { 0100 }
1562  { C3 } { 0106 }
1563  { C6 } { 0118 }
1564  { C7 } { 0112 }
1565  { C8 } { 010C }
1566  { CA } { 0179 }
1567  { CB } { 0116 }
1568  { CC } { 0122 }
1569  { CD } { 0136 }
1570  { CE } { 012A }
1571  { CF } { 013B }
1572  { D0 } { 0160 }
1573  { D1 } { 0143 }
1574  { D2 } { 0145 }
1575  { D4 } { 014C }
1576  { D8 } { 0172 }
1577  { D9 } { 0141 }
1578  { DA } { 015A }
1579  { DB } { 016A }
1580  { DD } { 017B }
1581  { DE } { 017D }
1582  { E0 } { 0105 }
1583  { E1 } { 012F }
1584  { E2 } { 0101 }
1585  { E3 } { 0107 }
1586  { E6 } { 0119 }
1587  { E7 } { 0113 }
1588  { E8 } { 010D }
1589  { EA } { 017A }

772
\_\_str\_declare\_eight\_bit\_encoding:nnnn { iso885914 } { 529 }

{ A1 } { 1E02 }
{ A2 } { 1E03 }
{ A4 } { 010A }
{ A5 } { 010B }
{ A6 } { 1E0A }
{ A8 } { 1E80 }
{ AA } { 1E82 }
{ AB } { 1E8B }
{ AC } { 1E8F }
{ AF } { 0178 }
{ B0 } { 1E1E }
{ B1 } { 1E1F }
{ B2 } { 0120 }
{ B3 } { 0121 }
{ B4 } { 1E40 }
{ B5 } { 1E41 }
{ B7 } { 1E56 }
{ B8 } { 1E81 }
{ B9 } { 1E57 }
{ BA } { 1E83 }
{ BB } { 1E60 }
{ BC } { 1E93 }
{ BD } { 1E84 }
{ BE } { 1E85 }
{ BF } { 1E61 }
{ DO } { 0174 }
{ D7 } { 1E6A }
{ DE } { 0176 }
{ F0 } { 0175 }
{ F7 } { 1E6B }
{ FE } { 0177 }
__str_declare_eight_bit_encoding:nnnn { iso885914 } { 383 }

{  
  A4
  A6
  A8
  B4
  B8
  BC
  BD
  BE
}

{  
  0160
  0161
  017D
  017E
  0152
  0153
  0178
}

__str_declare_eight_bit_encoding:nnnn { iso885915 } { 383 }

{  
  A4
  A6
  A8
  B4
  B8
  BC
  BD
  BE
}

{  
  0160
  0161
  017D
  017E
  0152
  0153
  0178
}

__str_declare_eight_bit_encoding:nnnn { iso885916 } { 558 }

{  
  A1
  A2
  A3
  A4
  A5
  A6
  A8
  AA
  AC
  AE
  AF
  B2
  B3
  B4
  B5
  B8
  B9
  BA
  BC
  BD
  BE
  BF
  C3
  C5
  C7
  D0
  D1
  D5
  D7
  D8
  DD
  DE
}
\langle/iso885916>
Chapter 55

\textbf{l3quark implementation}

The following test files are used for this code: \texttt{m3quark001.lvt}.

\section{Quarks}

\texttt{\quark\_new:N} Allocate a new quark.

\begin{verbatim}
\cs_new_protected:Npn \quark\_new:N #1
{\__kernel\_chk\_if\_free\_cs:N #1\cs\_gset\_nopar:Npn #1 {#1}}
\end{verbatim}

(End definition for \texttt{\quark\_new:N}. This function is documented on page 138.)

\texttt{\textbackslash q\_nil, \textbackslash q\_mark, \textbackslash q\_no\_value, \textbackslash q\_stop} Some “public” quarks. \texttt{\_stop} is an “end of argument” marker, \texttt{\_nil} is an empty value and \texttt{\_no\_value} marks an empty argument.

\begin{verbatim}
\quark\_new:N \_nil
\quark\_new:N \_mark
\quark\_new:N \_no\_value
\quark\_new:N \_stop
\end{verbatim}

(End definition for \texttt{\_nil} and others. These variables are documented on page 138.)

\texttt{\textbackslash q\_recursion\_tail, \textbackslash q\_recursion\_stop} Quarks for ending recursions. Only ever used there! \texttt{\_recursion\_tail} is appended to whatever list structure we are doing recursion on, meaning it is added as a proper list item with whatever list separator is in use. \texttt{\_recursion\_stop} is placed directly after the list.

\begin{verbatim}
\quark\_new:N \_recursion\_tail
\quark\_new:N \_recursion\_stop
\end{verbatim}

(End definition for \texttt{\_recursion\_tail} and \texttt{\_recursion\_stop}. These variables are documented on page 139.)

\texttt{\_s\_quark} Private scan mark used in l3quark. We don’t have l3scan yet, so we declare the scan mark here and add it to the scan mark pool later.

\begin{verbatim}
\cs\_new\_eq:N\_N \_s\_quark \scan\_stop:
\end{verbatim}
Private quark use for some tests.

When doing recursions, it is easy to spend a lot of time testing if the end marker has been found. To avoid this, a dedicated end marker is used each time a recursion is set up. Thus if the marker is found everything can be wrapper up and finished off. The simple case is when the test can guarantee that only a single token is being tested. In this case, there is just a dedicated copy of the standard quark test. Both a gobbling version and one inserting end code are provided.

See \texttt{\quark_if_recursion_tail_stop:nTF} for the details. Expanding \texttt{\_quark_if_recursion_tail:w} once in front of the tokens chosen here gives an empty result if and only if \texttt{#1} is exactly \texttt{\_quark_recursion_tail}. 

(End definition for \texttt{\quark_if_recursion_tail_stop:N}, \texttt{\quark_if_recursion_tail_stop:Do:N}, and \texttt{\_quark_if_recursion_tail:w}. These functions are documented on page 139.)
Analogues of the \quark_if_recursion_tail_stop... functions. Break the mapping using \#2.

\cs_new:Npn \quark_if_recursion_tail_break:NN #1#2
{\if_meaning:w \q_recursion_tail #1 \exp_after:wN #2 \fi:}
\cs_new:Npn \quark_if_recursion_tail_break:nN #1#2
{\tl_if_empty:oT { \__quark_if_recursion_tail:w {} #1 {} ?! \q_recursion_tail ??! } {#2}}

(End definition for \quark_if_recursion_tail_break:NN and \quark_if_recursion_tail_break:nN. These functions are documented on page 140.)

Here we test if we found a special quark as the first argument. We better start with \q_no_value as the first argument since the whole thing may otherwise loop if \#1 is wrongly given a string like aabc instead of a single token.\footnote{It may still loop in special circumstances however!}

\prg_new_conditional:Npnn \quark_if_nil:N #1 { p, T , F , TF }
{\if_meaning:w \q_nil #1 \prg_return_true: \else: \prg_return_false: \fi:}
\prg_new_conditional:Npnn \quark_if_no_value:N #1 { p, T , F , TF }
{\if_meaning:w \q_no_value #1 \prg_return_true: \else: \prg_return_false: \fi:}
\prg_generate_conditional_variant:Nnn \quark_if_no_value:N { c } { p , T , F , TF }

(End definition for \quark_if_nil:NTF and \quark_if_no_value:NTF. These functions are documented on page 138.)

Let us explain \quark_if_nil:nTF. Expanding \__quark_if_nil:w once is safe thanks to the trailing \q_nil ??!. The result of expanding once is empty if and only if both delimited arguments \#1 and \#2 are empty and \#3 is delimited by the last tokens ?!. Thanks to the leading {}, the argument \#1 is empty if and only if the argument of \quark_if_nil:n starts with \q_nil. The argument \#2 is empty if and only if this \q_nil is followed immediately by ? or by {}, coming either from the trailing tokens in the definition of \quark_if_nil:n, or from its argument. In the first case, \__quark_if:nil:w is followed by \{ \q_nil {} ?! \q_nil ??!, hence \#3 is delimited by the final ?!, and the test returns true as wanted. In the second case, the result is not empty since
The first ?! in the definition of \texttt{\quarkifnil:n} stop #3. The auxiliary here is the same as \texttt{\tlifempty:o}, with the same comments applying.

\begin{verbatim}
\prg_new_conditional:Npnn \quarkifnil:n \quarkifno_value:n #1 \quarkifrecursion_tail:n \quarkifrecursion_tail_stop:n \quarkifrecursion_tail_break:n \quarkifrecursion_tail_stop_do:n \quarkifrecursion_tail_break_do:n \quarkifrecursion_tail_stop:N \quarkifrecursion_tail_break:N \quarkifrecursion_tail_stop_do:N \quarkifrecursion_tail_break_do:N
\end{verbatim}

(End definition for \quarkifnil:nTF and others. These functions are documented on page 138.)

The function \texttt{\kernel_quark_new_test:N} defines #1 in a similar way as \texttt{\quarkifrecursion_tail...} functions (as described below), using \texttt{\__\langle(namespace)\__\_recursion_tail} as the test quark and \texttt{\__\langle(namespace)\__\_recursion_stop} as the delimiter quark, where the \langle(namespace)\rangle is determined as the first \_\_delimited part in #1.

There are six possible function types which this function can define, and which is defined depends on the signature of the function being defined:

:n gives an analogue of \texttt{\quarkifrecursion_tail_stop:n}
:nn gives an analogue of \texttt{\quarkifrecursion_tail_stop_do:nn}
:nN gives an analogue of \texttt{\quarkifrecursion_tail_break:nN}
:N gives an analogue of \texttt{\quarkifrecursion_tail_stop:N}
:Nn gives an analogue of \texttt{\quarkifrecursion_tail_stop_do:Nn}
:NN gives an analogue of \texttt{\quarkifrecursion_tail_break:NN}

Any other signature causes an error, as does a function without signature.
Similar to \_kernel_quark_new_test:N, but defines quark branching conditionals like \quark_if_nil:nTF that test for the quark \q__⟨namespace⟩⟨name⟩. The ⟨namespace⟩ and ⟨name⟩ are determined from the conditional \#1, which must take the rather rigid form \_⟨namespace⟩_quark_if_⟨name⟩:{arg spec}. There are only two cases for the ⟨arg spec⟩ here:

:n gives an analogue of \quark_if_nil:nTF

:N gives an analogue of \quark_if_nil:NTF

Any other signature causes an error, as does a function without signature. We use low-level emptiness tests as \l3tl is not available yet when these functions are used; thankfully we only care about whether strings are empty so a simple \if_meaning:w \q_nil ⟨string⟩ \q_nil suffices.
# conditional
\fi:
\fi:
\cs_generate_variant:Nn \_quark_new_conditional:Nnnn { Nxx }
\cs_new_protected:Npn \_quark_new_test_aux:nnNNnnnN #1 #2 #3 #4 #5
{\cs_if_exist_use:cTF { \_quark_new_#5_#2:Nnnn } { #4 }
{ \msg_error:nnxx { quark } { invalid-function }
{ \token_to_str:N #4 } {#2}
\use_none:nnn
\use_none:nnn
\}
(End definition for \_kernel_quark_new_test:N and others.)
\_quark_new_test_n:Nnnn
\_quark_new_test_nn:Nnnn
\_quark_new_test_N:Nnnn
\_quark_new_test_Nn:Nnnn
\_quark_new_test_NN:Nnnn
\_quark_new_test_NN:Nnnn
These macros implement the six possibilities mentioned above, passing the right arguments to \_quark_new_test_aux:nnNNnn, which defines some auxiliaries, and then to \_quark_new_test_define_tl:nnNNnn (:n(n) variants) or to \_quark_new_test_define_ifx:nnNNnn (:N(n)) which define the main conditionals.
\cs_new_protected:Npn \_quark_new_test_n:Nnnn #1 #2 #3 #4
{\_quark_new_test_aux_do:nnNNnnnN #4 #2 #3 { none } { } { } { }
\_quark_new_test_define_tl:nnNNnn #1 { }
}
\cs_new_protected:Npn \_quark_new_test_nn:Nnnn #1 #2 #3 #4
{\_quark_new_test_aux_do:nnNNnnnN #4 #2 #3 { i } { n } {#1} {#2}
\_quark_new_test_define_tl:nnNNnn #1 { \use_none:n }
}
\cs_new_protected:Npn \_quark_new_test_N:Nnnn #1 #2 #3 #4
{\_quark_new_test_aux_do:nnNNnnnN #4 #2 #3 { i } { n } {#1} {#2}
\_quark_new_test_define_break_tl:nnNNnn #1 { }
}
\cs_new_protected:Npn \_quark_new_test_Nn:Nnnn #1 #2 #3 #4
{\_quark_new_test_aux_do:nnNNnnnN #4 #2 #3 { i } { n } {#1} {#2}
\_quark_new_test_define_break_ifx:nnNNnn #1 { }
}
\cs_new_protected:Npn \_quark_new_test_NN:Nnnn #1 #2 #3 #4
{\_quark_new_test_aux_do:nnNNnnnN #4 #2 #3 { none } { } { } { }
\_quark_new_test_define_ifx:nnNNnn #1 { }
}
\cs_new_protected:Npn \_quark_new_test_NN:Nnnn #1 #2 #3 #4
{\_quark_new_test_aux_do:nnNNnnnN #4 #2 #3 { i } { n } {#1} {#2}
\_quark_new_test_define_ifx:nnNNnn #1 { }
\else: \exp_after:wN \use_none:n }
\}
\cs_new_protected:Npn \_quark_new_test_NN:Nnnn #1 #2 #3 #4
{\_quark_new_test_aux_do:nnNNnnnN #4 #2 #3 { i } { n } {#1} {#2}
\_quark_new_test_define_break_ifx:nnNNnn #1 { }
}
(End definition for \_quark_new_test_n:Nnnn and others.)
makes the control sequence names which will be used by \_\_quark_test_define_aux:NNNNNNn, and then later by \_\_quark_new_test_define_tl:nNNNn or \_\_quark_new_test_define_ifx:nNnnNn. The control sequences defined here are analogous to \_\_quark_if_recursion_tail:w and to \use_{(none|i)}_delimit_by_q_recursion_stop:(|n)w.

The name is composed by the name-space and the name of the quarks. Suppose \_\_kernel_quark_new_test:N was used with:

\__kernel_quark_new_test:N \__test_quark_tail:w

then the first auxiliary will be \_\_test_quark_recursion_tail:w, and the second one will be \_\_test_use_none_delimit_by_q_recursion_stop:w.

Note that the actual quarks are not defined here. They should be defined separately using \quark_new:N.

Finally, these two macros define the main conditional function using what’s been set up before.

(End definition for \_\_quark_new_test_do:nNNnnNNn and \_\_quark_test_define_aux:NNNNNNn.)
These macros implement the two possibilities for branching quark conditionals, passing the right arguments to \_\_quark\_new\_conditional\_aux\_do:Nnmmm, which defines some auxiliaries and defines the main conditionals.

\begin{verbatim}
\cs_new_protected:Npn \_\_quark\_new\_conditional\_n:Nnnn
{ \_\_quark\_new\_conditional\_aux\_do:NNnnn \use_i:nn }
\cs_new_protected:Npn \_\_quark\_new\_conditional\_N:Nnnn
{ \_\_quark\_new\_conditional\_aux\_do:NNnnn \use_ii:nn }
\end{verbatim}

\begin{verbatim}
\exp_args:Ncc \_\_quark\_new\_conditional\_define:NNNNn
{ __#4_if_quark_ #3:w }{ q__#4__#3} #2 #1
\end{verbatim}

Similar to the previous macros, but branching conditionals only require one auxiliary, so we take a shortcut. In \_\_quark\_new\_conditional\_define:NNNNn, #4 is \use_i:nn to define the n-type function (which needs an auxiliary) and is \use_ii:nn to define the N-type function.

\begin{verbatim}
\cs_new_protected:Npn \_\_quark\_new\_conditional\_aux\_do:NNnnn #1 #2 #3 #4
{ \exp_args:Ncc \_\_quark\_new\_conditional\_define:NNNNn
{ __ #4_if_quark_ #3:w }{ q__#4__#3} #2 #1 }
\end{verbatim}

\_\_quark\_module\_name:N takes a control sequence and returns its \textit{(module)} name, determined as the first non-empty non-single-character word, separated by \_ or ::. These rules give the correct result for public functions \textit{\langle module\rangle}..., private functions \textit{\_\_\_\langle module\rangle}..., and variables such as \l_\textit{\langle module\rangle}.... If no valid module is found the result is an empty string. The approach is to first cut off everything after the (first) : if any is present, then repeatedly grab \_\_delimited words until finding one of length at least 2 (we use low-level tests as \_\_kernel\_quark\_new-test:N is first used. If no \textit{\langle module\rangle} is found (such as in \_\_\_n) we get the trailing marker \use\_none:n \{}, which expands to nothing.
\cs_new:Npn \__quark_module_name_loop:w ##1 #2
{
  \use_i_ii:nnn \if_meaning:w \prg_do_nothing:
    ##1 \prg_do_nothing: \prg_do_nothing:
  \exp_after:wN \__quark_module_name_loop:w
  \else:
    \__quark_module_name_end:w ##1
  \fi:
}
\cs_new:Npn \__quark_module_name_end:w
  ##1 \fi: ##2 \s__quark { \fi: ##1 }
\exp_after:wN \__quark_tmp:w \tl_to_str:n { : _ }
(End definition for \__quark_module_name:N and others.)
\__quark_quark_conditional_name:N \__quark_quark_conditional_name:w
\__quark_quark_conditional_name:N determines the quark name that the quark con-
ditional function \#1 queries, as the part of the function name between _quark_if_
and the trailing :. Again we define it through \__quark_tmp:w, which receives : as \#1
and _quark_if_ as \#2. The auxiliary \__quark_quark_conditional_name:w returns
the part between the first _quark_if_ and the next :, and we apply this auxiliary to the
function name followed by : (in case the function name is lacking a signature), and
_quark_if_: so that \__quark_quark_conditional_name:N returns an empty string if
_quark_if_ is not present.
\cs_set:Npn \__quark_tmp:w #1 #2 \s__quark
{
\cs_new:Npn \__quark_quark_conditional_name:N ##1
{
  \exp_last_unbraced:Nf \__quark_quark_conditional_name:w
  { \cs_to_str:N ##1 } #1 #2 #1 \s__quark
}
\cs_new:Npn \__quark_quark_conditional_name:w
  ##1 #2 ##2 #1 ##3 \s__quark {##2}
\exp_after:wN \__quark_tmp:w \tl_to_str:n { : _quark_if_ }
(End definition for \__quark_quark_conditional_name:N and \__quark_quark_conditional_name:w.)

55.2 Scan marks
\g__scan_marks_tl The list of all scan marks currently declared. No \l3tl yet, so define this by hand.
\cs_gset:Np \g__scan_marks_tl { }
(End definition for \g__scan_marks_tl.)
\scan_new:N Check whether the variable is already a scan mark, then declare it to be equal to \scan_-_stop: globally.
\cs_new_protected:Npn \scan_new:N #1
{
  \tl_if_in:NnTF \g__scan_marks_tl { #1 }
  \{ 
  \}
}

784
\msg_error:nx { scanmark } { already-defined }
\token_to_str:N #1 }
\tl_gput_right:Nn \g__scan_marks_tl \{#1
\cs_new_eq:NN \scan_stop: #1 \scan_stop:
\}

(End definition for \scan_new:N. This function is documented on page 141.)

\s_stop We only declare one scan mark here, more can be defined by specific modules. Can’t use \scan_new:N yet because l3tl isn’t loaded, so define \s_stop by hand and add it to \g__scan_marks_tl. We also add \s__quark (declared earlier) to the pool here. Since it lives in a different namespace, a little l3docstrip cheating is necessary.
\cs_new_eq:NN \s_stop \scan_stop:
\cs_gset_nopar:Npx \g__scan_marks_tl
{ \exp_not:o \g__scan_marks_tl
\s_stop
⟨@@=quark⟩
\s__quark
⟨@@=scan⟩
}

(End definition for \s_stop. This variable is documented on page 141.)

\use_none_delimit_by_s_stop:w Similar to \use_none_delimit_by_q_stop:w.
\cs_new:Npn \use_none_delimit_by_s_stop:w #1 \s_stop { }

(End definition for \use_none_delimit_by_s_stop:w. This function is documented on page 141.)

}}
Chapter 56

\textbf{l3seq implementation}

The following test files are used for this code: \texttt{m3seq002,m3seq003}.

A sequence is a control sequence whose top-level expansion is of the form "\_\_seq \_\_seq_item:n \{\langle item1\rangle\} ... \_\_seq_item:n \{\langle item_n\rangle\}" with a leading scan mark followed by \texttt{n} items of the same form. An earlier implementation used the structure "\texttt{\seq_elw \langle item1\rangle \seq_\texttt{elt_end}; ... \seq_elw \langle item_n\rangle \seq_\texttt{elt_end};}". This allowed rapid searching using a delimited function, but was not suitable for items containing \{,\} and \# tokens, and also lead to the loss of surrounding braces around items

\_\_seq_item:n \*
\_\_seq_item:n \{\langle item\rangle\}

The internal token used to begin each sequence entry. If expanded outside of a mapping or manipulation function, an error is raised. The definition should always be set globally.

\_\_seq_push_item_def:n
\_\_seq_push_item_def:x

Saves the definition of \_\_seq_item:n and redefines it to accept one parameter and expand to \{\langle code\rangle\}. This function should always be balanced by use of \_\_seq_pop_item_def:n.

\_\_seq_pop_item_def:

Restores the definition of \_\_seq_item:n most recently saved by \_\_seq_push_item_def:n. This function should always be used in a balanced pair with \_\_seq_push_item_def:n.

\_\_seq This private scan mark.

\_\_seq \texttt{seq}

(End definition for \_\_seq)

\_\_seq_mark Private scan marks.

\_\_seq \texttt{seq mark}
\_\_seq \texttt{seq stop}

(End definition for \_\_seq_mark and \_\_seq_stop.)
The delimiter is always defined, but when used incorrectly simply removes its argument and hits an undefined control sequence to raise an error.

\begin{verbatim}
\cs_new:Npn \__seq_item:n #1
{\msg_expandable_error:nn { seq } { misuse } \use_none:n}
\end{verbatim}

(End definition for \__seq_item:n.)

Scratch space for various internal uses.

\begin{verbatim}
\tl_new:N \l__seq_internal_a_tl
\tl_new:N \l__seq_internal_b_tl
\end{verbatim}

(End definition for \l__seq_internal_a_tl and \l__seq_internal_b_tl.)

Scratch function for internal use.

\begin{verbatim}
\cs_new_eq:NN \__seq_tmp:w ?
\end{verbatim}

(End definition for \__seq_tmp:w.)

A sequence with no item, following the structure mentioned above.

\begin{verbatim}
\tl_const:Nn \c_empty_seq { \s__seq }
\end{verbatim}

(End definition for \c_empty_seq. This variable is documented on page 153.)

### 56.1 Allocation and initialisation

Sequences are initialized to \c_empty_seq.

\begin{verbatim}
\cs_new_protected:Npm \seq_new:N #1
{ \__kernel_chk_if_free_cs:N #1 \cs_gset_eq:NN #1 \c_empty_seq }
\cs_generate_variant:Nn \seq_new:N { c }
\end{verbatim}

(End definition for \seq_new:N. This function is documented on page 142.)

Clearing a sequence is similar to setting it equal to the empty one.

\begin{verbatim}
\cs_new_protected:Npm \seq_clear:N #1
{ \seq_set_eq:NN #1 \c_empty_seq \c_gset:Vn \c_empty_seq }
\cs_generate_variant:Nn \seq_clear:N { c }
\end{verbatim}

(End definition for \seq_clear:N and \seq_gclear:N. These functions are documented on page 142.)

Once again we copy code from the token list functions.

\begin{verbatim}
\cs_new_protected:Npm \seq_clear_new:N #1
{ \seq_if_exist:NTF #1 \seq_clear:N #1 \seq_new:N #1 \c_gset:Vn \c_empty_seq }
\cs_generate_variant:Nn \seq_clear_new:N { c }
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \seq_gclear_new:N #1
{ \seq_if_exist:NTF #1 \seq_gclear:N #1 \seq_new:N #1 \c_gset:Vn \c_empty_seq }
\cs_generate_variant:Nn \seq_gclear_new:N { c }
\end{verbatim}

787
Copying a sequence is the same as copying the underlying token list.

Setting a sequence from a comma-separated list is done using a simple mapping.

Almost identical to \seq_set_from_clist:N. These functions are documented on page 143.

(End definition for \seq_clear_new:N and \seq_gclear_new:N. These functions are documented on page 142.)
When the separator is empty, everything is very simple; just map \texttt{\_\_seq_wrap_item:n} through the items of the last argument. For non-trivial separators, the goal is to split a given token list at the marker, strip spaces from each item, and remove one set of outer braces if after removing leading and trailing spaces the item is enclosed within braces. After \texttt{\_\_tl_replace_all:Nnn}, the token list \texttt{\_\_seq_internal_a_tl} is a repetition of the pattern \texttt{\_\_seq_set_split_auxi:w \prg_do_nothing: (item with spaces) \_\_seq_set_split_end::}. Then, x-expansion causes \texttt{\_\_seq_set_split_auxi:w} to trim spaces, and leaves its result as \texttt{\_\_seq_set_split_auxii:w (trimmed item) \_\_seq_set_split_end::}. This is then converted to the \texttt{l3seq} internal structure by another x-expansion. In the first step, we insert \texttt{\_\_prg_do_nothing:} to avoid losing braces too early: that would cause space trimming to act within those lost braces. The second step is solely there to strip braces which are outermost after space trimming.

\begin{verbatim}
\cs_new_protected:Npn \seq_set_split:Nnn \{ \__seq_set_split:NNNnn \__kernel_tl_set:Nx \tl_trim_spaces:n \}
\cs_new_protected:Npn \seq_gset_split:Nnn \{ \__seq_set_split:NNNnn \__kernel_tl_gset:Nx \tl_trim_spaces:n \}
\cs_new_protected:Npn \seq_set_split_keep_spaces:Nnn \{ \__seq_set_split:NNNnn \__kernel_tl_set:Nx \exp_not:n \}
\cs_new_protected:Npn \seq_gset_split_keep_spaces:Nnn \{ \__seq_set_split:NNNnn \__kernel_tl_gset:Nx \exp_not:n \}
\cs_new_protected:Npn \__seq_set_split:NNNnn #1#2#3#4#5
\{ \tl_if_empty:nTF {#4}
\{ \tl_set:Nn \l__seq_internal_a_tl { \tl_map_function:nN {#5} \_\_seq_wrap_item:n } 
\}
\}
\tl_set:Nn \l__seq_internal_a_tl { \tl_replace_all:Nnn \l__seq_internal_a_tl {#4} 
{ \_\_seq_set_split:Nw \prg_do_nothing: #5
\_\_seq_set_split_end:: } 
}
\tl_replace_all:Nnn \l__seq_internal_a_tl {#4} 
{ \_\_seq_set_split:Nw \prg_do_nothing: #2
\_\_seq_set_split_end:: } 
\_\_kernel_tl_set:Nx \l__seq_internal_a_tl 
\}
\cs_new:Npn \_\_seq_set_split:NNNnn \_\_seq_set_split:end:
\{ \exp_not:N \_\_seq_set_split:w
\exp_args:No #1 \{#2\}
\exp_not:N \_\_seq_set_split:end:
\}
\cs_new:Npn \_\_seq_set_split:w \_\_seq_set_split:end:
\{ \_\_seq_wrap_item:n \{#1\} \}
\cs_generate_variant:Nn \seq_set_split:Nnn \{ \_\_seq_set_split:w \}
\cs_generate_variant:Nn \seq_gset_split:Nnn \{ \_\_seq_set_split:w \}
\end{verbatim}

789
When concatenating sequences, one must remove the leading \_\_\_\_\_\_\_\_\_\_ of the second sequence. The result starts with \_\_\_\_\_\_\_\_\_ (of the first sequence), which stops f-expansion.

\begin{verbatim}
\cs_new_protected:Npn \seq_concat:NNN \seq_gconcat:NNN \seq_concat:ccc \seq_gconcat:ccc
\end{verbatim}

When adding to the left of a sequence, remove \_\_\_\_\_\_\_\_\_\_. This is done by \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, which also stops f-expansion.

\begin{verbatim}
\cs_new_protected:Npn \seq_put_left:Nn \seq_gput_left:Nn \seq_put_left:NV \seq_gput_left:NV \seq_put_left:Nv \seq_gput_left:Nv \seq_put_left:No \seq_gput_left:No \seq_put_left:Nx \seq_gput_left:Nx \seq_put_left:cn \seq_gput_left:cn \seq_put_left:co \seq_gput_left:co \seq_put_left:cV \seq_gput_left:cV \seq_put_left:cv \seq_gput_left:cv \seq_put_left:co \seq_gput_left:co \seq_put_left:cx \seq_gput_left:cx \__seq_put_left_aux:w \__seq_put_left_aux:w \end{verbatim}

56.2 Appending data to either end

When adding to the left of a sequence, remove \_\_\_\_\_\_\_\_\_\_. This is done by \_\_\_\_\_\_\_\_\_\_\_\_\_, which also stops f-expansion.

\begin{verbatim}
\cs_new_protected:Npm \seq_set_split_keep_spaces:Nnn \seq_gset_split_keep_spaces:Nnn \seq_set_split:Nnn \seq_gset_split:Nnn \seq_set_split:Vn \seq_gset_split:Vn \seq_set_split:Vv \seq_gset_split:Vv \seq_set_split:Vo \seq_gset_split:Vo \seq_set_split:Vx \seq_gset_split:Vx \seq_set_split:Vw \seq_gset_split:Vw
\end{verbatim}

When concatenating sequences, one must remove the leading \_\_\_\_\_\_\_\_\_\_ of the second sequence. The result starts with \_\_\_\_\_\_\_\_\_\_ (of the first sequence), which stops f-expansion.

\begin{verbatim}
\cs_new_protected:Npn \seq_concat:NNN \seq_gconcat:NNN \seq_concat:ccc \seq_gconcat:ccc
\end{verbatim}

When adding to the left of a sequence, remove \_\_\_\_\_\_\_\_\_\_. This is done by \_\_\_\_\_\_\_\_\_\_\_\_\_, which also stops f-expansion.

\begin{verbatim}
\cs_new_protected:Npn \seq_put_left:Nn \seq_gput_left:Nn \seq_put_left:NV \seq_gput_left:NV \seq_put_left:Nv \seq_gput_left:Nv \seq_put_left:No \seq_gput_left:No \seq_put_left:Nx \seq_gput_left:Nx \seq_put_left:cn \seq_gput_left:cn \seq_put_left:co \seq_gput_left:co \seq_put_left:cV \seq_gput_left:cV \seq_put_left:cv \seq_gput_left:cv \seq_put_left:co \seq_gput_left:co \seq_put_left:cx \seq_gput_left:cx \__seq_put_left_aux:w \__seq_put_left_aux:w \end{verbatim}

56.2 Appending data to either end

When adding to the left of a sequence, remove \_\_\_\_\_\_\_\_\_\_. This is done by \_\_\_\_\_\_\_\_\_\_\_\_\_, which also stops f-expansion.

\begin{verbatim}
\cs_new_protected:Npm \seq_set_split_keep_spaces:Nnn \seq_gset_split_keep_spaces:Nnn \seq_set_split:Nnn \seq_gset_split:Nnn \seq_set_split:Vn \seq_gset_split:Vn \seq_set_split:Vv \seq_gset_split:Vv \seq_set_split:Vo \seq_gset_split:Vo \seq_set_split:Vx \seq_gset_split:Vx \seq_set_split:Vw \seq_gset_split:Vw
\end{verbatim}
Since there is no trailing marker, adding an item to the right of a sequence simply means wrapping it in \_\_\_seq_item:n.

\cs_new_protected:Npn \seq_put_right:Nn #1#2 { \tl_put_right:Nn #1 { \__seq_item:n {#2} } }
\cs_new_protected:Npn \seq_gput_right:Nn #1#2 { \tl_gput_right:Nn #1 { \__seq_item:n {#2} } }
\cs_generate_variant:Nn \seq_gput_right:Nn { NV , Nv , No , Nx }
\cs_generate_variant:Nn \seq_gput_right:Nn { c , cV , cv , co , cx }
\cs_generate_variant:Nn \seq_put_right:Nn { NV , Nv , No , Nx }
\cs_generate_variant:Nn \seq_put_right:Nn { c , cV , cv , co , cx }

56.3 Modifying sequences

This function converts its argument to a proper sequence item in an x-expansion context.

\cs_new:Npn \__seq_wrap_item:n #1 { \exp_not:n { \__seq_item:n {#1} } }

An internal sequence for the removal routines.

\seq_new:N \l__seq_remove_seq
\seq_remove_duplicates:N \seq_remove_duplicates:c
\seq_gremove_duplicates:N \seq_gremove_duplicates:c
\__seq_remove_duplicates:NN

Removing duplicates means making a new list then copying it.

\cs_new_protected:Npn \seq_remove_duplicates:N { \__seq_remove_duplicates:NN \seq_set_eq:NN }
\cs_new_protected:Npn \seq_gremove_duplicates:N { \__seq_remove_duplicates:NN \seq_gset_eq:NN }
\cs_new_protected:Npn \__seq_remove_duplicates:NN #1#2 { \seq_clear:N \l__seq_remove_seq \seq_map_inline:Nn #2 { \seq_if_in:NnF \l__seq_remove_seq {##1} { \seq_put_right:Nn \l__seq_remove_seq {##1} } } #1 #2 \l__seq_remove_seq }
\cs_generate_variant:Nn \seq_remove_duplicates:N { c }
\cs_generate_variant:Nn \seq_gremove_duplicates:N { c }

The idea of the code here is to avoid a relatively expensive addition of items one at a time to an intermediate sequence. The approach taken is therefore similar to that in \_\_\_seq_pop_right:NNN, using a “flexible” x-type expansion to do most of the work. As \tl_l_if_eq:nnT is not expandable, a two-part strategy is needed. First, the x-type expansion uses \str_l_if_eq:nnT to find potential matches. If one is found, the expansion is halted and the necessary set up takes place to use the \tl_l_if_eq:nnT test. The x-type is started
again, including all of the items copied already. This happens repeatedly until the entire sequence has been scanned. The code is set up to avoid needing and intermediate scratch list: the lead-off \texttt{x}-type expansion (\#1 \#2 \{\#2\}) ensures that nothing is lost.

\begin{verbatim}
cs_new_protected:Npn \seq_remove_all:Nn \seq_gremove_all:Nn

\cs_new_protected:Npn \__seq_remove_all_aux:NNn \__kernel_tl_set:Nx

\cs_new_protected:Npn \__seq_remove_all_aux:NNn \__kernel_tl_gset:Nx

\cs_new_protected:Npn \__seq_remove_all_aux:NNn \__seq_remove_all_aux:NNn #1\#2\#3

\__seq_push_item_def:n

\str_if_eq:nnT {##1} {#3}

\tl_set:Nn \l__seq_internal_b_tl {##1}

\tl_set:Nn \l__seq_internal_a_tl {#3}

\tl_set:Nn \l__seq_internal_a_tl \l__seq_internal_b_tl

\tl_if_eq:NNT \l__seq_internal_a_tl \l__seq_internal_b_tl

\use_none:nn

\__seq_wrap_item:n {##1}

\__seq_pop_item_def:
\end{verbatim}

(End definition for \seq_remove_all:Nn, \seq_gremove_all:Nn, and \__seq_remove_all_aux:NNn. These functions are documented on page 147.)

Previously, \texttt{\seq_reverse:N} was coded by collecting the items in reverse order after an \texttt{\exp_stop_f:} marker.

\begin{verbatim}
cs_new_protected:Npn \seq_reverse:N #1

\cs_set_eq:NN \@@_item:n \@@_reverse_item:nw

\tl_set:Nf #2 { #2 \exp_stop_f: }

\tl_set:Nf #2 \exp_stop_f: \#2

\tl_set:Nf \l__seq_reverse_item:nw \#2 \exp_stop_f:

\tl_set:Nf \l__seq_reverse_item:nw \#1 \exp_stop_f:

\tl_set:Nf \l__seq_reverse_item:nw \#1 \exp_stop_f:

\tl_set:Nf \l__seq_reverse_item:nw \#1
\end{verbatim}

At first, this seems optimal, since we can forget about each item as soon as it is placed after \texttt{\exp_stop_f:}. Unfortunately, \TeX’s usual tail recursion does not take place in this case: since the following \texttt{\__seq_reverse_item:nw} only reads tokens until \texttt{\exp_stop_f:}, and never reads the \texttt{\@@_item:n \#1} left by the previous call, \TeX cannot remove that previous call from the stack, and in particular must retain the various macro parameters in memory, until the end of the replacement text is reached. The stack is thus
only flushed after all the \_seq_reverse_item:nw are expanded. Keeping track of the arguments of all those calls uses up a memory quadratic in the length of the sequence. \TeX{} can then not cope with more than a few thousand items.

Instead, we collect the items in the argument of \exp_not:n. The previous calls are cleanly removed from the stack, and the memory consumption becomes linear.

\begin{verbatim}
\cs_new_protected:Npn \seq_reverse:N { \__seq_reverse:NN \__kernel_tl_set:Nx }
\cs_new_protected:Npn \seq_greverse:N { \__seq_reverse:NN \__kernel_tl_gset:Nx }
\cs_new_protected:Npn \__seq_reverse:NN #1 #2 { \cs_set_eq:NN \__seq_tmp:w \__seq_item:n
\cs_set_eq:NN \__seq_item:n \__seq_reverse_item:nwn #1 #2 \exp_not:n { } }
\cs_set_eq:NN \__seq_item:n \__seq_tmp:w
\cs_new:Npn \__seq_reverse_item:nwn #1 #2 \exp_not:n #3 { #2 \exp_not:n { \__seq_item:n {#1} #3 } }
\cs_generate_variant:Nn \seq_reverse:N { c }
\cs_generate_variant:Nn \seq_greverse:N { c }
\end{verbatim}

(End definition for \seq_reverse:N and others. These functions are documented on page 147.)

\seq_sort:Nn \seq_sort:cn \seq_gsort:Nn \seq_gsort:cn

Implemented in \texttt{l3sort}.

(End definition for \seq_sort:Nn and \seq_gsort:Nn. These functions are documented on page 147.)

### 56.4 Sequence conditionals

\seq_if_empty_p:N \seq_if_empty_p:c \seq_if_empty:NTF \seq_if_empty:cTF

Similar to token lists, we compare with the empty sequence.

\begin{verbatim}
\prg_new_conditional:Npnn \seq_if_empty:N #1 { p , T , F , TF } { \if_meaning:w #1 \c_empty_seq \prg_return_true: \else: \prg_return_false: \fi: }
\prg_generate_conditional_variant:Nnn \seq_if_empty:N { c } { p , T , F , TF }
\end{verbatim}

(End definition for \seq_if_empty:NTF. This function is documented on page 147.)

\seq_shuffle:N \seq_shuffle:c \seq_gshuffle:N \seq_gshuffle:c \__seq_shuffle:NN \__seq_shuffle_item:n \g__seq_internal_seq

We apply the Fisher–Yates shuffle, storing items in \texttt{toks} registers. We use the primitive \texttt{\texttt{\texttt{text_uniformdeviate:D}}} for speed reasons. Its non-uniformity is of order its argument divided by $2^{28}$, not too bad for small lists. For sequences with more than 13 elements there are more possible permutations than possible seeds ($13! > 2^{28}$) so the question of uniformity is somewhat moot. The integer variables are declared in \texttt{l3int: load-order issues}.

\begin{verbatim}
\cs_if_exist:NTF \text_uniformdeviate:D \texttt{793}
\end{verbatim}
{\seq_new:N \g__seq_internal_seq
\cs_new_protected:Npn \seq_shuffle:N \seq_set_eq:NN \_\_seq_shuffle_item:n \__seq_shuffle_item:n
\cs_new_protected:Npn \seq_gshuffle:N \seq_gset_eq:NN
\cs_new_protected:Npn \__seq_shuffle:NN #1 #2
{\int_compare:nNnTF {\seq_count:N #2} > \c_max_register_int
 {\msg_error:nnn {seq} {shuffle-too-large}}
 }\group_end:
\int_zero:N \l__seq_internal_a_int
\__seq_push_item_def:\cs_gset_eq:NN \__seq_item:n \__seq_shuffle_item:n #2
\__seq_pop_item_def:\seq_gset_from_inline_x:Nnn \g__seq_internal_seq {\int_step_function:nN {\l__seq_internal_a_int}}
{\tex_the:D \tex_toks:D ##1}\group_end:
#1 #2 \g__seq_internal_seq
\seq_gclear:N \g__seq_internal_seq
\cs_new_protected:Npn \__seq_shuffle_item:n
{\int_incr:N \l__seq_internal_a_int \int_set:Nn \l__seq_internal_b_int
 {1 + \tex_uniformdeviate:D \l__seq_internal_a_int}
 \tex_toks:D \l__seq_internal_a_int = \tex_toks:D \l__seq_internal_b_int
 \tex_toks:D \l__seq_internal_b_int}
\group_end:
\cs_new_protected:Npn \seq_shuffle:N \seq_shuffle:N #1
{\msg_error:nnn {kernel} {fp-no-random}}
{\seq_shuffle:N #1}
\cs_generate_variant:Nn \seq_shuffle:N {c}
\cs_generate_variant:Nn \seq_gshuffle:N {c}

(End definition for \seq_shuffle:N and others. These functions are documented on page 147.)

The approach here is to define \_\_seq_item:n to compare its argument with the test sequence. If the two items are equal, the mapping is terminated and \group_end:\prg_return_true: is inserted after skipping over the rest of the recursion. On the other hand, if there is no match then the loop breaks, returning \prg_return_false:. Everything is inside a group so that \_\_seq_item:n is preserved in nested situations.
The two pop functions share their emptiness tests. We also use a common emptiness test for all branching get and pop functions.

Getting an item from the left of a sequence is pretty easy: just trim off the first item after \texttt{\_\_seq_item:n} at the start. We append a \texttt{\q_no_value} item to cover the case of an empty sequence.
\texttt{\__seq_get_left:wnw} \#1 \texttt{\__seq_item:n} \{ \texttt{\q_no_value} \} \texttt{\s__seq_stop}

{(End definition for \seq_get_left:NN and \__seq_get_left:wnw. This function is documented on page 144.)}

\texttt{\seq_pop_left:NN} \texttt{\seq_pop_left:cN} \texttt{\seq_gpop_left:NN} \texttt{\__seq_pop_left:NNN} \texttt{\__seq_pop_left:wnwNNN}

The approach to popping an item is pretty similar to that to get an item, with the only difference being that the sequence itself has to be redefined. This makes it more sensible to use an auxiliary function for the local and global cases.

\texttt{\cs_new_protected:Npn} \texttt{\seq_gpop_left:NN} \texttt{\seq_gpop_left:cN} \texttt{\seq_gpop_left:NNN} \texttt{\__seq_gpop_left:wnwNNN}

{(End definition for \seq_pop_left:NN and others. These functions are documented on page 144.)}

\texttt{\seq_get_right:NN} \texttt{\seq_get_right:cN} \texttt{\__seq_get_right_loop:nw} \texttt{\__seq_get_right_end:NnN}

First remove \texttt{\s__seq} and prepend \texttt{\q_no_value}. The first argument of \texttt{\__seq_get-_right_loop:nw} is the last item found, and the second argument is empty until the end of the loop, where it is code that applies \texttt{\exp_not:n} to the last item and ends the loop.

{(End definition for \seq_get_left:NN and others. These functions are documented on page 144.)}
The approach to popping from the right is a bit more involved, but does use some of the same ideas as getting from the right. What is needed is a “flexible length” way to set a token list variable. This is supplied by the \{if_false: \fi: \} construct. Using an x-type expansion and a “non-expanding” definition for \__seq_item:n, the left-most \( n - 1 \) entries in a sequence of \( n \) items are stored back in the sequence. That needs a loop of unknown length, hence using the strange \if_false: way of including braces. When the last item of the sequence is reached, the closing brace for the assignment is inserted, and \tl_set:Nn #3 is inserted in front of the final entry. This therefore does the pop assignment. One more iteration is performed, with an empty argument and \use_none:nn, which finally stops the loop.

Getting from the left or right with a check on the results. The first argument to \__seq_pop_TF:NNN is left unused.
More or less the same for popping.

\seq_pop_right:NNTF
\seq_pop_right:cN
\seq_gpop_right:NN
\seq_gpop_right:cN

The idea here is to find the offset of the item from the left, then use a loop to grab the correct item. If the resulting offset is too large, then the argument delimited by \seq_item:n is \prg_break: instead of being empty, terminating the loop and returning nothing at all.

\cs_new:Npn \seq_item:Nn #1
\seq_item:cn
\__seq_item:wNn
\__seq_item:nN
\__seq_item:nwn
\__seq_item:n

(End definition for \seq_get_left:NNTF and \seq_get_right:NNTF. These functions are documented on page 146.)

798
Important, `\seq_item:Nn` only evaluates its argument once.

```latex
\cs_new:Npn \seq_item:Nn \seq_item:cn \seq_item:nw

\__seq_item:n \__seq_item:n \__seq_item:n \__seq_item:n
```

(End definition for `\seq_item:Nn` and others. This function is documented on page 145.)

### 56.6 Mapping over sequences

`\seq_map_break:` and `\seq_map_break:n` To break a function, the special token `\prg_break_point:Nn` is used to find the end of the code. Any ending code is then inserted before the return value of `\seq_map_break:n` is inserted.

```latex
\cs_new:Npn \seq_map_break: \seq_map_break:n

\__seq_map_function:Nw \__seq_map_function:cn \__seq_map_function:nw
```

The idea here is to apply the code of #2 to each item in the sequence without altering the definition of `\__seq_item:n`. The even-numbered arguments of `\__seq_map_function:Nw` delimited by `\__seq_item:n` are almost always empty, except at the end of the loop where it is `\prg_break:`. This allows to break the loop without needing to do a (relatively-expensive) quark test.

```latex
\cs_new:Npn \__seq_map_function:Nw \__seq_map_function:cn \__seq_map_function:nw

\__seq_map_function:nn \__seq_map_function:n \__seq_map_function:n \__seq_map_function:n
```

(End definition for `\seq_map_break:` and `\seq_map_break:n`. These functions are documented on page 149.)

16754 \cs_new:Npn \seq_map_break:n { \prg_break:Nn \seq_map_break: { } }
The definition of \_\_seq_item:n needs to be saved and restored at various points within
the mapping and manipulation code. That is handled here: as always, this approach uses
global assignments.

\begin{verbatim}
\cs_new_protected:Npn \_\_seq_push_item_def:n
\begin{verbatim}
\cs_gset:Npn \_\_seq_item:n ##1
\end{verbatim}
\end{verbatim}
\begin{verbatim}
\cs_new_protected:Npn \_\_seq_push_item_def:x
\begin{verbatim}
\cs_gset:Npx \_\_seq_item:n ##1
\end{verbatim}
\end{verbatim}
\begin{verbatim}
\cs_new_protected:Npn \_\_seq_push_item_def:
\begin{verbatim}
\int_gincr:N \g__kernel_prg_map_int
\cs_gset_eq:cN { __seq_map_ \int_use:N \g__kernel_prg_map_int :w }
\_\_seq_item:n
\end{verbatim}
\end{verbatim}
\begin{verbatim}
\cs_new_protected:Npn \_\_seq_pop_item_def:
\begin{verbatim}
\cs_gset_eq:Nc \_\_seq_item:n
\begin{verbatim}
{ __seq_map_ \int_use:N \g__kernel_prg_map_int :w }
\int_gdecr:N \g__kernel_prg_map_int
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}

\_\_seq_push_item_def:n
\_\_seq_push_item_def:x
\_\_seq_push_item_def:
\_\_seq_pop_item_def:

The idea here is that \_\_seq_item:n is already “applied” to each item in a sequence,
and so an in-line mapping is just a case of redefining \_\_seq_item:n.

\begin{verbatim}
\cs_new_protected:Npn \seq_map_inline:Nn #1#2
\begin{verbatim}
\_\_seq_push_item_def:n {#2}
\prg_break_point:Nn \seq_map_break: { \_\_seq_pop_item_def: }
\end{verbatim}
\end{verbatim}
\cs_generate_variant:Nn \seq_map_inline:Nn { c }
\end{verbatim}
\end{verbatim}
\end{verbatim}

\seq_map_inline:Nn
\seq_map_inline:cn

This is based on the function mapping but using the same tricks as described for \prop_map_tokens:Nn. The idea is to remove the leading \s__seq and apply the tokens such
that they are safe with the break points, hence the \use:n.

\begin{verbatim}
\cs_new:Npn \seq_map_tokens:Nn #1#2
\begin{verbatim}
\exp_last_unbraced:Nno \use_i:nn { \__seq_map_tokens:nw {#2} } #1
\prg_break:
\_\_seq_item:n { } \_\_seq_item:n { } \_\_seq_item:n { } \_\_seq_item:n { }
\prg_break_point:Nn \seq_map_break: { }
\prg_break_point:Nn \seq_map_break: { }
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}
\end{verbatim}

\seq_map_tokens:Nn
\seq_map_tokens:cn
\_\_seq_map_tokens:nw

(End definition for \seq_map_function:NN and \_\_seq_map_function:Nw. This function is documented
on page 148.)
\seq_map_variable:NNn
\seq_map_variable:Ncn
\seq_map_variable:cNn
\seq_map_variable:ccn

This is just a specialised version of the in-line mapping function, using an x-type expansion for the code set up so that the number of # tokens required is as expected.

\cs_new_protected:Npn \seq_map_variable:NNn #1#2#3
\__seq_push_item_def:x
\tl_set:Nn \exp_not:N #2 {##1}
\exp_not:n {#3}
\prg_break_point:Nn \seq_map_break: { \__seq_pop_item_def: }

\cs_new_protected:Npn \seq_map_variable:Ncn \seq_map_variable:cNn \seq_map_variable:ccn

(End definition for \seq_map_variable:NNn. This function is documented on page 148.)

\seq_map_indexed_function:NN
\seq_map_indexed_inline:Nn
\__seq_map_indexed:nNN
\__seq_map_indexed:Nw

Similar to \seq_map_function:NN but we keep track of the item index as a ;-delimited argument of \__seq_map_indexed:Nw.

\cs_new:Npn \seq_map_indexed_function:NN #1#2
\__seq_map_indexed:NN #1#2
\prg_break_point:Nn \seq_map_break: { }
\cs_new_protected:Npn \seq_map_indexed_inline:Nn \seq_map_indexed_inline:Nn \seq_map_indexed_inline:Nw

(End definition for \seq_map_variable:NNn. This function is documented on page 148.)
Very similar to \seq_set_filter:Nnn. We could actually merge the two within a single function, but it would have weird semantics.

\begin{verbatim}
\cs_new_protected:Npn \seq_set_map_x:NNn { \__seq_set_map_x:NNNn \__kernel_tl_set:Nx }
\cs_new_protected:Npn \seq_gset_map_x:NNn { \__seq_set_map_x:NNNn \__kernel_tl_gset:Nx }
\cs_new_protected:Npn \__seq_set_map:NNNn #1#2#3#4
{ \__seq_push_item_def:n { \exp_not:N \__seq_item:n {#4} } #1 #2 { #3 } \__seq_pop_item_def: }
\end{verbatim}

(End definition for \seq_set_map:NNn, \seq_gset_map:NNn, and \__seq_set_map:NNNn. These functions are documented on page 150.)

Similar to \seq_set_map_x:NNn, but prevents expansion of the <inline function>.

\begin{verbatim}
\cs_new_protected:Npn \seq_set_map:NNn { \__seq_set_map:NNNn \__kernel_tl_set:Nx }
\cs_new_protected:Npn \seq_gset_map:NNn { \__seq_set_map:NNNn \__kernel_tl_gset:Nx }
\cs_new_protected:Npn \__seq_set_map:NNNn #1#2#3#4
{ \__seq_push_item_def:n { \exp_not:n \__seq_item:n {#4} } #1 #2 { #3 } \__seq_pop_item_def: }
\end{verbatim}

(End definition for \seq_set_map:NNn, \seq_gset_map:NNn, and \__seq_set_map:NNNn. These functions are documented on page 150.)

Since counting the items in a sequence is quite common, we optimize it by grabbing 8 items at a time and correspondingly adding 8 to an integer expression. At the end of the loop, #9 is \__seq_count_end:w instead of being empty. It removes 8+ and instead
places the number of \_\_seq_item:n that \_\_seq_count:w grabbed before reaching the end of the sequence.

\cs_new:Npn \seq_count:N #1
\{ \int_eval:n
\{ \exp_after:wN \use_i:nn
\exp_after:wN \_\_seq_count:w
#1
\_\_seq_count_end:w \_\_seq_item:n 7
\_\_seq_count_end:w \_\_seq_item:n 6
\_\_seq_count_end:w \_\_seq_item:n 5
\_\_seq_count_end:w \_\_seq_item:n 4
\_\_seq_count_end:w \_\_seq_item:n 3
\_\_seq_count_end:w \_\_seq_item:n 2
\_\_seq_count_end:w \_\_seq_item:n 1
\_\_seq_count_end:w \_\_seq_item:n 0
\prg_break_point:
\}
\}
\cs_new:Npn \_\_seq_count:w
\#1 \_\_seq_item:n \#2 \_\_seq_item:n \#3 \_\_seq_item:n \#4 \_\_seq_item:n
\#5 \_\_seq_item:n \#6 \_\_seq_item:n \#7 \_\_seq_item:n \#8 \#9 \_\_seq_item:n
{ \#9 8 + \_\_seq_count:w }
\cs_new:Npn \_\_seq_count_end:w
\_\_seq_count:w
\#1\#2 \prg_break_point: {#1}
\cs_generate_variant:Nn \seq_count:N { c }

(End definition for \seq_count:N, \_\_seq_count:w, and \_\_seq_count_end:w. This function is documented on page 150.)

56.7 Using sequences

\seq_use:Nnnn \seq_use:cn \_\_seq_use:NNnn \_\_seq_use_setup:w \_\_seq_use:nwvnn \_\_seq_use:Nn \_\_seq_use:cn

See \clist_use:Nnnn for a general explanation. The main difference is that we use \_\_seq_item:n as a delimiter rather than commas. We also need to add \_\_seq_item:n at various places, and \s__seq.

\cs_new:Npn \seq_use:Nnnn #1#2#3#4
\{ \seq_if_exist:NTF #1
\{ \int_case:nnF { \seq_count:N #1 } 1#2#3#4
\{ 0 \} \}
\{ 1 \} \{ \exp_after:wN \_\_seq_use:NNnn #1 \? \} \}
\{ 2 \} \{ \exp_after:wN \_\_seq_use:NNnn #1 \#2 \}
\}
\}
\exp_after:wN \_\_seq_use_setup:w
\_\_seq_item:n
\s__seq_mark \_\_seq_use:nwvnn \#3 \}
\s__seq_mark \_\_seq_use:nwvnn \#4 \}
\s__seq_stop \}
\}

803
\section*{56.8 Sequence stacks}

The same functions as for sequences, but with the correct naming.

\verb|\seq_push:Nn|  
Pushing to a sequence is the same as adding on the left.

\verb|\seq_push:NV|  
\verb|\seq_push:Nv|  
\verb|\seq_push:Nx|  
\verb|\seq_push:cn|  
\verb|\seq_push:cV|  
\verb|\seq_push:co|  
\verb|\seq_push:cx|  

\verb|\seq_gpush:Nn|  
\verb|\seq_gpush:NV|  
\verb|\seq_gpush:Nv|  
\verb|\seq_gpush:Nx|  
\verb|\seq_gpush:cn|  
\verb|\seq_gpush:cV|  
\verb|\seq_gpush:co|  
\verb|\seq_gpush:cx|  

(End definition for \verb|\seq_push:Nn| and \verb|\seq_gpush:Nn|. These functions are documented on page 152.)

\verb|\seq_get:NN|  
\verb|\seq_get:ccN|  
\verb|\seq_pop:NN|  
\verb|\seq_pop:ccN|  
\verb|\seq_gpop:NN|  
\verb|\seq_gpop:ccN|

In most cases, getting items from the stack does not need to specify that this is from the left. So alias are provided.

(End definition for \verb|\seq_get:NN| and \verb|\seq_gpop:NN|. These functions are documented on page 152.)
More copies.

(End definition for \seq_get:NN, \seq_pop:NN, and \seq_gpop:NN. These functions are documented on page 151.)

56.9 Viewing sequences

Apply the general \_kernel_chk_tl_type:NnnT.

(End definition for \seq_get:NNTF, \seq_pop:NNTF, and \seq_gpop:NNTF. These functions are documented on page 151.)
56.10 Scratch sequences

\lst{\l_tmpa_seq} Temporary comma list variables.
\lst{\l_tmpb_seq} \seq_new:N \l_tmpa_seq
\lst{\g_tmpa_seq} \seq_new:N \l_tmpb_seq
\lst{\g_tmpb_seq} \seq_new:N \g_tmpa_seq
\seq_new:N \g_tmpb_seq

(End definition for \l_tmpa_seq and others. These variables are documented on page 154.)

{/package}
Chapter 57

\textbf{\l3int implementation}

The following test files are used for this code: \texttt{m3int001,m3int002,m3int03}.

\texttt{\c_max_register_int}

Done in \texttt{l3basics}.

(End definition for \texttt{\c_max_register_int}. This variable is documented on page 166.)

\texttt{\_\_int_to_roman:w \_\_int_compare:w}

Done in \texttt{l3basics}.

(End definition for \texttt{\_\_int_to_roman:w} and \texttt{\_\_int_compare:w}. This function is documented on page 167.)

\texttt{\_\_int_to_roman:w \_\_int_compare:w}

Done in \texttt{l3basics}.

(End definition for \texttt{\_\_int_to_roman:w} and \texttt{\_\_int_compare:w}. This function is documented on page 167.)

\texttt{\_\_int_to_roman:w \_\_int_compare:w}

Done in \texttt{l3basics}.

(End definition for \texttt{\_\_int_to_roman:w} and \texttt{\_\_int_compare:w}. This function is documented on page 167.)

\texttt{\int_value:w \_\_int_eval:w \_\_int_odd_end: \_\_int_odd:w \_\_int_case:w}

Here are the remaining primitives for number comparisons and expressions.

(End definition for \texttt{\int_value:w} and others. These functions are documented on page 167.)

\texttt{\s__int_mark \s__int_stop}

Scan marks used throughout the module.

(End definition for \texttt{\s__int_mark} and \texttt{\s__int_stop}.)

\texttt{\_\_int_use_none_delimit_by_s_stop:w}

Function to gobble until a scan mark.

(End definition for \texttt{\_\_int_use_none_delimit_by_s_stop:w}.)

\texttt{\_\_int_recursion_tail \_\_int_recursion_stop}

Quarks for recursion.

(End definition for \texttt{\_\_int_recursion_tail} and \texttt{\_\_int_recursion_stop}.)
Functions to query quarks.

17030 \_kernel_quark_new_test:N \_int_if_recursion_tail_stop_do:Nn
17031 \_kernel_quark_new_test:N \_int_if_recursion_tail_stop:N

(End definition for \_int_if_recursion_tail_stop_do:Nn and \_int_if_recursion_tail_stop:N.)

57.1 Integer expressions

\int_eval:n \int_eval:w

Wrapper for \_int_eval:w: can be used in an integer expression or directly in the input stream. It is very slightly faster to use \the rather than \number to turn the expression to a number. When debugging, we introduce parentheses to catch early termination (see \debug).

17032 \cs_new:Npn \int_eval:n #1
17033 \{ \tex_the:D \_int_eval:w #1 \_int_eval_end: \}
17034 \cs_new:Npn \int_eval:w \{ \tex_the:D \_int_eval:w \}

(End definition for \int_eval:n and \int_eval:w. These functions are documented on page 156.)

\int_sign:n \_int_sign:Nw

See \int_abs:n. Evaluate the expression once (and when debugging is enabled, check that the expression is well-formed), then test the first character to determine the sign. This is wrapped in \int_value:w... \exp_stop_f: to ensure a fixed number of expansions and to avoid dealing with closing the conditionals.

17035 \cs_new:Npn \int_sign:n #1
17036 \{ \int_value:w \exp_after:wN \_int_sign:Nw
17037 \int_value:w \_int_eval:w #1 \_int_eval_end: ;
17038 \exp_stop_f:
17039 \}
17040 \cs_new:Npn \_int_sign:Nw #1#2 ;
17041 \{ \int_value:w \_int_sign:Nw #1 #2 \\
17042 \exp_after:wN \_int_sign:Nw
17043 \_int_eval:w #1 \_int_eval_end: ;
17044 \exp_stop_f:
17045 \}
17046 \if_meaning:w 0 #1
17047 \int_value:w 0 #1
17048 \else:
17049 \_int_sign:Nw - #1 - \fi: 1
17050 \fi:
17051 \}

(End definition for \int_sign:n and \_int_sign:Nw. This function is documented on page 156.)

\int_abs:n \_int_abs:N \int_max:nn \int_min:nn \_int_maxmin:wwN

Functions for min, max, and absolute value with only one evaluation. The absolute value is obtained by removing a leading sign if any. All three functions expand in two steps.

17046 \cs_new:Npn \int_abs:n #1
17047 \{ \int_value:w \exp_after:wN \_int_abs:N
17048 \int_value:w \_int_eval:w #1 \_int_eval_end: ;
17049 \exp_stop_f:
17050 \}
17051 \cs_new:Npn \_int_abs:N #1
17052 \{ \int_value:w \exp_after:wN \_int_abs:N
17053 \int_value:w \_int_eval:w #1 \_int_eval_end: ;
17054 \exp_stop_f:
17055 \}
17056 \cs_new:Npn \_int_abs:N #1
17057 \{ \int_value:w \_int_abs:N \exp_after:wN \_int_abs:N
17058 \_int_eval:w #1 \_int_eval_end: ;
17059 \exp_stop_f:
17060 \}
17061 \cs_set:Npn \int_max:nn #1#2
17062 \{ \_int_maxmin:wwN \_int_maxmin:wwN
17063 \_int_eval:w #1 \_int_eval_end:
17064 \exp_stop_f:
17065 \}
17066 \cs_new:Npn \_int_maxmin:wwN #1
17067 \{ \_int_maxmin:wwN \_int_maxmin:wwN
17068 \_int_eval:w #1 \_int_eval_end:
17069 \exp_stop_f:
17070 \}
17071 \cs_new:Npn \_int_max:n #1
17072 \{ \_int_max:n \exp_after:wN \_int_max:n
17073 \_int_eval:w #1 \_int_eval_end: ;
17074 \exp_stop_f:
17075 \}
17076 \int_set:Npn \int_max:nn #1#2
17077 \{ \int_set:Npn \int_max:nn \_int_max:n #1
17078 \}

808
\input{tkz-base}

\newcommand{\int_div_truncate}{\int_div_round}
\newcommand{\int_mod}{\int_eval}
\newcommand{\__int_div_truncate}{\__int_mod}
\newcommand{\__int_mod}{\__int_maxmin}

\begin{verbatim}
\exp_stop_f:
\end{verbatim}

(End definition for \int_abs:n and others. These functions are documented on page 156.)

As \__int_eval:w rounds the result of a division we also provide a version that truncates the result. We use an auxiliary to make sure numerator and denominator are only evaluated once: this comes in handy when those are more expressions are expensive to evaluate (e.g., \tl_count:n). If the numerator \#1\#2 is 0, then we divide 0 by the denominator (this ensures that 0/0 is correctly reported as an error). Otherwise, shift the numerator \#1\#2 towards 0 by \((\#3\#4-1)/2\), which we round away from zero. It turns out that this quantity exactly compensates the difference between \$\varepsilon\$-\TeX{}’s rounding and the truncating behaviour that we want. The details are thanks to Heiko Oberdiek: getting things right in all cases is not so easy.

\input{tkz-base}
\begin{verbatim}
\exp_stop_f:
\end{verbatim}

809
For the sake of completeness:
\cs_new:Npn \int_div_round:nn #1 #2
{ \int_value:w \__int_eval:w ( #1 ) / ( #2 ) \__int_eval_end: }

Finally there’s the modulus operation.
\cs_new:Npn \int_mod:nn #1 #2
{ \int_value:w \__int_eval:w \exp_after:wN \__int_mod:ww
  \int_value:w \__int_eval:w #1 \exp_after:wN ;
  \int_value:w \__int_eval:w #2 ;
  \__int_eval_end: }
\cs_new:Npn \__int_mod:ww #1; #2;
{ #1 - ( \__int_div_truncate:NwNw #1 ; #2 ; ) * #2 }

(End definition for \int_div_truncate:nn and others. These functions are documented on page 156.)

\__kernel_int_add:nnn
Equivalent to \int_eval:n {#1+#2+#3} except that overflow only occurs if the final result overflows \([-2^{31} + 1, 2^{31} - 1]\). The idea is to choose the order in which the three numbers are added together. If #1 and #2 have opposite signs (one is in \([-2^{31} + 1, -1]\) and the other in \([0, 2^{31} - 1]\)) then #1+#2 cannot overflow so we compute the result as #1+#2+#3. If they have the same sign, then either #3 has the same sign and the order does not matter, or #3 has the opposite sign and any order in which #3 is not last will work. We use #1+#3+#2.
\cs_new:Npn \__kernel_int_add:nnn #1 #2 #3
{ \int_value:w \__int_eval:w #1
  \if_int_compare:w #2 < \c_zero_int \exp_after:wN \__int_eval_end: \fi:
    \if_int_compare:w #1 < \c_zero_int + #2 + #3 \else: + #3 + #2 \fi:
  \__int_eval_end: }

(End definition for \__kernel_int_add:nnn.)

57.2 Creating and initialising integers

\int_new:N \int_new:c
Two ways to do this: one for the format and one for the \LaTeX{} package. In plain \TeX{}, \newcount (and other allocators) are \texttt{outer}: to allow the code here to work in “generic” mode this is therefore accessed by name. (The same applies to \newbox, \newdimen and so on.)
\cs_new_protected:Npn \int_new:N #1
{ \if_int_compare:w \__kernel_chk_if_free_cs:N #1 \cs_end: \fi:
  \cs:w newcount \cs_end: #1 }
\cs_generate_variant:Nn \int_new:N { c }

(End definition for \int_new:N. This function is documented on page 157.)
As stated, most constants can be defined as \texttt{chardef} or \texttt{mathchardef} but that’s engine dependent. As a result, there is some set up code to determine what can be done. No full engine testing just yet so everything is a little awkward. We cannot use \texttt{int_gset:Nn} because (when \texttt{check-declarations} is enabled) this runs some checks that constants would fail.

\begin{verbatim}
\cs_new_protected:Npm \int_const:Nn \c__int_max_constdef_int
\int_compare:nNnTF {\#1} < \c_zero_int
{ \int_new:N \tex_global:D }
\int_compare:nNnTF {\#1} > \c__int_max_constdef_int
{ \__kernel_chk_if_free_cs:N \tex_global:D \__int_constdef:Nw }
\__int_eval:w \#1 \__int_eval_end:
\cs_generate_variant:Nn \int_const:Nn { c }
\if_int_odd:w 0 \cs_if_exist:NT \tex_luatexversion:D { 1 }
\cs_if_exist:NT \tex_omathchardef:D { 1 }
\cs_if_exist:NTF \tex_omathchardef:D { \cs_new_eq:NN \__int_constdef:Nw \tex_omathchardef:D }
{ \cs_new_eq:NN \__int_constdef:Nw \tex_mathchardef:D }
\__int_constdef:Nw \c__int_max_constdef_int 1114111 ~
\else:
\cs_new_eq:NN \__int_constdef:Nw \tex_mathchardef:D \c__int_max_constdef_int 32767 ~
\fi:
\end{verbatim}

(End definition for \texttt{int_const:Nn} and others. This function is documented on page 157.)

\begin{verbatim}
\int_zero:N \int_zero:c \int_gzero:N \int_gzero:c \int_zero_new:N \int_zero_new:c \int_gzero_new:N \int_gzero_new:c
\cs_new_protected:Npm \int_zero:N \c_zero_int
\cs_new_protected:Npm \int_gzero:N \c_zero_int
\cs_new_protected:Npm \int_zero_new:N \\c_zero_int
\cs_new_protected:Npm \int_gzero_new:N \\c_zero_int
\cs_generate_variant:Nn \int_zero:N { c }
\cs_generate_variant:Nn \int_gzero:N { c }
\cs_generate_variant:Nn \int_zero_new:N { c }
\cs_generate_variant:Nn \int_gzero_new:N { c }
\cs_new_protected:Npm \\c_zero_int
\cs_new_protected:Npm \\c_zero_int
\cs_new_protected:Npm \\c_zero_int
\cs_new_protected:Npm \\c_zero_int
\cs_generate_variant:Nn \int_zero:N { c }
\cs_generate_variant:Nn \int_gzero:N { c }
\cs_generate_variant:Nn \int_zero_new:N { c }
\cs_generate_variant:Nn \int_gzero_new:N { c }
\end{verbatim}

Functions that reset an \texttt{(integer)} register to zero.

(End definition for \texttt{int_zero:N} and \texttt{int_gzero:N}. These functions are documented on page 157.)

Create a register if needed, otherwise clear it.

811
57.3 Setting and incrementing integers

Adding and subtracting to and from a counter. Including here the optional by would slow down these operations by a few percent.

Incrementing and decrementing of integer registers is done with the following functions.

(End definition for \texttt{\texttt{int\_incr:}N} and others. These functions are documented on page 158.)
As integers are register-based \TeX{} issues an error if they are not defined. While the =

sign is optional, this version with = is slightly quicker than without, while adding

the optional space after = slows things down minutely.

\begin{tabular}{ll}
& \texttt{int_set:Nn} \\
& \texttt{int_set:cn} \\
& \texttt{int_gset:Nn} \\
& \texttt{int_gset:cn}
\end{tabular}

(End definition for \texttt{int_incr:N} and others. These functions are documented on page 158.)

\begin{tabular}{ll}
& \texttt{int_use:N} \\
& \texttt{int_use:c} \\
\end{tabular}

Here is how counters are accessed. We hand-code the \texttt{c} variant for some speed gain.

\begin{tabular}{ll}
& \texttt{int_set:Nn} \\
& \texttt{int_set:cn} \\
& \texttt{int_gset:Nn} \\
& \texttt{int_gset:cn}
\end{tabular}

(End definition for \texttt{int_set:Nn} and \texttt{int_gset:Nn}. These functions are documented on page 158.)

57.4 Using integers

\begin{tabular}{ll}
& \texttt{int_use:N} \\
& \texttt{int_use:c} \\
\end{tabular}

57.5 Integer expression conditionals

Those functions are used for comparison tests which use a simple syntax where only

one set of braces is required and additional operators such as \texttt{!=} and \texttt{>=} are supported.

The tests first evaluate their left-hand side, with a trailing \texttt{\_\_int_compare_error:}.

This marker is normally not expanded, but if the relation symbol is missing from the

test’s argument, then the marker inserts = (and itself) after triggering the relevant \TeX{}

error. If the first token which appears after evaluating and removing the left-hand side is

not a known relation symbol, then a judiciously placed \texttt{\_\_int_compare_error:Nw} gets

expanded, cleaning up the end of the test and telling the user what the problem was.

\begin{tabular}{ll}
& \texttt{int_new_protected:Nw} \\
\end{tabular}

\begin{tabular}{ll}
& \texttt{\_\_int_compare_error:} \\
& \texttt{\_\_int_compare_error:Nw}
\end{tabular}

813
Comparison tests using a simple syntax where only one set of braces is required, additional operators such as != and >= are supported, and multiple comparisons can be performed at once, for instance 0 < 5 <= 1. The idea is to loop through the argument, finding one operand at a time, and comparing it to the previous one. The looping auxiliary \_\_int_compare:Nw reads one (operand) and one (comparison) symbol, and leaves roughly

\( \text{(operand)} \ \text{prg\_return\_false:} \ \text{\textbackslash fi:} \) \reverse_if:N \if_int_compare:w \langle operand \rangle \langle comparison \rangle \\__int_compare:Nw

in the input stream. Each call to this auxiliary provides the second operand of the last call’s if_int_compare:w. If one of the (comparisons) is false, the true branch of the \text{T\text{}X} conditional is taken (because of reverse_if:N), immediately returning false as the result of the test. There is no \text{T\text{}X} conditional waiting the first operand, so we add an if_false: and expand by hand with \texttt{\textbackslash int\_value:w}, thus skipping prg_return_-false: on the first iteration.

Before starting the loop, the first step is to make sure that there is at least one relation symbol. We first let \text{T\text{}X} evaluate this left hand side of the (in)equality using \_\_int_eval:w. Since the relation symbols <, >, = and ! are not allowed in integer expressions, they would terminate the expression. If the argument contains no relation symbol, \_\_int_compare_error: is expanded, inserting = and itself after an error. In all cases, \_\_int_compare:w receives as its argument an integer, a relation symbol, and some more tokens. We then setup the loop, which is ended by the two odd-looking items e and {=nd_}, with a trailing \texttt{s\_\_int\_stop} used to grab the entire argument when necessary.

\begin{verbatim}
17226 \prg_new_conditional:Npnn \int_compare:n #1 { p , T , F , TF }
17227 { \exp_after:wN \_\_int_compare:w \int_value:w \_\_int_eval:w #1 \_\_int_compare_error: }
17228 \cs_new:Npn \_\_int_compare:w #1#2 \s__int_stop
17229 { \exp_after:wN \_\_int_compare:NNw \_\_int_to_roman:w - 0 #2 \s__int_mark }
\end{verbatim}

The goal here is to find an (operand) and a (comparison). The (operand) is already evaluated, but we cannot yet grab it as an argument. To access the following relation symbol, we remove the number by applying \_\_int_to_roman:w, after making sure that the argument becomes non-positive: its roman numeral representation is then empty. Then probe the first two tokens with \_\_int_compare:Nw to determine the relation symbol, building a control sequence from it (\texttt{\textbackslash token\_to\_str:N} gives better errors if #1 is not a character). All the extended forms have an extra = hence the test for that as a second token. If the relation symbol is unknown, then the control sequence is turned by \text{T\text{}X} into \texttt{\textbackslash can\_stop:}, ignored thanks to \texttt{\unexpanded}, and \_\_int_compare_error:Nw raises an error.

\begin{verbatim}
17238 \cs_new:Npn \_\_int_compare:Nw #1@ #2 \s__int_stop
17239 { \exp_after:wN \_\_int_compare:NNw \_\_int_to_roman:w - 0 #2 \s__int_mark }
\end{verbatim}

814
When the last \langle operand \rangle is seen, \_\_int_compare:NNw receives e and =nd as arguments, hence calling \_\_int_compare_end=:NNw to end the loop: return the result of the last comparison (involving the operand that we just found). When a normal relation is found, the appropriate auxiliary calls \_\_int_compare:nnN where #1 is \if_int_compare:w or \reverse_if:N \if_int_compare:w, #2 is the \langle operand \rangle, and #3 is one of \<, =, or \>. As announced earlier, we leave the \langle operand \rangle for the previous conditional. If this conditional is true the result of the test is known, so we remove all tokens and return false. Otherwise, we apply the conditional #1 to the \langle operand \rangle #2 and the comparison #3, and call \_\_int_compare:Nw to look for additional operands, after evaluating the following expression.

The actual comparisons are then simple function calls, using the relation as delimiter for a delimited argument and discarding \_\_int_compare_error:Nw \langle token \rangle responsible for error detection.
End definition for \int_compare:nTF and others. This function is documented on page 159.

\int_compare_p:nNn
\int_compare:nNn
More efficient but less natural in typing.

\prg_new_conditional:Npnn \int_compare:nNn #1#2#3 \{ p , T , F , TF \}
\{ \if_int_compare:w \__int_eval:w #1 #2 \__int_eval:w #3 \__int_eval_end: 
\prg_return_true:
\} \else:
\prg_return_false:
\fi:
\}

(End definition for \int_compare:nNnTF. This function is documented on page 159.)

\int_case:nn
\int_case:nNn
\__int_case:nnTF
\__int_case:nw
\__int_case_end:nw
For integer cases, the first task to fully expand the check condition. The over all idea
is then much the same as for \tl_case:nnTF as described in \l3tl.

\cs_new:Npn \int_case:nnTF #1 \{ \exp:w \exp_args:Nf \__int_case:nnTF { \int_eval:n {#1} } \}
\cs_new:Npn \int_case:nnT #1#2#3 \{ \exp:w \exp_args:Nf \__int_case:nnTF { \int_eval:n {#1} } {#2} {#3} { } \}
\cs_new:Npn \int_case:nnF #1#2 \{ \exp:w \exp_args:Nf \__int_case:nnTF { \int_eval:n {#1} } {#2} { } \}
\cs_new:Npn \int_case:nn \#1#2 \{ \exp:w \exp_args:Nf \__int_case:nnTF { \int_eval:n {#1} } {#2} { } { } \}
\cs_new:Npn \__int_case:nnTF #1#2#3#4 \{ \__int_case:nw {#1} {#2} \s__int_mark {#3} \s__int_mark {#4} \s__int_stop \}
\cs_new:Npn \__int_case:nw #1#2#3 \{ \__int_case_end:nw {#1} {#2} \s__int_mark {#3} \s__int_mark {#4} \s__int_stop \}
\cs_new:Npn \__int_case_end:nw #1#2#3 #4#5 #6 #7 \{ \exp_end: #1 #4 \}

(End definition for \int_case:nnTF and others. This function is documented on page 160.)

\int_if_odd_p:n
\int_if_odd:n
\int_if_even_p:n
\int_if_even:n
A predicate function.

\prg_new_conditional:Npnn \int_if_odd:n #1 \{ p , T , F , TF \}
\{ \if_int_odd:w \__int_eval:w #1 \__int_eval_end: \prg_return_true: \}
\}
\}
\}
\}
\}
\}

816
57.6 Integer expression loops

These are quite easy given the above functions. The \texttt{while} versions test first and then execute the body. The \texttt{do_while} does it the other way round.

\begin{verbatim}
\cs_new:Npn \int_while_do:nn #1#2
\{ \int_compare:nT {#1} \{ \int_while_do:nn {#1} {#2} \} \}
\cs_new:Npn \int_until_do:nn #1#2
\{ \int_compare:nF {#1} \{ \int_until_do:nn {#1} {#2} \} \}
\cs_new:Npn \int_do_while:nn #1#2
\{ \int_compare:nT {#1} \{ \int_do_while:nn {#1} {#2} \} \}
\cs_new:Npn \int_do_until:nn #1#2
\{ \int_compare:nF {#1} \{ \int_do_until:nn {#1} {#2} \} \}
\end{verbatim}

(End definition for \texttt{int\_if\_odd:nTF} and \texttt{int\_if\_even:nTF}. These functions are documented on page 160.)

\begin{verbatim}
\cs_new:Npn \int_while_do:nNnn #1#2#3#4
\{ \int_compare:nT {#1} \{ \int_while_do:nNnn {#1} {#2} {#3} {#4} \} \}
\cs_new:Npn \int_until_do:nNnn #1#2#3#4
\{ \int_compare:nF {#1} \{ \int_until_do:nNnn {#1} {#2} {#3} {#4} \} \}
\cs_new:Npn \int_do_while:nNnn #1#2#3#4
\{ \int_compare:nT {#1} \{ \int_do_while:nNnn {#1} {#2} {#3} {#4} \} \}
\cs_new:Npn \int_do_until:nNnn #1#2#3#4
\{ \int_compare:nF {#1} \{ \int_do_until:nNnn {#1} {#2} {#3} {#4} \} \}
\end{verbatim}

(End definition for \texttt{int\_while\_do:nn} and others. These functions are documented on page 161.)

As above but not using the more natural syntax.
57.7 Integer step functions

Before all else, evaluate the initial value, step, and final value. Repeating a function by steps first needs a check on the direction of the steps. After that, do the function for the start value then step and loop around. It would be more symmetrical to test for a step size of zero before checking the sign, but we optimize for the most frequent case (positive step).

\int_step_function:nnnN
\__int_step:wwwN
\__int_step:NwwwN
\int_step_function:nN
\int_step_function:nnN

(End definition for \int_while_do:nNnn and others. These functions are documented on page 161.)
The approach here is to build a function, with a global integer required to make the nesting safe (as seen in other in line functions), and map that function using \int_step_function:nnnN. We put a \prg_break_point:Nn so that map_break functions from other modules correctly decrement \g__kernel_prg_map_int before looking for their own break point. The first argument is \scan_stop:, so that no breaking function recognizes this break point as its own.
57.8 Formatting Integers

\int_to_arabic:n

Nothing exciting here.

\int_to_alph:n \int_to_Alph:n

These both use the above function with input functions that make sense for the alphabet in English.
\cs_new:Npn \int_to_alph:n #1
{ \int_to_symbols:nnn {#1} { 26 }
} \int_to_symbols:nnn 
{ 1 } { a } 
{ 2 } { b } 
{ 3 } { c } 
{ 4 } { d } 
{ 5 } { e } 
{ 6 } { f } 
{ 7 } { g } 
{ 8 } { h } 
{ 9 } { i } 
{ 10 } { j } 
{ 11 } { k } 
{ 12 } { l } 
{ 13 } { m } 
{ 14 } { n } 
{ 15 } { o } 
{ 16 } { p } 
{ 17 } { q } 
{ 18 } { r } 
{ 19 } { s } 
{ 20 } { t } 
{ 21 } { u } 
{ 22 } { v } 
{ 23 } { w } 
{ 24 } { x } 
{ 25 } { y } 
{ 26 } { z } 
\cs_new:Npn \int_to_Alph:n #1
{ \int_to_symbols:nnn {#1} { 26 }
} \int_to_symbols:nnn 
{ 1 } { A } 
{ 2 } { B } 
{ 3 } { C } 
{ 4 } { D } 
{ 5 } { E } 
{ 6 } { F } 
{ 7 } { G } 
{ 8 } { H } 
{ 9 } { I } 
{ 10 } { J } 
{ 11 } { K } 
{ 12 } { L } 
{ 13 } { M } 
{ 14 } { N } 
{ 15 } { O } 
{ 16 } { P } 
{ 17 } { Q } 
{ 18 } { R }
Converting from base ten (#1) to a second base (#2) starts with computing #1: if it is a complicated calculation, we shouldn’t perform it twice. Then check the sign, store it, either – or \texttt{\c_empty_tl}, and feed the absolute value to the next auxiliary function.

Here, the idea is to provide a recursive system to deal with the input. The output is built up after the end of the function. At each pass, the value in #1 is checked to see if it is less than the new base (#2). If it is, then it is converted directly, putting the sign back in front. On the other hand, if the value to convert is greater than or equal to the new base then the modulus and remainder values are found. The modulus is converted to a symbol and put on the right, and the remainder is carried forward to the next round.
Convert to a letter only if necessary, otherwise simply return the value unchanged. It would be cleaner to use \int_case:nn, but in our case, the cases are contiguous, so it is forty times faster to use the \if_case:w primitive. The first \exp_after:wN expands the conditional, jumping to the correct case, the second one expands after the resulting character to close the conditional. Since #1 might be an expression, and not directly a single digit, we need to evaluate it properly, and expand the trailing \fi::

\cs_new:Npn \__int_to_letter:n #1
\{\exp_after:wN \exp_after:wN\if_case:w \__int_eval:w #1 - 10 \__int_eval_end:\a\or: b\or: c\or: d\or: e\or: f\or: g\or: h\or: i\or: j\or: k\or: l\or: m\or: n\or: o\or: p\or: q\or: r\fi:\}

\cs_new:Npn \__int_to_Base:nnnN #1#2#3
\{\exp_after:wN \exp_after:wN \int_div_truncate:nn \int_compare:nNnTF {#1} < {#2}\{ \exp_last_unbraced:Nf \__int_to_Letter:n {#1} \}\exp_args:Nf \__int_to_Base:nnN \int_mod:nn {#1} {#2} \{#1\} \{#2\} \#3\}
\cs_new:Npn \__int_to_Base:nnnN #1#2#3 #4
\{\exp_args:Nf \__int_to_Base:nnN \int_div_truncate:nn {#2} {#3} \{#3\} \#4 #1\}
\cs_new:Npn \__int_to_Base:nn {#2} {#3} #4
\or: s
\or: t
\or: u
\or: v
\or: w
\or: x
\or: y
\or: z
\else: \int_value:w \__int_eval:w #1 \exp_after:wN \__int_eval_end:
\fi:
\}
\cs_new:Npn \__int_to_Letter:n #1
{
\exp_after:wN \exp_after:wN
\if_case:w \__int_eval:w #1 - 10 \__int_eval_end:
A
\or: B
\or: C
\or: D
\or: E
\or: F
\or: G
\or: H
\or: I
\or: J
\or: K
\or: L
\or: M
\or: N
\or: O
\or: P
\or: Q
\or: R
\or: S
\or: T
\or: U
\or: V
\or: W
\or: X
\or: Y
\or: Z
\else: \int_value:w \__int_eval:w #1 \exp_after:wN \__int_eval_end:
\fi:
\}

(End definition for \int_to_base:nn and others. These functions are documented on page 164.)

\int_to_bin:n Wrappers around the generic function.
\int_to_hex:n
\int_to_Hex:n
\int_to_oct:n
\int_to_base:nn {#1} { 2 }
\int_to_Base:nn {#1} { 16 }

824
The \__int_to_roman:w primitive creates tokens of category code 12 (other). Usually, what is actually wanted is letters. The approach here is to convert the output of the primitive into letters using appropriate control sequence names. That keeps everything expandable. The loop is terminated by the conversion of the $Q$.

The \__int_to_roman:w primitive creates tokens of category code 12 (other). Usually, what is actually wanted is letters. The approach here is to convert the output of the primitive into letters using appropriate control sequence names. That keeps everything expandable. The loop is terminated by the conversion of the $Q$.

\begin{verbatim}
\cs_new:Npn \int_to_roman:n #1
\exp_after:wN \__int_to_roman:N
\__int_to_roman:w \int_eval:n {#1} Q
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__int_to_Roman:N #1
\use:c { __int_to_Roman_#1:w }
\__int_to_Roman:N
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \int_to_Roman:n #1
\exp_after:wN \__int_to_Roman_aux:N
\__int_to_roman:w \int_eval:n {#1} Q
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__int_to_Roman_aux:N #1
\use:c { __int_to_Roman_#1:w }
\__int_to_Roman_aux:N
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__int_to_roman_i:w { i }
\cs_new:Npn \__int_to_roman_v:w { v }
\cs_new:Npn \__int_to_roman_x:w { x }
\cs_new:Npn \__int_to_roman_l:w { l }
\cs_new:Npn \__int_to_roman_c:w { c }
\cs_new:Npn \__int_to_roman_d:w { d }
\cs_new:Npn \__int_to_roman_m:w { m }
\cs_new:Npn \__int_to_roman_Q:w #1 { }
\cs_new:Npn \__int_to_Roman_i:w { I }
\cs_new:Npn \__int_to_Roman_v:w { V }
\cs_new:Npn \__int_to_Roman_x:w { X }
\cs_new:Npn \__int_to_Roman_l:w { L }
\cs_new:Npn \__int_to_Roman_c:w { C }
\cs_new:Npn \__int_to_Roman_d:w { D }
\cs_new:Npn \__int_to_Roman_m:w { M }
\cs_new:Npn \__int_to_Roman_Q:w #1 { }
\end{verbatim}

\begin{verbatim}
\__int_pass_signs:wn \__int_pass_signs_end:wn
\end{verbatim}

\section*{57.9 Converting from other formats to integers}

Called as \__int_pass_signs:wn \texttt{⟨signs and digits⟩} \s__int_stop \{⟨code⟩\}, this function leaves in the input stream any sign it finds, then inserts the \texttt{⟨code⟩} before the first
non-sign token (and removes \s__int_stop). More precisely, it deletes any + and passes any - to the input stream, hence should be called in an integer expression.

\cs_new:Npn \__int_pass_signs:wn #1 \exp_after:wN \__int_pass_signs:wn
\exp_after:wN \__int_pass_signs_end:wn \exp_after:wN #1 \fi:
\exp_after:wN \__int_pass_signs_end:wn #1 \s__int_stop #2 { #2 #1 }

\cs_new:Npn \__int_pass_signs_end:wn #1 \s__int_stop #2 { #2 #1 }

\cs_new:Npn \__int_from_alph:nN #1#2
\__int_from_alph:N #1 \exp_after:wN \__int_pass_signs:wn \tl_to_str:n {#1}
\__int_from_alph:N \q__int_recursion_tail \q__int_recursion_stop

\cs_new:Npn \__int_from_alph:N #1
\__int_from_alph:nN \__int_from_base:nnN \__int_from_base:N
\__int_from_base:nn
Leave the signs into the integer expression, then loop through characters, collecting the value found so far in the first argument of \__int_from_base:nnN. To convert a single character, \__int_from_base:N checks first for digits, then distinguishes lower from upper case letters, turning them into the appropriate number. Note that this auxiliary does not use \int_eval:n, hence is not safe for general use.

\cs_new:Npn \__int_from_base:nn #1#2
\__int_from_base:nnN \__int_from_base:nn
\__int_from_base:N

\cs_new:Npn \__int_from_base:nnN #1#2#3
{\__int_if_recursion_tail_stop_do:Nn #3 {#1}\exp_args:Nf \__int_from_base:nnN
{ \int_eval:n { #1 * #2 + \__int_from_base:N #3 } } {#2}}
\cs_new:Npn \__int_from_base:N #1
{\int_compare:nNnTF { '#1 } < { 58 } { #1 } { '#1 - \int_compare:nNnTF { '#1 } < { 91 } { 55 } { 87 } }}
\cs_new:Npn \int_from_base:nn, \__int_from_base:nnN, and \__int_from_base:N. This function is documented on page 165.
\cs_new:Npn \int_from_bin:n #1
{ \int_from_base:nn {#1} { 2 } }
\cs_new:Npn \int_from_hex:n #1
{ \int_from_base:nn {#1} { 16 } }
\cs_new:Npn \int_from_oct:n #1
{ \int_from_base:nn {#1} { 8 } }
\cs_new:Npn \int_from_roman:n #1
{\int_eval:n
{\int_from_roman:nn \__int_from_roman:NN \__int_from_roman_error:w #1}
\int_const:cn { c__int_from_roman_i_int } { 1 }
\int_const:cn { c__int_from_roman_v_int } { 5 }
\int_const:cn { c__int_from_roman_x_int } { 10 }
\int_const:cn { c__int_from_roman_l_int } { 50 }
\int_const:cn { c__int_from_roman_d_int } { 100 }
\int_const:cn { c__int_from_roman_m_int } { 500 }
\int_const:cn { c__int_from_roman_I_int } { 1000 }
\int_const:cn { c__int_from_roman_V_int } { 5 }
\int_const:cn { c__int_from_roman_X_int } { 10 }
\int_const:cn { c__int_from_roman_L_int } { 50 }
\int_const:cn { c__int_from_roman_C_int } { 100 }
\int_const:cn { c__int_from_roman_D_int } { 500 }
\int_const:cn { c__int_from_roman_M_int } { 1000 }
\cs_new:Npn \int_from_roman:n #1
{\int_eval:n
{\int_from_roman:nn \__int_from_roman:NN \__int_from_roman_error:w #1}
(End definition for \__int_from_roman:n, \__int_from_roman:NN, and \__int_from_roman_error:w. This function is documented on page 165.)

57.10 Viewing integer

\int_show:N \int_show:c \__int_show:nN

(Diagnostics.

\int_show:n

We don’t use the \TeX{} primitive \showthe{} to show integer expressions: this gives a more unified output.

\int_log:N \int_log:c

(Diagnostics.

828
Similar to \texttt{\int_show:n}.

\begin{verbatim}
\cs_new_protected:Npn \int_log:n { \msg_log_eval:Nn \int_eval:n }
\end{verbatim}

57.11 Random integers

\texttt{\int_rand:nn} Defined in \texttt{l3fp-random}.

57.12 Constant integers

\texttt{\c_zero_int} The zero is defined in \texttt{l3basics}.

\texttt{\c_one_int} \begin{verbatim}
\int_const:Nn \c_one_int { 1 }
\end{verbatim}

\texttt{\c_max_int} The largest number allowed is $2^{31} - 1$

\texttt{\c_max_char_int} \begin{verbatim}
\int_const:Nn \c_max_char_int { \if_int_odd:w 0 \\
\cs_if_exist:NT \tex_luatexversion:D { 1 } \\
\cs_if_exist:NT \tex_XeTeXversion:D { 1 } -
\else:
\texttt{10FFFF } \\
\else:
\texttt{FF}
\fi:}
\end{verbatim}

57.13 Scratch integers

\texttt{\l_tmpa_int} \texttt{\l_tmpb_int} \texttt{\g_tmpa_int} \texttt{\g_tmpb_int} We provide two local and two global scratch counters, maybe we need more or less.

\begin{verbatim}
\int_new:N \l_tmpa_int \\
\int_new:N \l_tmpb_int \\
\int_new:N \g_tmpa_int \\
\int_new:N \g_tmpb_int
\end{verbatim}

(End definition for \texttt{\l_tmpa_int} and \texttt{\l_tmpb_int}. These variables are documented on page 166.)
57.14 Integers for earlier modules

\l_int_internal_a_int
\l_int_internal_b_int
\int_new:N \l_int_internal_a_int
\int_new:N \l_int_internal_b_int

(End definition for \l_int_internal_a_int and \l_int_internal_b_int.)

\package
Chapter 58

\l3flag implementation

The following test files are used for this code: m3flag001.

58.1 Non-expandable flag commands

The height \( h \) of a flag (initially zero) is stored by setting control sequences of the form \( \texttt{\textbackslash flag \langle name \rangle \langle integer \rangle} \) to \texttt{\textbackslash relax} for \( 0 \leq \langle integer \rangle < h \). When a flag is raised, a “trap” function \( \texttt{\textbackslash flag \langle name \rangle} \) is called. The existence of this function is also used to test for the existence of a flag.

\texttt{\textbackslash flag\_new:n}

For each flag, we define a “trap” function, which by default simply increases the flag by 1 by letting the appropriate control sequence to \texttt{\textbackslash relax}. This can be done expandably!

\begin{verbatim}
\cs_new_protected:Npn \flag_new:n #1
\{ \cs_new:cpn { flag~#1 } ##1 ; \exp_after:wN \use_none:n \cs:w flag~#1~##1 \cs_end: \}
\end{verbatim}

(End definition for \texttt{\textbackslash flag\_new:n}. This function is documented on page 169.)

\texttt{\textbackslash flag\_clear:n} \texttt{\_\_flag\_clear:wn}

Undefine control sequences, starting from the 0 flag, upwards, until reaching an undefined control sequence. We don’t use \texttt{\cs_undefine:c} because that would act globally. When the option \texttt{check-declarations} is used, check for the function defined by \texttt{\textbackslash flag\_new:n}.

\begin{verbatim}
\cs_new_protected:Npn \flag_clear:n #1 \{ \_\_flag_clear:wn 0 ; {#1} \}
\cs_new_protected:Npn \_\_flag_clear:wn #1 ; #2
\{ \ifcs_exist:w flag~#2~#1 \cs_end: \cs_set_eq:cN { flag~#2~#1 } \tex_undefined:D \exp_after:wN \_\_flag_clear:wn \int_value:w \int_eval:w 1 + #1 \else: \use_i:nnn \fi: \{#2\}
\}
\end{verbatim}
58.2 Expandable flag commands

\flag_if_exist_nTF\flag_if_raised_nTF

A flag exist if the corresponding trap \flag \{flag name\}:n is defined.

\flag_if_exist_nTF { \flag_if_raised_nTF Test if the flag has a non-zero height, by checking the 0 control sequence.

\flag_if_raised_nTF \flag_if_raised_nTF

Extract the value of the flag by going through all of the control sequences starting from 0.

\flag_height_nTF

\flag_clear_nTF

As for other datatypes, clear the \{flag\} or create a new one, as appropriate.
\exp_after:wN \__flag_height_end:wn
\fi:
#1 ; {#2}
}
\cs_new:Npn \__flag_height_end:wn #1 ; #2 {#1}

(End definition for \flag_height:n, \__flag_height_loop:wn, and \__flag_height_end:wn. This function is documented on page 170.)

\flag_raise:n Simply apply the trap to the height, after expanding the latter.
\cs_new:Npn \flag_raise:n #1
{\cs:w flag~#1 \exp_after:wN \cs_end:
\int_value:w \flag_height:n {#1} ;
}

(End definition for \flag_raise:n. This function is documented on page 170.)
(/package)
Chapter 59

\texttt{l3clist} implementation

The following test files are used for this code: \texttt{m3clist002}.

\begin{verbatim}
\verb|\c_empty_clist| An empty comma list is simply an empty token list.
\end{verbatim}

\begin{verbatim}
\cs_new_eq:NN \c_empty_clist \c_empty_tl
\end{verbatim}

(End definition for \texttt{\c_empty_clist}. This variable is documented on page 180.)

\begin{verbatim}
\l__clist_internal_clist Scratch space for various internal uses. This comma list variable cannot be declared as such because it comes before \texttt{\clist_new:N}
\end{verbatim}

\begin{verbatim}
\tl_new:N \l__clist_internal_clist
\end{verbatim}

(End definition for \texttt{\l__clist_internal_clist}.)

\begin{verbatim}
\s__clist_mark \s__clist_stop Internal scan marks.
\end{verbatim}

\begin{verbatim}
\scan_new:N \s__clist_mark
\scan_new:N \s__clist_stop
\end{verbatim}

(End definition for \texttt{\s__clist_mark and \s__clist_stop}.)

\begin{verbatim}
\__clist_use_none_delimit_by_s_mark:w \__clist_use_none_delimit_by_s_stop:w \__clist_use_i_delimit_by_s_stop:nw
Functions to gobble up to a scan mark.
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__clist_use_none_delimit_by_s_mark:w #1 \s__clist_mark { }
\cs_new:Npn \__clist_use_none_delimit_by_s_stop:w #1 \s__clist_stop { }
\cs_new:Npn \__clist_use_i_delimit_by_s_stop:nw #1 #2 \s__clist_stop {#1}
\end{verbatim}

(End definition for \texttt{\__clist_use_none_delimit_by_s_mark:w, \__clist_use_none_delimit_by_s_stop:w, \__clist_use_i_delimit_by_s_stop:nw}.)

\begin{verbatim}
\__clist_tmp:w A temporary function for various purposes.
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__clist_tmp:w { }
\end{verbatim}

(End definition for \texttt{\__clist_tmp:w}.)
59.1 Removing spaces around items

\_\_clist\_trim\_next:w

Called as \exp:w \_\_clist\_trim\_next:w \prg\_do\_nothing: ⟨comma list⟩ ... it expands to {⟨trimmed item⟩} where the ⟨trimmed item⟩ is the first non-empty result from removing spaces from both ends of comma-delimited items in the ⟨comma list⟩. The \prg\_do\_nothing: marker avoids losing braces. The test for blank items is a somewhat optimized \tl\_if\_empty:oTF construction; if blank, another item is sought, otherwise trim spaces.

\cs\_new:Npn \_\_clist\_trim\_next:w { \tl\_if\_empty:oTF { \use\_none:nn #1 ? } { \_\_clist\_trim\_next:w \prg\_do\_nothing: } { \tl\_trim\_spaces\_apply:oN {#1} \exp\_end: } }

(\hspace{1em} End definition for \_\_clist\_trim\_next:w.)

\_\_clist\_sanitize:n
\_\_clist\_sanitize:Nn

The auxiliary \_\_clist\_sanitize:Nn receives a delimiter (\c\_empty_tl the first time, afterwards a comma) and that item as arguments. Unless we are done with the loop it calls \_\_clist\_wrap\_item:w to unbrace the item (using a comma delimiter is safe since #2 came from removing spaces from an argument delimited by a comma) and possibly re-brace it if needed.

\cs\_new:Npn \_\_clist\_sanitize:n #1 { \exp\_after:wN \_\_clist\_sanitize:Nn \exp\_after:wN \c\_empty_tl \exp:w \_\_clist\_trim\_next:w \prg\_do\_nothing: \_\_clist\_stop \prg\_break:, \prg\_break\_point: }

\cs\_new:Npn \_\_clist\_sanitize:Nn #1#2 { \_\_clist\_use\_none\_delimit\_by\_s\_stop:w \_\_clist\_stop #1 \_\_clist\_wrap\_item:w #2 \_\_clist\_sanitize:Nn \exp\_after:wN \exp\_after:wN \exp:w \_\_clist\_trim\_next:w \prg\_do\_nothing: }

(\hspace{1em} End definition for \_\_clist\_sanitize:n and \_\_clist\_sanitize:Nn.)

\_\_clist\_if\_wrap:nTF
\_\_clist\_if\_wrap:w

True if the argument must be wrapped to avoid getting altered by some clist operations. That is the case whenever the argument

- starts or end with a space or contains a comma,
- is empty, or
- consists of a single braced group.

If the argument starts or ends with a space or contains a comma then one of the three arguments of \_\_clist\_if\_wrap:w will have its end delimiter (partly) in one of the three copies of #1 in \_\_clist\_if\_wrap:nTF; this has a knock-on effect meaning that the result of the expansion is not empty; in that case, wrap. Otherwise, the argument is safe unless it starts with a brace group (or is empty) and it is empty or consists of a single n-type argument.

\prg\_new\_conditional:Npnn \_\_clist\_if\_wrap:n #1 { TF }
59.2 Allocation and initialisation

Internally, comma lists are just token lists.

Creating and initializing a constant comma list is done by sanitizing all items (stripping spaces and braces).

Clearing comma lists is just the same as clearing token lists.
Once again a copy from the token list functions.

\clist_set_eq:NN \clist_set_eq:cN \clist_set_eq:Nc \clist_set_eq:cc
\clist_gset_eq:NN \clist_gset_eq:cN \clist_gset_eq:Nc \clist_gset_eq:cc

Once again, these are simple copies from the token list functions.

\clist_set_from_seq:NN \clist_set_from_seq:cN \clist_set_from_seq:Nc \clist_set_from_seq:cc
\clist_gset_from_seq:NN \clist_gset_from_seq:cN \clist_gset_from_seq:Nc \clist_gset_from_seq:cc
\__clist_set_from_seq:NNNN \__clist_set_from_seq:n

Setting a comma list from a comma-separated list is done using a simple mapping. Safe
items are put in \exp_not:n, otherwise we put an extra set of braces. The first comma
must be removed, except in the case of an empty comma-list.

(End definition for \clist_clear:N and \clist_gclear:N. These functions are documented on page
172.)

(End definition for \clist_clear:N and \clist_gclear:N. These functions are documented on page
172.)

(End definition for \clist_set_eq:NN and \clist_gset_eq:NN. These functions are documented on page
172.)
Concatenating comma lists is not quite as easy as it seems, as there needs to be the correct addition of a comma to the output. So a little work to do.

\begin{verbatim}
\cs_new_protected:Npn \clist_concat:NNN { \__clist_concat:NNNN \__kernel_tl_set:Nx }
\cs_new_protected:Npn \clist_gconcat:NNN { \__clist_concat:NNNN \__kernel_tl_gset:Nx }
\cs_new_protected:Npn \__clist_concat:NNNN #1#2#3#4 {
#1 #2 {
\exp_not:o #3 
\clist_if_empty:NF #3 { \clist_if_empty:NF #4 { , } }
\exp_not:o #4 
} }
\cs_generate_variant:Nn \clist_concat:NNN { ccc }
\cs_generate_variant:Nn \clist_gconcat:NNN { ccc }
\cs_generate_variant:Nn \clist_set:Nn { NV , No , Nx , c , cV , co , cx }
\cs_generate_variant:Nn \clist_gset:Nn { NV , No , Nx , c , cV , co , cx }
\end{verbatim}

(Copies of the \cs functions defined in \texttt{l3basics}.

\begin{verbatim}
\prg_new_eq_conditional:NNn \clist_if_exist:N \cs_if_exist:N { TF , T , F , p }
\prg_new_eq_conditional:NNn \clist_if_exist:c \cs_if_exist:c { TF , T , F , p }
\end{verbatim}

(End definition for \clist_concat:NNN, \clist_gconcat:NNN, and \)__clist_concat:NNNN. These functions are documented on page 173.)

59.3 Adding data to comma lists

Everything is based on concatenation after storing in \_\_clist_internal_clist. This avoids having to worry here about space-trimming and so on.

\begin{verbatim}
\__clist_concat:NNNN

\end{verbatim}

(End definition for \clist_concat:NNN and \clist_gconcat:NNN. These functions are documented on page 173.)
59.4 Comma lists as stacks

Getting an item from the left of a comma list is pretty easy: just trim off the first item using the comma. No need to trim spaces as comma-list variables are assumed to have “cleaned-up” items. (Note that grabbing a comma-delimited item removes an outer pair of braces if present, exactly as needed to uncover the underlying item.)

An empty clist leads to \texttt{q\_no\_value}, otherwise grab until the first comma and assign to the variable. The second argument of \texttt{\_\_clist\_pop:wwNNN} is a comma list ending in a comma and \texttt{s\_clist\_mark}, unless the original clist contained exactly one item: then the argument is just \texttt{s\_clist\_mark}. The next auxiliary picks either \texttt{\exp\_not:n} or \texttt{\use\_none:n} as \texttt{#2}, ensuring that the result can safely be an empty comma list.
\begin{verbatim}
{ \_clist_pop:NN \_kernel_tl_set:Nx }
\cs_new_protected:Npn \clist_gpop:NN
{ \_clist_pop:NN \_kernel_tl_gset:Nx }
\cs_new_protected:Npn \_clist_pop:NN #1#2#3
{
    \if_meaning:w #2 \c_empty_clist
        \tl_set:Nn #3 { \q_no_value }
    \else:
        \exp_after:wN \_clist_pop:wwNNN #2 , \s__clist_mark \s__clist_stop #1#2#3
        \fi:
}
\cs_new_protected:Npn \_clist_pop:wwNNN #1 , #2 \s__clist_stop #3#4#5
{
    \tl_set:Nn #5 {#1}
    #3 #4
    { \__clist_pop:wN \prg_do_nothing: #2 \exp_not:o , \s__clist_mark \use_none:n \s__clist_stop }
}
\cs_new_protected:Npn \__clist_pop:wN #1 , \s__clist_mark #2 #3 \s__clist_stop { #2 {#1} }
\cs_generate_variant:Nn \clist_pop:NN { c }
\cs_generate_variant:Nn \clist_gpop:NN { c }
\end{verbatim}

The same, as branching code: very similar to the above.

\begin{verbatim}
\prg_new_protected_conditional:Npnn \clist_get:NN #1#2 { T , F , TF }
    {
        \if_meaning:w #1 \c_empty_clist
            \prg_return_false:
        \else:
            \exp_after:wN \__clist_get:wN #1 , \s__clist_stop #2
            \prg_return_true:
        \fi:
}
\prg_generate_conditional_variant:Nnn \clist_get:NN { c } { T , F , TF }
\prg_new_protected_conditional:Npnn \clist_pop:NN #1#2 { T , F , TF } { \_clist_pop:NN \_kernel_tl_set:Nx #1 #2 }
\prg_new_protected_conditional:Npnn \clist_gpop:NN #1#2 { T , F , TF } { \_clist_pop:NN \_kernel_tl_gset:Nx #1 #2 }
\cs_new_protected:Npn \__clist_pop:NN #1#2
{
    \if_meaning:w #2 \c_empty_clist
        \prg_return_false:
    \else:
        \exp_after:wN \_clist_pop:wwNNN #2 , \s__clist_mark \s__clist_stop #1#2#3
        \fi:
}
\prg_generate_conditional_variant:Nnn \clist_pop:NN { c } { T , F , TF }
\prg_generate_conditional_variant:Nnn \clist_gpop:NN { c } { T , F , TF }
\end{verbatim}

(End definition for \clist_pop:NN and others. These functions are documented on page 178.)
Modifying comma lists

An internal comma list and a sequence for the removal routines.

Removing duplicates means making a new list then copying it.

(End definition for \clist_remove_duplicates:N and \clist_gremove_duplicates:N. These functions are documented on page 179.)
The method used here for safe items is very similar to \tl_replace_all:Nnn. However, if the item contains commas or leading/trailing spaces, or is empty, or consists of a single brace group, we know that it can only appear within braces so the code would fail; instead just convert to a sequence and do the removal with l3seq code (it involves somewhat elaborate code to do most of the work expandably but the final token list comparisons non-expandably).

For “safe” items, build a function delimited by the \textlangle item\textrangle that should be removed, surrounded with commas, and call that function followed by the expanded comma list, and another copy of the \textlangle item\textrangle. The loop is controlled by the argument grabbed by \__clist_remove_all:w: when the item was found, the \s__clist_mark delimiter used is the one inserted by \__clist_tmp:w, and \__clist_use_none_delimit_by_s_stop:w is deleted. At the end, the final \textlangle item\textrangle is grabbed, and the argument of \__clist_remove_all:w contains \s__clist_mark: in that case, \__clist_remove_all:w removes the second \s__clist_mark (inserted by \__clist_tmp:w), and lets \__clist_use_none_delimit_by_s_stop:w act.

No brace is lost because items are always grabbed with a leading comma. The result of the first assignment has an extra leading comma, which we remove in a second assignment. Two exceptions: if the clist lost all of its elements, the result is empty, and we shouldn’t remove anything; if the clist started up empty, the first step happens to turn it into a single comma, and the second step removes it.
\begin{verbatim}
\catcode@=11
\def\__clist_remove_all: #1, \s__clist_mark #2, #3 \s__clist_stop { \exp_not:n {#1} }
\cs_new:Npn \__clist_remove_all:w #1, \s__clist_mark , #2, { \exp_not:n {#1} }
\cs_generate_variant:Nn \clist_remove_all:Nn { c, NV, cV }
\cs_new:Npn \__clist_remove_all:w #1, \s__clist_mark , #2, { \exp_not:n {#1} }
\cs_generate_variant:Nn \clist_gremove_all:Nn { c, NV, cV }
\end{verbatim}

(End definition for \clist_remove_all:Nn and others. These functions are documented on page 174.)

Use \clist_reverse:n and \clist_greverse:N. These functions are documented on page 174.

\begin{verbatim}
\cs_new_protected:Npn \clist_reverse:N #1 { \__kernel_tl_set:Nx #1 { \exp_args:No \clist_reverse:n {#1} } }
\cs_new_protected:Npn \clist_greverse:N #1 { \__kernel_tl_gset:Nx #1 { \exp_args:No \clist_reverse:n {#1} } }
\cs_generate_variant:Nn \clist_reverse:N { c }
\cs_generate_variant:Nn \clist_greverse:N { c }
\end{verbatim}

(End definition for \clist_reverse:N and \clist_greverse:N. These functions are documented on page 174.)

The reversed token list is built one item at a time, and stored between \s__clist_stop and \s__clist_mark, in the form of \texttt{?} followed by zero or more instances of \texttt{⟨item⟩,...,.⟨item⟩}. During the loop, the auxiliary \__clist_reverse:wwNww receives \texttt{⟨item⟩} as \#1, \texttt{⟨item⟩} as \#2, \texttt{⟨item⟩} as \#3, what remains until \s__clist_stop as \#4, and \texttt{⟨item⟩} as \#5. The auxiliary moves \#1 just before \#5, with a comma, and calls itself \#3. After the last item is moved, \__clist_reverse:wwNww receives \texttt{⟨item⟩} as its argument \#1, thus \__clist_reverse_end:ww as its argument \#3. This second auxiliary cleans up until the marker !, removes the trailing comma (introduced when the first item was moved after \s__clist_stop), and leaves its argument \#1 within \texttt{\exp_not:n}. There is also a need to remove a leading comma, hence \texttt{\exp_not:o} and \texttt{\use_none:n}.

\begin{verbatim}
\cs_new:Npn \clist_greverse:N { \__clist_reverse:wwNww ? \s__clist_mark \__clist_greverse:wwNww ! , \s__clist_mark \__clist_reverse:end:ww \s__clist_stop ? \s__clist_mark }
\cs_new:Npn \__clist_reverse_end:ww #1 ! #2 , \s__clist_mark { \exp_not:o { \use_none:n #2 } }
\end{verbatim}

(End definition for \clist_reverse:n, \__clist_reverse:wwNww, and \__clist_reverse_end:ww. This function is documented on page 174.)
Implemented in \texttt{l3sort}.

(End definition for \texttt{clist_sort:Nn} and \texttt{clist_gsort:Nn}. These functions are documented on page 175.)

\section{Comma list conditionals}

Simple copies from the token list variable material.

\begin{verbatim}
\prg_new_eq_conditional:Nnn \clist_if_empty:N \tl_if_empty:N 
\prg_new_eq_conditional:Nnn \clist_if_empty:c \tl_if_empty:c
\end{verbatim}

(End definition for \texttt{clist_if_empty:NTF}. This function is documented on page 175.)

As usual, we insert a token (here ?) before grabbing any argument: this avoids losing braces. The argument of \texttt{tl_if_empty:oTF} is empty if \#1 is ? followed by blank spaces (besides, this particular variant of the emptiness test is optimized). If the item of the comma list is blank, grab the next one. As soon as one item is non-blank, exit: the second auxiliary grabs \texttt{prg_return_false}: as \#2, unless every item in the comma list was blank and the loop actually got broken by the trailing \texttt{s__clist_mark prg_return_false}: item.

\begin{verbatim}
\prg_new_conditional:Npnn \clist_if_empty:n #1 { T, F, TF } 
\end{verbatim}

(End definition for \texttt{clist_if_empty:nTF}, \texttt{__clist_if_empty_n:w}, and \texttt{__clist_if_empty_n:wNw}. This function is documented on page 175.)

For “safe” items, we simply surround the comma list, and the item, with commas, then use the same code as for \texttt{tl_if_in:Nn}. For “unsafe” items we follow the same route as \texttt{seq_if_in:Nn}, mapping through the list a comparison function. If found, return \texttt{true} and remove \texttt{prg_return_false}:

\begin{verbatim}
\prg_new_protected_conditional:Nnn \clist_if_in:Nn \clist_if_in:No
\prg_new_protected_conditional:Nnn \clist_if_in:cn \clist_if_in:co
\prg_new_protected_conditional:Nnn \clist_if_in:nn \clist_if_in:no
\end{verbatim}

\verb|\__clist_if_in_return:nnN|
59.7 Mapping over comma lists

If the variable is empty, the mapping is skipped (otherwise, that comma-list would be seen as consisting of one empty item). Then loop over the comma-list, grabbing eight comma-delimited items at a time. The end is marked by \s__clist_stop, which may not appear in any of the items. Once the last group of eight items has been reached, we go through them more slowly using \__clist_map_function_end:w. The auxiliary function \__clist_map_function:Nw is also used in some other clist mappings.

(End definition for \clist_if_in:NNTF, \clist_if_in:nnTF, and \__clist_if_in_return:nnN. These functions are documented on page 173.)
The \texttt{\clist_map_function:nN}, \texttt{\clist_map_function_n:Nn}, and \texttt{\clist_map_unbrace:wn} functions are documented on page 176.

\texttt{\clist_map_inline:Nn}, \texttt{\clist_map_inline:cn}, and \texttt{\clist_map_inline:nn} functions are documented on page 176.

Since the mapping is non-expandable, we can perform the space-trimming needed by the \texttt{\n} version simply by storing the comma-list in a variable. We don’t need a different comma-list for each nesting level: the comma-list is expanded before the mapping starts.
\clist_map_variable:NNn
\clist_map_variable:cNn
\clist_map_variable:nNn
\__clist_map_variable:Nnn
The \texttt{N}-type version is a straightforward application of \texttt{\clist_map_tokens:Nn}, calling \texttt{\__clist_map_variable:Nnn} for each item to assign the variable and run the user's code. The \texttt{n}-type version is not implemented in terms of the \texttt{n}-type function \texttt{\clist_map_tokens:Nn}, because here we are allowed to clean up the \texttt{n}-type comma list non-expandably.

\cs_new_protected:Npn \clist_map_variable:NNn #1#2#3
{ \clist_map_tokens:Nn #1 { \__clist_map_variable:Nnn #2 {#3} } }
\cs_generate_variant:Nn \clist_map_variable:NNn { c }
\cs_new_protected:Npn \__clist_map_variable:Nnn #1#2#3
{ \tl_set:Nn #1 {#3} #2 }
\cs_new_protected:Npn \clist_map_variable:nNn #1
{ \clist_set:Nn \l__clist_internal_clist {#1} \clist_map_variable:NNn \l__clist_internal_clist }
\__clist_map_variable:Nnn}

(End definition for \texttt{\clist_map_variable:NNn}, \texttt{\clist_map_variable:nNn}, and \texttt{\__clist_map_variable:Nnn}. These functions are documented on page 176.)

\clist_map_tokens:Nn
\clist_map_tokens:cn
\__clist_map_tokens:nw
\__clist_map_tokens_end:w
Essentially a copy of \texttt{\clist_map_function:NN} with braces added.
\cs_new:Npn \clist_map_tokens:Nn #1#2
{ \clist_if_empty:NF #1 #2 { \__clist_map_tokens:nw {#2} #1 , \__clist_stop \__clist_stop \__clist_stop \__clist_stop , \__clist_stop \__clist_stop \__clist_stop \__clist_stop , \__clist_stop \__clist_stop \__clist_stop \__clist_stop , \prg_break_point:Nn \clist_map_break: { \int_gdecr:N \g_kernel_prg_map_int } } }
\__clist_map_tokens:nw {#1} #2, #3, #4, #5, #6, #7, #8, #9,
\__clist_map_tokens_end:w
\__clist_use_none_delimit_by_s_stop:w
\use:n \use:n \use:n \use:n \use:n \use:n \use:n \use:n \use:n \use:n \use:n \use:n \use:n \use:n \use:n \use:n \use:n \use:n \use:n 

(End definition for \texttt{\clist_map_tokens:Nn}, \texttt{\clist_map_tokens:cn}, and \texttt{\__clist_map_tokens:nw}. These functions are documented on page 176.)
Similar to `\clist_map_function:nN` but with a different way of grabbing items because we cannot use `\exp_after:wN` to pass the `⟨code⟩`.

```latex
\cs_new:Npn \clist_map_tokens:nn #1#2
{ \__clist_map_tokens_n:nw {#2} \prg_do_nothing: #1 , \s__clist_stop \clist_map_break: , \prg_break_point:Nn \clist_map_break: { } }
\cs_new:Npn \__clist_map_tokens_n:nw #1#2 ,
{ \tl_if_empty:oF { \use_none:nn #2 ? } { \__clist_use_none_delimit_by_s_stop:w #2 \s__clist_stop \tl_trim_spaces_apply:oN {#2} \use_ii_i:nn \__clist_map_unbrace:wn , {#1} } \__clist_map_tokens_n:nw {#1} \prg_do_nothing: }
```

```
(End definition for \clist_map_tokens:nn and \__clist_map_tokens_n:nw. This function is documented on page 176.)
```

The break statements use the general `\prg_map_break:Nn` mechanism.

```latex
\cs_new:Npn \clist_map_break: 
{ \prg_map_break:Nn \clist_map_break: { } }
\cs_new:Npn \clist_map_break:n
{ \prg_map_break:Nn \clist_map_break: }
```

```
(End definition for \clist_map_break: and \clist_map_break:n. These functions are documented on page 176.)
```

Counting the items in a comma list is done using the same approach as for other token count functions: turn each entry into a +1 then use integer evaluation to actually do the mathematics. In the case of an n-type comma-list, we could of course use `\clist_map_function:nN`, but that is very slow, because it carefully removes spaces. Instead, we loop manually, and skip blank items (but not `{}`, hence the extra spaces).

```latex
\cs_new:Npn \clist_count:N #1
{ \int_eval:n
{ 0 \clist_map_function:NN #1 \__clist_count:n }
}
```

```
(End definition for \clist_count:N, \__clist_count:nn, \__clist_count:n, and \__clist_count:w. This function is documented on page 176.)
```
### 59.8 Using comma lists

First check that the variable exists. Then count the items in the comma list. If it has
none, output nothing. If it has one item, output that item, brace stripped (note that
space-trimming has already been done when the comma list was assigned). If it has two,
place the ⟨separator between two⟩ in the middle.

Otherwise, \cs{__clist_use:nwwwwnwn} takes the following arguments: 1: a ⟨separator⟩,
2, 3, 4: three items from the comma list (or quarks), 5: the rest of the comma list, 6:
a ⟨continuation⟩ function (\use{ii} or \use{iii} with its ⟨separator⟩ argument), 7: junk,
and 8: the temporary result, which is built in a brace group following \cs{__clist_stop}.
The ⟨separator⟩ and the first of the three items are placed in the result, then we use
the ⟨continuation⟩, placing the remaining two items after it. When we begin this loop,
the three items really belong to the comma list, the first \cs{__clist_mark} is taken as a
delimiter to the \use{ii} function, and the continuation is \use{ii} itself. When we reach
the last two items of the original token list, \cs{__clist_mark} is taken as a third item,
and now the second \cs{__clist_mark} serves as a delimiter to \use{ii}, switching to the other ⟨continuation⟩, \use{iii}, which uses the ⟨separator between final two⟩.
\begin{verbatim}
\begin{verbatim}
\exp_after:wN \__clist_use:nwwwnn
\exp_after:wN { \exp_after:wN } #1 ,
\s__clist_mark , { \__clist_use:nwwwnn {#3} }
\s__clist_mark , { \__clist_use:nwn {#4} }
\s__clist_stop { }

\msg_expandable_error:nnn
{ kernel } { bad-variable } {#1}
\end{verbatim}
\end{verbatim}
\cs_generate_variant:Nn \clist_use:Nnnn { c }
\cs_new:Npn \__clist_use:wwn #1 , #2 , #3 { \exp_not:n { #1 #3 #2 } }
\cs_new:Npn \__clist_use:nwwwwnwn
#1#2 , #3 , #4 , #5 \s__clist_mark , #6#7 \s__clist_stop #8
\{ \exp_not:n { #6 {#3} , {#4} , #5 \s__clist_mark , {#6} \s__clist_stop { #8 #1 #2 } }\}
\cs_new:Npn \__clist_use:nwwn #1#2 , #3 \s__clist_stop #4
\{ \exp_not:n { #4 #1 #2 } \}
\cs_new:Npn \clist_use:Nn #1#2
\{ \clist_use:Nnnn #1 {#2} {#2} {#2} \}
\cs_generate_variant:Nn \clist_use:Nn { c }
\end{verbatim}

(End definition for \clist_use:Nnnn and others. These functions are documented on page 177.)

\clist_use:nnnn
\clist_use:nn
\__clist_use:Nw
\__clist_use_one:w
\__clist_use_end:w
\__clist_use_more:w

Items are grabbed by \__clist_use:Nw, which detects blank items with a \tl_if_empty:TF test (in which case it recurses). Non-blank items are either the end of the list, in which case the argument \texttt{#1} of \__clist_use:Nw is used to properly end the list, or are normal items, which must be trimmed and properly unbraced. As we find successive items, the long list of \__clist_use:Nw calls gets shortened and we end up calling \__clist_use_more:w once we have found 3 items. This auxiliary leaves the first-found item and the general separator, and calls \__clist_use:Nw to find more items. A subtlety is that we use \__clist_use_end:w both in the case of a two-item list and for the last two items of a general list: to get the correct separator, \__clist_use_end:w replaces the separator-of-two by the last-separator when called, namely as soon as we have found three items.
\end{verbatim}
\end{verbatim}

\cs_new:Npn \clist_use:nnnn \#1\#2\#3\#4
\{ \__clist_use:Nw \__clist_use_none_delimit_by_s_stop:w
\__clist_use:Nw \__clist_use_none_delimit_by_s_stop:w
\__clist_use:Nw \__clist_use_one:w
\__clist_use:Nw \__clist_use_end:w
\__clist_use_more:w ;
\{ #2 \} \{#3} \{#4\} ;
\prg_do_nothing: #1 , \s__clist_mark ,
\s__clist_stop
\}
\cs_new:Npn \__clist_use:Nnnn \#1\#2 \#3 \#4 ,
\{ \tl_if_empty:oTF \{ \use_none:nn \#4 \} \}
\{ \__clist_use:Nw \#1\#2 ; \}
\{ \__clist_use_none_delimit_by_s_stop:w \#4 \#1 \s__clist_mark
\end{verbatim}
\end{verbatim}

850
59.9 Using a single item

To avoid needing to test the end of the list at each step, we first compute the \langle length\rangle of the list. If the item number is 0, less than \langle length\rangle, or more than \langle length\rangle, the result is empty. If it is negative, but not less than \langle length\rangle, add \langle length\rangle + 1 to the item number before performing the loop. The loop itself is very simple, return the item if the counter reached 1, otherwise, decrease the counter and repeat.

\begin{verbatim}
\cs_new:Npn \clist_item:Nn #1#2 { \__clist_item:ffoN { \clist_count:N #1 } { \int_eval:n {#2} } \__clist_item_N_loop:nw }
\cs_new:Npn \__clist_item:nnnN #1#2#3#4 { \int_compare:nNnTF {#2} < 0 { \__clist_use_none_delimit_by_s_stop:w } { \exp_args:Nf #4 { \int_eval:n { #2 + 1 + #1 } } } { \__clist_item:nnnN \__clist_item:fofN \__clist_item:ff N } \__clist_item_N_loop:nw }
\cs_generate_variant:Nn \__clist_item:nnnN { ffo, ff }
\end{verbatim}
\int_compare:nNnTF {#1} = 0
\{ \__clist_use_i_delimit_by_s_stop:nw \{ \exp_not:n \{#2\} \}
\{ \exp_args:Nf \__clist_item_N_loop:nw \{ \int_eval:n \{ #1 - 1 \} \}
\}
\cs_generate_variant:Nn \clist_item:Nn { c }

(End definition for \clist_item:Nn, \__clist_item:nnN, and \__clist_item_N_loop:nw. This function is documented on page 179.)

\clist_item:nn
\__clist_item:nw
\__clist_item_n:nw
\__clist_item_n_loop:nw
\__clist_item_n_end:n
\__clist_item_n_strip:n
\__clist_item_n_strip:w

This starts in the same way as \clist_item:Nn by counting the items of the comma list. The final item should be space-trimmed before being brace-striped, hence we insert a couple of odd-looking \prg_do_nothing: to avoid losing braces. Blank items are ignored.

\cs_new:Npn \clist_item:nn #1#2
\{ \__clist_item:ffnN \{ \clist_count:n \{#1\} \}
\{ \int_eval:n \{#2\} \}
\} \__clist_item_n:nw
\cs_new:Npn \__clist_item_n_loop:nw #1 #2,
\{ \exp_args:No \tl_if_blank:nTF {#2}
\{ \__clist_item_n_loop:nw \{#1\} \prg_do_nothing: \}
\} \exp_args:No \__clist_item_n_end:n {#2}
\{ \__clist_item_n_loop:nw {#1} \prg_do_nothing: \}
\int_compare:nNnF {#1} = 0
\{ \exp_args:No \__clist_item_n_end:n \{#2\} \}
\{ \exp_args:Nf \__clist_item_n_loop:nw
\{ \int_eval:n \{ #1 - 1 \} \}
\prg_do_nothing:
\}
\}
\cs_new:Npn \__clist_item_n_end:n #1 #2 \s__clist_stop
\tl_trim_spaces_apply:nN \{#1\} \__clist_item_n_strip:n
\cs_new:Npn \__clist_item_n_strip:n \{ #1 \}
\cs_new:Npn \__clist_item_n_strip:w \{ #1 \}
\exp_not:n \{#1\}

(End definition for \clist_item:nn and others. This function is documented on page 179.)

\clist_rand_item:n
\clist_rand_item:N
\clist_rand_item:c
\__clist_rand_item:nn

The N-type function is not implemented through the n-type function for efficiency: for instance comma-list variables do not require space-trimming of their items. Even testing for emptiness of an n-type comma-list is slow, so we count items first and use that both for the emptiness test and the pseudo-random integer. Importantly, \clist_item:Nn and \clist_item:nn only evaluate their argument once.

\cs_new:Npn \clist_rand_item:n #1
\{ \exp_args:Nf \__clist_rand_item:nn \{ \clist_count:n \{#1\} \}
\}
\cs_new:Npn \__clist_rand_item:nn #1#2
\{ \int_compare:nNnTF {#1} = 0
\{ \clist_item:nn \{#2\} \int_rand:nn { 1 } \{#1\} \}
\}

852
\section{Viewing comma lists}

Apply the general \texttt{\_\_kernel\_chk\_tl\_type:NnnT} with \texttt{\exp\_not:o \#2} serving as a dummy code to prevent a check performed by this auxiliary.

\begin{verbatim}
\cs_new:Npn \clist_rnd_item:N #1 { \clist_if_empty:NF #1 { \clist_item:Nn #1 { \int_rnd:nn { 1 } { \clist_count:N #1 } } } }
\cs_generate_variant:Nn \clist_rnd_item:N { c }

\cs_new_protected:Npn \clist_rnd_item:n { \__clist_rnd_item:nn }
\cs_generate_variant:Nn \clist_rnd_item:n { c }

\cs_new_protected:Npn \clist_rnd_item:N { \__clist_rnd_item:nnc }
\cs_generate_variant:Nn \clist_rnd_item:N { c }
\end{verbatim}

A variant of the above: no existence check, empty first argument for the message.

\begin{verbatim}
\cs_new_protected:Npn \clist_rnd_item:n { \__clist_rnd_item:nn }
\cs_new_protected:Npn \clist_rnd_item:N { \__clist_rnd_item:nnc }
\end{verbatim}

(End definition for \texttt{\clist_rnd_item:n}, \texttt{\clist_rnd_item:N}, and \texttt{\__clist_rnd_item:nn}. These functions are documented on page 180.)
59.11 Scratch comma lists

\l_tmpa_clist Temporary comma list variables.
\l_tmpb_clist  \clist_new:N \l_tmpa_clist
\g_tmpa_clist  \clist_new:N \l_tmpb_clist
\g_tmpb_clist  \clist_new:N \g_tmpa_clist
\g_tmpb_clist  \clist_new:N \g_tmpb_clist

(End definition for \l_tmpa_clist and others. These variables are documented on page 180.)

⟨/package⟩
Chapter 60

l3token implementation

60.1 Internal auxiliaries

\__char_stop Internal scan mark.
\scan_new:N \__char_stop
(End definition for \__char_stop.)

\__char_no_value Internal recursion quarks.
\quark_new:N \__char_no_value
(End definition for \__char_no_value.)

\__char_quark_if_no_value:NTF Functions to query recursion quarks.
\__kernel_quark_new_conditional:Nn \__char_quark_if_no_value:NTF
(End definition for \__char_quark_if_no_value:NTF.)

60.2 Manipulating and interrogating character tokens

Simple wrappers around the primitives.
\cs_new_protected:Npn \char_set_catcode:nn #1#2
{ \tex_catcode:D \int_eval:n {#1} = \int_eval:n {#2} \exp_stop_f: }
\cs_new:Npn \char_value_catcode:n #1
{ \tex_the:D \tex_catcode:D \int_eval:n {#1} \exp_stop_f: }
\cs_new_protected:Npn \char_show_value_catcode:n #1
{ \exp_args:Nf \tl_show:n { \char_value_catcode:n {#1}} }
(End definition for \char_set_catcode:nn, \char_value_catcode:n, and \char_show_value_catcode:n. These functions are documented on page 184.)
End definition for \char_set_catcode_escape:N and others. These functions are documented on page 183.
\cs_new_protected:Npn \char_set_catcode_math_subscript:n \#1 \{ \char_set_catcode:nn \#1 \{ 8 \} \}
\cs_new_protected:Npn \char_set_catcode_ignore:n \#1 \{ \char_set_catcode:nn \#1 \{ 9 \} \}
\cs_new_protected:Npn \char_set_catcode_space:n \#1 \{ \char_set_catcode:nn \#1 \{ 10 \} \}
\cs_new_protected:Npn \char_set_catcode_letter:n \#1 \{ \char_set_catcode:nn \#1 \{ 11 \} \}
\cs_new_protected:Npn \char_set_catcode_other:n \#1 \{ \char_set_catcode:nn \#1 \{ 12 \} \}
\cs_new_protected:Npn \char_set_catcode_active:n \#1 \{ \char_set_catcode:nn \#1 \{ 13 \} \}
\cs_new_protected:Npn \char_set_catcode_comment:n \#1 \{ \char_set_catcode:nn \#1 \{ 14 \} \}
\cs_new_protected:Npn \char_set_catcode_invalid:n \#1 \{ \char_set_catcode:nn \#1 \{ 15 \} \}

(End definition for \char_set_catcode_escape:n and others. These functions are documented on page 184.)

\char_set_mathcode:nn
\char_value_mathcode:n
\char_show_value_mathcode:n
\char_set_lccode:nn
\char_value_lccode:n
\char_show_value_lccode:n
\char_set_uccode:nn
\char_value_uccode:n
\char_show_value_uccode:n
\char_set_sfcode:nn
\char_value_sfcode:n
\char_show_value_sfcode:n

Pretty repetitive, but necessary!

\l_char_active_seq
\l_char_special_seq

Two sequences for dealing with special characters. The first is characters which may be active, the second longer list is for “special” characters more generally. Both lists are escaped so that for example bulk code assignments can be carried out. In both cases, the order is by ASCII character code (as is done in for example \ExplSyntaxOn).

\seq_new:N \l_char_special_seq
\seq_set_split:Nnn \l_char_special_seq { \ }

857
60.3 Creating character tokens

Four simple functions with very similar definitions, so set up using an auxiliary. These are similar to LuaTeX’s `\letcharcode` primitive.

\char_set_active_eq:NN
\char_gset_active_eq:N
\char_set_active_eq:n
\char_gset_active_eq:nc
\char_set_catcode_active:N \^^@
\cs_set_protected:Npn __char_tmp:nN #1#2 {
  \cs_new_protected:cpn { #1 :nN } ##1 {
    \group_begin:
    \char_set_lccode:nn { \^^@ } { ##1 }
    \tex_lowercase:D { \group_end: #2 ^^@ }
  }
  \cs_new_protected:cpx { #1 :NN } ##1 { \exp_not:c { #1 : nN } { '##1 } }
}
\__char_tmp:nN { char_set_active_eq } \cs_set_eq:NN
\__char_tmp:nN { char_gset_active_eq } \cs_gset_eq:NN
\group_end:
\cs_generate_variant:Nn \char_set_active_eq:NN { Nc }
\cs_generate_variant:Nn \char_gset_active_eq:NN { Nc }
\cs_generate_variant:Nn \char_set_active_eq:nN { nc }
\cs_generate_variant:Nn \char_gset_active_eq:nN { nc }
\__char_int_to_roman:w
\char_generate:nn
\__char_generate_aux:nn
\__char_generate_auxii:nn
\l__char_tmp_tl \__char_generate_invalid_catcode:
\cs_new_eq:NN __char_int_to_roman:w \tex_romannumeral:D
\__char_int_to_roman:w
\char_generate:nn
\__char_generate_aux:nn
\__char_generate_auxii:nn
\l__char_tmp_tl \__char_generate_invalid_catcode:
\cs_new:Npn \char_generate:nn #1#2 {
  \exp:w \exp_after:wN \__char_generate_aux:nn \exp:w \int_value:w \int_eval:n {#1} \exp_after:wN ;
  \int_value:w \int_eval:n {#2} ;
}

For efficiency in 8-bit engines, we use the faster primitive approach to making roman numerals.

\char_generate:nn
\__char_generate_aux:nn
\__char_generate_auxii:nn
\l__char_tmp_tl \__char_generate_invalid_catcode:
\cs_new:Npn \char_generate:nn #1#2 {
  \exp:w \exp_after:wN \__char_generate_aux:nn \exp:w \int_value:w \int_eval:n {#1} \exp_after:wN ;
  \int_value:w \int_eval:n {#2} ;
}

The aim here is to generate characters of (broadly) arbitrary category code. Where possible, that is done using engine support (XeTeX, LuaTEX). There are though various issues which are covered below. At the interface layer, turn the two arguments into integers up-front so this is only done once.
Before doing any actual conversion, first some special case filtering. Spaces are out here as LuaTEX emulation only makes normal (charcode 32 spaces). However, ~@ is filtered out separately as that can’t be done with macro emulation either, so is flagged up separately. That done, hand off to the engine-dependent part.

\cs_new:Npn \__char_generate_aux:w #1 ; #2 ;
\begin{verbatim}
\if_int_odd:w \exp_end:
\else:
\if_int_odd:w 0 \exp_end:
\end{verbatim}
\tl_new:N \l__char_tmp_tl

Engine-dependent definitions are now needed for the implementation. For LuaTEX and Xe\TeX{} there is engine-level support. They can do cases that macro emulation can’t. All of those are filtered out here using a primitive-based boolean expression to avoid fixing the category code of the null character used in the false branch (for 8-bit engines). The final level is the basic definition at the engine level: the arguments here are integers so there is no need to worry about them too much. Older versions of Xe\TeX{} cannot generate active characters so we filter that: at some future stage that may change: the slightly odd ordering of auxiliaries reflects that.

\group_begin:
\char_set_catcode_active:N \^^L
\cs_set:Npn \^^L { }
\char_set_catcode_other:n { 0 }
\if_int_odd:w 0 \exp_end:
\sys_if_engine_luatex:T { 1 }
\sys_if_engine_xetex:T { 1 } \exp_end:
\cs_new:Npn \__char_generate_aux:nw {#1} {#2}
\exp_after:wN \exp_end:
\tex_Ucharcat:D #1 \exp_stop_f: #2 \exp_stop_f:
For engines where \texttt{Ucharcat} isn’t available or emulated, we have to work in macros, and cover only the 8-bit range. The first stage is to build up a \texttt{tl} containing \texttt{~\textasciitilde @} with each category code that can be accessed in this way, with an error set up for the other cases. This is all done such that it can be quickly accessed using a \texttt{if_case:w} low-level conditional. There are a few things to notice here. As \texttt{~\textasciitilde L} is \texttt{outer} we need to locally set it to avoid a problem. To get open/close braces into the list, they are set up using \texttt{if_false:} pairing and are then \texttt{x}-type expanded together into the desired form.

For making spaces, there needs to be an \texttt{o}-type expansion of a \texttt{use:n} (or some other tokenization) to avoid dropping the space.

Convert the above temporary list into a series of constant token lists, one for each character code, using \texttt{\tex_lowercase:D} to convert \texttt{~\textasciitilde @} in each case. The \texttt{x}-type expansion ensures that \texttt{\tex_lowercase:D} receives the contents of the token list. \texttt{~\textasciitilde L} is awkward.
hence this is done in three parts: up to \textasciitilde L, \textasciitilde L itself and above \textasciitilde L. Notice that at this stage \textasciitilde @ is active.

As \TeX{} is very unhappy if it finds an alignment character inside a primitive \texttt{\textbackslash halign} even when skipping false branches, some precautions are required. \TeX{} is happy if the token is hidden between braces within \texttt{\textbackslash if\_false: ... \textbackslash fi:].

This code converts a codepoint into the correct UTF-8 representation. In terms of the algorithm itself, see \url{https://en.wikipedia.org/wiki/UTF-8} for the octet pattern.
\_char_to_utfviii_bytes_outputii:nw
\{ \_char_to_utfviii_bytes_auxiii:n \{#1\} \}
\else:
\if_int_compare:w #1 < "10000 \exp_stop_f:
\__char_to_utfviii_bytes_outputii:nw
\{ \_char_to_utfviii_bytes_auxiii:NNn E {#1} \{ 64 * 64 \} \}
\else:
\__char_to_utfviii_bytes_outputii:nw
\{ \_char_to_utfviii_bytes_auxiii:n \}
\}
\__char_to_utfviii_bytes_outputii:nw
\{ \_char_to_utfviii_bytes_auxiii:n \{#1\} \}
\else:
\__char_to_utfviii_bytes_outputii:nw
\{ \_char_to_utfviii_bytes_auxiii:n \}
\}
\if_int_compare:w #1 < "10000 \exp_stop_f:
\__char_to_utfviii_bytes_outputii:nw
\{ \_char_to_utfviii_bytes_auxiii:NNn F {#1} \{ 64 * 64 * 64 \} \}
\else:
\__char_to_utfviii_bytes_outputii:nw
\{ \_char_to_utfviii_bytes_auxiii:n \}
\}
\__char_to_utfviii_bytes_outputii:nw
\{ \_char_to_utfviii_bytes_auxiii:n \}
\}
\__char_to_utfviii_bytes_outputii:nw
\{ \_char_to_utfviii_bytes_auxiii:n \}
\}
\__char_to_utfviii_bytes_outputii:nw
\{ \_char_to_utfviii_bytes_auxiii:n \}
\}
\__char_to_utfviii_bytes_outputii:nw
\{ \_char_to_utfviii_bytes_auxiii:n \}
\}
\__char_to_utfviii_bytes_outputii:nw
\{ \_char_to_utfviii_bytes_auxiii:n \}
\}
\__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_auxii:Nnn #1#2#3
{ #10 + \int_div_truncate:nn {#2} {#3} }
\cs_new:Npn \__char_to_utfviii_bytes_auxiii:n #1
{ \int_mod:nn {#1} { 64 } + 128 }
\cs_new:Npn \__char_to_utfviii_bytes_outputi:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_auxii:NNn #1 #2 #3
{ \int_eval:n {#1} } { } {#2} }
\cs_new:Npn \__char_to_utfviii_bytes_outputii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiii:nw #1 #2 #3 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiv:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputv:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputvi:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputvii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputivii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiviii:nw #1 #2 \__char_to_utfviii_bytes_end: { } { } { } { }
\cs_new:Npn \__char_to_utfviii_bytes_outputiv:nw #1 #2 \__char_to_utfviii_bytes_end: #3 #4 #5 #6 \
{ \__char_to_utfviii_bytes_output:fnn { \int_eval:n {#1} } { {#3} {#4} {#5} } {#2} }
\cs_new:Npn \__char_to_utfviii_bytes_output:nnn #1 #2 #3 
{ #3 \__char_to_utfviii_bytes_end: #2 {#1} }
\cs_generate_variant:Nn \__char_to_utfviii_bytes_output:nnn { f }
\cs_new:Npn \__char_to_utfviii_bytes_end: { }

(End definition for \char_to_utfviii_bytes:n and others. This function is documented on page 308.)

\char_to_nfd:N \__char_to_nfd:n \__char_to_nfd:Nw

Look up any nfd and recursively produce the result.

\cs_new:Npn \char_to_nfd:N #1 
{ \cs_if_exist:cTF { c__char_nfd_ \token_to_str:N #1 _ tl } 
{ \exp_after:wN \exp_after:wN \exp_after:wN \__char_to_nfd:Nw 
\exp_after:wN \exp_after:wN \exp_after:wN #1 
\cs:w c__char_nfd_ \token_to_str:N #1 _ tl \cs_end: 
\s__char_stop }
{ \exp_not:n {#1} } }
\cs_set_eq:NN \__char_to_nfd:n \char_to_nfd:N 
\cs_new:Npn \__char_to_nfd:Nw #1 #2 #3 \s__char_stop 
{ \exp_args:Ne \__char_to_nfd:n 
\char_generate:nn { '#2 } { \__char_change_case_catcode:N #1 } 
\tl_if_blank:nF {#3} 
{ \exp_args:Ne \__char_to_nfd:n 
\char_generate:nn { '#3 } { \char_value_catcode:n { '#3 } } }
}

(End definition for \char_to_nfd:N, \__char_to_nfd:n, and \__char_to_nfd:Nw. This function is documented on page 308.)

\char_lowercase:N \char_uppercase:N \char_titlecase:N \char_foldcase:N \__char_change_case:nNN \__char_change_case:NN \__char_change_case:NNNN \__char_change_case:NNNNw \__char_change_case_catcode:N \char_str_lowercase:N \char_str_uppercase:N \char_str_titlecase:N \char_str_foldcase:N \__char_str_change_case:nNN \__char_str_change_case:nN

To ensure that the category codes produced are predictable, every character is regenerated even if it is otherwise unchanged. This makes life a little interesting when we might have multiple output characters: we have to grab each of them and case change them in reverse order to maintain f-type expandability.

\cs_new:Npn \char_lowercase:N #1 
{ \__char_change_case:nNN { lower } \char_value_lccode:n #1 }
\cs_new:Npn \char_uppercase:N #1 
{ \__char_change_case:nNN { upper } \char_value_uccode:n #1 }
\cs_new:Npn \char_titlecase:N #1 
{ \tl_if_exist:cTF { c__char_titlecase_ \token_to_str:N #1 _tl } 
\exp_after:wN \exp_after:wN \exp_after:wN \__char_change_case:nNN 
\token_to_str:N #1 _ tl 
\cs_end: 
\__char_stop }

863
Same story for the string version, except category code is easier to follow. This of course makes this version significantly faster.

```latex
\cs_new:Npn \char_str_lowercase:N #1
{ \__char_str_change_case:nNN { lower } \char_value_lccode:n #1 }
\cs_new:Npn \char_str_uppercase:N #1
{ \__char_str_change_case:nNN { upper } \char_value_uccode:n #1 }
\cs_new:Npn \char_str_titlecase:N #1
{ \tl_if_exist:cTF { c__char_titlecase_ \token_to_str:N #1 _tl }
{ \tl_to_str:c { c__char_titlecase_ \token_to_str:N #1 _tl } }
{ \char_str_uppercase:N #1 } }
\cs_new:Npn \char_str_foldcase:N #1
{ \__char_str_change_case:nNN { fold } \char_value_lccode:n #1 }
\cs_new:Npn \__char_str_change_case:nN #1#2
{ \tl_if_exist:cTF { c__char_ #1 case_ \token_to_str:N #2 _tl }
\tl_to_str:c { c__char_ #1 case_ \token_to_str:N #2 _tl } }
\cs_new:Npn \__char_str_change_case:nNN #1#2#3
{ \exp_args:Nf \__char_str_change_case:nN { #2 } \token_to_str:N #3 #1 }
Create a space with category code 12: an “other” space.

\tl_const:Nx \c_catcode_other_space_tl { \char_generate:nn { ' ' } { 12 } }

\c_catcode_other_space_tl

60.4 Generic tokens

\langle @=token \rangle
\s__token_mark \s__token_stop

\internal scan marks.

\scan_new:N \s__token_mark
\scan_new:N \s__token_stop

\token_to_meaning:N \token_to_meaning:c \token_to_str:N \token_to_str:c
\c_group_begin_token \c_group_end_token \c_math_toggle_token \c_alignment_token \c_parameter_token \c_math_superscript_token \c_math_subscript_token \c_space_token \c_catcode_letter_token \c_catcode_other_token

\token_to_meaning:N \token_to_meaning:c \token_to_str:N \token_to_str:c

These are all defined in l3basics, as they are needed “early”. This is just a reminder!

(End definition for \token_to_meaning:N and \token_to_str:N. These functions are documented on page 187.)

We define these useful tokens. For the brace and space tokens things have to be done by hand: the formal argument spec. for \cs_new_eq:NN does not cover them so we do things by hand. (As currently coded it would work with \cs_new_eq:NN but that’s not really a good idea to show off: we want people to stick to the defined interfaces and that includes us.) So that these few odd names go into the log when appropriate there is a need to hand-apply the \__kernel_chk_if_free_cs:N check.

(End definition for \c_group_begin_token and \c_group_end_token.)

(End definition for \char_lowercase:N and others. These functions are documented on page 183.)

866
\cs_new_eq:NN \c_alignment_token * 
\cs_new_eq:NN \c_parameter_token # 
\cs_new_eq:NN \c_math_superscript_token ^ 
\char_set_catcode_math_subscript:N \* 
\cs_new_eq:NN \c_math_subscript_token * 
\__kernel_chk_if_free_cs:N \c_space_token 
\use:n { \tex_global:D \tex_let:D \c_space_token = ~ } ~ 
\cs_new_eq:NN \c_catcode_letter_token a 
\cs_new_eq:NN \c_catcode_other_token 1 
\group_end: 

(End definition for \c_group_begin_token and others. These functions are documented on page 186.) 

\c_catcode_active_tl 

Not an implicit token! 
\group_begin: 
\char_set_catcode_active:N \* 
\tl_const:Nn \c_catcode_active_tl { \exp_not:N * } 
\group_end: 

(End definition for \c_catcode_active_tl. This variable is documented on page 186.) 

\section{Token conditionals} 

\token_if_group_begin:N 
\token_if_group_begin:NTF 
Check if token is a begin group token. We use the constant \c_group_begin_token for this.

\token_if_math_toggle_p:N 
\token_if_math_toggle:NTF 
Check if token is a math shift token. We use the constant \c_math_toggle_token for this.

\token_if_group_end:N 
\token_if_group_end:NTF 
Check if token is an end group token. We use the constant \c_group_end_token for this.

\token_if_math_toggle_p:N 
\token_if_math_toggle:NTF 
(End definition for \c_catcode_active_tl. This variable is documented on page 186.) 

60.5 Token conditionals 

\token_if_group_begin:N 
\token_if_group_begin:NTF 
\prg_new_conditional:Nnn \token_if_group_begin:N #1 { p , T , F , TF } 
{ 
  \if_catcode:w \exp_not:N #1 \c_group_begin_token 
  \prg_return_true: \else: \prg_return_false: \fi: 
} 

(End definition for \c_group_begin_token. This function is documented on page 187.) 

\token_if_group_end:N 
\token_if_group_end:NTF 
\prg_new_conditional:Nnn \token_if_group_end:N #1 { p , T , F , TF } 
{ 
  \if_catcode:w \exp_not:N #1 \c_group_end_token 
  \prg_return_true: \else: \prg_return_false: \fi: 
} 

(End definition for \c_group_end_token. This function is documented on page 187.) 

\token_if_math_toggle_p:N 
\token_if_math_toggle:NTF 
\prg_new_conditional:Nnn \token_if_math_toggle:N #1 { p , T , F , TF } 
{ 
  \if_catcode:w \exp_not:N #1 \c_math_toggle_token 
  \prg_return_true: \else: \prg_return_false: \fi: 
} 

(End definition for \c_math_toggle_token. This function is documented on page 187.) 

867
Check if token is an alignment tab token. We use the constant \c_alignment_token for this.

\begin{verbatim}
\prg_new_conditional:Npn \token_if_alignment:N \#1 { p , T , F , TF }
\{ \if_catcode:w \exp_not:N \#1 \c_alignment_token \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}

(End definition for \token_if_alignment:NTF. This function is documented on page 188.)

Check if token is a parameter token. We use the constant \c_parameter_token for this.

\begin{verbatim}
\group_begin:
\cs_set_eq:NN \c_parameter_token \scan_stop:
\prg_new_conditional:Npn \token_if_parameter:N \#1 { p , T , F , TF }
\{ \if_catcode:w \exp_not:N \#1 \c_parameter_token \prg_return_true: \else: \prg_return_false: \fi: }
\group_end:
\end{verbatim}

(End definition for \token_if_parameter:NTF. This function is documented on page 188.)

Check if token is a math superscript token. We use the constant \c_math_superscript_token for this.

\begin{verbatim}
\prg_new_conditional:Npn \token_if_math_superscript:N \#1 { p , T , F , TF }
\{ \if_catcode:w \exp_not:N \#1 \c_math_superscript_token \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}

(End definition for \token_if_math_superscript:NTF. This function is documented on page 188.)

Check if token is a math subscript token. We use the constant \c_math_subscript_token for this.

\begin{verbatim}
\prg_new_conditional:Npn \token_if_math_subscript:N \#1 { p , T , F , TF }
\{ \if_catcode:w \exp_not:N \#1 \c_math_subscript_token \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}

(End definition for \token_if_math_subscript:NTF. This function is documented on page 188.)

Check if token is a space token. We use the constant \c_space_token for this.

\begin{verbatim}
\prg_new_conditional:Npn \token_if_space:N \#1 { p , T , F , TF }
\{ \if_catcode:w \exp_not:N \#1 \c_space_token \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}

(End definition for \token_if_space:NTF. This function is documented on page 188.)
Check if token is a letter token. We use the constant \c_catcode_letter_token for this.
\begin{verbatim}
\token_if_letter_p:N \token_if_letter:NTF
{ \prg_new_conditional:Nppnn \token_if_letter:N \c_catcode_letter_token \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}
(End definition for \token_if_letter:NTF. This function is documented on page 188.)

Check if token is an other char token. We use the constant \c_catcode_other_token for this.
\begin{verbatim}
\token_if_other_p:N \token_if_other:NTF
{ \prg_new_conditional:Nppnn \token_if_other:N \c_catcode_other_token \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}
(End definition for \token_if_other:NTF. This function is documented on page 188.)

Check if token is an active char token. We use the constant \c_catcode_active_tl for this. A technical point is that \c_catcode_active_tl is in fact a macro expanding to \exp_not:N *, where * is active.
\begin{verbatim}
\token_if_active_p:N \token_if_active:NTF
{ \prg_new_conditional:Nppnn \token_if_active:N \c_catcode_active_tl \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}
(End definition for \token_if_active:NTF. This function is documented on page 188.)

Check if the tokens \#1 and \#2 have same meaning.
\begin{verbatim}
\token_if_eq_meaning_p:NN \token_if_eq_meaning:NNTF
{ \prg_new_conditional:Nppnn \token_if_eq_meaning:NN \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}
(End definition for \token_if_eq_meaning:NNTF. This function is documented on page 189.)

Check if the tokens \#1 and \#2 have same category code.
\begin{verbatim}
\token_if_eq_catcode_p:NN \token_if_eq_catcode:NNTF
{ \prg_new_conditional:Nppnn \token_if_eq_catcode:NN \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}
(End definition for \token_if_eq_catcode:NNTF. This function is documented on page 188.)

Check if the tokens \#1 and \#2 have same character code.
\begin{verbatim}
\token_if_eq_charcode_p:NN \token_if_eq_charcode:NNTF
{ \prg_new_conditional:Nppnn \token_if_eq_charcode:NN \prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}
(End definition for \token_if_eq_charcode:NNTF. This function is documented on page 188.)
When a token is a macro, `\token_to_meaning:NN` always outputs something like `\long macro: #1 -> #1` so we could naively check to see if the meaning contains `->`. However, this can fail the five `..mark` primitives, whose meaning has the form `..mark: ⟨user material⟩`. The problem is that the `⟨user material⟩` can contain `->`. However, only characters, macros, and marks can contain the colon character. The idea is thus to grab until the first `:`, and analyse what is left. However, macros can have any combination of `\long`, `\protected` or `\outer` (not used in L\TeX3) before the string `macro:`. We thus only select the part of the meaning between the first `ma` and the first following `:`. If this string is `cro`, then we have a macro. If the string is `rk`, then we have a mark. The string can also be `cro parameter character` for a colon with a weird category code (namely the usual category code of `#`). Otherwise, it is empty.

This relies on the fact that `\long`, `\protected`, `\outer` cannot contain `ma`, regardless of the escape character, even if the escape character is `m`...

Both `ma` and `:` must be of category code 12 (other), so are detokenized.

```latex
\use:x
\begin{verbatim}
\prg_new_conditional:Nnn \exp_not:N \token_if_macro:N ##1 { p , T , F , TF }
{ \exp_not:N \exp_after:wN \exp_not:N \__token_if_macro_p:w
\exp_not:N \token_to_meaning:N ##1 \tl_to_str:n { ma : }
\s__token_stop }
\cs_new:Npn \exp_not:N \__token_if_macro_p:w ##1 \tl_to_str:n { ma } ##2 \c_colon_str ##3 \s__token_stop
{ \str_if_eq:nnTF { #2 } { cro } { \prg_return_true: } { \prg_return_false: } }
\end{verbatim}
\end{verbatim}
```

(End definition for `\token_if_macro:NTF` and `\__token_if_macro_p:w`. This function is documented on page 189.)

Check if token has same catcode as a control sequence. This follows the same pattern as for `\token_if_letter:N` etc. We use `\scan_stop:` for this.

```latex
\begin{verbatim}
\prg_new_conditional:Nnn \exp_not:N \token_if_cs:N #1 { p , T , F , TF }
{ \if_catcode:w \exp_not:N #1 \scan_stop:
\prg_return_true: \else: \prg_return_false: \fi: }
\end{verbatim}
```

(End definition for `\token_if_cs:NTF`. This function is documented on page 189.)

Check if token is expandable. We use the fact that \TeX temporarily converts `\exp_not:N ⟨token⟩` into `\scan_stop:` if ⟨token⟩ is expandable. An undefined token is not considered as expandable. No problem nesting the conditionals, since the third #1 is only skipped if it is non-expandable (hence not part of \TeX's conditional apparatus).

```latex
\begin{verbatim}
\prg_new_conditional:Nnn \exp_not:N \token_if_expandable:N #1 { p , T , F , TF }
\end{verbatim}
```

(End definition for `\token_if_expandable:NTF`. This function is documented on page 189.)
These auxiliary functions are used below to define some conditionals which detect whether the \meaning of their argument begins with a particular string. Each auxiliary takes an argument delimited by a string, a second one delimited by \s\token_stop, and returns the first one and its delimiter. This result is eventually compared to another string. Note that the “font” auxiliary is delimited by a space followed by “font”. This avoids an unnecessary check for the \font primitive below.

Each of these conditionals tests whether its argument’s \meaning starts with a given string. This is essentially done by having an auxiliary grab an argument delimited by the string and testing whether the argument was empty. Of course, a copy of this string must first be added to the end of the \meaning to avoid a runaway argument in case it does not contain the string. Two complications arise. First, the escape character is not fixed, and cannot be included in the delimiter of the auxiliary function (this function cannot be defined on the fly because tests must remain expandable): instead the first argument of the auxiliary (plus the delimiter to avoid complications with trailing spaces) is compared using \str_if_eq:eeTF to the result of applying \token_to_str:N to a control sequence. Second, the \meaning of primitives such as \dimen or \dimendef starts in the same way as registers such as \dimen123, so they must be tested for.

Characters used as delimiters must have catcode 12 and are obtained through \tl_to_str:n. This requires doing all definitions within x-expansion. The temporary function...
\_token\_tmp:w used to define each conditional receives three arguments: the name of the conditional, the auxiliary’s delimiter (also used to name the auxiliary), and the string to which one compares the auxiliary’s result. Note that the \meaning of a protected long macro starts with \protectlong macro, with no space after \protect but a space after \long, hence the mixture of \token\_to\_str:N and \tl\_to\_str:n.

For the first six conditionals, \cs\_if\_exist:cT turns out to be false (thanks to the leading space for font), and the code boils down to a string comparison between the result of the auxiliary on the \meaning of the conditional’s argument \#1, and \#3. Both are evaluated at run-time, as this is important to get the correct escape character.

The other five conditionals have additional code that compares the argument \#1 to two \TeX primitives which would wrongly be recognized as registers otherwise. Despite using \TeX’s primitive conditional construction, this does not break when \#1 is itself a conditional, because branches of the conditionals are only skipped if \#1 is one of the two primitives that are tested for (which are not \TeX conditionals).

\begin{verbatim}
\groupBegin
\cs_set_protected:Npn \_token\_tmp:w \#1\#2\#3
{ \use:x
  { \prg_new_conditional:Npnn \exp_not:c { token_if_ \#1 :N } \#1 
    { p , T , F , TF }
    { \cs_if_exist:cT { tex_ \#2 :D } 
      { \exp_not:N \if_meaning:w \#1 \exp_not:c { tex_ \#2 :D } 
        \exp_not:N \prg_return_false:
        \exp_not:N \else:
        \exp_not:N \if_meaning:w \#1 \exp_not:c { tex_ \#2 def:D } 
          \exp_not:N \prg_return_false:
        \exp_not:N \else:
          \exp_not:N \str_if_eq:eeTF
            { \exp_not:N \exp_after:wN 
              \exp_not:c { __token\_delimit\_by_ \#2 :w }
            } { \token_to_str:N \char" }
        \exp_not:N \prg_return_true:
      } 
      \exp_not:N \fi:
    } 
  } 
  \_token\_tmp:w { chardef } { char" } \{ \token\_to\_str:n \char" } 
  \_token\_tmp:w { mathchardef } { char" } \{ \token\_to\_str:n \mathchar" } 
  \_token\_tmp:w \{ long\_macro \} \{ macro \} \{ \token\_to\_str:n \{ long \} macro \} 
  \_token\_tmp:w \{ protected\_macro \} \{ macro \}
\end{verbatim}
We filter out macros first, because they cause endless trouble later otherwise.

Primitives are almost distinguished by the fact that the result of \token_to_meaning:N is formed from letters only. Every other token has either a space (e.g., the letter A), a digit (e.g., \count123) or a double quote (e.g., \char"A).

Ten exceptions: on the one hand, \tex_undefined:D is not a primitive, but its meaning is undefined, only letters; on the other hand, \space, \italiccorr, \hyphen, \firstmark, \topmark, \botmark, \splitfirstmark, \splitbotmark, and \nullfont are primitives, but have non-letters in their meaning.

We start by removing the two first (non-space) characters from the meaning. This removes the escape character (which may be nonexistent depending on \endlinechar), and takes care of three of the exceptions: \space, \italiccorr and \hyphen, whose meaning is at most two characters. This leaves a string terminated by some \__token_stop.

The meaning of each one of the five \...mark primitives has the form \langle letters \rangle \langle user material \rangle. In other words, the first non-letter is a colon. We remove everything after the first colon.

We are now left with a string, which we must analyze. For primitives, it contains only letters. For non-primitives, it contains either "., or a space, or a digit. Two exceptions remain: \tex_undefined:D, which is not a primitive, and \nullfont, which is a primitive.

Spaces cannot be grabbed in an undelimited way, so we check them separately. If there is a space, we test for \nullfont. Otherwise, we go through characters one by one, and stop at the first character less than ‘A’ (this is not quite a test for “only letters”, but is close enough to work in this context). If this first character is : then we have a primitive, or \tex_undefined:D, and if it is " or a digit, then the token is not a primitive.

For LuaTEX we use a different implementation which just looks at the command name of the token and compares it to a list of non-primitives. Again, \nullfont is a special case because it is the only primitive with the normally non-primitive \set_font command code.

In LaTeX some of the command names are different, so we check for both versions. The first one is always the LaTeX version.

\sys_if_engine_luatex:TF \token_if_primitive_p:N \token_if_primitive:NTF \__token_if_primitive_space:w \__token_if_primitive_nullfont:N \__token_if_primitive_loop:N \__token_if_primitive_undefined:N \__token_if_primitive_lua:N
local get_index = token.get_index
local get_mode = token.get_mode or token.get_index
local cmd = command_id
local set_font = cmd'get_font'
local biggest_char = token.biggest_char and token.biggest_char()
or status.getconstants().max_character_code
local mode_below_biggest_char = {}
local index_not_nil = {}
local mode_not_null = {}
local non_primitive = {
    [cmd'left_brace'] = true,
    [cmd'right_brace'] = true,
    [cmd'math_shift'] = true,
    [cmd'mac_param' or cmd'parameter'] = mode_below_biggest_char,
    [cmd'sup_mark' or cmd'superscript'] = true,
    [cmd'sub_mark' or cmd'subscript'] = true,
    [cmd'endv' or cmd'ignore'] = true,
    [cmd'spacer'] = true,
    [cmd'letter'] = true,
    [cmd'other_char'] = true,
    [cmd'tab_mark' or cmd'alignment_tab'] = mode_below_biggest_char,
    [cmd'char_given'] = true,
    [cmd'math_given' or 'math_char_given'] = true,
    [cmd'xmath_given' or 'math_char_xgiven'] = true,
    [cmd'set_font'] = mode_not_null,
    [cmd'undefined_cs'] = true,
    [cmd'call'] = true,
    [cmd'long_call' or cmd'protected_call'] = true,
    [cmd'outer_call' or cmd'tolerant_call'] = true,
    [cmd'long_outer_call' or cmd'tolerant_protected_call'] = true,
    [cmd'assign_glue' or cmd'register_glue'] = index_not_nil,
    [cmd'assign_mu_glue' or cmd'register_mu_glue'] = index_not_nil,
    [cmd'assign_toks' or cmd'register_toks'] = index_not_nil,
    [cmd'assign_int' or cmd'register_int'] = index_not_nil,
    [cmd'assign_attr' or cmd'register_attribute'] = true,
    [cmd'assign_dimen' or cmd'register_dimen'] = index_not_nil,
}

luacmd("__token_if_primitive_lua:N", function()
    local tok = get_next()
    local is_non_primitive = non_primitive[get_command(tok)]
    return put_next(
        is_non_primitive == true
        and false_tok
        or is_non_primitive == nil
        and true_tok
        or is_non_primitive == mode_not_null
        and (get_mode(tok) == 0 and true_tok or false_tok)
        or is_non_primitive == index_not_nil
        and (get_index(tok) and false_tok or true_tok)
        or is_non_primitive == mode_below_biggest_char
        and (get_mode(tok) > biggest_char and true_tok or false_tok)
    end, "global")
\prg_new_conditional:Nnn \token_if_primitive:N \ #1 \ { p , T , F , TF \}
\{ \__token_if_primitive_lua:N \ #1 \}
\}
\}
\tex_chardef:D \c__token_A_int = 'A ~ %
\use:x
\{ \prg_new_conditional:Nnn \exp_not:N \token_if_primitive:N \ #1 \ { p , T , F , TF \}
\{ \exp_not:N \token_if_macro:NTF \ #1 \ \exp_not:N \prg_return_false:
\{ \exp_not:N \exp_after:wN \exp_not:N \__token_if_primitive:NNw \exp_not:N \token_to_meaning:N \ #1 \ \tl_to_str:n \ { : : } \ s__token_stop \ #1 \}
\}
\cs_new:Npn \exp_not:N \__token_if_primitive:NNw \ #1 \ #2 \ #3 \ #4 \ s__token_stop
\{ \exp_not:N \tl_if_empty:oTF \{ \exp_not:N \__token_if_primitive_space:w \ #3 \ - \} \}
\{ \exp_not:N \__token_if_primitive_loop:N \ #3 \ \c_colon_str \ s__token_stop \}
\}
\{ \exp_not:N \__token_if_primitive_nullfont:N \}
\}
\cs_new:Npn \__token_if_primitive_space:w \ #1 ~ \{ \}
\cs_new:Npn \__token_if_primitive_nullfont:N \ #1
\{ \if_meaning:w \tex_nullfont:D \ #1 \ \prg_return_true: \else: \prg_return_false: \fi: \}
\cs_new:Npn \__token_if_primitive_loop:N \ #1
\{ \if_int_compare:w \ #1 < \c__token_A_int \ % \ \exp_after:wN \__token_if_primitive:Nw \ \exp_after:wN \ #1 \}
\else: \exp_after:wN \__token_if_primitive_loop:N \fi: \}
\cs_new:Npn \__token_if_primitive_loop:Nw \ #1 \ #2 \ s__token_stop
875
\token_case_catcode:Nn \token_case_catcode:NnT \token_case_charcode:Nn \token_case_charcode:NnT \token_case_meaning:Nn \token_case_meaning:NnT

\__token_case:NNnTF #1#2#3#4#5
\__token_case:NNw #1 #2 #3 #2 { }
\s__token_mark {#4}
\s__token_mark {#5}

\cs_new:Npn \__token_if_primitive_undefined:N \if:w \ifnum #1\else:\prg_return_false: \exp_after:wN \use_none:n \fi:\]
\cs_new:Npn \__token_if_primitive_undefined:N #1 \ifcs_exist:N #1 \prg_return_true: \else:\prg_return_false: \fi:

The aim here is to allow the case statement to be evaluated using a known number of
expansion steps (two), and without needing to use an explicit “end of recursion” marker.
That is achieved by using the test input as the final case, as this is always true. The
trick is then to tidy up the output such that the appropriate case code plus either the
true or false branch code is inserted.

\cs_new:Npn \token_case_catcode:Nn #1#2 { \exp:w \__token_case:NNnTF \token_if_eq_catcode:NNTF #1 {#2} { } { } }
\cs_new:Npn \token_case_catcode:NnT #1#2#3 { \exp:w \__token_case:NNnTF \token_if_eq_catcode:NNTF #1 {#2} {#3} { } }
\cs_new:Npn \token_case_catcode:NnF #1#2 { \exp:w \__token_case:NNnTF \token_if_eq_catcode:NNTF #1 {#2} { } }
\cs_new:Npn \token_case_meaning:Nn #1#2 { \exp:w \__token_case:NNnTF \token_if_eq_meaning:NNTF #1 {#2} { } { } }
\cs_new:Npn \token_case_meaning:NnT #1#2#3 { \exp:w \__token_case:NNnTF \token_if_eq_meaning:NNTF #1 {#2} {#3} { } }
\cs_new:Npn \token_case_meaning:NnF #1#2 { \exp:w \__token_case:NNnTF \token_if_eq_meaning:NNTF #1 {#2} { } }

\cs_new:Npn \__token_case:NNnTF #1#2#3#4#5 { \__token_case_end:sw #1#2#3#2 #5 #4 { } #5 { } #4 #5 { } #5 { } #4 { } #5 { } #4 { } #5 { }}

(End definition for \token_if_primitive:NTF and others. This function is documented on page 190.)
To tidy up the recursion, there are two outcomes. If there was a hit to one of the cases searched for, then #1 is the code to insert, #2 is the next case to check on and #3 is all of the rest of the cases code. That means that #4 is the true branch code, and #5 tidies up the spare \_\_token\_mark and the false branch. On the other hand, if none of the cases matched then we arrive here using the “termination” case of comparing the search with itself. That means that #1 is empty, #2 is the first \_\_token\_mark and so #4 is the false code (the true code is mopped up by #3).

(End definition for \token\_case\_catcode:NNwTF and others. These functions are documented on page 191.)

60.6 Peeking ahead at the next token

Peeking ahead is implemented using a two part mechanism. The outer level provides a defined interface to the lower level material. This allows a large amount of code to be shared. There are four cases:
1. peek at the next token;
2. peek at the next non-space token;
3. peek at the next token and remove it;
4. peek at the next non-space token and remove it.

\l\_\_peek\_search\_tl The token to search for as an explicit token: cf. \l\_\_peek\_search\_token.

(End definition for \l\_\_peek\_search\_token.)

\l\_\_peek\_search\_token The token to search for as an implicit token: cf. \l\_\_peek\_search\_tl.

(End definition for \l\_\_peek\_search\_token.)

\l\_peek\_token Storage tokens which are publicly documented: the token peeked.
\g\_peek\_token

(End definition for \l\_peek\_token and \g\_peek\_token. These variables are documented on page 191.)

\l\_peek_token
\g\_peek_token

877
\_\_peek_true:w Functions used by the branching and space-stripping code.
\_\_peek_true_aux:w
\_\_peek_false:w
\_\_peek_tmp:w
\cs_new:Npn \_\_peek_true:w { }
\cs_new:Npn \_\_peek_true_aux:w { }
\cs_new:Npn \_\_peek_false:w { }
\cs_new:Npn \_\_peek_tmp:w { }

(End definition for \_\_peek_true:w and others.)
\s__peek_mark Internal scan marks.
\s__peek_stop
\scan_new:N \s__peek_mark
\scan_new:N \s__peek_stop

(End definition for \s__peek_mark and \s__peek_stop.)
\_\_peek_use_none_delimit_by_s_stop:w Functions to gobble up to a scan mark.
\cs_new:Npn \_\_peek_use_none_delimit_by_s_stop:w #1 \s__peek_stop { }

(End definition for \_\_peek_use_none_delimit_by_s_stop:w.)
\peek_after:Nw \peek_gafter:Nw Simple wrappers for \futurelet: no arguments absorbed here.
\cs_new_protected:Npn \peek_after:Nw { \tex_futurelet:D \l_peek_token }
\cs_new_protected:Npn \peek_gafter:Nw { \tex_global:D \tex_futurelet:D \g_peek_token }

(End definition for \peek_after:Nw and \peek_gafter:Nw. These functions are documented on page 191.)
\_\_peek_true_remove:w A function to remove the next token and then regain control.
\cs_new_protected:Npn \_\_peek_true_remove:w
{ \tex_afterassignment:D \_\_peek_true_aux:w
  \cs_set:Npx \_\_peek_false:w { \exp_not:n {#1} }
  \group_align_safe_begin:
  \cs_set:Npn \_\_peek_true_aux:w { \peek_after:Nw \_\_peek_remove_spaces: }
  \_\_peek_true_aux:w
  \group_align_safe_end:
  \exp_after:N \_\_peek_remove_spaces: }

(End definition for \_\_peek_true_remove:w.)
\_\_peek_remove_spaces:n \_\_peek_remove_spaces:
Repeatedly use \_\_peek_true_remove:w to remove a space and call \_\_peek_true_-
aux:w.
\cs_new_protected:Npn \_\_peek_remove_spaces:n #1
{ \cs_set:Npx \_\_peek_false:w { \exp_not:n {#1} }
  \group_align_safe_begin:
  \cs_set:Npn \_\_peek_true_aux:w { \peek_after:Nw \_\_peek_remove_spaces: }
  \_\_peek_true_aux:w
  \group_align_safe_end:
  \exp_after:N \_\_peek_remove_spaces: }

(End definition for \_\_peek_remove_spaces:n)
Here we expand the input, removing spaces and \texttt{\textbackslash scan\_stop\_tokens} until we reach a non-expandable token. At that stage we re-insert the payload. To deal with the problem of \& tokens, we have to put the align-safe group in the correct place.

```latex
\cs_new_protected:Npn \peek_remove_filler:n #1
\exp_after:wN \peek_after:Nw \exp_after:wN \__peek_remove_filler:w
\exp:w \exp_end_continue_f:w
```

Here we can nest conditionals as \texttt{\l_peek_token} is only skipped over in the nested one if it’s a space: no problems with conditionals or outer tokens.

```latex
\if_catcode:w \exp_not:N \l_peek_token \c_space_token
\exp_after:wN \__peek_true_remove:w
\else:
\exp_after:wN \__peek_remove_filler:w
\fi:
```

To deal with undefined control sequences in the same way \TeX{} does, we need to check for expansion manually.

```latex
\if_meaning:w \exp_not:N \l_peek_token \l_peek_token
\exp_after:wN \__peek_false:w
\else:
\exp_after:wN \__peek_remove_filler_expand:w
\fi:
```

The generic functions store the test token in both implicit and explicit modes, and the \texttt{true} and \texttt{false} code as token lists, more or less. The two branches have to be absorbed
here as the input stream needs to be cleared for the peek function itself. Here, \#1 is \texttt{\_\_peek_true_remove:w} when removing the token and \texttt{\_\_peek_true_aux:w} otherwise.

```latex
\begin{verbatim}
cs_new_protected:Npn \_\_peek_token_generic_aux:NNNTF \#1\#2\#3\#4\#5
    \group_align_safe_begin:
cs_set_eq:NN \l__peek_search_token \#3
tl_set:Nn \l__peek_search_tl \{\#3\}
cs_set:Npx \_\_peek_true_aux:w
    \exp_not:N \group_align_safe_end:
    \exp_not:n \{\#4\}
}\cs_set_eq:NN \_\_peek_true:w \#1
\cs_set:Npx \_\_peek_false:w
    \exp_not:N \group_align_safe_end:
    \exp_not:n \{\#5\}
}\peek_after:Nw \#2
\end{verbatim}
```

\texttt{\_\_peek_token_remove_generic:NNTF} and \texttt{\_\_peek_token_remove_generic:NNF} for token removal there needs to be a call to the auxiliary function which does the work.

```latex
\begin{verbatim}
cs_new_protected:Npn \_\_peek_token_remove_generic:NNTF \_\_\_peek_true_remove:w
    \group_align_safe_begin:
cs_set_eq:NN \l__peek_search_token \#3
tl_set:Nn \l__peek_search_tl \{\#3\}
cs_set:Npx \_\_\_peek_true_remove:w
    \exp_not:N \group_align_safe_end:
    \exp_not:n \{\#4\}
}\cs_set_eq:NN \_\_\_peek_true:w \#1
\cs_set:Npx \_\_\_peek_false:w
    \exp_not:N \group_align_safe_end:
    \exp_not:n \{\#5\}
}\peek_after:Nw \#2
\end{verbatim}
```

\texttt{\_\_\_peek_token_remove_generic:NNTF} and \texttt{\_\_\_peek_token_remove_generic:NNF} (End definition for \texttt{\_\_\_peek_token_remove_generic:NNTF}.)

\texttt{\_\_\_peek_token_generic:NNTF} and \texttt{\_\_\_peek_token_remove_generic:NNTF} (End definition for \texttt{\_\_\_peek_token_remove_generic:NNTF} and \texttt{\_\_\_peek_token_remove_generic:NNTF}).

\texttt{\_\_\_peek_token:NN} \texttt{\_\_\_peek_token_remove:NN}

\texttt{\_\_\_peek_execute_branches_meaning:}

\texttt{\_\_\_peek_execute_branches_catcode:}
\texttt{\_\_\_peek_execute_branches_charcode:}
\texttt{\_\_\_peek_execute_branches_catcode_aux:}
\texttt{\_\_\_peek_execute_branches_catcode_auxii:N}
\texttt{\_\_\_peek_execute_branches_catcode_auxiii:}

The meaning test is straightforward.

```latex
\begin{verbatim}
cs_new:Npn \_\_\_peek_execute_branches_meaning: 
    \if_meaning:w \l_peek_token \l__peek_search_token
        \exp_after:wN \_\_\_peek_true:w
    \else:
        \exp_after:wN \_\_\_peek_false:w
    \fi:
\end{verbatim}
```

\texttt{\_\_\_peek_execute_branches_catcode:} and \texttt{\_\_\_peek_execute_branches_charcode:} (End definition for \texttt{\_\_\_peek_execute_branches_meaning:].)

The catcode and charcode tests are very similar, and in order to use the same auxiliaries we do something a little bit odd, firing \texttt{\_\_\_if_catcode:w} and \texttt{\_\_\_if_charcode:w} before finding the operands for those tests, which are only given in the \texttt{auxii:N} and \texttt{auxiii:auxiliaries}. For our purposes, three kinds of tokens may follow the peeking function:

---

880
• control sequences which are not equal to a non-active character token (e.g., macro, primitive);
• active characters which are not equal to a non-active character token (e.g., macro, primitive);
• explicit non-active character tokens, or control sequences or active characters set equal to a non-active character token.

The first two cases are not distinguishable simply using \TeX's \texttt{\futurelet}, because we can only access the \texttt{\meaning} of tokens in that way. In those cases, detected thanks to a comparison with \texttt{\scan_stop}, we grab the following token, and compare it explicitly with the explicit search token stored in \texttt{\l__peek_search_tl}. The \texttt{\exp_not:N} prevents outer macros (coming from non-\TeX3 code) from blowing up. In the third case, \texttt{\l__peek_token} is good enough for the test, and we compare it again with the explicit search token. Just like the peek token, the search token may be of any of the three types above, hence the need to use the explicit token that was given to the peek function.

\begin{verbatim}
\cs_new:Npn \__peek_execute_branches_catcode: 
{ \if_catcode:w \__peek_execute_branches_catcode_aux: }
\cs_new:Npn \__peek_execute_branches_charcode: 
{ \if_charcode:w \__peek_execute_branches_catcode_aux: }
\cs_new:Npn \__peek_execute_branches_catcode_aux: 
{ \if_catcode:w \exp_not:N \l_peek_token \scan_stop: 
\exp_after:wN \exp_after:wN 
\exp_after:wN \__peek_execute_branches_catcode_auxii:N 
\exp_after:wN \exp_not:N 
\else: 
\exp_after:wN \__peek_execute_branches_catcode_auxiii: 
\fi:}
\cs_new:Npn \__peek_execute_branches_catcode_auxii:N #1 
{ \exp_not:N #1 
\exp_after:wN \exp_not:N \l__peek_search_tl 
\exp_after:wN \__peek_true:w 
\exp_after:wN \__peek_false:w 
\fi: 
#1}
\cs_new:Npn \__peek_execute_branches_catcode_auxiii: 
{ \exp_not:N \l_peek_token 
\exp_after:wN \exp_not:N \l__peek_search_tl 
\exp_after:wN \__peek_true:w 
\exp_after:wN \__peek_false:w 
\fi:}
\end{verbatim}

(End definition for \texttt{\__peek_execute_branches_catcode:} and others.)

The public functions themselves cannot be defined using \texttt{\prg_new_conditional:Nnn}. Instead, the \texttt{TF}, \texttt{T}, \texttt{F} variants are defined in terms of corresponding variants of

\begin{verbatim}
\peek_catcode:NTF \peek_catcode_remove:NTF \peek_charcode:NTF \peek_charcode_remove:NTF \peek_meaning:NTF \peek_meaning_remove:NTF
\end{verbatim}

881
All tokens are N-type tokens, except in four cases: begin-group tokens, end-group tokens, space tokens with character code 32, and outer tokens. Since \l_peek_token might be outer, we cannot use the convenient \bool_if:NNTF function, and must resort to the old trick of using \ifodd to expand a set of tests. The false branch of this test is taken if the token is one of the first three kinds of non-N-type tokens (explicit or implicit), thus we call \__peek_false:w. In the true branch, we must detect outer tokens, without impacting performance too much for non-outer tokens. The first filter is to search for outer in the \meaning of \l_peek_token. If that is absent, \__peek_use_none_delimit_by_s_stop:w cleans up, and we call \__peek_true:w. Otherwise, the token can be a non-outer macro or a primitive mark whose parameter or replacement text contains outer, it can be the primitive outer, or it can be an outer token. Macros and marks would have ma in the part before the first occurrence of outer; the meaning of outer has nothing after outer, contrarily to outer macros; and that covers all cases, calling \__peek_true:w or \__peek_false:w as appropriate. Here, there is no \search token, so we feed a dummy \scan_stop: to the \__peek_token_generic:NNTF function.
\cs_new_protected:Npn \__peek_N_type:w ##1 #1 ##2 \s__peek_mark ##3
\{ ##3 {##1} {##2} \}
\exp_after:wN \__peek_tmp:w \tl_to_str:n { outer } \s__peek_stop
\group_end:
\cs_new_protected:Npn \__peek_N_type_aux:nnw #1 #2 #3 \fi:
\{ 
\fi:
\tl_if_in:noTF {#1} \{ \tl_to_str:n {ma} \}
\{ \__peek_true:w \}
\tl_if_empty:nTF {#2} \{ \__peek_true:w \} \{ \__peek_false:w \}
\}
\cs_new_protected:Npn \peek_N_type:TF
\{ \__peek_token_generic:NNTF
\__peek_execute_branches_N_type: \scan_stop:
\}
\cs_new_protected:Npn \peek_N_type:T
\{ \__peek_token_generic:NNT \__peek_execute_branches_N_type: \scan_stop: \}
\cs_new_protected:Npn \peek_N_type:F
\{ \__peek_token_generic:NNF \__peek_execute_branches_N_type: \scan_stop: \}

(End definition for \peek_N_type:TF and others. This function is documented on page 193.)

\end{document}
Chapter 61

l3prop implementation

The following test files are used for this code: m3prop001, m3prop002, m3prop003, m3prop004, m3show001.

\*package
\*prop

A property list is a macro whose top-level expansion is of the form

\__prop \__prop_pair:wn (key1) \__prop {value1}
...
\__prop_pair:wn (keyn) \__prop {value_n}

where \__prop is a scan mark (equal to \scan_stop:), and \__prop_pair:wn can be used to map through the property list.

\__prop The internal token used at the beginning of property lists. This is also used after each \langle key \rangle (see \__prop_pair:wn).

(End definition for \__prop.)

\__prop_pair:wn \__prop_pair:wn (key) \__prop {item}

The internal token used to begin each key–value pair in the property list. If expanded outside of a mapping or manipulation function, an error is raised. The definition should always be set globally.

(End definition for \__prop_pair:wn.)

\l__prop_internal_tl Token list used to store new key–value pairs to be inserted by functions of the \prop_put:Nnn family.

(End definition for \l__prop_internal_tl.)
\prop_split:NnTF \prop_split:NnTF \{\text{property list}\} \{\langle\text{key}\rangle\} \{\langle\text{true code}\rangle\} \{\langle\text{false code}\rangle\}

Splits the \langle\text{property list}\rangle at the \langle\text{key}\rangle, giving three token lists: the \langle\text{extract}\rangle of \langle\text{property list}\rangle before the \langle\text{key}\rangle, the \langle\text{value}\rangle associated with the \langle\text{key}\rangle and the \langle\text{extract}\rangle of the \langle\text{property list}\rangle after the \langle\text{value}\rangle. Both \langle\text{extracts}\rangle retain the internal structure of a property list, and the concatenation of the two \langle\text{extracts}\rangle is a property list. If the \langle\text{key}\rangle is present in the \langle\text{property list}\rangle then the \langle\text{true code}\rangle is left in the input stream, with \#1, \#2, and \#3 replaced by the first \langle\text{extract}\rangle, the \langle\text{value}\rangle, and the second extract. If the \langle\text{key}\rangle is not present in the \langle\text{property list}\rangle then the \langle\text{false code}\rangle is left in the input stream, with no trailing material. Both \langle\text{true code}\rangle and \langle\text{false code}\rangle are used in the replacement text of a macro defined internally, hence macro parameter characters should be doubled, except \#1, \#2, and \#3 which stand in the \langle\text{true code}\rangle for the three extracts from the property list. The \langle\text{key}\rangle comparison takes place as described for \str_if_eq:nn.

\s__prop
A private scan mark is used as a marker after each key, and at the very beginning of the property list.

\__prop_pair:wn
The delimiter is always defined, but when misused simply triggers an error and removes its argument.

\l__prop_internal_tl
Token list used to store the new key–value pair inserted by \prop_put:Nnn and friends.

\c_empty_prop
An empty prop.

61.1 Internal auxiliaries

\s__prop_mark
\s__prop_stop
Internal scan marks.

\q__prop_recursion_tail
\q__prop_recursion_stop
Internal recursion quarks.

\prop_if_recursion_tail_stop:n
\prop_if_recursion_tail_stop:o
Functions to query recursion quarks.
61.2 Allocation and initialisation

Property lists are initialized with the value `\c_empty_prop`.

\begin{verbatim}
\cs_new_protected:Npn \prop_new:N #1
{ \__kernel_chk_if_free_cs:N #1 \cs_gset_eq:NN #1 \c_empty_prop }
\cs_generate_variant:Nn \prop_new:N { c }
\end{verbatim}

(End definition for `\prop_new:N`. This function is documented on page 198.)

The same idea for clearing.

\begin{verbatim}
\cs_new_protected:Npn \prop_clear:N #1
{ \prop_set_eq:NN #1 \c_empty_prop }
\cs_generate_variant:Nn \prop_clear:N { c }
\cs_new_protected:Npn \prop_gclear:N #1
{ \prop_gset_eq:NN #1 \c_empty_prop }
\cs_generate_variant:Nn \prop_gclear:N { c }
\end{verbatim}

(End definition for `\prop_clear:N` and `\prop_gclear:N`. These functions are documented on page 198.)

Once again a simple variation of the token list functions.

\begin{verbatim}
\cs_new_protected:Npn \prop_clear_new:N #1
{ \prop_if_exist:NTF #1 { \prop_clear:N #1 } { \prop_new:N #1 } }
\cs_generate_variant:Nn \prop_clear_new:N { c }
\cs_new_protected:Npn \prop_gclear_new:N #1
{ \prop_if_exist:NTF #1 { \prop_gclear:N #1 } { \prop_new:N #1 } }
\cs_generate_variant:Nn \prop_gclear_new:N { c }
\end{verbatim}

(End definition for `\prop_clear_new:N` and `\prop_gclear_new:N`. These functions are documented on page 198.)

These are simply copies from the token list functions.

\begin{verbatim}
\cs_new_protected:Npn \prop_set_eq:NN \tl_set_eq:NN
\cs_new_protected:Npn \prop_set_eq:Nc \tl_set_eq:Nc
\cs_new_protected:Npn \prop_set_eq:cN \tl_set_eq:cN
\cs_new_protected:Npn \prop_set_eq:cc \tl_set_eq:cc
\cs_new_protected:Npn \prop_gset_eq:NN \tl_gset_eq:NN
\cs_new_protected:Npn \prop_gset_eq:Nc \tl_gset_eq:Nc
\cs_new_protected:Npn \prop_gset_eq:cN \tl_gset_eq:cN
\cs_new_protected:Npn \prop_gset_eq:cc \tl_gset_eq:cc
\end{verbatim}

(End definition for `\prop_set_eq:NN` and `\prop_gset_eq:NN`. These functions are documented on page 199.)

We can now initialize the scratch variables.

\begin{verbatim}
\prop_new:N \l_tmpa_prop
\prop_new:N \l_tmpb_prop
\prop_new:N \g_tmpa_prop
\prop_new:N \g_tmpb_prop
\prop_new:N \l__prop_internal_prop
\end{verbatim}

(End definition for `\l_tmpa_prop` and others. These variables are documented on page 205.)


\l__prop_internal_prop Property list used by `\prop_concat:NNN`, `\prop_set_from_keyval:Nn` and others.
\begin{verbatim}
\prop_new:N \l__prop_internal_prop
\end{verbatim}

886
Combine two property lists. We cannot use a simple \_\_\_prop_concat:NNN because there may be some duplicate keys between the two property lists.

\cs_new_protected:Npn \prop_concat:NNN { \__prop_concat:NNNN \prop_set_eq:NN }
\cs_generate_variant:Nn \prop_concat:NNN { ccc }
\cs_new_protected:Npn \prop_gconcat:NNN { \__prop_concat:NNNN \prop_gset_eq:NN }
\cs_generate_variant:Nn \prop_gconcat:NNN { ccc }
\cs_new_protected:Npn \__prop_concat:NNNN #1#2#3#4
{ \prop_set_eq:NN \l__prop_internal_prop #3
\prop_map_inline:Nn #4 \prop_put:Nnn \l__prop_internal_prop {##1} {##2} #1 #2 \l__prop_internal_prop }

\__prop_concat:NNNN
\prop_set_from_keyval:Nn \prop_set_from_keyval:cn \prop_gset_from_keyval:Nn \prop_gset_from_keyval:cn \prop_const_from_keyval:Nn \prop_const_from_keyval:cn \prop_put_from_keyval:Nn \prop_gput_from_keyval:Nn \prop_gput_from_keyval:cn \__prop_missing_eq:n

To avoid tracking throughout the loop the variable name and whether the assignment is local/global, do everything in a scratch variable and empty it afterwards to avoid wasting memory. Loop through items separated by commas, with \texttt{\_\_\_prop_do_nothing:} to avoid losing braces. After checking for termination, split the item at the first and then at the second = (which ought to be the first of the trailing = that we added). For both splits trim spaces and call a function (first \_\_\_prop_from_keyval_key:w then \_\_\_prop_from_keyval_value:w), followed by the trimmed material, \_\_\_prop_mark, the subsequent part of the item, and the trailing =’s and \_\_\_prop_stop. After finding the \langle key \rangle just store it after \_\_\_prop_stop. After finding the \langle value \rangle ignore completely empty items (both trailing = were used as delimiters and all parts are empty); if the remaining part #2 consists exactly of the second trailing = (namely there was exactly one = in the item) then output one key–value pair for the property list; otherwise complain about a missing or extra =.

\cs_new_protected:Npn \prop_set_from_keyval:Nn #1 #1
{ \prop_clear:N #1 \prop_put_from_keyval:Nn #1 #1 }
\cs_generate_variant:Nn \prop_set_from_keyval:Nn { c }
\cs_new_protected:Npn \prop_gset_from_keyval:Nn \prop_gset_from_keyval:cn \prop_gput_from_keyval:Nn \prop_gput_from_keyval:cn \__prop_missing_eq:n
61.3 Accessing data in property lists

This function is used by most of the module, and hence must be fast. It receives a ⟨property list⟩, a ⟨key⟩, a ⟨true code⟩ and a ⟨false code⟩. The aim is to split the ⟨property list⟩ at the given ⟨key⟩ into the ⟨extract1⟩ before the key–value pair, the ⟨value⟩ associated with the ⟨key⟩ and the ⟨extract2⟩ after the key–value pair. This is done using a delimited function, whose definition is as follows, where the ⟨key⟩ is turned into a string.

\begin{verbatim}
\cs_set:Npn \__prop_split_aux:w #1
  \__prop_pair:wn ⟨key⟩ \s__prop {#2} #3 \s__prop_mark #4 #5 \s__prop_stop
  { #4 {⟨true code⟩} } {⟨false code⟩}
\end{verbatim}

If the ⟨key⟩ is present in the property list, \__prop_split_aux:w’s #1 is the part before the ⟨key⟩, #2 is the ⟨value⟩, #3 is the part after the ⟨key⟩, #4 is \use_i:nn, and #5 is additional tokens that we do not care about. The ⟨true code⟩ is left in the input stream, and can use the parameters #1, #2, #3 for the three parts of the property list as desired. Namely, the original property list is in this case #1 \__prop_pair:wn ⟨key⟩ \s__prop ⟨#2⟩ #3.

If the ⟨key⟩ is not there, then the ⟨function⟩ is \use_ii:nn, which keeps the ⟨false code⟩.

\begin{verbatim}
\cs_new_protected:Npn \__prop_split:NnTF #1#2 #3
  \exp_args:NNo \__prop_split_aux:NnTF #1 { \tl_to_str:n {#2} } #3
\end{verbatim}

\begin{verbatim}
\cs_set:Npn \__prop_split_aux:w #1
\end{verbatim}
Deleting from a property starts by splitting the list. If the key is present in the property list, the returned value is ignored. If the key is missing, nothing happens.

```
cs_new_protected:Npn \prop_remove:Nn #1#2
{ \__prop_split:NnTF #1 {#2} { \tl_set:Nn #1 { ##1 ##3 } } { } }
cs_new_protected:Npn \prop_gremove:Nn #1#2
{ \__prop_split:NnTF #1 {#2} { \tl_gset:Nn #1 { ##1 ##3 } } { } }
cs_generate_variant:Nn \prop_remove:Nn { NV , c , cV } 
cs_generate_variant:Nn \prop_gremove:Nn { NV , c , cV } 
```

(End definition for \__prop_split:NnTF, \__prop_split_aux:NnTF, and \__prop_split_aux:w.)

Getting an item from a list is very easy: after splitting, if the key is in the property list, just set the token list variable to the return value, otherwise to \q_no_value.

```
cs_new_protected:Npn \prop_get:NnN #1#2#3
{ \__prop_split:NnTF #1 {#2} { \tl_set:Nn #3 {##2} \tl_set:Nn #1 { ##1 ##3 } } { \tl_set:Nn #3 { \q_no_value } } }
cs_generate_variant:Nn \prop_get:NnN { NV , Nv , No , c , cV , cv , co } 
```

(End definition for \prop_get:Nn. This function is documented on page 200.)

Popping a value also starts by doing the split. If the key is present, save the value in the token list and update the property list as when deleting. If the key is missing, save \q_no_value in the token list.

```
cs_new_protected:Npn \prop_pop:NnN #1#2#3
{ \__prop_split:NnTF #1 {#2} { \tl_set:Nn #3 {##2} } { \tl_set:Nn #3 { \q_no_value } } }
cs_generate_variant:Nn \prop_pop:NnN { NV , Nv , No , c , cV , cv , co } 
```

(End definition for \prop_pop:Nn. This function is documented on page 200.)
\prop_item:Nn

Getting the value corresponding to a key in a property list in an expandable fashion simply uses \prop_map_tokens:Nn to go through the property list. The auxiliary \_\prop_item:nnn receives the search string \#1, the key #2 and the value #3 and returns as appropriate.

\prop_count:N

Counting the key–value pairs in a property list is done using the same approach as for other count functions: turn each entry into a +1 then use integer evaluation to actually do the mathematics.

\prop_to_keyval:N

Each property name and value pair will be returned in the form \{\langle name\rangle\}={\langle value\rangle}. As one of the main use cases for this macro is to pass the ⟨property list⟩ on to a key–value parser, we have to make sure that the behaviour is as good as possible. Using a
space before the opening brace we get the correct brace stripping behaviour for most of
the key–value parsers available in L\TeX. If \texttt{\textexpanded:D} is available this function
makes use of it, so there are two different implementations here. They both start with
\texttt{\textkernelexpnot:w} to start the expansion context to expand in two steps. If the
\texttt{(property list)} is empty they just leave an empty set of braces in the input stream for
\texttt{\textkernelexpnot:w}.

\begin{verbatim}
\cs_if_exist:NTF \textexpanded:D
{ \__kernel_exp_not:w
\prop_if_empty:NTF #1
{ {} }
{ \exp_after:wN \exp_after:wN \exp_after:wN
\textexpanded:D
{ \__kernel_exp_not:w \use_none:n }
\prop_map_function:NN #1 \__prop_to_keyval:nn
\__kernel_exp_not:w
}
\cs_new:Npn \prop_to_keyval:N #1
{ \__kernel_exp_not:w
\prop_if_empty:NTF #1
{ {} }
{ \exp_after:wN \exp_after:wN \exp_after:wN
\textexpanded:D
{ \__kernel_exp_not:w \use_none:n }
\prop_map_function:NN #1 \__prop_to_keyval:nn
\__kernel_exp_not:w
}\__prop_to_keyval_exp_after:wN { \use_none:n }
\cs_new_eq:NN \__prop_to_keyval_exp_after:wN \exp_after:wN
\cs_new:Npn \__prop_to_keyval:nnw #1#2#3 \__prop_to_keyval_exp_after:wN #4
\end{verbatim}

The variant using \texttt{\textexpanded:D} can just iterate over the \texttt{(property list)} and remove
the leading comma afterwards. Only the value has to be protected in \texttt{\textkernelexpnot:w} as the
property name is always a string. After the loop the leading comma is removed by \use_none:n and afterwards
\texttt{\textkernelexpnot:w} eventually finds the
opening brace of its argument.

\begin{verbatim}
\cs_new:Npn \prop_to_keyval:N #1
{ \__kernel_exp_not:w
\prop_if_empty:NTF #1
{ {} }
{ \exp_after:wN \exp_after:wN \exp_after:wN
\textexpanded:D
{ \__kernel_exp_not:w \use_none:n }
\prop_map_function:NN #1 \__prop_to_keyval:nn
\__kernel_exp_not:w
}
\cs_new:Npn \__prop_to_keyval:nn #1#2
{ , - {#1} =\ { \__kernel_exp_not:w {#2} } }
\end{verbatim}

The other variant will iterate over the \texttt{(property list)} and has to output the result in a
\texttt{\_prop_to_keyval:nn} #1\texttt{\_prop_to_keyval_exp_after:wN}. As a result this is considerably
slower than the \texttt{\textexpanded:D} using variant as it has to read the entire contents
of the \texttt{(property list)} for each item. Since the marker is just \texttt{\exp_after:wN} with another
name, after the loop the leading comma is gobbled by \use_none:n, leaving the result
as the argument to \texttt{\textkernelexpnot:w}.

\begin{verbatim}
\cs_new:Npn \prop_to_keyval:N #1
{ \__kernel_exp_not:w
\prop_if_empty:NTF #1
{ {} }
{ \exp_after:wN \exp_after:wN \exp_after:wN
\textexpanded:D
{ \__kernel_exp_not:w \use_none:n }
\prop_map_function:NN #1 \__prop_to_keyval:nn
\__kernel_exp_not:w
}\__prop_to_keyval_exp_after:wN \__prop_to_keyval:nn #1\__prop_to_keyval_exp_after:wN\exp_after:wN
\cs_new:Npn \__prop_to_keyval:nnw #1#2#3 \__prop_to_keyval_exp_after:wN #4
\end{verbatim}

891
Popping an item from a property list, keeping track of whether the key was present or not, is implemented as a conditional. If the key was missing, neither the property list, nor the token list are altered. Otherwise, \texttt{prg\_return\_true:} is used after the assignments.

\begin{verbatim}
\prop_pop:NnNTF \prop_gpop:NnNTF
\end{verbatim}

Since the branches of \texttt{\_\_prop\_split:NnTF} are used as the replacement text of an internal macro, and since the \langle key \rangle and new \langle value \rangle may contain arbitrary tokens, it is not safe to include them in the argument of \texttt{\_\_prop\_split:NnTF}. We thus start by storing in \texttt{\_\_prop\_internal\_tl} tokens which (after x-expansion) encode the key–value pair. This variable can safely be used in \texttt{\_\_prop\_split:NnTF}. If the \langle key \rangle was absent, append the new key–value to the list. Otherwise concatenate the extracts \texttt{##1} and \texttt{##3} with the new key–value pair \texttt{\_\_prop\_internal\_tl}. The updated entry is placed at the same spot as the original \langle key \rangle in the property list, preserving the order of entries.

\begin{verbatim}
\prop_put:Nnn \prop_gput:Nnn
\end{verbatim}

(End definition for \prop\_to\_keyval:N and others. This function is documented on page 201.)
Adding conditionally also splits. If the key is already present, the three brace groups given by \_\_prop_split:NnTF are removed. If the key is new, then the value is added, being careful to convert the key to a string using \tl_to_str:n.

```latex
\cs_new_protected:Npn \prop_put_if_new:Nnn { \__prop_put_if_new:NNnn \__kernel_tl_set:Nx }
\cs_new_protected:Npn \prop_gput_if_new:Nnn { \__prop_put_if_new:NNnn \__kernel_tl_gset:Nx }
\cs_new_protected:Npn \__prop_put_if_new:NNnn \l__prop_internal_tl
\tl_set:Nn \l__prop_internal_tl { \exp_not:N \__prop_pair:wn \tl_to_str:n {#3} \s__prop \exp_not:n { {#4} } }
\__prop_split:NnTF \__kernel_tl_set:Nx \_\_prop_put_if_new:Nnn \__kernel_tl_gset:Nx \_\_prop splitting \_\_prop_put_if_new:Nn #1#2#3#4
\cs_generate_variant:Nn \prop_put_if_new:Nnn { c }
\cs_generate_variant:Nn \prop_gput_if_new:Nnn { c }

(End definition for \prop_put_if_new:Nnn and others. These functions are documented on page 199.)

61.4 Property list conditionals

```latex
\prg_new_eq_conditional:NNn \prop_if_exist:N \cs_if_exist:N { TF , T , F , p }
\prg_new_eq_conditional:NNn \prop_if_exist:c \cs_if_exist:c { TF , T , F , p }
(End definition for \prop_if_exist:NTF. This function is documented on page 201.)

\prg_new_conditional:Npnn \prop_if_empty:N #1 { p , T , F , TF }
\tl_if_eq:NNTF #1 \c_empty_prop \prg_return_true: \prg_return_false:
\prg_generate_conditional_variant:Nnn \prop_if_empty:N { c } { p , T , F , TF }

(End definition for \prop_if_empty:NTF. These functions are documented on page 200.)

Copies of the cs functions defined in \l3basics.

Same test as for token lists.
Testing expandably if a key is in a property list requires to go through the key-value pairs one by one. This is rather slow, and a faster test would be

\begin{verbatim}
\prg_new_protected_conditional:Npnn \prop_if_in:Nn #1 #2 
{ \@@_split:NnTF #1 {#2} 
{ \prg_return_true: } 
{ \prg_return_false: } 
}
\end{verbatim}

but \_prop_split:NnTF is non-expandable. Instead, we use \prop_map_tokens:Nn to compare the search key to each key in turn using \str_if_eq:ee, which is expandable.

\begin{verbatim}
\prg_new_protected_conditional:Npnn \prop_if_in:NnTF #1#2 { p , T , F , TF } 
{ \exp_args:NNo \prop_map_tokens:Nn #1 
{ \exp_after:wN \__prop_if_in:nnn \exp_after:wN { \tl_to_str:n {#2} } } 
\prg_return_false: 
}
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__prop_if_in:nnn #1#2#3 
{ \str_if_eq:eeT {#1} {#2} 
{ \prop_map_break:n { \use_i:nn \prg_return_true: } } 
} 
\end{verbatim}

\begin{verbatim}
\prg_generate_conditional_variant:Nnn \prop_if_in:Nn 
{ NV , No , c , cV , co } { p , T , F , TF } 
\end{verbatim}

(End definition for \prop_if_empty:NTF. This function is documented on page 202.)

Getting the value corresponding to a key, keeping track of whether the key was present or not, is implemented as a conditional (with side effects). If the key was absent, the token list is not altered.

\begin{verbatim}
\prg_new_protected_conditional:Npnn \prop_get:NnN #1#2#3 { T , F , TF } 
{ \__prop_split:NnTF #1 {#2} 
{ \tl_set:Nn #3 {##2} 
\prg_return_true: } 
} 
\end{verbatim}

\begin{verbatim}
\prg_generate_conditional_variant:Nnn \prop_get:NnN 
{ NV , No , c , cV , co } { T , F , TF } 
\end{verbatim}

(End definition for \prop_get:NnNTF and \_prop_if_in:nnn. This function is documented on page 202.)

61.5 Recovering values from property lists with branching

\begin{verbatim}
\prg_new_protected_conditional:Npnn \prop_get:NVVNTF #1#2#3 { p , T , F , TF } 
{ \prg_return_true: } 
\end{verbatim}

(End definition for \prop_get:NnNTF. This function is documented on page 202.)
61.6 Mapping over property lists

The even-numbered arguments of \_\_prop_map_function:Nw are keys, hence have string
catcodes, except at the end where they are \fi: \prop_map_break:. The \fi: ends the
\if_false: #\langle even\rangle \fi: construction and we jump out of the loop. No need for any
quark test.

\cs_new:Npn \prop_map_function:NN #1#2
{ \exp_after:wN \use_i_ii:nnn \exp_after:wN \__prop_map_function:Nw \exp_after:wN #2
#1 \__prop_pair:wn \fi: \prop_map_break: \s__prop { } \__prop_pair:wn \fi: \prop_map_break: \s__prop { } \__prop_pair:wn \fi: \prop_map_break: \s__prop { } \__prop_pair:wn \fi: \prop_map_break: \s__prop { } \prg_break_point:Nn \prop_map_break: { }
}
\cs_generate_variant:Nn \prop_map_function:NN { Nc , c , cc }

(End definition for \prop_map_function:NN and \_\_prop_map_function:Nw. This function is docu-
mented on page 203.)

\cs_new_protected:Npn \prop_map_inline:Nn #1#2
{ \cs_gset_eq:cN { __prop_map_ \int_use:N \g__kernel_prg_map_int :wn } \__prop_pair:wn \int_gincr:N \g__kernel_prg_map_int \cs_gset_protected:Npn \__prop_pair:wn ##1 \s__prop ##2 {#2} #1 \prg_break_point:Nn \prop_map_break: 
{ \int_gdecr:N \g__kernel_prg_map_int \cs_gset_eq:Nc \__prop_pair:wn { __prop_map_ \int_use:N \g__kernel_prg_map_int :wn }

highlighted text
The mapping is very similar to \prop_map_function:NN. The \use_i:nn removes the leading \s__prop. The odd construction \use:n {#1} allows #1 to contain any token without interfering with \prop_map_break:. The loop stops when the ⟨key⟩ between \__prop_pair:wn and \s__prop is \fi: \prop_map_break: instead of being a string.

\cs_new:Npn \prop_map_tokens:Nn #1#2
\exp_last_unbraced:Nno \use_i:nn { \__prop_map_tokens:nw {#2} } #1
\__prop_pair:wn \fi: \prop_map_break: \s__prop { }
\__prop_pair:wn \fi: \prop_map_break: \s__prop { }
\__prop_pair:wn \fi: \prop_map_break: \s__prop { }
\__prop_pair:wn \fi: \prop_map_break: \s__prop { }
\__prop_pair:wn \fi: \prop_map_break: \s__prop { }
\__prop_pair:wn \fi: \prop_map_break: \s__prop { }
\__prop_pair:wn \fi: \prop_map_break: \s__prop { }
\prg_break_point:Nn \prop_map_break: { }
\cs_generate_variant:Nn \prop_map_tokens:Nn { c }

(End definition for \prop_map_tokens:Nn and \__prop_map_tokens:nw. This function is documented on page 203.)

The break statements are based on the general \prg_map_break:Nn.
\cs_new:Npn \prop_map_break:
\prg_map_break:Nn \prop_map_break:
\cs_new:Npn \prop_map_break:n
\prg_map_break:Nn \prop_map_break:n

(End definition for \prop_map_break: and \prop_map_break:n. These functions are documented on page 204.)

61.7 Viewing property lists

Apply the general \__kernel_chk_tl_type:NnnT. Contrarily to sequences and comma lists, we use \msg_show_item:nn to format both the key and the value for each pair.
\cs_new_protected:Npn \prop_show:N { \__prop_show:NN \msg_show:nxxxx }
\cs_new_protected:Npn \prop_show:c { \__prop_show:NN \msg_show:nxxxx }
\cs_generate_variant:Nn \prop_show:N { c }
\cs_new_protected:Npn \prop_log:N { \__prop_show:NN \msg_log:nxxxx }
\cs_new_protected:Npn \prop_log:c { \__prop_show:NN \msg_log:nxxxx }
\__prop_show_validate:w
\cs_generate_variant:Nn \prop_log:N { c }
\cs_new_protected:Npn \__prop_show:NN #1#2
{ \__kernel_chk_tl_type:NnnT #2 { prop } { #1 { prop } { show } { \token_to_str:N #2 } { \prop_map_function:NN #2 \msg_show_item:nn } { } { } } }
\cs_new:Npn \__prop_show_validate:w #1 \__prop_pair:wn #2 \s__prop #3
{ \quark_if_recursion_tail_stop:n {#2} \exp_not:N \__prop_pair:wn \tl_to_str:n {#2} \s__prop \exp_not:n { {#3} } \__prop_show_validate:w }

(End definition for \prop_show:N and others. These functions are documented on page 204.)

(/package)
Chapter 62

\texttt{l3skip} implementation

\section*{62.1 Length primitives renamed}

Primitives renamed.

\begin{verbatim}
\if_dim:w \__dim_eval:w \__dim_eval_end: \cs_new_eq:NN \if_dim:w \tex_ifdim:D
\cs_new_eq:NN \__dim_eval:w \tex_dimexpr:D
\cs_new_eq:NN \__dim_eval_end: \tex_relax:D
\end{verbatim}

(End definition for \texttt{\if_dim:w}, \texttt{\__dim_eval:w}, and \texttt{\__dim_eval_end:}. This function is documented on page 220.)

\section*{62.2 Internal auxiliaries}

Internal scan marks.

\begin{verbatim}
\scan_new:N \s__dim_mark
\scan_new:N \s__dim_stop
\end{verbatim}

(End definition for \texttt{\s__dim_mark} and \texttt{\s__dim_stop}.)

Functions to gobble up to a scan mark.

\begin{verbatim}
\cs_new:Npn \__dim_use_none_delimit_by_s_stop:w #1 \s__dim_stop { }
\end{verbatim}

(End definition for \texttt{\__dim_use_none_delimit_by_s_stop:w}.)

\section*{62.3 Creating and initialising \texttt{dim} variables}

Allocating (\texttt{dim}) registers ...

\begin{verbatim}
\dim_new:N \dim_new:c
\end{verbatim}

\begin{verbatim}
\cs_generate_variant:Nn \dim_new:N { c }
\end{verbatim}
Contrarily to integer constants, we cannot avoid using a register, even for constants. We cannot use \texttt{\dim_gset:Nn} because debugging code would complain that the constant is not a global variable. Since \texttt{\dim_const:Nn} does not need to be fast, use \texttt{\dim_eval:n} to avoid needing a debugging patch that wraps the expression in checking code.

\begin{verbatim}
\cs_new_protected:Npn \dim_const:Nn #1#2
  { \dim_new:N #1 \tex_global:D #1 = \dim_eval:n {#2} \scan_stop: }
\end{verbatim}

Reset the register to zero. Using \texttt{\c_zero_skip} deals with the case where the variable passed is incorrectly a skip (for example a \LaTeX\,2\,\varepsilon\,length). Besides, these functions are then simply copied for \texttt{\skip_zero:N} and related functions.

\begin{verbatim}
\cs_new_protected:Npn \dim_zero:N #1 { #1 = \c_zero_skip }
\cs_new_protected:Npn \dim_gzero:N #1 { \tex_global:D #1 = \c_zero_skip }
\end{verbatim}

Create a register if needed, otherwise clear it.

\begin{verbatim}
\cs_new_protected:Npn \dim_zero_new:N #1
  { \dim_if_exist:NTF #1 { \dim_zero:N #1 } { \dim_new:N #1 } }
\cs_new_protected:Npn \dim_gzero_new:N #1
  { \dim_if_exist:NTF #1 { \dim_gzero:N #1 } { \dim_new:N #1 } }
\end{verbatim}

Copies of the \texttt{cs} functions defined in \texttt{l3basics}.

\begin{verbatim}
\prg_new_eq_conditional:NNn \dim_if_exist:N \cs_if_exist:N { TF , T , F , p }
\prg_new_eq_conditional:NNn \dim_if_exist:c \cs_if_exist:c { TF , T , F , p }
\end{verbatim}

62.4 Setting \texttt{dim} variables

Setting dimensions is easy enough but when debugging we want both to check that the variable is correctly local/global and to wrap the expression in some code. The \texttt{\scan_stop:} deals with the case where the variable passed is a skip (for example a \LaTeX\,2\,\varepsilon\,length).


62.5 Utilities for dimension calculations

Functions for min, max, and absolute value with only one evaluation. The absolute value is evaluated by removing a leading - if present.

\cs_new_protected:Npn \dim_abs:n #1
\__dim_abs:N
\dim_max:nn
\dim_min:nn
\__dim_maxmin:wwN

(End definition for \dim_add:Nn and others. These functions are documented on page 207.)

900
With dimension expressions, something like 10 pt * ( 5 pt / 10 pt ) does not work. Instead, the ratio part needs to be converted to an integer expression. Using \int_value:w forces everything into sp, avoiding any decimal parts.

\dim_ratio:nn
\__dim_ratio:n

With dimension expressions, something like 10 pt * ( 5 pt / 10 pt ) does not work. Instead, the ratio part needs to be converted to an integer expression. Using \int_value:w forces everything into sp, avoiding any decimal parts.

\dim_compare_p:nNn
\dim_compare:nNnTF
\__dim_compare:w
\__dim_compare:wNN
\__dim_compare=:w
\__dim_compare!=:w
\__dim_compare:<:w
\__dim_compare>:w
\__dim_compare_error:

This code is adapted from the \int_compare:nTF function. First make sure that there is at least one relation operator, by evaluating a dimension expression with a trailing \__dim_compare_error:. Just like for integers, the looping auxiliary \__dim_compare:wNN closes a primitive conditional and opens a new one. It is actually easier to
grab a dimension operand than an integer one, because once evaluated, dimensions all end with \textit{pt} (with category other). Thus we do not need specific auxiliaries for the three “simple” relations $<$, $=$, and $>$.

\begin{verbatim}
\prg_new_conditional:Npnn \dim_compare:n #1 { p , T , F , TF } 
\{ \exp_after:wN \__dim_compare:w 
\dim_use:N \__dim_eval:w #1 \__dim_compare_error: \}\cs_new:Npn \__dim_compare:w #1 \__dim_compare_error: 
\{ \exp_after:wN \if_false: \exp:w \exp_end_continue_f:w 
\__dim_compare:wNN #1 ? { = \__dim_compare_end:w \else: } \s__dim_stop 
\}\exp_args:Nno \use:nn \{ \cs_new:Npn \__dim_compare:wNN #1 \__dim_compare_error: 
\{ \if_meaning:w = #3 \use:c { \__dim_compare_#2:w } \fi: 
#1 pt \exp_stop_f: \prg_return_false: \exp_after:wN \__dim_compare:wNN \dim_use:N \__dim_eval:w #3 
\}\cs_new:cpn { \__dim_compare_! :w } #1 \reverse_if:N #2 ! #3 = { #1 #2 = #3 } \cs_new:cpn { \__dim_compare_ = :w } #1 \__dim_eval:w = { #1 \__dim_eval:w } \cs_new:cpn { \__dim_compare_< :w } 
#1 \reverse_if:N #2 < #3 = { #1 #2 > #3 } \cs_new:cpn { \__dim_compare_> :w } 
#1 \reverse_if:N #2 > #3 = { #1 #2 < #3 } \cs_new:Npn \__dim_compare_end:nw \prg_return_false: \else: \prg_return_true: \fi: \cs_new_protected:Npn \__dim_compare_error: 
\{ \if_int_compare:w \c_zero_int \c_zero_int \fi: = \__dim_compare_error: \}
\end{verbatim}

(\textit{End definition for} \texttt{dim_compare:nTF} \textit{and others. This function is documented on page 209.})

For dimension cases, the first task to fully expand the check condition. The over all idea is then much the same as for \texttt{str_case:nnTF} as described in \texttt{l3basics}.

\begin{verbatim}
\cs_new:Npn \dim_case:nn \dim_case:nnTF \__dim_case:nnTF \__dim_case:nw \__dim_case_end:nw 
\{ \exp:w \exp_args:Nf \__dim_case:nnTF \{ \dim_eval:n \{#1\} \}
\}\cs_new:Npn \__dim_case_end:nw \__dim_compare_error: 
\{ \if_int_compare:w \c_zero_int \c_zero_int \fi: = \__dim_compare_error: \}
\end{verbatim}
\begin{quote}
\texttt{\{ \\
\exp:w \\
\exp_args:Nf \_dim_case:nnTF \{ \dim_eval:n {#1} \} {#2} {#3} \} \\
\cs_new:Npn \_dim_case:nnF #1#2 \\
{ \\
\exp:w \\
\exp_args:Nf \_dim_case:nnTF \{ \dim_eval:n {#1} \} {#2} { } \\
\} \\
\cs_new:Npn \_dim_case:nn #1#2 \\
{ \\
\exp:w \\
\exp_args:Nf \_dim_case:nnTF \{ \dim_eval:n {#1} \} {#2} { } { } \\
\} \\
\cs_new:Npn \_dim_case:nnTF #1#2#3#4 \\
{ \_dim_case:nw {#1} {#2} {#1} { } \s__dim_mark {#3} \s__dim_mark {#4} \s__dim_stop } \\
\cs_new:Npn \_dim_case:nw #1#2#3 \\
{ \dim_compare:nNnTF {#1} = {#2} \\
{ \_dim_case_end:nw {#3} } \\
{ \_dim_case:nw {#1} } \\
} \\
\cs_new:Npn \_dim_case_end:nw #1#2#3 \s__dim_mark #4#5 \s__dim_stop \\
{ \exp_end: #1 #4 } \\
\end{quote}

(End definition for \_dim_case:nnTF and others. This function is documented on page 210.)

## 62.7 Dimension expression loops

while_do and do_while functions for dimensions. Same as for the int type only the names have changed.

\begin{quote}
\texttt{\cs_new:Npn \_dim_while_do:nn \_dim_until_do:nn \_dim_do_while:nn \_dim_do_until:nn \\
\{ \\
\dim_compare:nT {#1} \\
{ \\
\_dim_compare:nF {#1} \\
{ \\
\_dim_compare:nF {#1} \} {#2} \\
\} \\
\} \\
\cs_new:Npn \_dim_compare:nNnTF \#1 = \#2 \\
\{ \_dim_case_end:nw \#3 \} \\
\{ \_dim_case:nw \#1 \} \\
\} \\
\cs_new:Npn \_dim_compare:nw \#1\#2\#3 \#4\#5 \#6 \_dim_stop \\
\{ \_exp_end: \#1 \#4 \} \\
\end{quote}
\cs_new:Npn \dim_do_until:nn #1#2
\{#2\dim_compare:nF {#1}{\dim_do_until:nn {#1} {#2}}\}
(End definition for \dim_while_do:nn and others. These functions are documented on page 211.)
\dim_while_do:nNnn \dim_until_do:nNnn \dim_do_while:nNnn \dim_do_until:nNnn while_do and do_while functions for dimensions. Same as for the int type only the names have changed.
\cs_new:Npn \dim_while_do:nNnn #1#2#3#4
\{ \dim_compare:nNnT {#1} #2 {#3} \#4 \dim_while_do:nNnn {#1} #2 {#3} {#4} \}
\cs_new:Npn \dim_until_do:nNnn #1#2#3#4
\{ \dim_compare:nNnF {#1} #2 {#3} \#4 \dim_until_do:nNnn {#1} #2 {#3} {#4} \}
\cs_new:Npn \dim_do_while:nNnn #1#2#3#4
\{ #4 \dim_compare:nNnT {#1} #2 {#3} \dim_do_while:nNnn {#1} #2 {#3} {#4} \}
\cs_new:Npn \dim_do_until:nNnn #1#2#3#4
\{ #4 \dim_compare:nNnF {#1} #2 {#3} \dim_do_until:nNnn {#1} #2 {#3} {#4} \}
(End definition for \dim_while_do:nNnn and others. These functions are documented on page 211.)

62.8 Dimension step functions

Before all else, evaluate the initial value, step, and final value. Repeating a function by steps first needs a check on the direction of the steps. After that, do the function for the start value then step and loop around. It would be more symmetrical to test for a step size of zero before checking the sign, but we optimize for the most frequent case (positive step).
\cs_new:Npn \dim_step_function:nnnN \__dim_step:wwwN \__dim_step:NnnnN
\texttt{\_\_dim\_eval:w} \#2 \texttt{\exp\_after:w} ;
\texttt{\_\_dim\_eval:w} \#3 ;
}
\cs_new:Npn \__dim_step:wwwN #1; #2; #3; #4
{
\dim_compare:nNnTF {#2} > \c_zero_dim
{ \__dim_step:NnnnN > }
{ \__dim_compare:nNnTF {#2} = \c_zero_dim
  { \msg\_expandable\_error:nnn \{ kernel \} \{ zero\_step \} \{#4\}
    \use\_none:nnnn
  }
  { \__dim\_step:NnnnN < }
}
{ \#1} { \#2} { \#3} \#4
}
\cs_new:Npn \__dim_step:NnnnN #1#2#3#4#5
{
\int_gincr:N \g__kernel_prg\_map\_int
\exp_args:NNc \__dim\_step:NNnnnn
\cs_gset_protected:Npx
{ __dim\_map_ \int_use:N \g__kernel_prg\_map\_int:w }
{#1}{\#2}{\#3}
{\tl_set:Nn \exp_not:N #4 {##1}\exp_not:n {#5}}
}
(End definition for \texttt{\_\_dim\_step\_function:nnnN}, \texttt{\_\_dim\_step:wwwN}, and \texttt{\_\_dim\_step:NnnnN}. This function is documented on page 211.)

The approach here is to build a function, with a global integer required to make the nesting safe (as seen in other in line functions), and map that function using \texttt{\_\_dim\_step\_function:nnnN}. We put a \texttt{\prg\_break\_point:Nn} so that map_break functions from other modules correctly decrement \texttt{\g__kernel\_prg\_map\_int} before looking for their own break point. The first argument is \texttt{\scan\_stop:}, so that no breaking function recognizes this break point as its own.

\cs_new_protected:Npn \dim_step_inline:nnnn
{
\int_gincr:N \g__kernel_prg\_map\_int
\exp_args:NNc \__dim\_step:NNnnnn
\cs_gset_protected:Npn
{ __dim\_map_ \int_use:N \g__kernel_prg\_map\_int:w }
}\cs_new_protected:Npn \dim_step_variable:nnnNn #1#2#3#4#5
{
\int_gincr:N \g__kernel_prg\_map\_int
\exp_args:NNc \__dim\_step:NNnnnn
\cs_gset_protected:Npx
{ __dim\_map_ \int_use:N \g__kernel_prg\_map\_int:w }
{#1}{#2}{#3}
{\tl_set:Nn \exp_not:N #4 {##1}\exp_not:n {#5}}
}
62.9 Using dim expressions and variables

\texttt{\textbackslash dim_eval:n} Evaluating a dimension expression expandably.

\begin{verbatim}
\cs_new:Npn \dim_eval:n #1 { \dim_use:N \__dim_eval:w #1 \__dim_eval_end: }
\end{verbatim}

(End definition for \texttt{\dim_eval:n}. This function is documented on page 212.)

\texttt{\textbackslash dim_sign:n} See \texttt{\dim_abs:n}. Contrarily to \texttt{\int_sign:n} the case of a zero dimension cannot be distinguished from a positive dimension by looking only at the first character, since \texttt{0.2pt} and \texttt{0pt} start the same way. We need explicit comparisons. We start by distinguishing the most common case of a positive dimension.

\begin{verbatim}
\cs_new:Npn \dim_sign:n #1 { \int_value:w \exp_after:wN \__dim_sign:Nw \dim_use:N \__dim_eval:w #1 \__dim_eval_end: ; \exp_stop_f: }
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__dim_sign:Nw #1#2 ; {
  \if_dim:w #1#2 > \c_zero_dim 1 \else:\n    \if_meaning:w - #1 -1 \else:\n      0 \fi:\fi:
  \fi:
}
\end{verbatim}

(End definition for \texttt{\dim_sign:n} and \texttt{\__dim_sign:Nw}. This function is documented on page 212.)

\texttt{\textbackslash dim_use:N} Accessing a \texttt{(dim)}. We hand-code the \texttt{c} variant for some speed gain.

\begin{verbatim}
\cs_new_eq:NN \dim_use:N \tex_the:D \cs_new:Npn \dim_use:c #1 { \tex_the:D \cs:w #1 \cs_end: }
\end{verbatim}

(End definition for \texttt{\dim_use:N}. This function is documented on page 212.)

\texttt{\textbackslash dim_to_decimal:n} A function which comes up often enough to deserve a place in the kernel. Evaluate the dimension expression \texttt{#1} then remove the trailing \texttt{pt}. When debugging is enabled, the argument is put in parentheses as this prevents the dimension expression from terminating early and leaving extra tokens lying around. This is used a lot by low-level manipulations.

\begin{verbatim}
\cs_new_protected:Npn \__dim_step:NNnnnn #1#2#3#4#5#6 { #1 #2 ##1 {#6} \dim_step_function:nnnN {#3} {#4} {#5} #2 \prg_break_point:Nn \scan_stop: { \int_gdecr:N \g__kernel_prg_map_int } }
\end{verbatim}

(End definition for \texttt{\dim_step_inline:nnnn}, \texttt{\dim_step_variable:nnnNn}, and \texttt{\__dim_step:NNnnnn}. These functions are documented on page 211.)
\cs_new:Npn \dim_to_decimal:n #1
{|}
\exp_after:wN \__dim_to_decimal:w \dim_use:N \__dim_eval:w #1 \__dim_eval_end:
|}
\use:x |}
\cs_new:Npn \exp_not:N \__dim_to_decimal:w
{|##1 . ##2 \tl_to_str:n { pt }}
{|\int_compare:nNnTF {#2} > \c_zero_int
{| \{ \#1 \ . \#2 \}
| \{ \#1 \}
|}
|}
(End definition for \dim_to_decimal:n and \__dim_to_decimal:w. This function is documented on page 212.)
\dim_to_decimal_in_bp:n Conversion to big points is done using a scaling inside \__dim_eval:w as \TeX does that using 64-bit precision. Here, 800/803 is the integer fraction for 72/72.27. This is a common case so is hand-coded for accuracy (and speed).
\cs_new:Npn \dim_to_decimal_in_bp:n #1
{| \dim_to_decimal:n { ( #1 ) * 800 / 803 } }
(End definition for \dim_to_decimal_in_bp:n. This function is documented on page 213.)
\dim_to_decimal_in_sp:n Another hard-coded conversion: this one is necessary to avoid things going off-scale.
\cs_new:Npn \dim_to_decimal_in_sp:n #1
{| \int_value:w \__dim_eval:w #1 \__dim_eval_end: }
(End definition for \dim_to_decimal_in_sp:n. This function is documented on page 213.)
\dim_to_decimal_in_unit:nn An analogue of \dim_ratio:nn that produces a decimal number as its result, rather than a rational fraction for use within dimension expressions.
\cs_new:Npn \dim_to_decimal_in_unit:nn #1#2
{| \dim_to_decimal:n
{| lpt * \dim_ratio:nn \{#1\} \{#2\}
|}
|}
(End definition for \dim_to_decimal_in_unit:nn. This function is documented on page 213.)
\dim_to_fp:n Defined in l3fp-convert, documented here.
(End definition for \dim_to_fp:n. This function is documented on page 213.)
62.10 Viewing \texttt{dim} variables

\begin{verbatim}
\dim_show:N \dim_show:c
\end{verbatim}

Diagnostics.
\begin{verbatim}
cs_new_eq:NN \dim_show:N \_kernel_register_show:N
\cs_generate_variant:Nn \dim_show:N {c}
\end{verbatim}

(End definition for \texttt{\dim_show:N}. This function is documented on page 213.)

\begin{verbatim}
\dim_show:n
\end{verbatim}

Diagnostics. We don’t use the TeX primitive \texttt{\showthe} to show dimension expressions: this gives a more unified output.
\begin{verbatim}
cs_new_protected:Npn \dim_show:n
\end{verbatim}

(End definition for \texttt{\dim_show:n}. This function is documented on page 214.)

\begin{verbatim}
\dim_log:N \dim_log:c \dim_log:n
\end{verbatim}

Diagnostics. Redirect output of \texttt{\dim_show:n} to the log.
\begin{verbatim}
cs_new_eq:NN \dim_log:N \_kernel_register_log:N
\cs_new_eq:NN \dim_log:c \_kernel_register_log:c
\cs_new_protected:Npn \dim_log:n
\end{verbatim}

(End definition for \texttt{\dim_log:N} and \texttt{\dim_log:n}. These functions are documented on page 214.)

62.11 Constant dimensions

\begin{verbatim}
\c_zero_dim \c_max_dim
\end{verbatim}

Constant dimensions.
\begin{verbatim}
dim_const:Nn \c_zero_dim { 0 pt }
dim_const:Nn \c_max_dim { 16383.99999 pt }
\end{verbatim}

(End definition for \texttt{\c_zero_dim} and \texttt{\c_max_dim}. These variables are documented on page 214.)

62.12 Scratch dimensions

\begin{verbatim}
\l_tmpa_dim \l_tmpb_dim \g_tmpa_dim \g_tmpb_dim
\end{verbatim}

We provide two local and two global scratch registers, maybe we need more or less.
\begin{verbatim}
dim_new:N \l_tmpa_dim
dim_new:N \l_tmpb_dim
dim_new:N \g_tmpa_dim
dim_new:N \g_tmpb_dim
\end{verbatim}

(End definition for \texttt{\l_tmpa_dim} and others. These variables are documented on page 214.)

62.13 Creating and initialising \texttt{skip} variables

\begin{verbatim}
\s__skip_stop
\end{verbatim}

Internal scan marks.
\begin{verbatim}
\scan_new:N \s__skip_stop
\end{verbatim}

(End definition for \texttt{\s__skip_stop}.)

908
Allocation of a new internal registers.

Contrarily to integer constants, we cannot avoid using a register, even for constants. See \cs_new_protected:Npn \skip_new:N for why we cannot use \skip_gset:Nn.

Reset the register to zero.

Create a register if needed, otherwise clear it.

Copies of the cs functions defined in l3basics.

(End definition for \skip_new:N. This function is documented on page 214.)

(End definition for \skip_const:Nn. This function is documented on page 215.)

(End definition for \skip_zero:N and \skip_gzero:N. These functions are documented on page 215.)

(End definition for \skip_zero_new:N and \skip_gzero_new:N. These functions are documented on page 215.)

(End definition for \skip_if_exist:NTF. This function is documented on page 215.)
62.14 Setting skip variables

\skip_set:Nn \skip_set:cn \skip_gset:Nn \skip_gset:cn

Much the same as for dimensions.

\skip_set_eq:NN \skip_set_eq:cN \skip_set_eq:Nc \skip_set_eq:cc
\skip_gset_eq:NN \skip_gset_eq:cN \skip_gset_eq:Nc \skip_gset_eq:cc

All straightforward.

\skip_add:Nn \skip_add:cn \skip_gadd:Nn \skip_gadd:cn
\skip_sub:Nn \skip_sub:cn \skip_gsub:Nn \skip_gsub:cn

Using by here deals with the (incorrect) case \skip123.

\skip_if_eq_p:nn \skip_if_eq:nn TF

Comparing skips means doing two expansions to make strings, and then testing them. As a result, only equality is tested.

\skip_if_eq:nnTF

(End definition for \skip_set:Nn and \skip_gset:Nn. These functions are documented on page 215.)

\skip_set_eq:NN \skip_set_eq:cN \skip_set_eq:Nc \skip_set_eq:cc
\skip_gset_eq:NN \skip_gset_eq:cN \skip_gset_eq:Nc \skip_gset_eq:cc

(End definition for \skip_set_eq:NN and \skip_gset_eq:NN. These functions are documented on page 215.)

\skip_add:Nn \skip_add:cn \skip_gadd:Nn \skip_gadd:cn
\skip_sub:Nn \skip_sub:cn \skip_gsub:Nn \skip_gsub:cn

(End definition for \skip_add:Nn and others. These functions are documented on page 215.)

62.15 Skip expression conditionals

\skip_if_eq_p:nn \skip_if_eq:nnTF

Comparing skips means doing two expansions to make strings, and then testing them. As a result, only equality is tested.

\skip_if_eq:nnTF

(End definition for \skip_if_eq:nnTF. This function is documented on page 216.)
With $\varepsilon$-\TeX{}, we have an easy access to the order of infinities of the stretch and shrink components of a skip. However, to access both, we either need to evaluate the expression twice, or evaluate it, then call an auxiliary to extract both pieces of information from the result. Since we are going to need an auxiliary anyways, it is quicker to make it search for the string `fil` which characterizes infinite glue.

\begin{verbatim}
\cs_set_protected:Npn \__skip_tmp:w #1
\prg_new_conditional:Npnn \skip_if_finite:n ##1 { p , T , F , TF }
\exp_after:wN \__skip_if_finite:wwNw
\skip_use:N \tex_glueexpr:D ##1 ; \prg_return_false:
#1 ; \prg_return_true: \s__skip_stop
\}
\cs_new:Npn \__skip_if_finite:wwNw ##1 #1 ##2 ; ##3 ##4 \s__skip_stop {##3}
\}
\exp_args:No \__skip_tmp:w { \tl_to_str:n { fil } }
\end{verbatim}

(End definition for `\skip_if_finite:nTF` and `\__skip_if_finite:wNw`. This function is documented on page 216.)

\section{Using skip expressions and variables}

\subsection{Evaluating a skip expression expandably}

\begin{verbatim}
\cs_new:Npn \skip_eval:n #1
{ \skip_use:N \tex_glueexpr:D #1 \scan_stop: }
\end{verbatim}

(End definition for `\skip_eval:n`. This function is documented on page 216.)

\subsection{Accessing a \langle skip\rangle.}

\begin{verbatim}
\cs_new_eq:NN \skip_use:N \dim_use:N
\cs_new_eq:NN \skip_use:c \dim_use:c
\end{verbatim}

(End definition for `\skip_use:N`. This function is documented on page 216.)

\section{Inserting skips into the output}

\subsection{Inserting skips.}

\begin{verbatim}
\cs_new_eq:NN \skip_horizontal:N \tex_hskip:D
\cs_new:Npn \skip_horizontal:n #1
{ \skip_horizontal:N \tex_glueexpr:D #1 \scan_stop: }
\cs_generate_variant:Nn \skip_horizontal:N { c }
\end{verbatim}

(End definition for `\skip_horizontal:N` and others. These functions are documented on page 217.)

911
62.18 Viewing skip variables

\skip_show:N  \skip_show:c
\skip_show:n  \skip_show:c
\skip_show:n

Diagnostics.
\cs_new_eq:NN \skip_show:N \__kernel_register_show:N
\cs_generate_variant:Nn \skip_show:N { c }
(End definition for \skip_show:N. This function is documented on page 216.)

\skip_show:n

Diagnostics. We don’t use the \TeX\ primitive \showthe\ to show skip expressions: this
gives a more unified output.
\cs_new_protected:Npn \skip_show:n
\{ \msg_show_eval:Nn \skip_eval:n \}
(End definition for \skip_show:n. This function is documented on page 216.)

\skip_log:N  \skip_log:c  \skip_log:n
\skip_log:n

Diagnostics. Redirect output of \skip_show:n to the log.
\cs_new_eq:NN \skip_log:N \__kernel_register_log:N
\cs_new_eq:NN \skip_log:c \__kernel_register_log:c
\cs_new_protected:Npn \skip_log:n
\{ \msg_log_eval:Nn \skip_eval:n \}
(End definition for \skip_log:N and \skip_log:n. These functions are documented on page 217.)

62.19 Constant skips

\c_zero_skip  \c_max_skip

Skips with no rubber component are just dimensions but need to terminate correctly.
\skip_const:Nn \c_zero_skip { \c_zero_dim }
\skip_const:Nn \c_max_skip { \c_max_dim }
(End definition for \c_zero_skip and \c_max_skip. These functions are documented on page 217.)

62.20 Scratch skips

\l_tmpa_skip  \l_tmpb_skip  \g_tmpa_skip  \g_tmpb_skip

We provide two local and two global scratch registers, maybe we need more or less.
\skip_new:N \l_tmpa_skip
\skip_new:N \l_tmpb_skip
\skip_new:N \g_tmpa_skip
\skip_new:N \g_tmpb_skip
(End definition for \l_tmpa_skip and others. These variables are documented on page 217.)

62.21 Creating and initialising muskip variables

\muskip_new:N  \muskip_new:c
\muskip_new:N

And then we add muskips.
\cs_new_protected:Npn \muskip_new:N \#1
\{ \__kernel_chk_if_free_cs:N \#1
\cs:w newmuskip \cs_end: \#1
\}
\cs_generate_variant:Nn \muskip_new:N { c }
(End definition for \muskip_new:N. This function is documented on page 218.)
\textbf{\muskip\_const:Nn}\ See \textbf{\skip\_const:Nn}.
\begin{verbatim}
\cs_new_protected:Npn \muskip\_const:Nn \#1\#2
\{ \muskip\_new:N \#1
\tex\_global:D \#1 = \muskip\_eval:n \{\#2\} \scan\_stop:\
\}
\cs\_generate\_variant:Nn \muskip\_const:Nn \{ \cs\_const:Nn \}
\end{verbatim}

(End definition for \textbf{\muskip\_const:Nn}. This function is documented on page 218.)

\textbf{\muskip\_zero:N} \textbf{\muskip\_zero:c} \textbf{\muskip\_gzero:N} \textbf{\muskip\_gzero:c} \reset\ the\ register\ to\ zero.
\begin{verbatim}
\cs_new_protected:Npn \muskip\_zero:N \#1
\{ \#1 = \c\_zero\_muskip \}
\cs_new_protected:Npn \muskip\_gzero:N \#1
\{ \tex\_global:D \#1 = \c\_zero\_muskip \}
\cs\_generate\_variant:Nn \muskip\_zero:N \{ \cs\_const:Nn \}
\cs\_generate\_variant:Nn \muskip\_gzero:N \{ \cs\_const:Nn \}
\end{verbatim}

(End definition for \textbf{\muskip\_zero:N} \textbf{\_and} \textbf{\muskip\_gzero:N}. These functions are documented on page 218.)

\textbf{\muskip\_zero\_new:N} \textbf{\muskip\_zero\_new:c} \textbf{\muskip\_gzero\_new:N} \textbf{\muskip\_gzero\_new:c} \Create\ a\ register\ if\ needed,\ otherwise\ clear\ it.
\begin{verbatim}
\cs_new_protected:Npn \muskip\_zero\_new:N \#1
\{ \muskip\_if\_exist:NTF \#1 \{ \muskip\_zero:N \#1 \} \{ \muskip\_new:N \#1 \} \}
\cs_new_protected:Npn \muskip\_gzero\_new:N \#1
\{ \muskip\_if\_exist:NTF \#1 \{ \muskip\_gzero:N \#1 \} \{ \muskip\_new:N \#1 \} \}
\cs\_generate\_variant:Nn \muskip\_zero\_new:N \{ \cs\_const:Nn \}
\cs\_generate\_variant:Nn \muskip\_gzero\_new:N \{ \cs\_const:Nn \}
\end{verbatim}

(End definition for \textbf{\muskip\_zero\_new:N} \textbf{\_and} \textbf{\muskip\_gzero\_new:N}. These functions are documented on page 218.)

\textbf{\muskip\_if\_exist\_p:N} \textbf{\muskip\_if\_exist\_p:c} \textbf{\muskip\_if\_exist:NTF} \textbf{\muskip\_if\_exist:cTF} \Copies\ of\ the\ cs\ functions\ defined\ in\ \texttt{l3basics}.
\begin{verbatim}
\prg\_new\_eq\_conditional:NNn \muskip\_if\_exist:N \cs\_if\_exist:N
\{ \text{TF, T, F, p} \}
\prg\_new\_eq\_conditional:NNn \muskip\_if\_exist:c \cs\_if\_exist:c
\{ \text{TF, T, F, p} \}
\end{verbatim}

(End definition for \textbf{\muskip\_if\_exist:NTF}. This function is documented on page 218.)

\subsection{Setting \texttt{muskip} variables}

This should be pretty familiar.
\begin{verbatim}
\cs_new_protected:Npn \muskip\_set:Nn \#1\#2
\{ \#1 = \tex\_muexpr:D \#2 \scan\_stop: \}
\cs\_generate\_variant:Nn \muskip\_set:Nn \{ \cs\_const:Nn \}
\end{verbatim}

(End definition for \textbf{\muskip\_set:Nn} \textbf{\_and} \textbf{\muskip\_gset:Nn}. These functions are documented on page 219.)
All straightforward.

Using by here deals with the (incorrect) case \muskip123.

62.23 Using muskip expressions and variables

Evaluating a muskip expression expandably.

Accessing a (muskip).

62.24 Viewing muskip variables

Diagnostics.

Diagnostics. We don’t use the \TeX\ primitive \showthe\ to show muskip expressions: this gives a more unified output.
\texttt{\textbackslash muskip\_log:N} \linebreak \texttt{\textbackslash muskip\_log:c} \linebreak \texttt{\textbackslash muskip\_log:n} \texttt{Diagnostics. Redirect output of \texttt{\textbackslash muskip\_show:n} to the log.}\linebreak \texttt{\cs\new_eq:NN \textbackslash muskip\_log:N \textbackslash \_kernel\_register\_log:N} \linebreak \texttt{\cs\new_eq:NN \textbackslash muskip\_log:c \textbackslash \_kernel\_register\_log:c} \linebreak \texttt{\cs\new_protected:Npn \textbackslash muskip\_log:n} \linebreak \texttt{\{ \texttt{\msg\_log\_eval:Nn \textbackslash muskip\_eval:n} \}} \linebreak \texttt{(End definition for \textbackslash muskip\_log:N and \textbackslash muskip\_log:n. These functions are documented on page 220.)}

\texttt{\textbackslash c\_zero\_muskip} \linebreak \texttt{\textbackslash c\_max\_muskip} \texttt{Constant muskips given by their value.}\linebreak \texttt{\muskip\const:Nn \textbackslash c\_zero\_muskip \{ 0 \text{ mu} \}} \linebreak \texttt{\muskip\const:Nn \textbackslash c\_max\_muskip \{ 16383.99999 \text{ mu} \}} \linebreak \texttt{(End definition for \textbackslash c\_zero\_muskip and \textbackslash c\_max\_muskip. These functions are documented on page 220.)}

\texttt{\textbackslash l\_tmpa\_muskip} \linebreak \texttt{\textbackslash l\_tmpb\_muskip} \texttt{We provide two local and two global scratch registers, maybe we need more or less.}\linebreak \texttt{\muskip\new:N \textbackslash l\_tmpa\_muskip} \linebreak \texttt{\muskip\new:N \textbackslash l\_tmpb\_muskip} \linebreak \texttt{\muskip\new:N \textbackslash g\_tmpa\_muskip} \linebreak \texttt{\muskip\new:N \textbackslash g\_tmpb\_muskip} \linebreak \texttt{(End definition for \textbackslash l\_tmpa\_muskip and others. These variables are documented on page 220.)}

\texttt{/package}
Chapter 63

\l3keys Implementation

63.1 Low-level interface

The low-level key parser’s implementation is based heavily on \expkv. Compared to \keyval it adds a number of additional “safety” requirements and allows to process the parsed list of key–value pairs in a variety of ways. The net result is that this code needs around one and a half the amount of time as \keyval to parse the same list of keys. To optimise speed as far as reasonably practical, a number of lower-level approaches are taken rather than using the higher-level expl3 interfaces.

\l3keys
\l3prop
\keyval
\keyval_parse
\keyval_parse:N
\keyval_parse:NN

The general behavior of the \l3keys module is to throw an error on blank key names. However to support the usage of \keyval_parse:nnn in the \l3prop module we allow this error to be switched off temporarily and just ignore blank names.

\l3kernel_keyval_allow_blank_keys_bool

This temporary macro will be used since some of the definitions will need an active comma or equals sign. Inside of this macro #1 will be the active comma and #2 will be the active equals sign.

The main function starts the first of two loops. The outer loop splits the key–value list at active commas, the inner loop will do so at other commas. The use of \s__keyval_mark here prevents loss of braces from the key argument.
\cs_if_exist:NTF \tex_expanded:D 
  \cs_new:Npn \keyval_parse:nnn ##1 ##2 ##3 
  \__kernel_exp_not:w \tex_expanded:D 
  \cs_new:Npn \keyval_parse:nnn ##1 ##2 ##3 
  \group_align_safe_begin: 
  \__keyval_loop_active:nnw {##1} {##2} 
  \s__keyval_mark ##3 \s__keyval_tail #1 
  \group_align_safe_end: 
\cs_new_eq:NN \keyval_parse:NNn \keyval_parse:nnn

(End definition for \keyval_parse:nnn and \keyval_parse:NNn. These functions are documented on page 234.)

\__keyval_loop_active:nnw
First a fast test for the end of the loop is done, it’ll gobble everything up to a \s__keyval_tail. The loop ending macro will gobble everything to the last comma in this definition. If the end isn’t reached yet, start the second loop splitting at other commas, the next iteration of this first loop will be inserted by the end of \__keyval_loop_other:nnw.

\cs_new:Npn \__keyval_loop_active:nnw ##1 ##2 ##3 #1 
  \__keyval_if_recursion_tail:w ##3 
  \__keyval_end_loop_active:w \s__keyval_tail 
  \__keyval_loop_other:nnw {##1} {##2} ##3 , \s__keyval_tail , 
\cs_new_eq:NN \__keyval_loop_active:nnw \__keyval_loop_active:nnw

(End definition for \__keyval_loop_active:nnw.)

\__keyval_split_other:w
\__keyval_split_active:w
These two macros allow to split at the first equals sign of category 12 or 13. At the same time they also execute branching by inserting the first token following \s__keyval_mark that followed the equals sign. Hence they also test for the presence of such an equals sign simultaneously.

\cs_new:Npn \__keyval_split_other:w ##1 = ##2 \s__keyval_mark #3 
  \cs_new:Npn \__keyval_split_active:w ##1 #2 \s__keyval_mark #3 
  \cs_new:Npn \__keyval_split_active:w #1 #2 \s__keyval_mark #3 
\cs_new:Npn \__keyval_split_active:w \s__keyval_mark #3 
\cs_new:Npn \__keyval_split_active:w \s__keyval_mark #2 
\cs_new:Npn \__keyval_split_active:w \s__keyval_mark #2 
\cs_new:Npn \__keyval_split_active:w \s__keyval_mark #2 
\cs_new:Npn \__keyval_split_active:w \s__keyval_mark #2 

(End definition for \__keyval_split_other:w and \__keyval_split_active:w.)

\__keyval_loop_other:nnw
The second loop uses the same test for its end as the first loop, next it splits at the first active equals sign using \__keyval_split_active:w. The \s__keyval_nil prevents
accidental brace stripping and acts as a delimiter in the next steps. First testing for an active equals sign will reduce the number of necessary expansion steps for the expected average use case of other equals signs and hence perform better on average.

\cs_new:Npn \__keyval_loop_other:nnw ##1 ##2 ##3 ,
\{ 
  \__keyval_if_recursion_tail:w ##3 
  \__keyval_end_loop_other:w \s__keyval_nil
  \s__keyval_mark \__keyval_split_active:w ##3 \s__keyval_nil
  \s__keyval_mark \__keyval_split_active_auxi:w
    \#2 \s__keyval_mark \__keyval_clean_up_active:w
    \{##1} {##2}
  \s__keyval_mark \__keyval_split_active_auxii:w
\}

(End definition for \__keyval_loop_other:nnw.)

\__keyval_split_active_auxi:w
\__keyval_split_active_auxii:w
\__keyval_split_active_auxiii:w
\__keyval_split_active_auxiv:w
\__keyval_split_active_auxv:w

After \__keyval_split_active:w the following will only be called if there was at least one active equals sign in the current key–value pair. Therefore this is the execution branch for a key–value pair with an active equals sign. \#1 will be everything up to the first active equals sign. First it tests for other equals signs in the key name, which will eventually throw an error via \__keyval_misplaced_equal_after_active_error:w. If none was found we forward the key to \__keyval_split_active_auxii:w.

\cs_new:Npn \__keyval_split_active_auxii:w
\{ \s__keyval_stop \s__keyval_mark \__keyval_misplaced_equal_after_active_error:w
  \s__keyval_nil #2 \s__keyval_mark \__keyval_clean_up_active:w
  \{ \__keyval_trim:nN {##1} \__keyval_split_active_auxiii:w {##2} \s__keyval_nil \}
\}

\__keyval_split_active_auxiii:w gets the correct key name with a leading \s__keyval_mark as \#1. It has to sanitise the remainder of the previous test and trims the key name which will be forwarded to \__keyval_split_active_auxiiw.

\cs_new:Npn \__keyval_split_active_auxiiw
\{ \s__keyval_nil \s__keyval_misplaced_equal_after_active_error:w
  \s__keyval_mark \__keyval_split_other:w #1 \s__keyval_nil
  \s__keyval_nil \__keyval_clean_up_active:w \s__keyval_mark \__keyval_split_active_auxii:w
\}

Next we test for a misplaced active equals sign in the value, if none is found \__keyval_split_active_auxiv:w will be called.

\cs_new:Npn \__keyval_split_active_auxiv:w
\{ \s__keyval_nil \s__keyval_misplaced_equal_in_split_error:w
  \s__keyval_nil \s__keyval_nil \__keyval_clean_up_active:w \s__keyval_mark \__keyval_split_active_auxiv:w
\}

This runs the last test after sanitising the remainder of the previous one. This time test for a misplaced equals sign of category 12 in the value. Finally the last auxiliary macro will be called.
This last macro in this execution branch sanitises the last test, trims the value and passes it to `\__keyval_pair:nnnn`.

```latex
\cs_new:Npn \__keyval_split_active_auxi:w ##1 \s__keyval_nil \s__keyval_mark \__keyval_misplaced_equal_in_split_error:w \s__keyval_stop \s__keyval_mark \__keyval_pair:nnnn
```

(End definition for `\__keyval_split_active_auxi:w` and others.)

The following is the branch taken if the key–value pair doesn’t contain an active equals sign. The remainder of that test will be cleaned up by `\__keyval_clean_up_active:w` which will then split at an equals sign of category other.

```
\cs_new:Npn \__keyval_clean_up_active:w
#1 \s__keyval_nil \s__keyval_mark \__keyval_split_active_auxi:w \s__keyval_stop
{ \__keyval_split_other:w ##1 \s__keyval_nil \s__keyval_mark \__keyval_split_other_auxi:w \__keyval_clean_up_other:w}
```

(End definition for `\__keyval_clean_up_active:w`.)

This is executed if the key–value pair doesn’t contain an active equals sign but at least one other. `#1` of `\__keyval_split_other_auxi:w` will contain the complete key name, which is trimmed and forwarded to the next auxiliary macro.

```
\cs_new:Npn \__keyval_split_other_auxi:w #1 \s__keyval_stop
{ \__keyval_split_other_auxii:w #1 \s__keyval_nil \s__keyval_mark \__keyval_split_other_auxiii:w}
```

We know that the value doesn’t contain misplaced active equals signs but we have to test for others. Also we need to sanitise the previous test, which is done here and not earlier to avoid superfluous argument grabbing.

```
\cs_new:Npn \__keyval_split_other_auxii:w
#1 #2 \s__keyval_nil = \s__keyval_mark \__keyval_clean_up_other:w
{ \__keyval_split_other:w #1 \s__keyval_nil \s__keyval_mark \__keyval_misplaced_equal_in_split_error:w \s__keyval_stop \s__keyval_mark \__keyval_pair:nnnn
```

(End definition for `\__keyval_split_other_auxi:w`, `\__keyval_split_other_auxii:w`, and `\__keyval_split_other_auxiii:w`.)
\__keyval_clean_up_other:w  \__keyval_clean_up_other:w is the last branch that might exist. It is called if no equals sign was found, hence the only possibilities left are a blank list element, which is to be skipped, or a lonely key. If it’s no empty list element this will trim the key name and forward it to \__keyval_key:nn.

\cs_new:Npn \__keyval_clean_up_other:w  
  \__keyval_if_blank:w \#1 \s__keyval_nil \s__keyval_stop \s__keyval_blank_true:w
  \s__keyval_mark \s__keyval_stop
  \__keyval_trim:nN { \#1 } \__keyval_key:nn
\__keyval_end_loop_other:w
\__keyval_end_loop_active:w

\__keyval_misplaced_equal_after_active_error:w  \__keyval_misplaced_equal_in_split_error:w

All these two macros do is gobble the remainder of the current other loop execution and throw an error. Afterwards they have to insert the next loop iteration.

\cs_new:Npn \__keyval_misplaced_equal_after_active_error:w  
  \s__keyval_mark \#1 \s__keyval_nil \s__keyval_mark \#2 \s__keyval_mark \s__keyval_split_active_auxii:w
  \s__keyval_mark \#3 \s__keyval_nil
  \__keyval_clean_up_active:w
  \msg_expandable_error:nn { keyval } { misplaced-equals-sign }
  \__keyval_loop_active:nnw
\cs_new:Npn \__keyval_misplaced_equal_in_split_error:w  
  \s__keyval_mark \#1 \s__keyval_nil \s__keyval_mark \#2 \s__keyval_mark \#3 \s__keyval_mark \#4 \#5
  \msg_expandable_error:nn { keyval } { misplaced-equals-sign }
  \__keyval_loop_other:nnw

\__keyval_end_loop_other:w  \__keyval_end_loop_active:w

All that’s left for the parsing loops are the macros which end the recursion. Both just gobble the remaining tokens of the respective loop including the next recursion call. \__keyval_end_loop_other:w also has to insert the next iteration of the active loop.

\cs_new:Npn \__keyval_end_loop_other:w  
  \s__keyval_tail \__keyval_split_active:w
  \s__keyval_mark \s__keyval_tail
  \s__keyval_nil \s__keyval_mark
  \__keyval_split_active_auxii:w
  \__keyval_clean_up_active:w
  \\msg \__keyval_loop_active:nnw
\cs_new:Npn \__keyval_end_loop_active:w  
  \s__keyval_tail \__keyval_loop_other:nnw \#1 \s__keyval_nil \s__keyval_tail , \s__keyval_tail ,
  \s__keyval_end_loop_active:w

920
The parsing loops are done, so here ends the definition of \_\_keyval\_tmp:w, which will finally set up the macros.

\begin{verbatim}
\char_set_catcode_active:n \{ ', \}
\char_set_catcode_active:n \{ '=' \}
\_\_keyval\_tmp:w , =
\end{verbatim}

\_\_keyval\_pair:nnnn
\_\_keyval\_key:nn

These macros will be called on the parsed keys and values of the key–value list. All arguments are completely trimmed. They test for blank key names and call the functions passed to \keyval\_parse:nnn inside of \exp\_not:n with the correct arguments. Afterwards they insert the next iteration of the other loop.

\begin{verbatim}
\cs_set_protected:Npn \_\_keyval\_tmp:w #1 #2
{ \cs_new:Npn \_\_keyval\_pair:nnnn ##1 ##2 ##3 ##4
{ \__keyval\_if\_blank:w \s__keyval\_mark \s__keyval\_nil \s__keyval\_stop \__keyval\_blank\_key\_error:w
\s__keyval\_stop
#1 \exp_not:n { ##4 {##2} {##1} }
#2 \__keyval\_loop\_other:nnw {##3} {##4} }
\cs_new:Npn \_\_keyval\_key:nn ##1 ##2
{ \__keyval\_if\_blank:w \s__keyval\_mark ##1 \s__keyval\_nil \s__keyval\_stop \__keyval\_blank\_key\_error:w
\s__keyval\_stop
#1 \exp_not:n { ##2 {##1} }
#2 \__keyval\_loop\_other:nnw {##2} }
\cs_if_exist:NTF \tex\_expanded:D
{ \_\_keyval\_tmp:w { } { } }
{ \_\_keyval\_tmp:w \group\_align\_safe\_end: \group\_align\_safe\_begin: }
\end{verbatim}

\_\_keyval\_if\_empty:w
\_\_keyval\_if\_blank:w
\_keyval\_if\_recursion\_tail:w

All these tests work by gobbling tokens until a certain combination is met, which makes them pretty fast. The test for a blank argument should be called with an arbitrary token following the argument. Each of these utilize the fact that the argument will contain a leading \s__keyval\_mark.

\begin{verbatim}
\cs_new:Npn \_\_keyval\_if\_empty:w \_\_keyval\_mark \_\_keyval\_stop \{ \}
\cs_new:Npn \_\_keyval\_if\_blank:w \_\_keyval\_mark \_\_keyval\_stop \{ \_\_keyval\_if\_empty:w \_\_keyval\_mark
\cs_new:Npn \_\_keyval\_if\_recursion\_tail:w \_\_keyval\_mark \_\_keyval\_tail \{ \}
\end{verbatim}

(End definition for \_\_keyval\_pair:nnnn and \_\_keyval\_key:nn.)
These macros will be called if the tests above didn’t gobble them, they execute the branching.

\cs_new:Npn \__keyval_blank_true:w \s__keyval_mark \s__keyval_stop \__keyval_trim:nN #1 \__keyval_loop_other:nnw

\cs_new:Npn \__keyval_blank_key_error:w \s__keyval_mark \s__keyval_stop \__keyval_loop_other:nnw
\bool_if:NTF \l__kernel_keyval_allow_blank_keys_bool
#1
{ \msg_expandable_error:nn { keyval } { blank-key-name } }
\__keyval_loop_other:nnw

(End definition for \__keyval_blank_true:w and \__keyval_blank_key_error:w.)

Two messages for the low level parsing system.

\msg_new:nnn { keyval } { misplaced-equals-sign } { Misplaced~‘=’~in~key-value~input~\msg_line_context: }
\msg_new:nnn { keyval } { blank-key-name } { Blank~key~name~in~key-value~input~\msg_line_context: }
\prop_gput:Nnn \g_msg_module_name_prop { keyval } { LaTeX3 }
\prop_gput:Nnn \g_msg_module_type_prop { keyval } { }

And an adapted version of \_\_tl_trim_spaces:nn which is a bit faster for our use case, as it can strip the braces at the end. This is pretty much the same concept, so I won’t comment on it here. The speed gain by using this instead of \_\_tl_trim_spaces_apply:nn is about 10% of the total time for \keyval_parse:NNn with one key and one key–value pair, so I think it’s worth it.

\group_begin:
\cs_set_protected:Npn \__keyval_tmp:w #1
{ \cs_new:Npn \__keyval_trim:nN ##1
{ \__keyval_trim_auxi:w
  #1
  \s__keyval_nil
  \s__keyval_mark \l__kernel_keyval_allow_blank_keys_bool
  { #1 }
  \__keyval_trim_auxii:w
  \__keyval_trim_auxiii:w
  #1 \s__keyval_nil
  \__keyval_trim_auxiv:w
}
\__keyval_trim_auxi:w
\__keyval_trim_auxii:w
\__keyval_trim_auxiii:w
\__keyval_trim_auxiv:w
}
\cs_new:Npn \__keyval_trim_auxi:w \__keyval_trim_auxii:w \s__keyval_mark \s__keyval_mark #1 \__keyval_trim_auxiii:w
\__keyval_trim_auxii:w \s__keyval_mark \s__keyval_mark \s__keyval_mark #1
\__keyval_trim_auxiii:w
\__keyval_trim_auxiv:w
\__keyval_trim_auxiv:w

922
This is the one macro which differs from the original definition.

\cs_new:Npn \__keyval_trim_auxiii:w ##1 #1 \s__keyval_nil #2
\{ #2
  #1 \s__keyval_nil
  \__keyval_trim_auxiii:w
\}

\cs_new:Npn \__keyval_trim_auxiv:w
\s__keyval_mark ##1 \s__keyval_nil
\__keyval_trim_auxiii:w \s__keyval_nil \__keyval_trim_auxiii:w #2
\{ #2 { #1 } \}
\__keyval_tmp:w { ~ }
group_end:

(End definition for \__keyval_trim:nN and others.)

63.2 Constants and variables

\c__keys_code_root_str
\c__keys_check_root_str
\c__keys_default_root_str
\c__keys_groups_root_str
\c__keys_inherit_root_str
\c__keys_type_root_str
\c__keys_props_root_str
\l_keys_choice_int
\l_keys_choice_tl
\l__keys_groups_clist
\l_keys_key_str
\l_keys_key_tl

Various storage areas for the different data which make up keys.

\str_const:Nn \c__keys_code_root_str { key~code~>~ }
\str_const:Nn \c__keys_check_root_str { key~check~>~ }
\str_const:Nn \c__keys_default_root_str { key~default~>~ }
\str_const:Nn \c__keys_groups_root_str { key~groups~>~ }
\str_const:Nn \c__keys_inherit_root_str { key~inherit~>~ }
\str_const:Nn \c__keys_type_root_str { key~type~>~ }
\str_const:Nn \c__keys_props_root_str { key~prop~>~ }

(End definition for \c__keys_code_root_str and others.)

(End definition for \c__keys_props_root_str.)

Publicly accessible data on which choice is being used when several are generated as a set.

\int_new:N \l_keys_choice_int
\tl_new:N \l_keys_choice_tl

(End definition for \l_keys_choice_int and \l_keys_choice_tl. These variables are documented on page 228.)

Used for storing and recovering the list of groups which apply to a key: set as a comma list but at one point we have to use this for a token list recovery.

\clist_new:N \l__keys_groups_clist

(End definition for \l__keys_groups_clist.)

The name of a key itself: needed when setting keys. The \tl version is deprecated but has to be handled manually.

\str_new:N \l_keys_key_str
\tl_new:N \l_keys_key_tl
\l__keys_module_str  The module for an entire set of keys.
\l__keys_module_str

\l__keys_no_value_bool  A marker is needed internally to show if only a key or a key plus a value was seen: this is recorded here.
\l__keys_no_value_bool

\l__keys_only_known_bool  Used to track if only “known” keys are being set.
\l__keys_only_known_bool

\l__keys_path_str  The “path” of the current key is stored here: this is available to the programmer and so is public. The older version is deprecated but has to be handled manually.
\l__keys_path_str
\l__keys_path_tl
\l__keys_path_tl

\l__keys_inherit_str
\l__keys_inherit_str

\l__keys_relative_tl  The relative path for passing keys back to the user. As this can be explicitly no-value, it must be a token list.
\l__keys_relative_tl
\l__keys_relative_tl
\tl_new:N \l__keys_relative_tl
\tl_set:Nn \l__keys_relative_tl { \q__keys_no_value }
\l__keys_relative_tl

\l__keys_property_str  The “property” begin set for a key at definition time is stored here.
\l__keys_property_str

\l__keys_selective_bool  Two flags for using key groups: one to indicate that “selective” setting is active, a second to specify which type (“opt-in” or “opt-out”).
\l__keys_selective_bool
\l__keys_filtered_bool
\l__keys_filtered_bool

\l__keys_selective_seq  The list of key groups being filtered in or out during selective setting.
\l__keys_selective_seq
\l__keys_unused_clist

Used when setting only some keys to store those left over.

(End definition for \l__keys_unused_clist.)

\l_keys_value_tl

The value given for a key: may be empty if no value was given.

(End definition for \l_keys_value_tl. This variable is documented on page 230.)

\l_keys_tmp_bool \l__keys_tmpa_tl \l__keys_tmpb_tl

Scratch space.

(End definition for \l__keys_tmp_bool, \l__keys_tmpa_tl, and \l__keys_tmpb_tl.)

\l_keys_precompile_bool \l_keys_precompile_tl

For digesting keys.

(End definition for \l__keys_precompile_bool and \l__keys_precompile_tl.)

\l_keys_usage_load_prop \l_keys_usage_preamble_prop

Global data for document-level information.

(End definition for \l_keys_usage_load_prop and \l_keys_usage_preamble_prop. These variables are documented on page 230.)

63.2.1 Internal auxiliaries

\s__keys_nil \s__keys_mark \s__keys_stop

Internal scan marks.

(End definition for \s__keys_nil, \s__keys_mark, and \s__keys_stop.)

\q__keys_no_value

Internal quarks.

(End definition for \q__keys_no_value.)

\__keys_quark_if_no_value:NTF

Branching quark conditional.

(End definition for \__keys_quark_if_no_value:NTF.)

\__keys_precompile:n

An auxiliary to allow cleaner showing of code.

(End definition for \__keys_precompile:n.)
63.3 The key defining mechanism

The public function for definitions is just a wrapper for the lower level mechanism, more or less. The outer function is designed to keep a track of the current module, to allow safe nesting. The module is set removing any leading / (which is not needed here).

\keys_define:nn
\keys_define:nnn
\keys_define:onn

The outer functions here record whether a value was given and then converge on a common internal mechanism. There is first a search for a property in the current key name, then a check to make sure it is known before the code hands off to the next step.

\keys_property_find:n
\keys_property_find_auxi:w
\keys_property_find_auxii:w
\keys_property_find_auxiii:w
\keys_property_find_auxiv:w

Searching for a property means finding the last . in the input, and storing the text before and after it. Everything is turned into strings, so there is no problem using an x-type expansion. Since \keys_trim_spaces:n will turn its argument into a string anyway, this function uses \cs_set_nopar:Npx instead of \tl_set:Nx to gain some speed.
Two possible cases. If there is a value for the key, then just use the function. If not, then check to make sure there is no need for a value with the property. If there should be one then complain, otherwise execute it. There is no need to check for a : as if it was missing the earlier tests would have failed.

(End definition for \_keys_property_find:n and others.)
63.4 Turning properties into actions

Boolean keys are really just choices, but all done by hand. The second argument here is the scope: either empty or \texttt{g} for global.

To make a choice from a key, two steps: set the code, and set the unknown key. As multichoices and choices are essentially the same bar one function, the code is given together.

(End definition for \texttt{\_keys_bool_set:Nn} and \texttt{\_keys_bool_set:w}.)
Auto-generating choices means setting up the root key as a choice, then defining each choice in turn.

Setting the code for a key first logs if appropriate that we are defining a new key, then saves the code.
Creating control sequences is a bit more tricky than other cases as we need to pick up the \texttt{p} argument. To make the internals look clearer, the trailing \texttt{n} argument here is just for appearance.

Assigning a key to one or more groups uses comma lists. As the list of groups only exists if there is anything to do, the setting is done using a scratch list. For the usual grouping reasons we use the low-level approach to undefining a list. We also use the low-level approach for the other case to avoid tripping up the \texttt{check-declarations} code.
\__keys_inherit:n

Inheritance means ignoring anything already said about the key: zap the lot and set up.

\cs_new_protected:Npn \__keys_inherit:n #1
{\__keys_undefine:
\cs_set_nopar:cpn { \c__keys_inherit_root_str \l_keys_path_str } {#1}}

(End definition for \__keys_inherit:n.)

\__keys_initialise:n

A set up for initialisation: just run the code if it exists.
\cs_new_protected:Npn \__keys_initialise:n #1
{\cs_if_exist:cTF { \c__keys_inherit_root_str \__keys_parent:o \l_keys_path_str }
\{ \__keys_execute_inherit: \}
\{ \str_clear:N \l__keys_inherit_str
\cs_if_exist:cT { \c__keys_code_root_str \l_keys_path_str }
\{ \__keys_execute:nn \l_keys_path_str {#1} \}
\} }

(End definition for \__keys_initialise:n.)

\__keys_legacy_if_set:nn
\__keys_legacy_if_inverse:nn
\__keys_legacy_if_inverse:nnn

Much the same as expl3 booleans, except we assume that the switch exists.
\cs_new_protected:Npn \__keys_legacy_if_set:nn #1#2
{\__keys_legacy_if_set:nnn {#1} {#2} { true } { false } }
\cs_new_protected:Npn \__keys_legacy_if_set_inverse:nn #1#2
{\__keys_legacy_if_set:nnn {#1} {#2} { false } { true } }
\cs_new_protected:Npn \__keys_legacy_if_set:nnnn #1#2#3#4
{ }
\__keys_choice_make:
\__keys_cmd_set:nx { \l_keys_path_str / true }
{ \exp_not:c { legacy_if_#2 set_ #3 :n } { \exp_not:n {#1} } }
\__keys_cmd_set:nx { \l_keys_path_str / false }
{ \exp_not:c { legacy_if_#2 set_ #4 :n } { \exp_not:n {#1} } }
\__keys_cmd_set:nn { \l_keys_path_str / unknown }
{ \msg_error:nmx { keys } { boolean-values-only }
\l_keys_key_str }
\__keys_default_set:n { true }
\cs_if_exist:cF { if#1 }
{ \cs:w newif \exp_after:wN \cs_end:
\cs:w if#1 \cs_end: }
}
To create a meta-key, simply set up to pass data through.

```latex
\cs_new_protected:Npn \__keys_meta_make:n #1
\begin{verbatim}
\exp_after:wN \keys_set:nn \exp_after:wN \l__keys_module_str \{#1\}
\end{verbatim}
```

(End definition for \__keys_meta_make:n and \__keys_meta_make:nn.)

Much the same as other variables, but needs a dedicated auxiliary.

```latex
\cs_new_protected:Npn \__keys_prop_put:Nn #1#2
\begin{verbatim}
\prop_if_exist:NF #1 { \prop_new:N #1 }
\exp_after:wN \__keys_find_key_module:WNN \l_keys_path_str \s__keys_stop
\l__keys_tmpa_tl \l__keys_tmpb_tl
\__keys_cmd_set:nx \l_keys_path_str
\begin{verbatim}
\exp_not:c { prop_ #2 put:Nnn }
\exp_not:N #1
\exp_not:n { {##1} }
\end{verbatim}
\end{verbatim}
```

(End definition for \__keys_prop_put:Nn and \__keys_prop_put:nn.)

Undefining a key has to be done without \cs_undefine:c as that function acts globally.

```latex
\cs_new_protected:Npn \__keys_undefine:
\begin{verbatim}
\clist_map_inline:nn
\begin{verbatim}
\cs_set_eq:cN { \tl_use:c { c__keys_ ##1_root_str } \l_keys_path_str }
\tex_undefined:D
\end{verbatim}
\end{verbatim}
```

(End definition for \__keys_undefine:)
Validating key input is done using a second function which runs before the main key code. Setting that up means setting it equal to a generic stub which does the check. This approach makes the lookup very fast at the cost of one additional csname per key that needs it. The cleanup here has to know the structure of the following code.

\cs_new_protected:Npn \__keys_value_requirement:nn #1#2
\str_case:nnF (#2)
{ true }

{ \cs_set_eq:cc
  { \c__keys_check_root_str \l_keys_path_str }
  { __keys_check_ #1 : }
}
{ false }

{ \cs_if_eq:ccT
  { \c__keys_check_root_str \l_keys_path_str }
  { __keys_check_ #1 : }
  { \cs_set_eq:cN
    { \c__keys_check_root_str \l_keys_path_str }
    \tex_undefined:D }
}
{ \msg_error:nnx { keys }{ boolean-values-only }
  .value_ #1 :n }

\cs_new_protected:Npn \__keys_check_forbidden:
{ \bool_if:NF \l__keys_no_value_bool
  { \msg_error:nxx { keys }{ value-forbidden }
    \l_keys_path_str \l_keys_value_tl
    \use_none:nnn }
}
\cs_new_protected:Npn \__keys_check_required:
{ \bool_if:NT \l__keys_no_value_bool
  { \msg_error:nx { keys }{ value-required }
    \l_keys_path_str \l_keys_value_tl
    \use_none:nnn }
}

(End definition for \__keys_value_requirement:nn, \__keys_check_forbidden:, and \__keys_check_required:)
Save the relevant data.

```latex
\cs_new_protected:Npn \__keys_usage:n #1
\str_case:nnF {#1}
\{ 
\{ general \}
\{ \__keys_usage:NN \l_keys_usage_load_prop \c_false_bool \__keys_usage:NN \l_keys_usage_preamble_prop \c_false_bool \}
\{ load \}
\{ \__keys_usage:NN \l_keys_usage_load_prop \c_true_bool \__keys_usage:NN \l_keys_usage_preamble_prop \c_false_bool \}
\{ preamble \}
\{ \__keys_usage:NN \l_keys_usage_load_prop \c_false_bool \__keys_usage:NN \l_keys_usage_preamble_prop \c_true_bool \}
\}
\msg_error:nnnn { keys }
\{ choice-unknown \}
\{ .usage:n \}
\{#1\}
\}
```

(End definition for \__keys_usage:n, \__keys_usage:NN, and \__keys_usage:w.)

Setting a variable takes the type and scope separately so that it is easy to make a new variable if needed.

```latex
\cs_new_protected:Npn \__keys_variable_set:NnnN #1#2#3#4
\prop_get:NVNF #1 \l__keys_module_str \l__keys_tmpa_tl
\tl_clear:N \l__keys_tmpa_tl
\tl_set:Nx \l__keys_tmpb_tl \exp_after:wN \__keys_usage:w \l_keys_path_str \s__keys_stop
\bool_if:NTF #2
\clist_put_right:NV \l__keys_tmpa_tl \l__keys_tmpb_tl
\clist_remove_all:NV \l__keys_tmpa_tl \l__keys_tmpb_tl
\prop_put:NVV #1 \l__keys_module_str
\l__keys_tmpa_tl
\}
```

63.5 Creating key properties

The key property functions are all wrappers for internal functions, meaning that things stay readable and can also be altered later on.

Importantly, while key properties have “normal” argument specs, the underlying code always supplies one braced argument to these. As such, argument expansion is handled by hand rather than using the standard tools. This shows up particularly for the two-argument properties, where things would otherwise go badly wrong.

One function for this.

(End definition for \_keys_variable_set:NnnN and \_keys_variable_set_required:NnnN.)

One function for this.

(End definition for .bool_set:N and .bool_gset:N. These functions are documented on page 223.)
Making a choice is handled internally, as it is also needed by \texttt{\_generate\_choices:n}.

\begin{verbatim}
\cs_new_protected:cpn { \_keys_props_root_str .choice: }
  { \_keys_choice_make: }
\end{verbatim}

(End definition for \texttt{\_choice:]. This function is documented on page 223.)

For auto-generation of a series of mutually-exclusive choices. Here, \#1 consists of two separate arguments, hence the slightly odd-looking implementation.

\begin{verbatim}
\cs_new_protected:cpn { \_keys_props_root_str .choices:nn } #1
  \{ \__keys_choices_make:nn #1 \}
\cs_new_protected:cpn { \_keys_props_root_str .choices:Vn } #1
  \{ \exp_args:NV \__keys_choices_make:nn #1 \}
\cs_new_protected:cpn { \_keys_props_root_str .choices:on } #1
  \{ \exp_args:No \__keys_choices_make:nn #1 \}
\cs_new_protected:cpn { \_keys_props_root_str .choices:xn } #1
  \{ \exp_args:Nx \__keys_choices_make:nn #1 \}
\end{verbatim}

(End definition for \texttt{\_choices:nn}. This function is documented on page 223.)

Creating code is simply a case of passing through to the underlying \texttt{\_cmd_set:n} function.

\begin{verbatim}
\cs_new_protected:cpn { \_keys_props_root_str .code:n } #1
  \{ \__keys_cmd_set:nn \l_keys_path_str {#1} \}
\end{verbatim}

(End definition for \texttt{\_code:n}. This function is documented on page 224.)

Creating code is simply a case of passing through to the underlying \texttt{\_cmd_set:n} function.

\begin{verbatim}
\cs_new_protected:cpn { \_keys_props_root_str .clist_set:N } #1
  \{ \__keys_variable_set:NnnN #1 { clist } { } n \}
\cs_new_protected:cpn { \_keys_props_root_str .clist_set:c } #1
  \{ \__keys_variable_set:cnnN {#1} { clist } { } n \}
\cs_new_protected:cpn { \_keys_props_root_str .clist_gset:N } #1
  \{ \__keys_variable_set:NnnN #1 { clist } { g } n \}
\cs_new_protected:cpn { \_keys_props_root_str .clist_gset:c } #1
  \{ \__keys_variable_set:cnnN {#1} { clist } { g } n \}
\end{verbatim}

(End definition for \texttt{\_clist_set:N} and \texttt{\_clist_gset:N}. These functions are documented on page 223.)

\begin{verbatim}
\cs_new_protected:cpn { \_keys_props_root_str .cs_set:Np } #1
  \{ \__keys_cs_set:NNpn \cs_set:Npn #1 { } \}
\cs_new_protected:cpn { \_keys_props_root_str .cs_set:cp } #1
  \{ \__keys_cs_set:Ncpn \cs_set:Npn #1 { } \}
\cs_new_protected:cpn { \_keys_props_root_str .cs_set_protected:Np } #1
  \{ \__keys_cs_set:NNpn \cs_set_protected:Npn #1 { } \}
\cs_new_protected:cpn { \_keys_props_root_str .cs_set_protected:cp } #1
  \{ \__keys_cs_set:NNpn \cs_set_protected:Ncpn #1 { } \}
\cs_new_protected:cpn { \_keys_props_root_str .cs_gset:Np } #1
  \{ \__keys_cs_set:NNpn \cs_gset:Npn #1 { } \}
\cs_new_protected:cpn { \_keys_props_root_str .cs_gset:cp } #1
  \{ \__keys_cs_set:NNpn \cs_gset:Ncpn #1 { } \}
\end{verbatim}
(End definition for .cs_set:Np and others. These functions are documented on page 224.)
\begin{verbatim}
$.default:n$
Expansion is left to the internal functions.
\begin{verbatim}
$.default:V$
$.default:o$
$.default:x$
\end{verbatim}
(End definition for .default:n. This function is documented on page 224.)
\begin{verbatim}
$.dim_set:N$
$.dim_set:c$
$.dim_gset:N$
$.dim_gset:c$
\end{verbatim}
Setting a variable is very easy: just pass the data along.
\begin{verbatim}
$.fp_set:N$
$.fp_set:c$
$.fp_gset:N$
$.fp_gset:c$
\end{verbatim}
Setting a variable is very easy: just pass the data along.
\begin{verbatim}
$.groups:n$
A single property to create groups of keys.
$.inherit:n$
Nothing complex: only one variant at the moment!
\end{verbatim}
(End definition for .inherit:n. This function is documented on page 225.)
The standard hand-off approach.

Setting a variable is very easy: just pass the data along.

Making a meta is handled internally.

Meta with path: potentially lots of variants, but for the moment not so many defined.

The same idea as .choice: and .choices:nn, but where more than one choice is allowed.
Setting a variable is very easy: just pass the data along.

.\cs_new_protected:cpn { \c__keys_props_root_str .muskip_set:N } #1
\cs_new_protected:cpn { \c__keys_props_root_str .muskip_set:c } #1
\cs_new_protected:cpn { \c__keys_props_root_str .muskip_gset:N } #1
\cs_new_protected:cpn { \c__keys_props_root_str .muskip_gset:c } #1

(End definition for .muskip_set:N and .muskip_gset:N. These functions are documented on page 226.)

.\cs_new_protected:cpn { \c__keys_props_root_str .prop_put:N } #1
\cs_new_protected:cpn { \c__keys_props_root_str .prop_put:c } #1
\cs_new_protected:cpn { \c__keys_props_root_str .prop_gput:N } #1
\cs_new_protected:cpn { \c__keys_props_root_str .prop_gput:c } #1

(End definition for .prop_put:N and .prop_gput:N. These functions are documented on page 226.)

.\cs_new_protected:cpn { \c__keys_props_root_str .skip_set:N } #1
\cs_new_protected:cpn { \c__keys_props_root_str .skip_set:c } #1
\cs_new_protected:cpn { \c__keys_props_root_str .skip_gset:N } #1
\cs_new_protected:cpn { \c__keys_props_root_str .skip_gset:c } #1

(End definition for .skip_set:N and .skip_gset:N. These functions are documented on page 226.)

.\cs_new_protected:cpn { \c__keys_props_root_str .str_set:N } #1
\cs_new_protected:cpn { \c__keys_props_root_str .str_set:c } #1
\cs_new_protected:cpn { \c__keys_props_root_str .str_set_x:N } #1
\cs_new_protected:cpn { \c__keys_props_root_str .str_set_x:c } #1
\cs_new_protected:cpn { \c__keys_props_root_str .str_gset:N } #1
\cs_new_protected:cpn { \c__keys_props_root_str .str_gset:c } #1
\cs_new_protected:cpn { \c__keys_props_root_str .str_gset_x:N } #1
\cs_new_protected:cpn { \c__keys_props_root_str .str_gset_x:c } #1

(End definition for .str_set:N and .skip_gset:N. These functions are documented on page 226.)
Setting a variable is very easy: just pass the data along.

(End definition for .str_set:N and others. These functions are documented on page 226.)

.undefine: Another simple wrapper.

(End definition for .undefine:. This function is documented on page 227.)

These are very similar, so both call the same function.

(End definition for .value_forbidden:n and .value_required:n. These functions are documented on page 227.)
A simple wrapper allowing for nesting.

Setting known keys simply means setting the appropriate flag, then running the standard code. To allow for nested setting, any existing value of `\l__keys_unused_clist` is saved on the stack and reset afterwards. Note that for speed/simplicity reasons we use a tl operation to set the clist here!
The idea of setting keys in a selective manner again uses flags wrapped around the basic code. The comments on `\keys_set_known:nnn` also apply here. We have a bit more shuffling to do to keep everything nestable.

```
\keys_set_filter:nnn \keys_set_filter:nnV \keys_set_filter:nnv \keys_set_filter:nnn
\keys_set_filter:nnv \keys_set_filter:nnV \keys_set_filter:nnv \keys_set_filter:nnn
\keys_set_filter:nnn \keys_set_groups:nnn \keys_set_groups:nnV \keys_set_groups:nnv
\keys_set_groups:nnv \keys_set_groups:nnn \keys_set_groups:nnn
\keys_set_groups:nnn \keys_set_selective:nnn \keys_set_selective:nnn
```

(End definition for `\keys_set_known:nnn` and others. These functions are documented on page 231.)

942
\keys_precompile:nnN A simple wrapper.
A shared system once again. First, set the current path and add a default if needed. There are then checks to see if the a value is required or forbidden. If everything passes, move on to execute the code.

The key path here can be fully defined, after which there is a search for the key and module names: the user may have passed them with part of what is actually the module (for our purposes) in the key name. As that happens on a per-key basis, we use the stack approach to restore the module name without a group.

This function uses \cs_set_nopar:Npx internally for performance reasons, the argument #1 is already a string in every usage, so turning it into a string again seems unnecessary.
If selective setting is active, there are a number of possible sub-cases to consider. The key name may not be known at all or if it is, it may not have any groups assigned. There is then the question of whether the selection is opt-in or opt-out.

In the case where selective setting requires a comparison of the list of groups which apply to a key with the list of those which have been set active. That requires two mappings, and again a different outcome depending on whether opt-in or opt-out is set. We cannot replace the clist mapping by \clist_if_in:NnTF because catcodes may not be the same; they cannot be normalized easily in the clist because of the remote possibility that some items need braces if they involve commas or leading/trailing spaces.
\__keys_value_or_default:n
If a value is given, return it as #1, otherwise send a default if available.

\__keys_default_inherit:
\cs_new_protected:Npn \__keys_value_or_default:n #1
\bool_if:NTF \l__keys_no_value_bool
{ \cs_if_exist:cTF { \c__keys_default_root_str \l_keys_path_str } }\tl_set_eq:Nc \l_keys_value_tl \c__keys_default_root_str \l_keys_path_str
\tl_clear:N \l_keys_value_tl \cs_if_exist:cT { \c__keys_inherit_root_str \__keys_parent:o \l_keys_path_str } \__keys_default_inherit:
\tl_set:Nn \l_keys_value_tl {\c__keys_default_root_str \#1} \l_keys_key_str
\cs_new_protected:Npn \__keys_default_inherit:
\clist_map_inline:cn { \c__keys_inherit_root_str \__keys_parent:o \l_keys_path_str } \cs_if_exist:cT { \c__keys_default_root_str \l_keys_key_str }\tl_set_eq:Nc \l_keys_value_tl \c__keys_default_root_str \#1 \l_keys_key_str
\l_keys_value_tl \c__keys_default_root_str \#1 \l_keys_key_str
Actually executing a key is done in two parts. First, look for the key itself, then look for the unknown key with the same path. If both of these fail, complain. What exactly happens if a key is unknown depends on whether unknown keys are being skipped or if an error should be raised.

To deal with the case where there is no hit, we leave \_keys\_execute\_unknown: in the input stream and clean it up using the break function: that avoids needing a boolean.

947
A key’s code is in the control sequence with csname \c__keys_code_root_str #1. We expand it once to get the replacement text (with argument #2) and call \use:n with this replacement as its argument. This ensures that any undefined control sequence error in the key’s code will lead to an error message of the form \langle argument\rangle...\langle control sequence\rangle in which one can read the (undefined) \langle control sequence\rangle in full, rather than an error message that starts with the potentially very long key name, which would make the (undefined) \langle control sequence\rangle be truncated or sometimes completely hidden. See https://github.com/latex3/latex2e/issues/351.

\cs_new:Npn \__keys_execute:nn #1#2
{ \__keys_execute:no {#1} { \prg_do_nothing: #2 } }
\cs_new:Npn \__keys_execute:no #1#2
{ \exp_args:NNo \exp_args:No \use:n
{ \cs:w \c__keys_code_root_str #1 \exp_after:wN \cs_end:
\exp_after:wN {#2} }
}

When there is no relative path, things here are easy: just save the key name and value. When we are working with a relative path, first we need to turn it into a string: that can’t happen earlier as we need to store \q__keys_no_value. Then, use a standard delimited approach to fish out the partial path.

\cs_new_protected:Npn \__keys_store_unused:
{ \__keys_quark_if_no_value:NTF \l__keys_relative_tl
 { \clist_put_right:Nx \l__keys_unused_clist
   { \l_keys_key_str
     \bool_if:NF \l__keys_no_value_bool
     { = { \exp_not:o \l_keys_value_tl } } } }
 { \tl_if_empty:NTF \l__keys_relative_tl
   { \clist_put_right:Nx \l__keys_unused_clist
     { \l_keys_path_str
       \bool_if:NF \l__keys_no_value_bool
       { = { \exp_not:o \l_keys_value_tl } } } }
   { \__keys_store_unused_aux: }
 }
\cs_new_protected:Npn \__keys_store_unused_aux:
{ }
\__keys_choice_find:n
\__keys_choice_find:nn
\__keys_multichoice_find:n

Executing a choice has two parts. First, try the choice given, then if that fails call the
unknown key. That always exists, as it is created when a choice is first made. So there
is no need for any escape code. For multiple choices, the same code ends up used in a
mapping.

\cs_new:Npn \__keys_choice_find:n #1
\{  
\str_if_empty:NTF \l__keys_inherit_str  
\{ \__keys_choice_find:nn \l_keys_path_str #1 \}  
\}
\cs_new:Npn \__keys_choice_find:nn #1#2
\{  
\cs_if_exist:cTF { \c__keys_code_root_str #1 / \__keys_trim_spaces:n {#2} }  
\{ \__keys_execute:nn { #1 / \__keys_trim_spaces:n {#2} } #2 \}  
\{ \__keys_execute:nn { #1 / unknown } #2 \}  
\}
\cs_new:Npn \__keys_multichoice_find:n #1
\{  
\cs_if_exist:cTF { \c__keys_code_root_str #1 / \__keys_trim_spaces:n {#2} }  
\{ \__keys_execute:nn { #1 / \__keys_trim_spaces:n {#2} } (#2) \}  
\{ \__keys_execute:nn { #1 / unknown } (#2) \}  
\}
(End definition for \__keys_execute: and others.)
\__keys_parent:o
\__keys_parent_auxi:w
\__keys_parent_auxii:w
\__keys_parent_auxiii:n
\__keys_parent_auxiv:w

\__keys_trim_spaces:n
\__keys_trim_spaces_auxi:w
\__keys_trim_spaces_auxii:w
\__keys_trim_spaces_auxiii:w

Used to strip off the ending part of the key path after the last /.

Space stripping has to allow for the fact that the key here might have several parts, and
spaces need to be stripped from each part. Since the key name is turned into a string
groups can’t be stripped accidentally and the precautions of \tl_trim_spaces:n aren’t
necessary, in this case it is much faster to just directly strip spaces around /.
\keys_if_exist_p:nn
\keys_if_exist:nnTF
A utility for others to see if a key exists.

\keys_if_choice_exist_p:nn
\keys_if_choice_exist:nnTF
Just an alternative view on \keys_if_exist:nnTF.

\keys_show:nn
\keys_log:nn
\keys_show:Nnn
\keys_show:Nn
\keys_show:n
\keys_show:w
\keys_show:Nw
To show a key, show its code using a message.

(End definition for \keys_trim_spaces:n and others.)
63.8 Messages

For when there is a need to complain.

\msg_new:nnnn { keys } { bad-relative-key-path }
\{ The-key-`#1'-is-not-inside-the-`#2'-path. \}
\msg_new:nnnn { keys } { boolean-values-only }
\{ Key-`#1'-accepts-boolean-values-only. \}
\msg_new:nnnn { keys } { boolean-values-only }
\{ The-key-`#1'-only-accepts-the-values-`true'-and-`false'. \}

(End definition for \keys_show:nn and others. These functions are documented on page 233.)
Key-"#1"-accepts-only-a-fixed-set-of-choices.

The-key-"#1"-only-accepts-predefined-values,-
and-"#2"-is-not-one-of-these.

The-key-"#1"-is-unknown-and-is-being-ignored.

The-module-"#2"-does-not-have-a-key-called-"#1".\%
Check-that-you-have-spelled-the-key-name-correctly.

The-key-"#1"-cannot-be-defined-as-a-choice-as-the-parent-key-"#2"-is-
itself-a-choice.

The-key-"#1"-should-be-given-without-a-value.\%
The-value-"#2"-was-present:-the-key-will-be-ignored.

The-key-"#1"-requires-a-value. \%
No-value-was-present:-the-key-will-be-ignored.

The-key-"#1"-is-undefined. \%
has-the-properties: \#2 .

msg_new:nnnn { keys } { show-key }

\tl_if_empty:nTF \{#2} 
{ is-undefined. } 
{ has-the-properties: \#2 . } 

\prop_gput:Nnn \g_msg_module_name_prop { keys } { \LaTeX3 } 
\prop_gput:Nnn \g_msg_module_type_prop { keys } { } 

Chapter 64

l3intarray implementation

There are two implementations for this module: One \fontdimen based one for more traditional \TeX\ engines and a Lua based one for engines with Lua support. Both versions do not allow negative array sizes.

64.1 Lua implementation

First, let’s look at the Lua variant:

We select the Lua version if the Lua helpers were defined. This can be detected by the presence of \_\_intarray_gset_count:W.

64.1.1 Allocating arrays

Used to differentiate intarrays in Lua and to record an invalid index. Used as marker for intarrays in Lua. Followed by an unbraced number identifying the array and a single space. This format is used to make it easy to scan from Lua.
\intarray_new:Nn \intarray_new:cn \__intarray_new:N

Declare #1 as a tokenlist with the scanmark and a unique number. Pass the array’s size to the Lua helper. Every \texttt{intarray} must be global; it’s enough to run this check in \intarray_new:Nn.\intarray_new:Nn\__intarray_new:N

\cs_new_protected:Npn \__intarray_new:N #1
\__kernel_chk_if_free_cs:N #1 \int_gincr:N \g__intarray_table_int \cs_gset_nopar:Npx #1 { \__intarray:w \int_use:N \g__intarray_table_int \c_space_tl }
\cs_new_protected:Npn \intarray_new:Nn #1#2
\__intarray_new:N #1 \__intarray_gset_count:Nw #1 \int_eval:n {#2} \scan_stop:
\int_compare:nNnT { \intarray_count:N #1 } < 0
\msg_error:nnx { kernel } { negative-array-size } { \intarray_count:N #1 }
\cs_generate_variant:Nn \intarray_new:Nn { c }

\__intarray_table

Internal helper to scan an intarray token, extract the associated Lua table and return an error if the input is invalid.
\__intarray_table do
local tables = get_luadata and get_luadata’\__intarray’ or {[0] = {}}
function __intarray_table()
local t = scan_token()
if t ~= intarray_marker then
put_next(t)
tex.error’LaTeX Error: intarray expected’
\end{verbatim}

(End definition for \__intarray:w.)
\intarray_new:Nn \intarray_new:cn

Before we get to the first command implemented in Lua, we first need some definitions. Since \texttt{token.create} only works correctly if \texttt{TeX} has seen the tokens before, we first run a short \texttt{TeX} sequence to ensure that all relevant control sequences are known.
\begin{verbatim}
local scan_token = token.scan_token
local put_next = token.put_next
local intarray_marker = token_create_safe’\__intarray:w’
local use_none = token_create_safe’use_none:n’
local use_i = token_create_safe’use:n’
local expand_after_scan_stop = {token_create_safe’exp_after:wN’,
token_create_safe’scan_stop:’}
local comma = token_create(string.byte’,’)

local __intarray_table do
local tables = get_luadata and get_luadata’\__intarray’ or {[0] = {}}
function __intarray_table()
local t = scan_token()
if t ~= intarray_marker then
put_next(t)
tex.error’LaTeX Error: intarray expected’
\end{verbatim}

(End definition for \intarray_new:Nn and \__intarray_new:N. This function is documented on page 236.)

Before we get to the first command implemented in Lua, we first need some definitions. Since \texttt{token.create} only works correctly if \texttt{TeX} has seen the tokens before, we first run a short \texttt{TeX} sequence to ensure that all relevant control sequences are known.
local i = scan_int()
local current_table = tables[i]
if current_table then return current_table end
current_table = {}  
tagged{22035}
tables[i] = current_table
return current_table
end

return tables[0]
end

Since in \LaTeX{} this is loaded in the format, we want to preserve any intarrays which are
created while format building for the actual run.

To do this, we use the \texttt{register\_luadata} mechanism from \texttt{l3luatex}: Directly before
the format get dumped, the following function gets invoked and serializes all existing
tables into a string. This string gets compiled and dumped into the format and is made
available at the beginning of regular runs as \texttt{get\_luadata'@@'}.\ntagged{22037}

if register\_luadata then
  register\_luadata('_\texttt{\_intarray}', function()

  local \texttt{t} = "\{0\}={}{{},"

  for \texttt{i}=1, \#tables do
    \texttt{t} = string.format("\%s\%s", \texttt{t}, table.concat(tables[i], ','))
  end

  return \texttt{t} .. ")"
end
end

\langle \texttt{lua} \rangle

⟨∗tex⟩
\cs_generate_variant:Nn \intarray\_count:N { c }
\langle tex \rangle

\langle \texttt{lua} \rangle

\langle tex \rangle

(End definition for \_\texttt{\_intarray\_table}.)

\texttt{\_\intarray\_count:N}

\texttt{\_\intarray\_count:c}
\texttt{\_\intarray\_gset\_count:Nw}

Set and get the size of an array. “Setting the size” means in this context that we add
zeros until we reach the desired size.\ntagged{22047}

local \texttt{sprint} = tex\_sprint

luacmd('\_\texttt{\_intarray\_gset\_count:Nw}', function()
  local \texttt{t} = \_\texttt{\_intarray\_table}()
  local \texttt{n} = scan\_int()

  for \texttt{i}=#\texttt{t}+1, \texttt{n} do \texttt{t}[\texttt{i}] = 0 end

  end, 'protected', 'global')

luacmd('\texttt{\intarray\_count:N}', function()
  \texttt{sprint}(-2, \_\texttt{\_intarray\_table}())

  return '\(\texttt{lua}\)'\ntagged{22059}

\langle tex \rangle
\cs_generate_variant:Nn \intarray\_count:N { c }
\langle\texttt{tex}\rangle

(End definition for \texttt{\_\intarray\_count:N} and \\texttt{\_\intarray\_gset\_count:Nw}. This function is documented
on page 236.)
64.1.2 Array items

\__intarray_gset:wF
\__intarray_gset:w

The setter provided by Lua. The argument order somewhat emulates the \fontdimen:

First the array index, followed by the intarray and then the new value. This has been
chosen over a more conventional order to provide a delimiter for the numbers.

\lua
\luacmd{\__intarray_gset:wF', function() \local i = scan_int() \local t = \__intarray_table() \if t[i] then \t[i] = scan_int() \put_next(use_none) \else \tex.count.l__intarray_bad_index_int = i \scan_int() \put_next(use_i) \end \end, 'protected', 'global')\}
\luacmd{\__intarray_gset:w', function() \local i = scan_int() \local t = \__intarray_table() \t[i] = scan_int() \end, 'protected', 'global')\}

(End definition for \__intarray_gset:wF and \__intarray_gset:w.)

\intarray_gset:Nnn
\intarray_gset:cnn
\__kernel_intarray_gset:Nnn

The \__kernel_intarray_gset:Nnn function does not use \int_eval:n, namely its
arguments must be suitable for \int_value:w. The user version checks the position and
value are within bounds.

\tex
\cs_new_protected:Npn \__kernel_intarray_gset:Nnn #1#2#3 \{ \__intarray_gset:w #2 #1 #3 \scan_stop: \}
\cs_new_protected:Npn \intarray_gset:Nnn #1#2#3 \{
\__intarray_gset:wF \int_eval:n {#2} #1 \int_eval:n{#3} \}
\msg_error:nnxxx { kernel } { out-of-bounds } { \token_to_str:N #1 } { \int_use:N \l__intarray_bad_index_int } \{ \intarray_\}
\cs_generate_variant:Nn \intarray_gset:Nnn { c }\}
\}(End definition for \intarray_gset:Nnn and \__kernel_intarray_gset:Nnn. This function is docu-
mented on page 236.)

\intarray_gzero:N
\intarray_gzero:c

Set the appropriate array entry to zero. No bound checking needed.

\lua
\luacmd{intarray_gzero:N', function() \local t = \__intarray_table() \for i=1, #t do \t[i] = 0 \end\}

957
End definition for \intarray_gzero:N. This function is documented on page 237.

Importantly, \intarray_item:Nn only evaluates its argument once.
\intarray_item:Nn #1 { \int_rand:n { \intarray_count:N #1 + 1 } } \intarray_rand_item:N \cs_generate_variant:Nn \intarray_rand_item:N { c }

(End definition for \intarray_rand_item:N. This function is documented on page 237.)

64.1.3 Working with contents of integer arrays

We use the \kernel_intarray_gset:Nnn which does not do bounds checking and instead automatically resizes the array. This is not implemented in Lua to ensure that the clist parsing is consistent with the clist module.

\cs_new_protected:Npn \intarray_const_from_clist:Nn #1#2
\__intarray_new:N #1
\int_zero:N \l__intarray_loop_int
\clist_map_inline:nn {#2}
\int_incr:N \l__intarray_loop_int
\__kernel_intarray_gset:Nnn #1 \l__intarray_loop_int { \int_eval:n {##1} } \cs_generate_variant:Nn \intarray_const_from_clist:Nn { c }

(End definition for \intarray_const_from_clist:Nn. This function is documented on page 237.)

\intarray_to_clist:N
\intarray_to_clist:c
\__intarray_to_clist:Nn
\__intarray_to_clist:w

The \kernel_intarray_to_clist:N auxiliary allows to choose the delimiter and is also used in \intarray_show:N. Here we just pass the information to Lua and let table.concat do the actual work. We discard the category codes of the passed delimiter but this is not an issue since the delimiter is always just a comma or a comma and a space. In both cases \sprint(2, ...) provides the right catcodes.

\cs_new:Npn \intarray_to_clist:N #1 { \__intarray_to_clist:w #1 { , } }
\cs_generate_variant:Nn \intarray_to_clist:N { c }
\lucmd{\__intarray_to_clist:w', function()}
\lucmd{\__intarray_to_clist:Nn', function()}
\lucmd{\__intarray_to_clist:Nc', function()}
\lucmd{\__intarray_to_clist:Nw', function()}
\lucmd{'global'}

(End definition for \intarray_to_clist:N, \__intarray_to_clist:Nn, and \__intarray_to_clist:w. This function is documented on page 302.)

\kernel_intarray_range_to_clist:Nnn
\__intarray_range_to_clist:w

Loop through part of the array.

\cs_new:Npn \__kernel_intarray_range_to_clist:Nnn #1#2#3
\__intarray_range_to_clist:w #1
\int_eval:n {#2} ~ \int_eval:n {#3} ~
\cs_generate_variant:Nn \intarray_range_to_clist:Nnn { c }
\lucmd{\__intarray_range_to_clist:Nnn', function()}
\lucmd{\__intarray_range_to_clist:Nn', function()}
\lucmd{\__intarray_range_to_clist:Nn', function()}
\lucmd{\__intarray_range_to_clist:Nw', function()}
\lucmd{'global'}

\local t = \__intarray_table()
local from = scan_int()
local to = scan_int()
sprint(-2, concat(t, ',', from, to))
end, 'global')

(End definition for \_\_kernel_intarray_range_to_clist:Nnn and \_\_intarray_range_to_clist:w)

Loop through part of the array. We allow additional commas at the end.

\cs_new_protected:Npn \_\_kernel_intarray_gset_range_from_clist:Nnn #1#2#3
  \_\_intarray_gset_range:w \int_eval:w #2 #1 #3 , , \scan_stop:
\end{verbatim}

(End definition for \_\_kernel_intarray_gset_range_from_clist:Nnn and \_\_intarray_gset_range:nNw.)

In order to allow some code sharing later we provide the \_\_intarray_gset_overflow_test:nw name here. It doesn’t actually test anything since the Lua implementation accepts all integers which could be tested with \text_ifabsnum:D.

\cs_new_protected:Npn \_\_intarray_gset_overflow_test:nw #1
\end{verbatim}

(End definition for \_\_intarray_gset_overflow_test:nw.)

64.2 Font dimension based implementation

Go to the false branch of the conditional above.
64.2.1 Allocating arrays

\__intarray_entry:w \__intarray_count:w

We use these primitives quite a lot in this module.

\cs_new_eq:NN \__intarray_entry:w \tex_fontdimen:D
\cs_new_eq:NN \__intarray_count:w \tex_hyphenchar:D

(End definition for \__intarray_entry:w and \__intarray_count:w.)

\c__intarray_sp_dim

Used to convert integers to dimensions fast.

\dim_const:Nn \c__intarray_sp_dim { 1 sp }

(End definition for \c__intarray_sp_dim.)

\g__intarray_font_int

Used to assign one font per array.

\int_new:N \g__intarray_font_int

(End definition for \g__intarray_font_int.)

\intarray_new:Nn \intarray_new:cn \__intarray_new:N

Declare #1 to be a font (arbitrarily \textsf{cmr10} at a never-used size). Store the array’s size as the \hyphenchar of that font and make sure enough \fontdimen are allocated, by setting the last one. Then clear any \fontdimen that \textsf{cmr10} starts with. It seems \LaTeX{}’s \textsf{cmr10} has an extra \fontdimen parameter number 8 compared to other engines (for a math font we would replace 8 by 22 or some such). Every \texttt{intarray} must be global; it’s enough to run this check in \texttt{intarray_new:Nn}.

\cs_new_protected:Npn \__intarray_new:N #1
\__kernel_chk_if_free_cs:N #1
\int_gincr:N \g__intarray_font_int
\tex_global:D \tex_font:D #1
= \textsf{cmr10-at-}\ texttt{\g__intarray_font_int} \c__intarray_sp_dim \scan_stop:
\int_step_inline:nn { 8 } { \__kernel_intarray_gset:Nnn #1 {##1} \c_zero_int }
\}
\cs_new_protected:Npn \intarray_new:Nn #1#2
\__intarray_new:N #1\__intarray_count:w #1 = \int_eval:n {#2} \scan_stop:
\int_compare:nNnT { \intarray_count:N #1 } < 0
\{ \msg_error:nnx { kernel } { negative-array-size } { \intarray_count:N #1 } \}
\int_compare:nNnT { \intarray_count:N #1 } > 0
\{ \__kernel_intarray_gset:Nnn #1 { \intarray_count:N #1 } { 0 } \}
\}
\cs_generate_variant:Nn \intarray_new:Nn { c }

(End definition for \intarray_new:Nn and \__intarray_new:N. This function is documented on page 236.)

\intarray_count:N \intarray_count:c

Size of an array.

\cs_new:Npn \intarray_count:N #1 { \int_value:w \__intarray_count:w #1 }
\cs_generate_variant:Nn \intarray_count:N { c }

(End definition for \intarray_count:N. This function is documented on page 236.)
64.2.2 Array items

\__intarray_signed_max_dim:n

Used when an item to be stored is larger than \c\_max\_dim in absolute value; it is replaced by ±\c\_max\_dim.

\cs_new:Npn \__intarray_signed_max_dim:n #1
\{ \int_value:w \int_compare:nNnT \{#1\} < 0 \{ - \} \c\_max\_dim \}

(End definition for \__intarray_signed_max_dim:n.)

\__intarray_bounds:NNnTF
\__intarray_bounds_error:NNnw

The functions \intarray_gset:Nnn and \intarray_item:Nn share bounds checking. The T branch is used if #3 is within bounds of the array #2.

\cs_new:Npn \__intarray_bounds:NNnTF #1#2#3
\{\if_int_compare:w 1 > #3 \exp_stop_f: \__intarray_bounds_error:NNnw #1 #2 {#3} \else: \if_int_compare:w #3 > \intarray_count:N #2 \exp_stop_f: \__intarray_bounds_error:NNnw #1 #2 {#3} \fi: \fi: \use_i:nn \}
\cs_new:Npn \__intarray_bounds_error:NNnw #1#2#3#4 \use_i:nn #5#6
\{ #4 \#1 \{ \text{kernel} \} \{ \text{out-of-bounds} \} \{ \text{token_to_str:N} \} \{ \text{intarray_count:N} \} \{ \text{intarray_count:N} \} \#6 \}

(End definition for \__intarray_bounds:NNnTF and \__intarray_bounds_error:NNnw.)

\intarray_gset:Nnn
\intarray_gset:cnn
\__kernel_intarray_gset:Nnn
\__intarray_gset:Nnn
\__intarray_gset_overflow:Nnn

Set the appropriate \fontdimen. The \__kernel_intarray_gset:Nnn function does not use \int\_eval:n, namely its arguments must be suitable for \int\_value:w. The user version checks the position and value are within bounds.

\cs_new_protected:Npn \__kernel_intarray_gset:Nnn #1#2#3
\{ \__intarray_entry:w #2 #1 #3 \c\__intarray_sp_dim \}
\cs_new_protected:Npn \intarray_gset:Nnn #1#2#3
\{ \exp_after:wN \__intarray_gset:Nww \exp_after:wN #1 \int_value:w \int_eval:n {#2} \exp_after:wN ; \int_value:w \int_eval:n {#3} ; \}
\cs_generate_variant:Nn \intarray_gset:Nnn { c }
\cs_new_protected:Npn \__intarray_gset:Nww #1#2 ; #3 ;
\{ \__intarray_bounds:NNnTF \msg_error:nnnn \#1 \{ \#2 \} \{ \#3 \} \__kernel_intarray_gset:Nnn #1 #2 #3 \}
\}

962
\cs_if_exist:NTF \tex_ifabsnum:D
  {
  \cs_new_protected:Npn \_\intarray_gset_overflow_test:nw #1
    {\tex_ifabsnum:D #1 > \c_max_dim
      \exp_after:wN \_\intarray_gset_overflow:NNnn
      \fi:}
  }
\cs_new_protected:Npn \_\intarray_gset_overflow_test:nw #1
  {\if_int_compare:w \int_abs:n {#1} > \c_max_dim
    \exp_after:wN \_\intarray_gset_overflow:NNnn
    \fi:}
\cs_new_protected:Npn \_\intarray_gset_overflow:NNnn #1#2#3#4
  {\msg_error:nnxxxx { kernel } { overflow }{ \token_to_str:N #2 } {#3} {#4} }{ \__intarray_signed_max_dim:n {#4} } #1 #2 {#3} { \__intarray_signed_max_dim:n {#4} }
\intarray_gzero:N
\intarray_gzero:c
Set the appropriate \fontdimen to zero. No bound checking needed. The \prg_replicate:nn possibly uses quite a lot of memory, but this is somewhat comparable to the size of the array, and it is much faster than an \int_step_inline:nn loop.
\cs_new_protected:Npn \intarray_gzero:N #1
  {\int_zero:N \l__intarray_loop_int
   \prg_replicate:nn { \intarray_count:N #1 } { \int_incr:N \l__intarray_loop_int
     \__intarray_entry:w \l__intarray_loop_int #1 \c_zero_dim } }
\cs_generate_variant:Nn \intarray_gzero:N { c }
\intarray_item:Nn
\intarray_item:cn
\__kernel_intarray_item:Nn
\__intarray_item:Nn
Get the appropriate \fontdimen and perform bound checks. The \_\kernel_\intarray_item:Nn function omits bound checks and omits \int_eval:n, namely its argument must be a \TeX integer suitable for \int_value:w.
\cs_new:Npn \_\kernel_intarray_item:Nn #1#2
  {\int_value:w \_\intarray_entry:w #2 #1 }\cs_new:Npn \intarray_item:Nn #1#2
  {\exp_after:wN \_\intarray_item:Nw \exp_after:wN #1 \int_value:w \int_eval:n {#2} ; }
\cs_generate_variant:Nn \intarray_item:Nn { c }

(End definition for \intarray_gset:Nnn and others. This function is documented on page 236.)
\intarray_gzero:N
\intarray_gzero:c
(End definition for \intarray_gzero:N. This function is documented on page 237.)
\cs_new:Npn \__intarray_item:Nw #1#2 ;
{
  \__intarray_bounds:NNnTF \msg_expandable_error:nnff \msg_error:ff #1 \{#2\}
  \__kernel_intarray_item:Nn #1 \{#2\}
  \l_zero:N \l__intarray_loop_int
}
\cs_new:Npn \__intarray_bounds:NNnTF \msg_expandable_error:nnff #1 \{#2\}
{
  \__kernel_intarray_item:Nn #1 \{#2\}
  \l_zero:N \l__intarray_loop_int
}
\begin{Verbatim}
(End definition for \intarray_item:Nn, \__kernel_intarray_item:Nn, and \__intarray_item:Nn. This function is documented on page 237.)
\end{Verbatim}

\intarray_rand_item:N
\intarray_rand_item:c
Importantly, \intarray_item:Nn only evaluates its argument once.
\cs_new:Npn \intarray_rand_item:N #1
{
  \intarray_item:Nn #1 \{ \int_rand:n \{ \intarray_count:N #1 \} \}
}
\cs_generate_variant:Nn \intarray_rand_item:N { c }
\begin{Verbatim}
(End definition for \intarray_rand_item:N. This function is documented on page 237.)
\end{Verbatim}

64.2.3 Working with contents of integer arrays

\intarray_const_from_clist:N
\intarray_const_from_clist:c
\__intarray_const_from_clist:nN
Similar to \intarray_new:Nn (which we don’t use because when debugging is enabled function checks the variable name starts with g_). We make use of the fact that T\LaTeX allows allocation of successive \texttt{\fontdimen} as long as no other font has been declared: no need to count the comma list items first. We need the code in \intarray_gset:Nnn that checks the item value is not too big, namely \__intarray_gset_overflow_test:nw, but not the code that checks bounds. At the end, set the size of the intarray.
\cs_new_protected:Npn \intarray_const_from_clist:Nn #1#2
{
  \__intarray_new:N #1
  \l_zero:N \l__intarray_loop_int
  \clist_map_inline:nn {#2}
  { \exp_args:Nf \__intarray_const_from_clist:nN { \int_eval:n {##1} } #1 }
  \__intarray_count:w #1 \l__intarray_loop_int
}
\cs_generate_variant:Nn \intarray_const_from_clist:Nn { c }
\cs_new_protected:Npn \__intarray_const_from_clist:nN #1#2
{
  \int_incr:N \l__intarray_loop_int
  \__intarray_gset_overflow_test:nw {#1}
  \__kernel_intarray_gset:Nnn #2 \l__intarray_loop_int \{#1\}
}
\begin{Verbatim}
(End definition for \intarray_const_from_clist:N and \__intarray_const_from_clist:nN. This function is documented on page 237.)
\end{Verbatim}

\intarray_to_clist:N
\intarray_to_clist:c
\__intarray_to_clist:Nn
\__intarray_to_clist:w
Loop through the array, putting a comma before each item. Remove the leading comma with \texttt{f}-expansion. We also use the auxiliary in \intarray_show:N with argument comma, space.
\cs_new:Npn \intarray_to_clist:Nn #1
{
  \__intarray_to_clist:Nn #1 \{ , \}
}
\cs_generate_variant:Nn \intarray_to_clist:Nn { c }
\cs_new:Npn \__intarray_to_clist:Nn #1
{
  \int_compare:nNnF \intarray_count:N #1 \l_zero_int
  \{ \int_incr:N \l__intarray_loop_int
  \__intarray_to_clist:w #1 \l__intarray_loop_int \{#1\} \}
}
\begin{Verbatim}
(End definition for \intarray_to_clist:Nn and \__intarray_to_clist:Nn. This function is documented on page 237.)
\end{Verbatim}
\cs_new:Npn \intarray_to_clist:w #1
{
  \exp_last_unbraced:Nf \use_none:n
}\end{Verbatim}

964
\__kernel_intarray_range_to_clist:Nnn
\__intarray_range_to_clist:ww
Loop through part of the array.
\cs_new:Npn \__kernel_intarray_range_to_clist:Nnn #1#2#3
{ \exp_last_unbraced:Nf \use_none:n
  \exp_after:wN \__intarray_range_to_clist:ww
  \int_value:w \int_eval:w #2 \exp_after:wN ;
  \int_value:w \int_eval:w #3 ;
  #1 \prg_break_point:
}
\cs_new:Npn \__intarray_range_to_clist:ww #1 ; #2 ; #3
{ \if_int_compare:w #1 > #2 \exp_stop_f:
  \prg_break:n
  \fi:
  \__kernel_intarray_item:Nn #3 {#1}
  \exp_after:wN \__intarray_range_to_clist:Nnn \int_value:w \int_eval:w #1 + \c_one_int ; #2 ; #3
}
(End definition for \intarray_to_clist:N, \__intarray_to_clist:Nn, and \__intarray_to_clist:w.
This function is documented on page 302.)

\__kernel_intarray_gset_range_from_clist:Nnn
\__intarray_gset_range:Nw
Loop through part of the array.
\cs_new_protected:Npn \__kernel_intarray_gset_range_from_clist:Nnn #1#2#3
{ \int_set:Nn \l__intarray_loop_int {#2}
  \__intarray_gset_range:Nw #1 #3 , , \prg_break_point:
}
\cs_new_protected:Npn \__intarray_gset_range:Nw #1 #2 ,
{ \if_catcode:w \scan_stop: \tl_to_str:n {#2} \scan_stop:
  \prg_break:n
  \fi:
  \__kernel_intarray_gset:Nn #1 \l__intarray_loop_int {#2}
  \int_incr:N \l__intarray_loop_int
  \__intarray_gset_range:Nw #1
}
64.3 Common parts

Convert the list to a comma list (with spaces after each comma)

\intarray_gset_rand:Nn
\intarray_gset_rand:cn
\intarray_gset_rand:Nnn
\intarray_gset_rand:cnn
\__intarray_gset_rand:Nnn
\__intarray_gset_rand:Nff
\__intarray_gset_rand:auxi:Nnn
\__intarray_gset_rand:auxii:Nnn
\__intarray_gset_all_same:Nn

We only perform the bounds checks once. This is done by two \__intarray_gset_-overflow_test:nw, with an appropriate empty argument to avoid a spurious “at position #1” part in the error message. Then calculate the number of choices: this is at most $2^{30} - 1 - \left(-2^{30} - 1\right) + 1 = 2^{31} - 1$, which just barely does not overflow. For small ranges use \__kernel_randint:n (making sure to subtract 1 before adding the random number to the \langle min\rangle, to avoid overflow when \langle min\rangle or \langle max\rangle are $\pm\max_int$), otherwise \__kernel_randint:nn. Finally, if there are no random numbers do not define any of the auxiliaries.
\begin{verbatim}
}\cs_generate_variant:Nn \__intarray_gset_rand:Nnn { Nff }
\cs_new_protected:Npn \__intarray_gset_rand_auxi:Nnnn #1 { } {#2} {#3}
\cs_new_protected:Npn \__intarray_gset_rand_auxii:Nnnn #1 { } {#4} {#3}
\cs_new_protected:Npn \__intarray_gset_rand_auxiii:Nnnn #1#2#3#4
\exp_args:NNf \__intarray_gset_all_same:Nn #1
{ \int_compare:nNnTF {#2} > \c__kernel_randint_max_int
{ \exp_stop_f:
\int_eval:n { \__kernel_randint:nn {#3} {#4} }
}
{ \exp_stop_f:
\int_eval:n { \__kernel_randint:n {#2} - 1 + #3 }
}
}\cs_new_protected:Npn \intarray_gset_rand:Nnn #1#2#3
{\msg_error:nnn { kernel } { fp-no-random }
{ \intarray_gset_rand:Nnn #1 {#2} {#3} }
}\cs_generate_variant:Nn \intarray_gset_rand:Nnn { c }
\end{verbatim}

(End definition for \intarray_gset_rand:Nn and others. These functions are documented on page 302.)
Chapter 65

\texttt{I3fp implementation}

Nothing to see here: everything is in the subfiles!
Chapter 66

l3fp-aux implementation

66.1 Access to primitives

Largely for performance reasons, we need to directly access primitives rather than use \int_eval:n. This happens a lot, so we use private names. The same is true for \romannumeral, although it is used much less widely.

(End definition for \__fp_int_eval:w, \__fp_int_eval_end:, and \__fp_int_to_roman:w.)

66.2 Internal representation

Internally, a floating point number \(X\) is a token list containing

\[s__fp \__fp_chk:w \langle \text{case} \rangle \langle \text{sign} \rangle \langle \text{body} \rangle;\]

Let us explain each piece separately.

Internal floating point numbers are used in expressions, and in this context are subject to \f-expansion. They must leave a recognizable mark after \f-expansion, to prevent the floating point number from being re-parsed. Thus, \s__fp is simply another name for \relax.

When used directly without an accessor function, floating points should produce an error: this is the role of \__fp_chk:w. We could make floating point variables be protected to prevent them from expanding under \x-expansion, but it seems more convenient to treat them as a subcase of token list variables.

The (decimal part of the) IEEE-754-2008 standard requires the format to be able to represent special floating point numbers besides the usual positive and negative cases. We distinguish the various possibilities by their \langle \text{case} \rangle, which is a single digit:

0 zeros: +0 and -0,

1 “normal” numbers (positive and negative),
Table 3: Internal representation of floating point numbers.

<table>
<thead>
<tr>
<th>Representation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0 _<em>fp</em>...</td>
<td>Positive zero.</td>
</tr>
<tr>
<td>0 2 _<em>fp</em>...</td>
<td>Negative zero.</td>
</tr>
<tr>
<td>1 0 {\text{exponent}} {X_1} {X_2} {X_3} {X_4} ;</td>
<td>Positive floating point.</td>
</tr>
<tr>
<td>1 2 {\text{exponent}} {X_1} {X_2} {X_3} {X_4} ;</td>
<td>Negative floating point.</td>
</tr>
<tr>
<td>2 0 _<em>fp</em>...</td>
<td>Positive infinity.</td>
</tr>
<tr>
<td>2 2 _<em>fp</em>...</td>
<td>Negative infinity.</td>
</tr>
<tr>
<td>3 1 _<em>fp</em>...</td>
<td>Quiet \text{n}an.</td>
</tr>
<tr>
<td>3 1 _<em>fp</em>...</td>
<td>Signalling \text{n}an.</td>
</tr>
</tbody>
</table>

2 infinities: +\text{inf} and -\text{inf},
3 quiet and signalling \text{n}an.

The \langle sign \rangle is 0 (positive) or 2 (negative), except in the case of \text{n}an, which have \langle sign \rangle = 1. This ensures that changing the \langle sign \rangle digit to 2 – \langle sign \rangle is exactly equivalent to changing the sign of the number.

Special floating point numbers have the form
\verb|\s__fp \__fp_chk:w (case) (sign) __fp_... ;|
where \verb|__fp_...| is a scan mark carrying information about how the number was formed (useful for debugging).

Normal floating point numbers ((case) = 1) have the form
\verb|\s__fp \__fp_chk:w 1 (sign) \{\text{exponent}\} \{X_1\} \{X_2\} \{X_3\} \{X_4\} ;|

Here, the \langle exponent \rangle is an integer, between \(-10000\) and \(10000\). The body consists in four blocks of exactly 4 digits, \(0000 \leq \langle X_i \rangle \leq 9999\), and the floating point is
\((-1)^{\langle sign \rangle/2} \langle X_1 \rangle \langle X_2 \rangle \langle X_3 \rangle \langle X_4 \rangle \cdot 10^{\langle exponent \rangle - 16}\)
where we have concatenated the 16 digits. Currently, floating point numbers are normalized such that the \langle exponent \rangle is minimal, in other words, \(1000 \leq \langle X_1 \rangle \leq 9999\).
Calculations are done in base \(10000\), i.e. one myriad.

66.3 Using arguments and semicolons

This function removes an argument (typically a digit) and replaces it by \verb|\exp_stop_f:|, a marker which stops \text{f}-type expansion.
\verb|\_fp_use_none_stop_f:n|

Those functions place a semicolon after one or two arguments (typically digits).
\verb|\_fp_use_s:n| and \verb|\_fp_use_s:nn|

970
Those functions select specific arguments among a set of arguments delimited by a semicolon.

\__fp_use_none_until_s:w
\__fp_use_i_until_s:ww
\__fp_use_ii_until_s:nww

\__cs_new:Npn \__fp_use_none_until_s:w #1; { } \__cs_new:Npn \__fp_use_i_until_s:ww #1#2; {#1} \__cs_new:Npn \__fp_use_ii_until_s:nww #1#2#3; {#2}
(End definition for \__fp_use_none_until_s:w, \__fp_use_i_until_s:ww, and \__fp_use_ii_until_s:nww.)

\__fp_reverse_args:Nww

Many internal functions take arguments delimited by semicolons, and it is occasionally useful to swap two such arguments.
\cs_new:Npn \__fp_reverse_args:Nww #1 #2; #3; { #1 #3; #2; }
(End definition for \__fp_reverse_args:Nww.)

\__fp_rrot:www

Rotate three arguments delimited by semicolons. This is the inverse (or the square) of the Forth primitive ROT, hence the name.
\cs_new:Npn \__fp_rrot:www #1; #2; #3; { #2; #3; #1; }
(End definition for \__fp_rrot:www.)

\__fp_use_i:ww
\__fp_use_i:www

Many internal functions take arguments delimited by semicolons, and it is occasionally useful to remove one or two such arguments.
\cs_new:Npn \__fp_use_i:ww #1; #2; #3; { #1; }
\cs_new:Npn \__fp_use_i:www #1; #2; #3; { #1; }
(End definition for \__fp_use_i:ww and \__fp_use_i:www.)

66.4 Constants, and structure of floating points

\__fp_misused:n

This receives a floating point object (floating point number or tuple) and generates an error stating that it was misused. This is called when for instance an fp variable is left in the input stream and its contents reach \TeX{}’s stomach.
\cs_new_protected:Npn \__fp_misused:n #1 \{ \msg_error:nnx { fp } { misused } { \fp_to_tl:n {#1} } \}
(End definition for \__fp_misused:n.)

\s__fp
\__fp_chk:w

Floating points numbers all start with \s__fp \__fp_chk:w, where \s__fp is equal to the \TeX{} primitive \relax, and \__fp_chk:w is protected. The rest of the floating point number is made of characters (or \relax). This ensures that nothing expands under f-expansion, nor under x-expansion. However, when typeset, \s__fp does nothing, and \__fp_chk:w is expanded. We define \__fp_chk:w to produce an error.
\scan_new:N \s__fp
\cs_new_protected:Npn \__fp_misused:n \{ \msg_error:nnx { fp } { misused } { \fp_to_tl:n {#1} } \}
(End definition for \s__fp and \__fp_chk:w.)

\s__fp_expr_mark
\s__fp_expr_stop

Aliases of \texttt{\lrelax:D}, used to terminate expressions.
\scan_new:N \s__fp_expr_mark
\scan_new:N \s__fp_expr_stop
(End definition for \s__fp_expr_mark and \s__fp_expr_stop.)
\texttt{\_fp_mark} Generic scan marks used throughout the module.
\texttt{\_fp_stop} (End definition for \_fp_mark and \_fp_stop.)
\texttt{\_fp_use_i_delimit_by_s_stop:nw} Functions to gobble up to a scan mark.
\texttt{\_fp_invalid} (End definition for \_fp_use_i_delimit_by_s_stop:nw.)
\texttt{\_fp_underflow} \texttt{\_fp_overflow} \texttt{\_fp_division} \texttt{\_fp_exact} A couple of scan marks used to indicate where special floating point numbers come from.
\texttt{\_fp_invalid} \texttt{\_fp_underflow} \texttt{\_fp_overflow} \texttt{\_fp_division} \texttt{\_fp_exact} (End definition for \_fp_invalid and others.)
\texttt{\c_zero_fp} \texttt{\c_minus_zero_fp} \texttt{\c_inf_fp} \texttt{\c_minus_inf_fp} \texttt{\c_nan_fp} The special floating points. We define the floating points here as “exact”.
\texttt{\c_zero_fp} \texttt{\c_minus_zero_fp} \texttt{\c_inf_fp} \texttt{\c_minus_inf_fp} \texttt{\c_nan_fp} (End definition for \c_zero_fp and others. These variables are documented on page 246.)
\texttt{\c__fp_prec_int} \texttt{\c__fp_half_prec_int} \texttt{\c__fp_block_int} The number of digits of floating points.
\texttt{\c__fp_prec_int} \texttt{\c__fp_half_prec_int} \texttt{\c__fp_block_int} (End definition for \c__fp_prec_int, \c__fp_half_prec_int, and \c__fp_block_int.)
\texttt{\c__fp_myriad_int} Blocks have 4 digits so this integer is useful.
\texttt{\c__fp_myriad_int} (End definition for \c__fp_myriad_int.)
\texttt{\c__fp_minus_min_exponent_int} \texttt{\c__fp_max_exponent_int} Normal floating point numbers have an exponent between \texttt{\c__fp_minus_min_exponent_int} and \texttt{\c__fp_max_exponent_int} inclusive. Larger numbers are rounded to \pm\infty. Smaller numbers are rounded to \pm0. It would be more natural to define a \texttt{\c__fp_min_exponent_int} with the opposite sign but that would waste one \TeX count.
\texttt{\c__fp_minus_min_exponent_int} \texttt{\c__fp_max_exponent_int} (End definition for \c__fp_minus_min_exponent_int and \c__fp_max_exponent_int.)
\texttt{\c__fp_max_exp_exponent_int} If a number’s exponent is larger than that, its exponential overflows/underflows.
\texttt{\c__fp_max_exp_exponent_int} (End definition for \c__fp_max_exp_exponent_int.)

972
\_fp\_overflowing\_fp

A floating point number that is bigger than all normal floating point numbers. This replaces infinities when converting to formats that do not support infinities.

```
\tl\_const:Nx \_fp\_overflowing\_fp
\{\n  \_fp\_chk:w 1 0
  \{ \int\_eval:n { \_fp\_max\_exponent\_int + 1 } \}
  \{1000\} \{0000\} \{0000\} \{0000\} ;
\}
```

(End definition for \_fp\_overflowing\_fp.)

\_fp\_zero\_fp:N
\_fp\_inf\_fp:N

In case of overflow or underflow, we have to output a zero or infinity with a given sign.

```
\cs\_new:Npn \_fp\_zero\_fp:N #1
\{ \_fp\_chk:w 0 #1 \_fp\_underflow ; \}
\cs\_new:Npn \_fp\_inf\_fp:N #1
\{ \_fp\_chk:w 2 #1 \_fp\_overflow ; \}
```

(End definition for \_fp\_zero\_fp:N and \_fp\_inf\_fp:N.)

\_fp\_exponent:w

For normal numbers, the function expands to the exponent, otherwise to 0. This is used in l3str-format.

```
\cs\_new:Npn \_fp\_exponent:w \_fp\_chk:w #1
\{\n  \if\_meaning:w 1 #1\n    \exp\_after:wN \_fp\_use\_ii\_until\_s:nnw
  \else:\n    \exp\_after:wN \_fp\_use\_i\_until\_s:nw
    \exp\_after:wN 0
  \fi:
\}
```

(End definition for \_fp\_exponent:w.)

\_fp\_neg\_sign:N

When appearing in an integer expression or after \int\_value:w, this expands to the sign opposite to #1, namely 0 (positive) is turned to 2 (negative), 1 (nan) to 1, and 2 to 0.

```
\cs\_new:Npn \_fp\_neg\_sign:N #1
\{ \_fp\_int\_eval:w 2 - #1 \_fp\_int\_eval\_end: \}
```

(End definition for \_fp\_neg\_sign:N.)

\_fp\_kind:w

Expands to 0 for zeros, 1 for normal floating point numbers, 2 for infinities, 3 for nan, 4 for tuples.

```
\cs\_new:Npn \_fp\_kind:w #1
\{ \_fp\_if\_type\_fp:NTwFw
  #1 \_fp\_use\_ii\_until\_s:nw
\}
```

(End definition for \_fp\_kind:w.)
66.5 Overflow, underflow, and exact zero

Expects the sign and the exponent in some order, then the significand (which we don’t touch). Outputs the corresponding floating point number, possibly underflowed to ±0 or overflowed to ±∞. The functions \_fp_underflow:w and \_fp_overflow:w are defined in l3fp-traps.

\cs_new:Npn \_fp_sanitize:Nw \#1 \#2; 
\{ \if_case:w
\else:
\fi:
\exp_after:wN \_fp_overflow:w
\exp_after:wN \_fp_underflow:w
\fi:
\exp_after:wN \_fp_sanitize_zero:w
\fi:
\s__fp \_fp_chk:w \#1 \#2
\}
\cs_new:Npn \_fp_sanitize:wN \#1; \#2 { \_fp_sanitize:Nw \#2 \#1; }
\cs_new:Npn \_fp_sanitize_zero:w \s__fp \_fp_chk:w \#1 \#2 \#3; 
\{ \c_zero_fp \}

(End definition for \_fp_sanitize:Nw, \_fp_sanitize:wN, and \_fp_sanitize_zero:w.)

66.6 Expanding after a floating point number

\_fp_exp_after_o:w (floating point)
\_fp_exp_after_f:nw \{tokens\} (floating point)

Places \{tokens\} (empty in the case of \_fp_exp_after_o:w) between the \{floating point\} and the following tokens, then hits those tokens with o or f-expansion, and leaves the floating point number unchanged.

We first distinguish normal floating points, which have a significand, from the much simpler special floating points.

\cs_new:Npn \_fp_exp_after_o:w \_fp_exp_after_normal:nNW \{\}
\cs_new:Npn \_fp_exp_after_special:nNW \{\}
\cs_new:Npn \_fp_exp_chk:w \#1 \#2 \#3; 
\{ \c_zero_fp \}

(End definition for \_fp_exp_after_o:w, \_fp_exp_after_normal:nNW, \_fp_exp_after_special:nNW, and \_fp_exp_chk:w.)
Special floating point numbers are easy to jump over since they contain few tokens.

\__fp_exp_after_special:nNNw \{ \langle after \rangle \langle case \rangle \langle sign \rangle \langle scan mark \rangle \};

Floating point tuples take the form \s__fp_tuple \__fp_tuple_chk:w \{ \langle fp 1 \rangle \langle fp 2 \rangle \ldots \}; where each \langle fp \rangle is a floating point number or tuple, hence ends with \}; itself. When a tuple is typeset, \__fp_tuple_chk:w produces an error, just like usual floating point numbers. Tuples may have zero or one element.

For normal floating point numbers, life is slightly harder, since we have many tokens to jump over. Here it would be slightly better if the digits were not braced but instead were delimited arguments (for instance delimited by ,). That may be changed some day.

66.7 Other floating point types
\__fp_tuple_count:w
\__fp_array_count:n
\__fp_tuple_count_loop:Nw

Count the number of items in a tuple of floating points by counting semicolons. The technique is very similar to \tl_count:n, but with the loop built-in. Checking for the end of the loop is done with the \use_none:n \#1 construction.

\cs_new:Npn \__fp_array_count:n \#1
{ \__fp_tuple_count:w \s__fp_tuple \__fp_tuple_chk:w {#1} ; }

\cs_new:Npn \__fp_tuple_count:w \s__fp_tuple \__fp_tuple_chk:w #1 ;
{ \int_value:w \__fp_int_eval:w 0
\__fp_tuple_count_loop:Nw #1 { ? \prg_break: } ;
\prg_break_point:
\__fp_int_eval_end:
}

\cs_new:Npn \__fp_tuple_count_loop:Nw #1#2;
{ \use_none:n #1 + 1 \__fp_tuple_count_loop:Nw }

(End definition for \__fp_tuple_count:w, \__fp_array_count:n, and \__fp_tuple_count_loop:Nw.)

\__fp_if_type_fp:NTwFw

Used as \__fp_if_type_fp:NTwFw ⟨marker⟩ \{ ⟨true code⟩ \} \s__fp \{ ⟨false code⟩ \} \s__fp_stop, this test whether the ⟨marker⟩ is \s__fp or not and runs the appropriate ⟨code⟩. The very unusual syntax is for optimization purposes as that function is used for all floating point operations.

\cs_new:Npn \__fp_if_type_fp:NTwFw #1 \s__fp #2 #3 \s__fp_stop {#2}

(End definition for \__fp_if_type_fp:NTwFw.)

\__fp_array_if_all_fp:nTF
\__fp_array_if_all_fp_loop:w

True if all items are floating point numbers. Used for min.

\cs_new:Npn \__fp_array_if_all_fp:nTF #1
{ \__fp_array_if_all_fp_loop:w #1 { \s__fp \prg_break: } ;
\prg_break_point: \use_i:nnn
}

\cs_new:Npn \__fp_array_if_all_fp_loop:w #1#2
{ \__fp_if_type_fp:NTwFw #1 \__fp_array_if_all_fp_loop:w
\s__fp { \prg_break:n \use_iii:nnn }
\s__fp_stop
}

(End definition for \__fp_array_if_all_fp:nTF and \__fp_array_if_all_fp_loop:w.)

\__fp_type_from_scan:N
\__fp_type_from_scan_other:N
\__fp_type_from_scan:w

Used as \__fp_type_from_scan:N ⟨token⟩. Grabs the pieces of the stringified ⟨token⟩ which lies after the first \s__fp. If the ⟨token⟩ does not contain that string, the result is _?._

\cs_new:Npn \__fp_type_from_scan:N #1
{ \__fp_if_type_fp:NTwFw #1 { }
\s__fp { \__fp_type_from_scan_other:N #1 }
\s__fp_stop
}
\cs_new:Npx \__fp_type_from_scan_other:N #1
{ \exp_not:N \exp_after:wN \exp_not:N \__fp_type_from_scan:w

976
(End definition for \_fp_type_from_scan:N, \_fp_type_from_other:N, and \_fp_type_from-_scan:w.)

Arguments are \{type marker\} \{function\} \{recovery\}. This gives the function obtained by placing the type after \@@. If the function is not defined then \{recovery\} \{function\} is used instead; however that test is not run when the \{type marker\} is \s__fp.

The \Nnw function simply dispatches to the appropriate \_fp_exp_after..._f:nw with "..." (either empty or \_\{type\}) extracted from \#1, which should start with \s__fp. If it doesn’t start with \s__fp the function \_fp_exp_after:_f:nw defined in l3fp-parse gives an error; another special \{type\} is stop, useful for loops, see below. The \nw function has an important optimization for floating points numbers; it also fetches its type marker \#2 from the floating point.
The loop works by using the \( n \) argument of \texttt{\_\_fp\_exp\_after\_any\_f:nw} to place the loop macro after the next item in the tuple and expand it.

\begin{verbatim}
\_\_fp\_exp\_after\_array\_f:w
\_\_fp\_exp\_after\_tuple\_o:w
\_\_fp\_exp\_after\_tuple\_f:nw
\_\_fp\_exp\_after\_array\_f:w
\end{verbatim}

66.8 Packing digits

When a positive integer \( \#1 \) is known to be less than \( 10^8 \), the following trick splits it into two blocks of 4 digits, padding with zeros on the left.

\begin{verbatim}
\cs_new:Npn \_\_fp\_after\_tuple\_o:w
{ \_\_fp\_after\_tuple\_f:nw \exp_after:wN \exp_stop_f: }
\cs_new:Npn \_\_fp\_after\_tuple\_f:nw
#1 \_\_fp\_tuple \_\_fp\_tuple\_chk:w #2 ;
\cs_new:Npn \_\_fp\_after\_array\_f:w
{ \_\_fp\_exp\_after\_array\_f:w }
\end{verbatim}

\begin{verbatim}
\exp_after:wN \_\_fp\_tuple
\exp_after:wN \_\_fp\_tuple\_chk:w
\exp_after:wN { \_\_fp\_exp\_after\_array\_f:w #2 \_\_fp\_expr\_stop
\exp_after:wN \exp_end_continue_f:w
\exp:w \exp_end_continue_f:w #1
\end{verbatim}

(End definition for \texttt{\_\_fp\_exp\_after\_tuple\_o:w}, \texttt{\_\_fp\_exp\_after\_tuple\_f:nw}, and \texttt{\_\_fp\_exp\_after\_array\_f:w}.)
The \exp_after:wN \int_value:w \__fp_int_eval:w, which starts a first computation, whose initial value is $-5\,0000$ (the “leading shift”). In that computation appears an \exp_after:wN \int_value:w \__fp_int_eval:w, which triggers the nested computation \int_value:w \__fp_int_eval:w with starting value $4\,9995\,0000$ (the “middle shift”). That, in turn, expands \exp_after:wN \int_value:w \__fp_int_eval:w which triggers the third computation. The third computation’s value is $5\,0000\,0000 + 12345 \times 8899$, which has 9 digits. Adding $5 \cdot 10^8$ to the product allowed us to know how many digits to expect as long as the numbers to multiply are not too big; it also works to some extent with negative results. The pack function puts the last 4 of those 9 digits into a brace group, moves the semi-colon delimiter, and inserts a $+$, which combines the carry with the previous computation. The shifts nicely combine into $5\,0000\,0000 / 10^4 + 4\,9995\,0000 = 5\,0000\,0000$. As long as the operands are in some range, the result of this second computation has 9 digits. The corresponding pack function, expanded after the result is computed, braces the last 4 digits, and leaves $+\langle5\text{ digits}\rangle$ for the initial computation. The “leading shift” cancels the combination of the other shifts, and the \post_processing:w takes care of packing the last few digits.

Admittedly, this is quite intricate. It is probably the key in making l3fp as fast as other pure \TeX floating point units despite its increased precision. In fact, this is used so much that we provide different sets of packing functions and shifts, depending on ranges of input.

This set of shifts allows for computations involving results in the range $[-4\cdot10^8, 5\cdot10^8 - 1]$. Shifted values all have exactly 9 digits.

This set of shifts allows for computations involving results in the range $[-5\cdot10^8, 6\cdot10^8 - 1]$ (actually a bit more). Shifted values all have exactly 10 digits. Note that the upper bound is due to \TeX’s limit of $2^{31} - 1$ on integers. The shifts are chosen to be roughly the mid-point of $10^9$ and $2^{31}$, the two bounds on 10-digit integers in \TeX.

(End definition for \__fp_pack:NNNNNw and others.)
This set of shifts allows for computations with results in the range \([-1 \cdot 10^9, 147483647]\); the end-point is \(2^{31} - 1 - 2 \cdot 10^9 \approx 1.47 \cdot 10^8\). Shifted values all have exactly 10 digits.

\[
\begin{align*}
\int_{const:Nn} \c{\_fp_Bigg\_leading\_shift\_int} \{ - 20000 \} \\
\int_{const:Nn} \c{\_fp_Bigg\_middle\_shift\_int} \{ 20000 \ast 9999 \} \\
\int_{const:Nn} \c{\_fp_Bigg\_trailing\_shift\_int} \{ 20000 \ast 10000 \} \\
\cs{n}{\_fp\_pack\_Bigg:NNNNNNw} \{ #1#2 #3#4#5#6 #7; \}
\end{align*}
\]

(End definition for \(\_fp\_pack\_Bigg:NNNNNNw\) and others.)

\[
\begin{align*}
\cs{n}{\_fp\_pack\_twice\_four:wNNNNNNNN} \\
\cs{n}{\_fp\_pack\_eight:wNNNNNNNN}
\end{align*}
\]

Grabs two sets of 4 digits and places them before the semi-colon delimiter. Putting several copies of this function before a semicolon packs more digits since each takes the digits packed by the others in its first argument.

\[
\begin{align*}
\cs{n}{\_fp\_pack\_twice\_four:wNNNNNNNN} \{ #1; #2#3#4#5 #6#7#8#9 \} \\
\cs{n}{\_fp\_pack\_eight:wNNNNNNNN} \{ #1; #2#3#4#5 #6#7#8#9 \}
\end{align*}
\]

(End definition for \(\_fp\_pack\_twice\_four:wNNNNNNNN\).)

Addition and multiplication of significands are done in two steps: first compute a (more or less) exact result, then round and pack digits in the final (braced) form. These functions take care of the packing, with special attention given to the case where rounding has caused a carry. Since rounding can only shift the final digit by 1, a carry always produces an exact power of 10. Thus, \(\_fp\_basics\_pack\_high\_carry:w\) is always followed by four times \(\{0000\}\).

This is used in \texttt{l3fp-basics} and \texttt{l3fp-extended}.

\[
\begin{align*}
\cs{n}{\_fp\_basics\_pack\_low:NNNNw} \\
\cs{n}{\_fp\_basics\_pack\_high:NNNNw} \\
\cs{n}{\_fp\_basics\_pack\_high\_carry:v}
\end{align*}
\]
This is used in \texttt{l3fp-basics} for additions and divisions. Their syntax is confusing, hence the name.

\begin{verbatim}
\cs_new:Npn \__fp_basics_pack_weird_low:NNNNw #1 #2#3#4 #5;
  \if_meaning:w 2 #1 + 1 \__fp_int_eval_end: #2#3#4; {#5} ;
\}
\cs_new:Npn \__fp_basics_pack_weird_high:NNNNNNNNw
  1 #1#2#3#4 #5#6#7#8 #9; { ; {#1#2#3#4} {#5#6#7#8} {#9} }
\end{verbatim}

(End definition for \__fp_basics_pack_weird_low:NNNNw and \__fp_basics_pack_weird_high:NNNNNNNNw.)

\section{Decimate (dividing by a power of \texttt{10})}

\begin{verbatim}
\__fp_decimate:nNnnnn \{\langle shift\rangle\} \{f_1\} \{\langle X_1\rangle\} \{\langle X_2\rangle\} \{\langle X_3\rangle\} \{\langle X_4\rangle\}
\end{verbatim}

Each $\langle X_i\rangle$ consists in 4 digits exactly, and $1000 \leq \langle X_1\rangle < 9999$. The first argument determines by how much we shift the digits. $\langle f_1\rangle$ is called as follows:

$\langle f_1\rangle \langle\text{rounding}\rangle \{\langle X_1\rangle\} \{\langle X_2\rangle\} \{\text{extra-digits}\}$ ;

where $0 \leq \langle X'_i\rangle < 10^8 - 1$ are 8 digit integers, forming the truncation of our number. In other words,

\[
\left( \sum_{i=1}^{4} \langle X_i \rangle \cdot 10^{-4i} \cdot 10^{-\langle\text{shift}\rangle} \right) - \left( \langle X'_1 \rangle \cdot 10^{-8} + \langle X'_2 \rangle \cdot 10^{-16} \right) = 0.\langle\text{extra-digits}\rangle \cdot 10^{-16} \in [0, 10^{-16}].
\]

To round properly later, we need to remember some information about the difference. The $\langle\text{rounding}\rangle$ digit is 0 if and only if the difference is exactly 0, and 5 if and only if the difference is exactly $0.5 \cdot 10^{-16}$. Otherwise, it is the (non-0, non-5) digit closest to $10^{17}$ times the difference. In particular, if the shift is 17 or more, all the digits are dropped, \langle\text{rounding}\rangle is 1 (not 0), and $\langle X'_1 \rangle$ and $\langle X'_2 \rangle$ are both zero.

If the shift is 1, the $\langle\text{rounding}\rangle$ digit is simply the only digit that was pushed out of the brace groups (this is important for subtraction). It would be more natural for the $\langle\text{rounding}\rangle$ digit to be placed after the $\langle X'_i\rangle$, but the choice we make involves less reshuffling.

Note that this function treats negative $\langle\text{shift}\rangle$ as 0.
Each of the auxiliaries see the function \( f_1 \), followed by 4 blocks of 4 digits.

(End definition for \( \_\_f_\_d_\_c\_e\_m\_a\_t\_i:\_N_\_n_\_n_\_n_\_n_\_n_\_n_\_n \).

If the \( \langle \text{shift} \rangle \) is zero, or too big, life is very easy.

22761 \cs_new:Npn \__fp_decimate_:Nnnnn #1 #2#3#4#5
22762 { #1 0 {#2#3} {#4#5} ; }
22763 \cs_new:Npn \__fp_decimate_tiny:Nnnnn #1 #2#3#4#5
22764 { #1 1 { 0000 0000 } { 0000 0000 } 0 {#2#3#4#5} ; }

(End definition for \( \_\_f_\_d_\_c\_e\_m\_a\_t\_i:\_N_\_n_\_n_\_n_\_n_\_n_\_n_\_n \) and \( \_\_f_\_d_\_c\_e\_m\_a\_t\_i\_t\_y:N_\_n_\_n_\_n_\_n_\_n_\_n_\_n \).

Shifting happens in two steps: compute the \( \langle \text{rounding} \rangle \) digit, and repack digits into two blocks of 8. The sixteen functions are very similar, and defined through \( \_\_f_\_p\_t\_m:p:w \). The arguments are as follows: \#1 indicates which function is being defined; after one step of expansion, \#2 yields the “extra digits” which are then converted by \( \_\_f_\_p\_r\_o\_u\_n\_d\_d_\_i\_g\_t:Nw \) to the \( \langle \text{rounding} \rangle \) digit (note the + separating blocks of digits to avoid overflowing \TeX's integers). This triggers the \-expansion of \( \_\_f_\_p\_d_\_c_\_e\_m\_a\_t\_i\_e\_n\_y:p_\_t\_m:w \), responsible for building two blocks of 8 digits, and removing the rest. For this to work, \#3 alternates between braced and unbraced blocks of 4 digits, in such a way that the 5 first and 5 next token groups yield the correct blocks of 8 digits.

22765 \cs_new:Npn \_\_f_\_p_\_t_\_m:p:w #1 #2 #3
22766 { \cs_new:cpn { __fp_decimate_ #1 :Nnnnn } ##1 ##2##3##4##5
22767 { \exp_after:wN ##1 \int_value:w \exp_after:wN \_\_f_\_p\_r\_o\_u\_n\_d\_d_\_i\_g\_t:Nw #2 ;
22768 \_\_f_\_p\_d_\_c_\_e\_m\_a\_t\_i\_e\_n\_y:p_\_t\_m:w #3 ;
22769 }
22770 }
22771 \_\_f_\_p_\_t_\_m:p:w {i} { \use_none:nnn #50}{ 0{#2}#3{#4}#5 }
22772 \_\_f_\_p_\_t_\_m:p:w {ii} { \use_none:nn #5 }{ 00{#2}#3{#4}#5 }
22773 \_\_f_\_p_\_t_\_m:p:w {iii} { \use_none:n #5 }{ 000{#2}#3{#4}#5 }
22774 \_\_f_\_p_\_t_\_m:p:w {iv} { \use_none:nnn #5 }{ (0000)\#2(3)#4 #5 }
22775 \_\_f_\_p_\_t_\_m:p:w {v} { \use_none:nn #4#5 }{ 0(0000)\#2(3)#4 #5 }
22776 \_\_f_\_p_\_t_\_m:p:w {vi} { \use_none:n #4#5 }{ 00(0000)\#2(3)#4 #5 }
22777 \_\_f_\_p_\_t_\_m:p:w {vii} { \use_none:nn #4#5 }{ 000(0000)\#2(3)#4 #5 }
22778 \_\_f_\_p_\_t_\_m:p:w {viii} { \use_none:n #4#5 }{ (0000)(0000)\#2#3 #4 #5 }
22779 \_\_f_\_p_\_t_\_m:p:w {ix} { \use_none:nnn #3#5*5}{ 0(0000)(0000)\#2#3 #4 #5 }
22780 \_\_f_\_p_\_t_\_m:p:w {x} { \use_none:nn #3#5*5 }{ 00(0000)(0000)\#2#3 #4 #5 }
22781 \_\_f_\_p_\_t_\_m:p:w {xi} { \use_none:n #3#5*5 }{ 000(0000)(0000)\#2#3 #4 #5 }
22782 \_\_f_\_p_\_t_\_m:p:w {xii} { \use_none:nnn #3#5*5 }{ (0000)(0000)(0000)\#2#3 #4 #5 }
22783 \_\_f_\_p_\_t_\_m:p:w {xiii} { \use_none:nnn #3#5*5 }{ 0(0000)(0000)(0000)\#2#3 #4 #5 }
22784 \_\_f_\_p_\_t_\_m:p:w {xiv} { \use_none:n #3#5*5 }{ 00(0000)(0000)(0000)\#2#3 #4 #5 }
22785 \_\_f_\_p_\_t_\_m:p:w {xv} { \use_none:nn #3#5*5 }{ 000(0000)(0000)(0000)\#2#3 #4 #5 }
22786 \_\_f_\_p_\_t_\_m:p:w {xvi} { #3#5*5 }{ (0000)(0000)(0000)(0000)\#2#3 #4 #5 }

(End definition for \( \_\_f_\_p\_d_\_c_\_e\_m\_a\_t\_i\_a\_u\_x:i:N_\_n_\_n_\_n_\_n_\_n_\_n_\_n \) and others.)

\[ \text{No, the argument spec is not a mistake: the function calls an auxiliary to do half of the job.} \]
The computation of the (rounding) digit leaves an unfinished \texttt{int_value:w}, which expands the following functions. This allows us to repack nicely the digits we keep. Those digits come as an alternation of unbraced and braced blocks of 4 digits, such that the first 5 groups of token consist in 4 single digits, and one brace group (in some order), and the next 5 have the same structure. This is followed by some digits and a semicolon.

\begin{verbatim}
\cs_new:Npn \__fp_decimate_pack:nnnnnnnnnnw #1#2#3#4#5
{ \__fp_decimate_pack:nnnnnnnw { #1#2#3#4#5 } }
\cs_new:Npn \__fp_decimate_pack:nnnnnnnw #1 #2#3#4#5#6
{ {#1} {#2#3#4#5#6} }
\end{verbatim}

(End definition for \texttt{\_fp_decimate_pack:nnnnnnnnnnw}.)

66.10 Functions for use within primitive conditional branches

The functions described in this section are not pretty and can easily be misused. When correctly used, each of them removes one \texttt{\if:w} as part of its parameter text, and puts one back as part of its replacement text.

Many computation functions in \texttt{l3fp} must perform tests on the type of floating points that they receive. This is often done in an \texttt{\if_case:w} statement or another conditional statement, and only a few cases lead to actual computations: most of the special cases are treated using a few standard functions which we define now. A typical use context for those functions would be

\begin{verbatim}
\if_case:w \integer \exp_stop_f:
 \__fp_case_return_o:Nw \fp_var
\or: \__fp_case_use:nw \{ \texttt{some computation} \}
\or: \__fp_case_return_same_o:w
\or: \__fp_case_return:nw \{ \texttt{something} \}
\fi:
\texttt{junk}
\texttt{floating point}
\end{verbatim}

In this example, the case 0 returns the floating point \texttt{\fp_var}, expanding once after that floating point. Case 1 does \texttt{\texttt{some computation}} using the \texttt{\texttt{floating point}} (presumably compute the operation requested by the user in that non-trivial case). Case 2 returns the \texttt{\texttt{floating point}} without modifying it, removing the \texttt{\texttt{junk}} and expanding once after. Case 3 closes the conditional, removes the \texttt{\texttt{junk}} and the \texttt{\texttt{floating point}}, and expands \texttt{\texttt{something}} next. In other cases, the \texttt{\texttt{junk}} is expanded, performing some other operation on the \texttt{\texttt{floating point}}. We provide similar functions with two trailing \texttt{\texttt{floating points}}.

\begin{verbatim}
\__fp_case_use:nw
\end{verbatim}

This function ends a \TeX conditional, removes junk until the next floating point, and places its first argument before that floating point, to perform some operation on the floating point.

\begin{verbatim}
\cs_new:Npn \__fp_case_use:nw #1\#2 \fi: \#3 \a__fp \{ \fi: \#1 \a__fp \}
\end{verbatim}

(End definition for \texttt{\_fp_case_use:nw}.)
\__fp_case_return:nw  This function ends a \TeX conditional, removes junk and a floating point, and places its first argument in the input stream. A quirk is that we don’t define this function requiring a floating point to follow, simply anything ending in a semicolon. This, in turn, means that the \langle junk\rangle may not contain semicolons.
\cs_new:Npn \__fp_case_return:nw #1#2 \fi: #3 ; { \fi: #1 }
(End definition for \__fp_case_return:nw.)

\__fp_case_return_o:Nw  This function ends a \TeX conditional, removes junk and a floating point, and returns its first argument (an \langle fp var\rangle) then expands once after it.
\cs_new:Npn \__fp_case_return_o:Nw #1#2 \fi: #3 \s__fp #4 ; { \fi: \exp_after:wN #1 }
(End definition for \__fp_case_return_o:Nw.)

\__fp_case_return_same_o:w  This function ends a \TeX conditional, removes junk, and returns the following floating point, expanding once after it.
\cs_new:Npn \__fp_case_return_same_o:w #1 \fi: #2 \s__fp #3 ; { \fi: \__fp_exp_after_o:w \s__fp #3 ; }
(End definition for \__fp_case_return_same_o:w.)

\__fp_case_return_o:Nww  Same as \__fp_case_return_o:Nw but with two trailing floating points.
\cs_new:Npn \__fp_case_return_o:Nww #1#2 \fi: #3 \s__fp #4 ; #5 ; { \fi: \exp_after:wN #1 }
(End definition for \__fp_case_return_o:Nww.)

\__fp_case_return_i_o:ww \__fp_case_return_ii_o:ww  Similar to \__fp_case_return_same_o:w, but this returns the first or second of two trailing floating point numbers, expanding once after the result.
\cs_new:Npn \__fp_case_return_i_o:ww #1 \fi: #2 \s__fp #3 ; \s__fp #4 ; { \fi: \__fp_exp_after_o:w } \cs_new:Npn \__fp_case_return_ii_o:ww #1 \fi: #2 \s__fp ; { \fi: \__fp_exp_after_o:w }
(End definition for \__fp_case_return_i_o:ww and \__fp_case_return_ii_o:ww.)

66.11 Integer floating points
\__fp_int_p:w \__fp_int:w  Tests if the floating point argument is an integer. For normal floating point numbers, this holds if the rounding digit resulting from \__fp_decimate:nNnnnn is 0.
\prg_new_conditional:Npnn \__fp_int:w \s__fp \__fp_chk:w #1 #2 #3 #4; { TF , T , F , p }
{ \if_case:w #1 \exp_stop_f: \prg_return_true: \or: \if_charcode:w 0 \__fp_decimate:nNnnnn \{ \c__fp_prec_int - \#3 \} \__fp_use_i_until_s:nw #4 \__fp_return_true: \else: \__fp_return_false: }
984
\fi:
\else: \prg_return_false:
\fi:
\fi:
}
(End definition for \_fp_int:wTF.)

66.12 Small integer floating points

Tests if the floating point argument is an integer or $\pm \infty$. If so, it is clipped to an integer in the range $[-10^8, 10^8]$ and fed as a braced argument to the (true code). Otherwise, the (false code) is performed.

First filter special cases: zeros and infinities are integers, \texttt{nan} is not. For normal numbers, decimate. If the rounding digit is not 0 run the (false code). If it is, then the integer is #2 #3; use #3 if #2 vanishes and otherwise $10^8$.

\begin{verbatim}
\cs_new:Npn \__fp_small_int:wTF \s__fp \__fp_chk:w #1#2
\if_case:w #1 \exp_stop_f:
\__fp_case_return:nw { \__fp_small_int_true:wTF 0 ; }
\or: \exp_after:wN \__fp_small_int_normal:NnwTF
\or:
\__fp_case_return:nw
\{ 
\exp_after:wN \__fp_small_int_true:wTF \int_value:w
\if_meaning:w 2 #2 - \fi: 1 0000 0000 ;
\}
\else: \__fp_case_return:nw \use_ii:nn
\fi:
#2
\}
\cs_new:Npn \__fp_small_int_true:wTF #1; #2#3 { #2 {#1} }
\cs_new:Npn \__fp_small_int_normal:NnwTF #1#2#3;
\{ 
\__fp_decimate:nNnnnn { \c__fp_prec_int - #2 }
\__fp_small_int_test:NnnwNw
#3 #1
\}
\cs_new:Npn \__fp_small_int_test:NnnwNw #1#2#3#4; #5
\{ 
\if_meaning:w 0 #1 
\exp_after:wN \__fp_small_int_true:wTF
\int_value:w \if_meaning:w 2 #5 - \fi:
\if_int_compare:w #2 > \c_zero_int
  1 0000 0000
\else:
#3
\fi:
\exp_after:wN ;
\else:
\exp_after:wN \use_ii:nn
\fi:
#5
\}
(End definition for \_fp_small_int:wTF and others.)
\end{verbatim}

985
66.13 Fast string comparison

\__fp_str_if_eq:nn
A private version of the low-level string comparison function.
\cs_new_eq:NN \__fp_str_if_eq:nn \tex_strcmp:D
(End definition for \__fp_str_if_eq:nn.)

66.14 Name of a function from its \l3fp-parse name

\__fp_func_to_name:N
\__fp_func_to_name_aux:w
The goal is to convert for instance \__fp_sin_o:w to \sin. This is used in error messages hence does not need to be fast.
\cs_new:Npn \__fp_func_to_name:N #1
\exp_last_unbraced:Nf
\__fp_func_to_name_aux:w { \cs_to_str:N #1 } X
\cs_set_protected:Npn \__fp_tmp:w #1 #2
\exp_args:Nff \__fp_tmp:w { \tl_to_str:n { __fp_ } }
\tl_to_str:n { _o: } （End definition for \__fp_func_to_name:N and \__fp_func_to_name_aux:w.）

66.15 Messages

Using a floating point directly is an error.
\msg_new:nnnn { fp } { misused }{ A floating-point with value '#1' was misused. }
\prop_gput:Nnn \g_msg_module_name_prop { fp } { LaTeX3 }
\prop_gput:Nnn \g_msg_module_type_prop { fp } { }
⟨/package⟩
Chapter 67

l3fp-traps Implementation

Exceptions should be accessed by an n-type argument, among
• invalid_operation
• division_by_zero
• overflow
• underflow
• inexact (actually never used).

67.1 Flags

Flags to denote exceptions.

<table>
<thead>
<tr>
<th>Flag</th>
</tr>
</thead>
<tbody>
<tr>
<td>flag fp_invalid_operation</td>
</tr>
<tr>
<td>flag fp_division_by_zero</td>
</tr>
<tr>
<td>flag fp_overflow</td>
</tr>
<tr>
<td>flag fp_underflow</td>
</tr>
</tbody>
</table>

(End definition for flag fp_invalid_operation and others. These variables are documented on page 248.)

67.2 Traps

Exceptions can be trapped to obtain custom behaviour. When an invalid operation or a division by zero is trapped, the trap receives as arguments the result as an n-type floating point number, the function name (multiple letters for prefix operations, or a single symbol for infix operations), and the operand(s). When an overflow or underflow is trapped, the trap receives the resulting overly large or small floating point number if it is not too big, otherwise it receives $+\infty$. Currently, the inexact exception is entirely ignored.

The behaviour when an exception occurs is controlled by the definitions of the functions

• \_\_fp_invalid_operation::nw,
Rather than changing them directly, we provide a user interface as \(\text{\texttt{fp\_trap:nn}}\) \(|\langle\text{exception}\rangle|\langle\text{way of trapping}\rangle|\), where the \langle\text{way of trapping}\rangle is one of \text{error}, \text{flag}, \text{or none}.

We also provide \(\text{\texttt{__fp\_invalid\_operation\_o:nw}}\), defined in terms of \(\text{\texttt{__fp\_invalid\_operation:nnw}}\).

\(\text{\texttt{\textbackslash fp\_trap:nn}}\)

\begin{verbatim}
\cs_new_protected:Npn \fp_trap:nn #1#2
{ \cs_if_exist_use:cF { __fp_trap_#1_set_#2: } 
  \clist_if_in:nnTF
  { invalid_operation, division_by_zero, overflow, underflow }
  {#1}
  { \msg_error:nnxx { fp } { unknown-fpu-trap-type } {#1} {#2} }
  { \msg_error:nnx { fp } { unknown-fpu-exception } {#1} }
}
\end{verbatim}

(End definition for \(\text{\texttt{\textbackslash fp\_trap:nn}}\). This function is documented on page 248.)

We provide three types of trapping for invalid operations: either produce an error and raise the relevant flag; or only raise the flag; or don’t even raise the flag. In most cases, the function produces as a result its first argument, possibly with post-expansion.

\begin{verbatim}
\cs_new_protected:Npn \__fp_trap_invalid_operation_set_error:
\__fp_trap_invalid_operation_set_flag:
\__fp_trap_invalid_operation_set_none:
\__fp_trap_invalid_operation_set:N
{ \exp_args:Nno \use:n
  \cs_set:Npn \__fp_invalid_operation:nnw ##1##2##3; }
{ \fp_to_tl:n { #3; } {#2} }
\end{verbatim}
We provide three types of trapping for invalid operations and division by zero: either produce an error and raise the relevant flag; or only raise the flag; or don’t even raise the flag. In all cases, the function must produce a result, namely its first argument, ±∞ or nan.

\cs_new_protected:Npn \_\_fp_trap_division_by_zero_set_error:
\{ \cs_set:Npn \_\_fp_trap_division_by_zero_o:Nww ##1##2; ##3; \}
\{ \_\_fp_error:nffn \ invalid-ii \}
\{ \fp_to_tl:n \ { ##2; } \} \{ \fp_to_tl:n \ { ##3; } \} \{##1\}
\flag_raise_if_clear:n \ fp_invalid_operation
\exp_after:wN \c_nan_fp
\}
\exp_args:Nno \use:n
\{ \cs_set:Npn \_\_fp_trap_division_by_zero_tl_o:ff #1##2 \}
\{ #1 \}
\_\_fp_error:nffn \ invalid \ {##1} \ {##2} \{}
\flag_raise_if_clear:n \ fp_invalid_operation
\exp_after:wN \c_nan_fp
\}
\exp_args:Nno \use:n
\{ \cs_set:Npn \_\_fp_trap_division_by_zero_set:N#1 \}
\exp_args:Nno \use:n
\{ \cs_set:Npn \_\_fp_trap_division_by_zero_o:Nww ##1##2; ##3; \}
\{ \_\_fp_error:nffn \ invalid-ii \}
\{ \fp_to_tl:n \ { ##2; } \} \{ \fp_to_tl:n \ { ##3; } \} \{##1\}
\flag_raise_if_clear:n \ fp_invalid_operation
\exp_after:wN \c_nan_fp
\}
(End definition for \_\_fp_trap_invalid_operation_set_error: and others.)
Just as for invalid operations and division by zero, the three different behaviours are obtained by feeding \texttt{\textbackslash prg\_do\_nothing:}, \texttt{\textbackslash use\_none:mmmm} or \texttt{\textbackslash use\_none:mmmmmm} to an auxiliary, with a further auxiliary common to overflow and underflow functions. In most cases, the argument of the \texttt{\_\_fp\_overflow:w} and \texttt{\_\_fp\_underflow:w} functions will be an (almost) normal number (with an exponent outside the allowed range), and the error message thus displays that number together with the result to which it overflowed or underflowed. For extreme cases such as \texttt{10 ** 1e9999}, the exponent would be too large for \LaTeX, and \texttt{\_\_fp\_overflow:w} receives \texttt{\pm \infty} (\texttt{\_\_fp\_underflow:w} would receive \texttt{\pm 0}); then we cannot do better than simply say an overflow or underflow occurred.

\begin{verbatim}
\cs_new_protected:Npn \_\_fp\_trap\_overflow\_set\_error:N #1
{ \_\_fp\_trap\_overflow\_set:NnNn #1 { overflow } \_\_fp\_inf\_fp:N { inf } }
\end{verbatim}

\begin{verbatim}
(End definition for \_\_fp\_trap\_division\_by\_zero\_set\_error: and others.)
\end{verbatim}

\begin{verbatim}
\_\_fp\_invalid\_operation:nnw
\_\_fp\_invalid\_operation:o:Nw
\_\_fp\_invalid\_operation_tl:o:ff
\_\_fp\_division\_by\_zero:o:Nw
\_\_fp\_division\_by\_zero:o:mmm
\_\_fp\_overflow:w
\_\_fp\_underflow:w
\end{verbatim}

Initialize the control sequences (to log properly their existence). Then set invalid operations to trigger an error, and division by zero, overflow, and underflow to act silently on their flag.

\begin{verbatim}
\cs_new:Npn \_\_fp\_invalid\_operation:nnw #1#2
{ }
\end{verbatim}

(End definition for \_\_fp\_trap\_overflow\_set\_error: and others.)
67.3 Errors

Some messages.

67.4 Messages
* - none

\msg_new:nnn { fp } { flow }
{ An ~ #3 ~ occurred. }
\msg_new:nnn { fp } { flow-to }
{ #1 ~ #3 ed ~ to ~ #2 . }
\msg_new:nnn { fp } { zero-div }
{ Division-by-zero-in- #1 (#2) }
\msg_new:nnn { fp } { zero-div-ii }
{ Division-by-zero-in- (#1) #3 (#2) }
\msg_new:nnn { fp } { invalid }
{ Invalid-operation- #1 (#2) }
\msg_new:nnn { fp } { invalid-ii }
{ Invalid-operation- (#1) #3 (#2) }
\msg_new:nnn { fp } { unknown-type }
{ Unknown-type-for-’#1’ }

\package
Chapter 68

l3fp-round implementation

\__fp_parse_word_trunc:N
\__fp_parse_word_floor:N
\__fp_parse_word_ceil:N

\cs_new:Npn \__fp_parse_word_trunc:N
\__fp_parse_function:NNN \__fp_round_o:Nw \__fp_round_to_zero:NNN
\cs_new:Npn \__fp_parse_word_floor:N
\__fp_parse_function:NNN \__fp_round_o:Nw \__fp_round_to_ninf:NNN
\cs_new:Npn \__fp_parse_word_ceil:N
\__fp_parse_function:NNN \__fp_round_o:Nw \__fp_round_to_pinf:NNN

(End definition for \__fp_parse_word_trunc:N, \__fp_parse_word_floor:N, and \__fp_parse_word_ceil:N.)

\__fp_parse_word_round:N
\__fp_parse_round:Nw

\cs_new:Npn \__fp_parse_word_round:N \#1\#2
\__fp_parse_function:NNN
\__fp_round_o:Nw \__fp_round_to_nearest:NNN \#1
\#2
\cs_new:Npn \__fp_parse_round:Nw \#1 \#2 \__fp_round_to_nearest:NNN \#3\#4
\{ \#2 \#1 \#3 \}

(End definition for \__fp_parse_word_round:N and \__fp_parse_round:Nw.)

68.1 Rounding tools

\c__fp_five_int

This is used as the half-point for which numbers are rounded up/down.

\int_const:Nn \c__fp_five_int \{ 5 \}

(End definition for \c__fp_five_int.)

Floating point operations often yield a result that cannot be exactly represented in a significand with 16 digits. In that case, we need to round the exact result to a representable number. The IEEE standard defines four rounding modes:
• Round to nearest: round to the representable floating point number whose absolute difference with the exact result is the smallest. If the exact result lies exactly at the mid-point between two consecutive representable floating point numbers, round to the floating point number whose last digit is even.

• Round towards negative infinity: round to the greatest floating point number not larger than the exact result.

• Round towards zero: round to a floating point number with the same sign as the exact result, with the largest absolute value not larger than the absolute value of the exact result.

• Round towards positive infinity: round to the least floating point number not smaller than the exact result.

This is not fully implemented in l3fp yet, and transcendental functions fall back on the “round to nearest” mode. All rounding for basic algebra is done through the functions defined in this module, which can be redefined to change their rounding behaviour (but there is not interface for that yet).

The rounding tools available in this module are many variations on a base function \_\_fp\_round:NNN, which expands to 0\exp_stop_f: or 1\exp_stop_f: depending on whether the final result should be rounded up or down.

- \_\_fp\_round:NNN \langle sign \rangle \langle digit_1 \rangle \langle digit_2 \rangle can expand to 0\exp_stop_f: or 1\exp_stop_f:.
- \_\_fp\_round_s:NNNw \langle sign \rangle \langle digit_1 \rangle \langle digit_2 \rangle \langle more digits \rangle can expand to 0\exp_stop_f:; or 1\exp_stop_f:.
- \_\_fp\_round_neg:NNN \langle sign \rangle \langle digit_1 \rangle \langle digit_2 \rangle can expand to 0\exp_stop_f: or 1\exp_stop_f:.

See implementation comments for details on the syntax.
\cs_new:Npn \__fp_round_to_ninf:NNN #1 #2 #3 
{ 
  \if_meaning:w 2 #1 
    \if_int_compare:w #3 > \c_zero_int 
      \__fp_round_return_one: 
    \fi: 
  \fi: 
  \c_zero_int 
} 
\cs_new:Npn \__fp_round_to_zero:NNN #1 #2 #3 { \c_zero_int } 
\cs_new:Npn \__fp_round_to_pinf:NNN #1 #2 #3 
{ 
  \if_meaning:w 0 #1 
    \if_int_compare:w #3 > \c_zero_int 
      \__fp_round_return_one: 
    \fi: 
  \fi: 
  \c_zero_int 
} 
\cs_new:Npn \__fp_round_to_nearest:NNN #1 #2 #3 
{ 
  \if_int_compare:w #3 > \c__fp_five_int 
    \__fp_round_return_one: 
  \else: 
    \if_meaning:w 5 #3 
      \if_int_odd:w #2 \exp_stop_f: 
        \__fp_round_return_one: 
      \fi: 
    \fi: 
    \c_zero_int 
  \fi: 
  \c_zero_int 
} 
\cs_new:Npn \__fp_round_to_nearest_ninf:NNN #1 #2 #3 
{ 
  \if_int_compare:w #3 > \c__fp_five_int 
    \__fp_round_return_one: 
  \else: 
    \if_meaning:w 5 #3 
      \if_meaning:w 2 #1 
        \__fp_round_return_one: 
      \fi: 
    \fi: 
    \c_zero_int 
  \fi: 
  \c_zero_int 
} 
\cs_new:Npn \__fp_round_to_nearest_zero:NNN #1 #2 #3 
{ 
  \if_int_compare:w #3 > \c__fp_five_int 
    \__fp_round_return_one: 
  \fi: 
  \c_zero_int 
} 
\cs_new:Npn \__fp_round_to_nearest_pinf:NNN #1 #2 #3 
{ 

\if_int_compare:w #3 > \c\_fp\_five\_int
\_fp\_round\_return\_one:
\else:
\if_meaning:w 5 #3
 \if_meaning:w 0 #1
 \_fp\_round\_return\_one:
 \fi:
 \fi:
 \fi:
\fi:
\c\_zero\_int
\}
\cs\_new\_eq:NN \_fp\_round:NNN \_fp\_round\_to\_nearest:NNN

(End definition for \_fp\_round:NNN and others.)

\_fp\_round\_s:NNNw
\_fp\_round\_s:NNNw \{final\_sign\} \{digit\} \{more\_digits\} ;
Similar to \_fp\_round:NNN, but with an extra semicolon, this function expands to 0\exp\_stop\_f;; if rounding \{final\_sign\}\{digit\}\{more\_digits\} to an integer truncates, and to 1\exp\_stop\_f;; otherwise. The \{more\_digits\} part must be a digit, followed by something that does not overflow a \int\_use:N \_fp\_int\_eval:w construction. The only relevant information about this piece is whether it is zero or not.
\cs\_new\_Npn \_fp\_round\_s:NNNw #1 #2 #3 #4;
\{ \exp\_after:wN \_fp\_round\_s:NNN
\exp\_after:wN #1
\exp\_after:wN #2
\int\_value:w \_fp\_int\_eval:w
\if_int_odd:w 0 \if_meaning:w 0 #3 1 \fi:
 \if_meaning:w 5 #3 1 \fi:
 \exp\_stop\_f:
 \if_int_compare:w \_fp\_int\_eval:w #4 > \c\_zero\_int
 1 +
 \fi:
 \fi:
 \fi:
 #3
\};

(End definition for \_fp\_round\_s:NNNw.)

\_fp\_round\_digit:Nw
\_fp\_round\_digit:Nw \{digit\} \{intexpr\} ;
This function should always be called within an \int\_value:w or \_fp\_int\_eval:w expansion; it may add an extra \_fp\_int\_eval:w, which means that the integer or integer expression should not be ended with a synonym of \relax, but with a semi-colon for instance.
\cs\_new\_Npn \_fp\_round\_digit:Nw #1 #2;
\{ \if_int_odd:w \if_meaning:w 0 #1 1 \else:
 \if_meaning:w 5 #1 1 \else:
 0 \fi: \fi: \exp\_stop\_f:
 \if_int_compare:w \_fp\_int\_eval:w #2 > \c\_zero\_int
 \_fp\_int\_eval:w 1 +
 \fi:
 \fi:
This expands to 0\exp_stop_f: or 1\exp_stop_f: after doing the following test. Starting from a number of the form \( \text{(final sign)}0.(15 \text{ digits})\text{(digit)} \) with exactly 15 (non-all-zero) digits before \( \text{(digit)} \), subtract from it \( \text{(final sign)}0.0\ldots0\text{(digit)} \), where there are 16 zeros. If in the current rounding mode the result should be rounded down, then this function returns 0\exp_stop_f:. Otherwise, \textit{i.e.}, if the result is rounded back to the first operand, then this function returns 0\exp_stop_f:.

It turns out that this negative “round to nearest” is identical to the positive one. And this is the default mode.

68.2 The \textit{round} function

First check that all arguments are floating point numbers. The \texttt{trunc}, \texttt{ceil} and \texttt{floor} functions expect one or two arguments (the second is 0 by default), and the \texttt{round} function also accepts a third argument (\texttt{nan} by default), which changes \#1 from \texttt{\_fp_round_-to_nearest:NNN} to one of its analogues.
\begin{verbatim}
{ \if_case:w \__fp_int_eval:w \__fp_array_count:n {#2} \__fp_int_eval_end:
  \__fp_round_no_arg_o:Nw #1 \exp:w
  \or: \__fp_round:Nwn #1 #2 {0} \exp:w
  \or: \__fp_round:Nww #1 #2 \exp:w
  \else: \__fp_round:Nwww #1 #2 @ \exp:w
  \fi:
  \exp_after:wN \exp_end:
}\end{verbatim}

(End definition for \__fp_round_no_arg_o:Nw and \__fp_round_aux_o:Nw.)

\__fp_round_no_arg_o:Nw
\begin{verbatim}
\cs_new:Npn \__fp_round_no_arg_o:Nw #1
{ \cs_if_eq:NNTF #1 \__fp_round_to_nearest:NNN
  { \__fp_error:nnnn { num-args } { round () } { 1 } { 3 }
  { \__fp_error:nffn { num-args }
    \__fp_round_name_from_cs:N #1 () { 1 } { 2 }
  \exp_after:wN \c_nan_fp
}
\end{verbatim}

(End definition for \__fp_round_no_arg_o:Nw.)

\__fp_round:Nwww
\begin{verbatim}
\cs_new:Npn \__fp_round:Nwww #1#2 ; #3 ; \s__fp \__fp_chk:w #4#5#6 ; #7 @
{ \cs_if_eq:NNTF #1 \__fp_round_to_nearest:NNN
  { \tl_if_empty:nTF {#7}
    { \exp_args:Nc \__fp_round:Nww
      \if_meaning:w 0 #4 _zero \else:
        \if_case:w #5 \exp_stop_f: _pinf \or: \else: _ninf \fi: \fi:
        :NNN
      \else: \__fp_round_to_nearest_zero:NNN
    \exp_after:wN \c_nan_fp
  }
  { \__fp_error:nnnn { num-args }
    \__fp_round_name_from_cs:N #1 () { 1 } { 3 }
    \exp_after:wN \exp_end:
  }
}
\end{verbatim}

Having three arguments is only allowed for round, not trunc, ceil, floor, so check for that case. If all is well, construct one of \__fp_round_to_nearest:NNN, \__fp_round_to_nearest_zero:NNN, \__fp_round_to_nearest_ninf:NNN, \__fp_round_to_nearest_pinf:NNN and act accordingly.

\__fp_round:Nwww
\begin{verbatim}
\cs_new:Npn \__fp_round:Nwww #1#2 ; #3 ; \s__fp \__fp_chk:w #4#5#6 ; #7 @
{ \cs_if_eq:NNTF #1 \__fp_round_to_nearest:NNN
  { \tl_if_empty:nTF {#7}
    { \exp_args:Nc \__fp_round:Nww
      \if_meaning:w 0 #4 _zero \else:
        \if_case:w #5 \exp_stop_f: _pinf \or: \else: _ninf \fi: \fi:
        :NNN
      \else: \__fp_round_to_nearest_zero:NNN
    \exp_after:wN \c_nan_fp
  }
  { \__fp_error:nnnn { num-args }
    \__fp_round_name_from_cs:N #1 () { 1 } { 2 }
    \exp_after:wN \exp_end:
  }
}
\end{verbatim}

\end{verbatim}
If the number of digits to round to is an integer or infinity all is good; if it is nan then
just produce a nan; otherwise invalid as we have something like round(1,3.14) where
the number of digits is not an integer.

\exp_after:wN \c_nan_fp
}

(End definition for \_\_fp_round:Nww.)

\_\_fp_round_name_from_cs:N
\cs_new:Npn \_\_fp_round_name_from_cs:N #1
{\cs_if_eq:NNTF #1 \_\_fp_round_to_zero:NNN { trunc } }
\cs_if_eq:NNTF #1 \_\_fp_round_to_ninf:NNN { floor } }
\cs_if_eq:NNTF #1 \_\_fp_round_to_pinf:NNN { ceil } }
{ round }
\}
(End definition for \_\_fp_round_name_from_cs:N.)
\_\_fp_round:Nww
\_\_fp_round:Nwn
\_\_fp_round_normal:Nwnnnnn
\_\_fp_round_normal_end:wwnnn
\_\_fp_round_special:NwNNn
\_\_fp_round_special_aux:Nw

\_\_fp_round:Nww
\_\_fp_round:Nwn
\_\_fp_round_normal:Nwnnnnn
\_\_fp_round_normal_end:wwnnn
\_\_fp_round_special:NwNNn
\_\_fp_round_special_aux:Nw

\if:w 3 \_\_fp_kind:w #3 ;
\exp_after:wN \use_i:nn
\else:
\exp_after:wN \use_ii:nn
\fi:
{ \exp_after:wN \c_nan_fp }
{ \_\_fp invalid_operation_tl_o:ff
{ \_\_fp_round_name_from_cs:N #1 }
{ \_\_fp_array_to_clist:n { #2; #3; } }
\}
\)
\cs_new:Npn \_\_fp_round:Nww #1 #2 ; #3 ;
{ \_\_fp_small_int:wTF #3 ; { \_\_fp_round:Nwn #1 #2 ; #5 }
{ \_\_fp_round:Nww #1 #2 ; #3 ;
{ \_\_fp_small_int:wTF #3 ; { \_\_fp_round:Nww #1 #2 ; }
{ \_\_fp_round:Nww #1 #2 ; #3 ;
\exp_after:wN \_\_fp_to_zero:Nww
{ \_\_fp_round:Nww #1 #2 ; #3 ;
\exp_after:wN \_\_fp_to_zero:Nww
\exp_after:wN \_\_fp_to_zero:Nww
{ \_\_fp_round:Nww #1 #2 ; #3 ; #5;
\int_value:w #5
\else:
\exp_after:wN \_\_fp_exp_after_o:w
\fi:
\s__fp \_\_fp_chk:w #2 #3 #4; #5
\}
\}
\}
\}
\}
\}
\)
\cs_new:Npn \_\_fp_round:Nww #1 \s__fp \_\_fp_chk:w #2 #3 #4; #5
{ \if meaning:w 1 #2
\exp_after:wN \_\_fp_round_normal:NwNNn
\exp_after:wN \_\_fp_round_normal:NwNNn
\int_value:w #5
\else:
\exp_after:wN \_\_fp_exp_after_o:w
\fi:
\s__fp \_\_fp chk:w #2 #3 #4;
\}
\cs_new:Npn \_\_fp_round_normal:NwNNn #1 #2 \s__fp \_\_fp chk:w 1 #3 #4 #5;
{
\_fp_decimate:nNnnn { \c__fp_prec_int - \#4 - \#2 }
\_fp_round_normal:NnnwNnn \#5 \#1 \#3 \{\#4\} \{\#2\}
\cs_new:Npn \_fp_round_normal:NnnwNnn \#1\#2\#3\#4; \#5\#6 
\{ \exp_after:wN \_fp_round_normal:NNwNnn 
\int_value:w \_fp_int_eval:w 
\if_int_compare:w \#2 > \c_zero_int 
 1 \int_value:w \#2 
\exp_after:wN \_fp_round_pack:Nw 
\int_value:w \_fp_int_eval:w \#3 + 
\else: 
  \if_int_compare:w \#3 > \c_zero_int 
 1 \int_value:w \#3 + 
  \fi: 
\fi: 
\exp_after:wN \_fp_round_normal_end:wwNnn \#2; \#6 
\cs_new:Npn \_fp_round_pack:Nw \#1 
{ \if_meaning:w 2 \#1 + 1 \fi: \_fp_int_eval_end: } 
\cs_new:Npn \_fp_round_normal:NNwNnn \#1 \#2 
{ \if_meaning:w 0 \#2 
 \exp_after:wN \_fp_round_special:NwwNnn 
 \exp_after:wN \#1 
 \exp_after:wN \_fp_pack_twice_four:wNNNNNNNN 
 \_fp_pack_twice_four:wNNNNNNNN 
 \_fp_round_normal_end:wwNnn 
 \#2 
 \} 
\cs_new:Npn \_fp_round_normal_end:wwNnn \#1; \#2; \#3 \#4 \#5 
{ \exp_after:wN \_fp_round_special:NwwNnn \#1; \#2; \#3 \#4 \#5 \#6 
 \if_meaning:w 0 \#1 
 \_fp_case_return:nw 
 \{ \exp_after:wN \_fp_zero_fp:N \exp_after:wN \#4 \} 
 \else: 
 \exp_after:wN \_fp_round_special_aux:Nw 
 \exp_after:wN \#4 
 \int_value:w \_fp_int_eval:w \#1 
 \if_meaning:w \#1 \#1 -\#6 \else: +\#5 \fi: 
 \fi: 
 \} 
\cs_new:Npn \_fp_round_special:NwwNnn \#1 \#2 \#3 \#4 \#5 \#6 
{ \exp_after:wN \_fp_round_special:NwwNnn \#1; \#2; \#3; \#4; \#5; \#6 
 \} 
\cs_new:Npn \_fp_round_special:NwwNnn \#1 \#2 \#3 \#4 \#5 \#6 
{ \exp_after:wN \_fp_round_special:NwwNnn \#1; \#2; \#3; \#4; \#5; \#6 
 \} 
\cs_new:Npn \_fp_round_special:NwwNnn \#1 \#2 \#3 \#4 \#5 \#6 
{ \exp_after:wN \_fp_round_special:NwwNnn \#1; \#2; \#3; \#4; \#5; \#6 
 \} 
1000
\cs_new:Npn \_fp_round_special_aux:Nw \#1#2;
{
\exp_after:wN \_fp_exp_after_o:w \exp:w \exp_end_continue_f:w
\_fp_sanitize:Nw \#1#2; \{1000\}{0000}\{0000\}{0000};
}

(End definition for \_fp_round:Nww and others.)

\package
69.1 Work plan

The task at hand is non-trivial, and some previous failed attempts show that the code leads to unreadable logs, so we had better get it (almost) right the first time. Let us first describe our goal, then discuss the design precisely before writing any code.

In this file at least, a (floating point object) is a floating point number or tuple. This can be extended to anything that starts with \texttt{\_fp} or \texttt{\_fp\_type} and ends with \texttt{;} with some internal structure that depends on the \texttt{type}.

\_fp\_parse:n

\begin{verbatim}
\_fp\_parse:n \{fexpr\}
\end{verbatim}

Evaluates the (floating point expression) and leaves the result in the input stream as a floating point object. This function forms the basis of almost all public l3fp functions. During evaluation, each token is fully f-expanded.

\_fp\_parse_o:n does the same but expands once after its result.

\textbf{\TeX} hackers note: Registers (integers, toks, etc.) are automatically unpacked, without requiring a function such as \texttt{\int\_use:N}. Invalid tokens remaining after f-expansion lead to unrecoverable low-level \TeX errors.

Floating point expressions are composed of numbers, given in various forms, infix operators, such as +, **, or , (which joins two numbers into a list), and prefix operators, such as the unary -, functions, or opening parentheses. Here is a list of precedences which control the order of evaluation (some distinctions are irrelevant for the order of evaluation, but serve as signals), from the tightest binding to the loosest binding.

\begin{itemize}
  \item 16 Function calls.
  \item 13/14 Binary ** and \~{} (right to left).
  \item 12 Unary +, -, ! (right to left).
  \item 11 Juxtaposition (implicit *) with no parenthesis.
\end{itemize}
10 Binary * and /.
9 Binary + and −.
7 Comparisons.
6 Logical and, denoted by &&.
5 Logical or, denoted by ||.
4 Ternary operator ?; piece ?.
3 Ternary operator ?; piece :.
2 Commas.
1 Place where a comma is allowed and generates a tuple.

Start and end of the expression.

\int_const:Nn \c__fp_prec_func_int { 16 }
\int_const:Nn \c__fp_prec_hatii_int { 14 }
\int_const:Nn \c__fp_prec_hat_int { 13 }
\int_const:Nn \c__fp_prec_not_int { 12 }
\int_const:Nn \c__fp_prec_juxt_int { 11 }
\int_const:Nn \c__fp_prec_times_int { 10 }
\int_const:Nn \c__fp_prec_plus_int { 9 }
\int_const:Nn \c__fp_prec_comp_int { 7 }
\int_const:Nn \c__fp_prec_and_int { 6 }
\int_const:Nn \c__fp_prec_or_int { 5 }
\int_const:Nn \c__fp_prec_quest_int { 4 }
\int_const:Nn \c__fp_prec_colon_int { 3 }
\int_const:Nn \c__fp_prec_comma_int { 2 }
\int_const:Nn \c__fp_prec_tuple_int { 1 }
\int_const:Nn \c__fp_prec_end_int { 0 }

(End definition for \c__fp_prec_func_int and others.)

### 69.1.1 Storing results

The main question in parsing expressions expandably is to decide where to put the intermediate results computed for various subexpressions.

One option is to store the values at the start of the expression, and carry them together as the first argument of each macro. However, we want to f-expand tokens one by one in the expression (as \int_eval:n does), and with this approach, expanding the next unread token forces us to jump with \exp_after:wN over every value computed earlier in the expression. With this approach, the run-time grows at least quadratically in the length of the expression, if not as its cube (inserting the \exp_after:wN is tricky and slow).

A second option is to place those values at the end of the expression. Then expanding the next unread token is straightforward, but this still hits a performance issue: for long expressions we would be reaching all the way to the end of the expression at every step of the calculation. The run-time is again quadratic.

A variation of the above attempts to place the intermediate results which appear when computing a parenthesized expression near the closing parenthesis. This still lets
us expand tokens as we go, and avoids performance problems as long as there are enough parentheses. However, it would be better to avoid requiring the closing parenthesis to be present as soon as the corresponding opening parenthesis is read: the closing parenthesis may still be hidden in a macro yet to be expanded.

Hence, we need to go for some fine expansion control: the result is stored \textit{before} the start!

Let us illustrate this idea in a simple model: adding positive integers which may be resulting from the expansion of macros, or may be values of registers. Assume that one number, say, 12345, has already been found, and that we want to parse the next number. The current status of the code may look as follows.

\begin{verbatim}
\exp_after:wN \add:ww \int_value:w 12345 \exp_after:wN ; \\
\exp:w \operand:w ⟨stuff⟩
\end{verbatim}

One step of expansion expands \texttt{\exp_after:wN}, which triggers the primitive \texttt{\int_value:w}, which reads the five digits we have already found, 12345. This integer is unfinished, causing the second \texttt{\exp_after:wN} to expand, and to trigger the construction \texttt{\exp:w}, which expands \texttt{\operand:w}, defined to read what follows and make a number out of it, then leave \texttt{\exp_end:}, the number, and a semicolon in the input stream. Once \texttt{\operand:w} is done expanding, we obtain essentially

\begin{verbatim}
\exp_after:wN \add:ww \int_value:w 12345 ; \\
\exp:w \exp_end: 333444 ;
\end{verbatim}

where in fact \texttt{\exp_after:wN} has already been expanded, \texttt{\int_value:w} has already seen 12345, and \texttt{\exp:w} is still looking for a number. It finds \texttt{\exp_end:}, hence expands to nothing. Now, \texttt{\int_value:w} sees the \texttt{;}, which cannot be part of a number. The expansion stops, and we are left with

\begin{verbatim}
\add:ww 12345 ; 333444 ;
\end{verbatim}

which can safely perform the addition by grabbing two arguments delimited by \texttt{;}.

If we were to continue parsing the expression, then the following number should also be cleaned up before the next use of a binary operation such as \texttt{\add:ww}. Just like \texttt{\int_value:w} expanded what follows once, we need \texttt{\add:ww} to do the calculation, and in the process to expand the following once. This is also true in our real application: all the functions of the form \texttt{\__fp_..._o:ww} expand what follows once. This comes at the cost of leaving tokens in the input stack, and we need to be careful not to waste this memory. All of our discussion above is nice but simplistic, as operations should not simply be performed in the order they appear.

\subsection*{69.1.2 Precedence and infix operators}

The various operators we will encounter have different precedences, which influence the order of calculations: $1 + 2 \times 3 = 1 + (2 \times 3)$ because $\times$ has a higher precedence than $+$. The true analog of our macro \texttt{\operand:w} must thus take care of that. When looking for an operand, it needs to perform calculations until reaching an operator which has lower precedence than the one which called \texttt{\operand:w}. This means that \texttt{\operand:w} must know what the previous binary operator is, or rather, its precedence: we thus rename it \texttt{\operand:Nw}. Let us describe as an example how we plan to do the calculation $41 - 2^3 + 4 + 5$. More precisely we describe how to perform the first operation in this expression. Here, we abuse notations: the first argument of \texttt{\operand:Nw} should be an integer

\begin{verbatim}
\add:ww 12345 ; 333444 ;
\end{verbatim}
constant (\lc__fp_prec_plus_int, \ldots) equal to the precedence of the given operator, not directly the operator itself.

- Clean up 41 and find -. We call \operand:Nw - to find the second operand.
- Clean up 2 and find ^.
- Compare the precedences of - and ^. Since the latter is higher, we need to compute the exponentiation. For this, find the second operand with a nested call to \operand:Nw ^.
- Clean up 3 and find *.
- Compare the precedences of ^ and *. Since the former is higher, \operand:Nw ^ has found the second operand of the exponentiation, which is computed: $2^3 = 8$.
- We now have 41-8+4+5, and \operand:Nw - is still looking for a second operand for the subtraction. Is it 8?
- Compare the precedences of - and *. Since the latter is higher, we are not done with 8. Call \operand:Nw * to find the second operand of the multiplication.
- Clean up 4, and find +.
- Compare the precedences of * and +. Since the former is higher, \operand:Nw * has found the second operand of the multiplication, which is computed: $8 \times 4 = 32$.
- We now have 41-32+5, and \operand:Nw - is still looking for a second operand for the subtraction. Is it 32?
- Compare the precedences of - and +. Since they are equal, \operand:Nw - has found the second operand for the subtraction, which is computed: $41 - 32 = 9$.
- We now have 9+5.

The procedure above stops short of performing all computations, but adding a surrounding call to \operand:Nw with a very low precedence ensures that all computations are performed before \operand:Nw is done. Adding a trailing marker with the same very low precedence prevents the surrounding \operand:Nw from going beyond the marker.

The pattern above to find an operand for a given operator, is to find one number and the next operator, then compare precedences to know if the next computation should be done. If it should, then perform it after finding its second operand, and look at the next operator, then compare precedences to know if the next computation should be done. This continues until we find that the next computation should not be done. Then, we stop.

We are now ready to get a bit more technical and describe which of the l3fp-parse functions correspond to each step above.

First, \__fp_parse_operand:Nw is the \operand:Nw function above, with small modifications due to expansion issues discussed later. We denote by \langle precedence \rangle the argument of \__fp_parse_operand:Nw, that is, the precedence of the binary operator whose operand we are trying to find. The basic action is to read numbers from the input stream. This is done by \__fp_parse_one:Nw. A first approximation of this function is that it reads one \langle number \rangle, performing no computation, and finds the following binary \langle operator \rangle. Then it expands to
expanding the \texttt{infix} auxiliary before leaving the above in the input stream.

We now explain the \texttt{infix} auxiliaries. We need some flexibility in how we treat the case of equal precedences: most often, the first operation encountered should be performed, such as 1-2-3 being computed as (1-2)-3, but 2^{-3^4} should be evaluated as \(2^{-3^4}\) instead. For this reason, and to support the equivalence between \(\ast\ast\) and \(^\ast\) more easily, each binary operator is converted to a control sequence \texttt{\_fp\_parse\_infix\_}(\operator):N when it is encountered for the first time. Instead of passing both precedences to a test function to do the comparison steps above, we pass the \langle\texttt{precedence}\rangle (of the earlier operator) to the \texttt{infix} auxiliary for the following \langle\operator\rangle, to know whether to perform the computation of the \langle\operator\rangle. If it should not be performed, the \texttt{infix} auxiliary expands to

\begin{verbatim}
\use_none:n \_fp\_parse\_infix\_\langle\operator\rangle:N
\end{verbatim}

and otherwise it calls \texttt{\_fp\_parse\_operand:Nw} with the precedence of the \langle\operator\rangle to find its second operand \langle\texttt{number}\rangle and the next \langle\operator\rangle, and expands to

\begin{verbatim}
\__fp_parse_apply_binary:NwNwN \langle\operator\rangle \langle\texttt{number}\rangle @ \_fp\_parse\_infix\_\langle\operator\rangle:N
\end{verbatim}

The \texttt{infix} function is responsible for comparing precedences, but cannot directly call the computation functions, because the first operand \langle\texttt{number}\rangle is before the \texttt{infix} function in the input stream. This is why we stop the expansion here and give control to another function to close the loop.

A definition of \texttt{\_fp\_parse\_operand:Nw} (\texttt{precedence}) with some of the expansion control removed is

\begin{verbatim}
\exp_after:wN \_fp\_parse\_continue:NwN \langle\operator\rangle \langle\texttt{number}\rangle @ \_fp\_parse\_infix\_\langle\operator\rangle:N
\end{verbatim}

This expands \texttt{\_fp\_parse\_one:Nw} (\texttt{precedence}) completely, which finds a number, wraps the next \langle\operator\rangle into an \texttt{infix} function, feeds this function the \langle\texttt{precedence}\rangle, and expands it, yielding either

\begin{verbatim}
\_fp\_parse\_continue:NwN (\texttt{precedence}) \langle\texttt{number}\rangle 0
\use_none:n \_fp\_parse\_infix\_\langle\operator\rangle:N
\end{verbatim}
or
\begin{verbatim}
\_fp\_parse\_continue:NwN (\texttt{precedence}) \langle\texttt{number}\rangle 0
\_fp\_parse\_apply\_binary:NwNwN \langle\operator\rangle \langle\texttt{number}\rangle 0 \_fp\_parse\_infix\_\langle\operator\rangle:N
\end{verbatim}
The definition of \texttt{\_fp\_parse\_continue:NwN} is then very simple:

\begin{verbatim}
\cs_new:Npn \_fp\_parse\_continue:NwN \#1\#2\#3 \{ \#3 \#1 \#2 0 \}
\end{verbatim}

1006
In the first case, \#3 is \use_none:n, yielding
\use_none:n \langle \text{precedence} \rangle \langle \text{number} \rangle @ 
\_\_fp_parse_infix_ \langle \text{operator} \rangle :N
then \langle \text{number} \rangle @ \_\_fp_parse_infix_ \langle \text{operator} \rangle :N. In the second case, \#3 is \_\_fp_parse_apply_binary:NwNwN, whose role is to compute \langle \text{number} \rangle \langle \text{operator} \rangle \langle \text{number2} \rangle and to prepare for the next comparison of precedences: first we get
\_\_fp_parse_apply_binary:NwNwN
\langle \text{precedence} \rangle \langle \text{number} \rangle @ 
\langle \text{operator} \rangle \langle \text{number2} \rangle @ \_\_fp_parse_infix_ \langle \text{operator} \rangle :N
then
\exp_after:wN \_\_fp_parse_continue:NwN
\exp_after:wN \_\_fp_parse_continue:w
\exp:w \exp_end_continue_f:w 
\_\_fp_ \langle \text{operator} \rangle :o:w \langle \text{number} \rangle @ \langle \text{operator} \rangle @ \langle \text{number2} \rangle \_\_fp_parse_infix_ \langle \text{operator} \rangle :N \langle \text{precedence} \rangle
where \_\_fp_ \langle \text{operator} \rangle :o:w \langle \text{number} \rangle @ \langle \text{operator} \rangle \langle \text{number2} \rangle \langle \text{precedence} \rangle, continuing the loop.

We have introduced the most important functions here, and the next few paragraphs we describe various subtleties.

### 69.1.3 Prefix operators, parentheses, and functions

Prefix operators (unary -, +, !) and parentheses are taken care of by the same mechanism, and functions (\text{sin}, \text{exp}, etc.) as well. Finding the argument of the unary -, for instance, is very similar to grabbing the second operand of a binary infix operator, with a subtle precedence explained below. Once that operand is found, the operator can be applied to it (for the unary -, this simply flips the sign). A left parenthesis is just a prefix operator with a very low precedence equal to that of the closing parenthesis (which is treated as an infix operator, since it normally appears just after numbers), so that all computations are performed until the closing parenthesis. The prefix operator associated to the left parenthesis does not alter its argument, but it removes the closing parenthesis (with some checks).

Prefix operators are the reason why we only summarily described the function \_\_fp_parse_one:Nw earlier. This function is responsible for reading in the input stream the first possible \langle \text{number} \rangle and the next infix \langle \text{operator} \rangle. If what follows \_\_fp_parse_one:Nw \langle \text{precedence} \rangle is a prefix operator, then we must find the operand of this prefix operator through a nested call to \_\_fp_parse_operand:Nw with the appropriate precedence, then apply the operator to the operand found to yield the result of \_\_fp_parse_one:Nw. So far, all is simple.

The unary operators +, -, ! complicate things a little bit: \(-3**2\) should be \(-3^2\) = \(-9\), and not \((-3)^2\) = 9. This would easily be done by giving \(-\) a lower precedence, equal to that of the infix + and -. Unfortunately, this fails in cases such as \(3**-2*4\), yielding \(3^{-2} \times 4\) instead of the correct \(3^{2} \times 4\). A second attempt would be to call \_\_fp_parse_operand:Nw with the \langle \text{precedence} \rangle of the previous operator, but \(0^2-2+3\) is then
parsed as $0 > -(2+3)$: the addition is performed because it binds more tightly than the comparison which precedes $\sim$. The correct approach is for a unary $\sim$ to perform operations whose precedence is greater than both that of the previous operation, and that of the unary $\sim$ itself. The unary $\sim$ is given a precedence higher than multiplication and division. This does not lead to any surprising result, since $-(x/y) = (-x)/y$ and similarly for multiplication, and it reduces the number of nested calls to \__fp_parse_operand:Nw.

Functions are implemented as prefix operators with very high precedence, so that their argument is the first number that can possibly be built.

Note that contrarily to the \textit{infix} functions discussed earlier, the \textit{prefix} functions do perform tests on the previous \textit{precedence} to decide whether to find an argument or not, since we know that we need a number, and must never stop there.

69.1.4 Numbers and reading tokens one by one

So far, we have glossed over one important point: what is a “number”? A number is typically given in the form \langle significand \rangle e \langle exponent \rangle, where the \langle significand \rangle is any non-empty string composed of decimal digits and at most one decimal separator (a period), the exponent \langle e \langle exponent \rangle \rangle is optional and is composed of an exponent mark e followed by a possibly empty string of signs + or - and a non-empty string of decimal digits. The \langle significand \rangle can also be an integer, dimension, skip, or muskip variable, in which case dimensions are converted from points (or mu units) to floating points, and the \langle exponent \rangle can also be an integer variable. Numbers can also be given as floating point variables, or as named constants such as nan, inf or pi. We may add more types in the future.

When \__fp_parse_one:Nw is looking for a “number”, here is what happens.

- If the next token is a control sequence with the meaning of \scan_stop:, it can be: \s__fp, in which case our job is done, as what follows is an internal floating point number, or \s__fp_expr_mark, in which case the expression has come to an early end, as we are still looking for a number here, or something else, in which case we consider the control sequence to be a bad variable resulting from c-expansion.

- If the next token is a control sequence with a different meaning, we assume that it is a register, unpack it with \tex_the:D, and use its value (in pt for dimensions and skips, mu for muskips) as the \langle significand \rangle of a number: we look for an exponent.

- If the next token is a digit, we remove any leading zeros, then read a significand larger than 1 if the next character is a digit, read a significand smaller than 1 if the next character is a period, or we have found a significand equal to 0 otherwise, and look for an exponent.

- If the next token is a letter, we collect more letters until the first non-letter: the resulting word may denote a function such as asin, a constant such as pi or be unknown. In the first case, we call \__fp_parse_operand:Nw to find the argument of the function, then apply the function, before declaring that we are done. Otherwise, we are done, either with the value of the constant, or with the value nan for unknown words.

- If the next token is anything else, we check whether it is a known prefix operator, in which case \__fp_parse_operand:Nw finds its operand. If it is not known, then either a number is missing (if the token is a known infix operator) or the token is simply invalid in floating point expressions.
Once a number is found, \_\_fp\_parse\_one:Nw also finds an infix operator. This goes as follows.

- If the next token is a control sequence, it could be the special marker \s\_\fp\_expr\_mark, and otherwise it is a case of juxtaposing numbers, such as 2\c\_zero\_int, with an implied multiplication.

- If the next token is a letter, it is also a case of juxtaposition, as letters cannot be proper infix operators.

- Otherwise (including in the case of digits), if the token is a known infix operator, the appropriate \_\_fp\_infix\_⟨operator⟩:N function is built, and if it does not exist, we complain. In particular, the juxtaposition \c\_zero\_int 2 is disallowed.

In the above, we need to test whether a character token #1 is a digit:

\if_int_compare:w 9 < 1 \token_to_str:N #1 \exp_stop_f:
   is a digit
\else:
   not a digit
\fi:

To exclude 0, replace 9 by 10. The use of \token_to_str:N ensures that a digit with any catcode is detected. To test if a character token is a letter, we need to work with its character code, testing if ‘#1 lies in [65, 90] (uppercase letters) or [97, 112] (lowercase letters).

\if_int_compare:w \_\_fp\_int\_eval:w
   ( ‘#1 \if_int_compare:w ‘#1 > ‘Z - 32 \fi: ) / 26 = 3 \exp_stop_f:
   is a letter
\else:
   not a letter
\fi:

At all steps, we try to accept all category codes: when #1 is kept to be used later, it is almost always converted to category code other through \token_to_str:N. More precisely, catcodes \{3, 6, 7, 8, 11, 12\} should work without trouble, but not \{1, 2, 4, 10, 13\}, and of course \{0, 5, 9\} cannot become tokens.

Floating point expressions should behave as much as possible like \texttt{\en-US-\LaTeX X}-based integer expressions and dimension expressions. In particular, f-expansion should be performed as the expression is read, token by token, forcing the expansion of protected macros, and ignoring spaces. One advantage of expanding at every step is that restricted expandable functions can then be used in floating point expressions just as they can be in other kinds of expressions. Problematically, spaces stop f-expansion: for instance, the macro \texttt{\en-US-\LaTeX X} below would not be expanded if we simply performed f-expansion.

\DeclareDocumentCommand{\test}{m} { \fp_eval:n {#1} }
\ExplSyntaxOff
\test { 1 + \en-US-\LaTeX X }

Of course, spaces typically do not appear in a code setting, but may very easily come in document-level input, from which some expressions may come. To avoid this problem, at every step, we do essentially what \use: would do: take an argument, put it back.
in the input stream, then \texttt{f}-expand it. This is not a complete solution, since a macro’s expansion could contain leading spaces which would stop the \texttt{f}-expansion before further macro calls are performed. However, in practice it should be enough: in particular, floating point numbers are correctly expanded to the underlying \texttt{s\_fp} \ldots structure. The \texttt{f}-expansion is performed by \texttt{\_fp_parse_expand:w}.

### 69.2 Main auxiliary functions

\begin{verbatim}
\_fp_parse_operand:Nw
\exp:w \_fp_parse_operand:Nw \texttt{\textlangle precedence\textrangle} \_fp_parse_expand:w
\end{verbatim}

Reads the \texttt{\ldots}, performing every computation with a precedence higher than \texttt{\langle precedence\rangle}, then expands to

\texttt{\langle result\rangle \_fp_parse_infix\_\langle operation\rangle:N \ldots}

where the \texttt{\langle operation\rangle} is the first operation with a lower precedence, possibly \texttt{end}, and the \texttt{\ldots} start just after the \texttt{\langle operation\rangle}.

(End definition for \texttt{\_fp_parse_operand:Nw}.)

\begin{verbatim}
\_fp_parse_infix+:N
\_fp_parse_infix+:N \texttt{\textlangle precedence\textrangle} \ldots
\end{verbatim}

If + has a precedence higher than the \texttt{\langle precedence\rangle}, cleans up a second \texttt{\langle operand\rangle} and finds the \texttt{\langle operation\rangle} which follows, and expands to

\texttt{\_fp_parse_apply_binary:NwNwN + \langle operand\rangle \_fp_parse_infix\_\langle operation\rangle:N \ldots}

Otherwise expands to

\texttt{\_use_none:n \_fp_parse_infix+:N \ldots}

A similar function exists for each infix operator.

(End definition for \texttt{\_fp_parse_infix+:N}.)

\begin{verbatim}
\_fp_parse_one:Nw
\_fp_parse_one:Nw \texttt{\textlangle precedence\textrangle} \ldots
\end{verbatim}

Cleans up one or two operands depending on how the precedence of the next operation compares to the \texttt{\langle precedence\rangle}. If the following \texttt{\langle operation\rangle} has a precedence higher than \texttt{\langle precedence\rangle}, expands to

\texttt{\langle operand\rangle \_fp_parse_apply_binary:NwNwN \langle operation\rangle \langle operand\rangle \_fp_parse_infix\_\langle operation\rangle:N \ldots}

and otherwise expands to

\texttt{\_use_none:n \_fp_parse_infix\_\langle operation\rangle:N \ldots}

(End definition for \texttt{\_fp_parse_one:Nw}.)
69.3 Helpers

\__fp_parse_expand:w
\exp:w \__fp_parse_expand:w \langle tokens \rangle

This function must always come within a \exp:w expansion. The \langle tokens \rangle should be the part of the expression that we have not yet read. This requires in particular closing all conditionals properly before expanding.

\cs_new:Npn \__fp_parse_expand:w #1 { \exp_end_continue_f:w #1 }

(End definition for \__fp_parse_expand:w.)

\__fp_parse_return_semicolon:w

This very odd function swaps its position with the following \fi: and removes \__fp_parse_expand normally responsible for expansion. That turns out to be useful.

\cs_new:Npn \__fp_parse_return_semicolon:w #1 \fi: \__fp_parse_expand:w { \fi: ; #1 }

(End definition for \__fp_parse_return_semicolon:w.)

\__fp_parse_digits_vii:N
\__fp_parse_digits_vi:N
\__fp_parse_digits_v:N
\__fp_parse_digits_iv:N
\__fp_parse_digits_iii:N
\__fp_parse_digits_ii:N
\__fp_parse_digits_i:N
\__fp_parse_digits_:N

These functions must be called within an \int_value:w or \__fp_int_eval:w construction. The first token which follows must be f-expanded prior to calling those functions.

The functions read tokens one by one, and output digits into the input stream, until meeting a non-digit, or up to a number of digits equal to their index. The full expansion is

\langle digits \rangle ; \langle filling 0 \rangle ; \langle length \rangle

where \langle filling 0 \rangle is a string of zeros such that \langle digits \rangle \langle filling 0 \rangle has the length given by the index of the function, and \langle length \rangle is the number of zeros in the \langle filling 0 \rangle string. Each function puts a digit into the input stream and calls the next function, until we find a non-digit. We are careful to pass the tested tokens through \token_to_str:N to normalize their category code.

\cs_set_protected:Npn \__fp_tmp:w #1 #2 #3
\cs_new:cpn { __fp_parse_digits_ #1 :N } ##1
\if_int_compare:w 9 < 1 \token_to_str:N ##1 \exp_stop_f:
\token_to_str:N ##1 \exp_after:wN #2 \exp:w
\else:
\__fp_parse_return_semicolon:w #3 ##1
\fi:
\__fp_parse_expand:w

(End definition for \__fp_parse_digits_vii:N and others.)
### 69.4 Parsing one number

This function finds one number, and packs the symbol which follows in an `\_fp_parse_infix...` csname. #1 is the previous (precedence), and #2 the first token of the operand. We distinguish four cases: #2 is equal to `\scan_stop:` in meaning, #2 is a different control sequence, #2 is a digit, and #2 is something else (this last case is split further later). Despite the earlier f-expansion, #2 may still be expandable if it was protected by `\exp_not:N`, as may happen with the \LaTeX{}2e command `\protect`. Using a well placed `\reverse_if:N`, this case is sent to `\_fp_parse_one_fp:NN` which deals with it robustly.

```latex
\__fp_parse_one:Nw \#1 \#2
\{\if_catcode:w \scan_stop: \exp_not:N \#2
\exp_after:wN \if_meaning:w \exp_not:N \#2 \#2 \else:
\exp_after:wN \reverse_if:N
\fi:
\if_meaning:w \scan_stop: \#2
\exp_after:wN \exp_after:wN
\exp_after:wN \__fp_parse_one_fp:NN
\else:
\exp_after:wN \exp_after:wN
\exp_after:wN \__fp_parse_one_register:NN
\fi:
\else:
\if_int_compare:w 9 < 1 \token_to_str:N \#2 \exp_stop_f:
\exp_after:wN \exp_after:wN
\exp_after:wN \__fp_parse_one_digit:NN
\else:
\exp_after:wN \exp_after:wN
\exp_after:wN \__fp_parse_one_other:NN
\fi:
\fi:
\#1 \#2
\}
```

(End definition for `\_fp_parse_one:Nw`.)

This function receives a (precedence) and a control sequence equal to `\scan_stop:` in meaning. There are three cases.

- `\_fp` starts a floating point number, and we call `\_fp_exp_after_f:nw`, which f-expands after the floating point.

- `\_fp_expr_mark` is a premature end, we call `\_fp_exp_after_expr_mark_-f:nw`, which triggers an fp-early-end error.

- For a control sequence not containing `\_fp`, we call `\_fp_exp_after_-f:nw`, causing a bad-variable error.

This scheme is extensible: additional types can be added by starting the variables with a scan mark of the form `\_fp_⟨type⟩` and defining `\_fp_exp_after_⟨type⟩_f:nw`. In all cases, we make sure that the second argument of `\_fp_parse_infix:NN` is correctly expanded. A special case only enabled in \LaTeX{}2e is that if `\protect` is encountered then
the error message mentions the control sequence which follows it rather than `\protect` itself. The test for \LaTeX{} 2e uses `\unexpandable@\protect` rather than `\protect` because `\protect` is often `\scan_stop:` hence “does not exist”.

```latex
\cs_new:Npn \__fp_parse_one_fp:NN #1 #2
\exp_after:wN \__fp_parse_infix:NN \exp_after:wN #1 \exp:w \__fp_parse_expand:w
\exp:w \exp_end_continue_f:w
\cs_new:Npn \__fp_exp_after_expr_mark_f:nw #1
\int_case:nnF { \exp_after:wN \use_i:nnn \use_none:nnn #1 }
\c__fp_prec_comma_int { }
\c__fp_prec_tuple_int { }
\c__fp_prec_end_int
\exp_after:wN \c__fp_empty_tuple_fp
\exp:w \exp_end_continue_f:w
\msg_expandable_error:nn { fp } { early-end }
\exp_after:wN \c_nan_fp \exp:w \exp_end_continue_f:w
\cs_new:cpn { __fp_exp_after_?_f:nw } #1#2
\msg_expandable_error:nnn { kernel } { bad-variable } {#2}
\exp_after:wN \c_nan_fp \exp:w \exp_end_continue_f:w #1
\cs_set_protected:Npn \__fp_tmp:w #1
\cs_if_exist:NT #1
\cs_gset:cpn { __fp_exp_after_?_f:nw } ##1##2
\exp_after:wN \c_nan_fp \exp:w \exp_end_continue_f:w #1
\str_if_eq:nnTF {##2} { \protect } { \use_i:nn } { \use:n }(#2) { \protect }
\cs_if_eq:NNTF ##2 #1 { \use_i:nn } { \use:n }
\msg_expandable_error:nn { fp }
\{ robust-cmd
\}
\msg_expandable_error:nnn { kernel } { bad-variable } {#2}
```

1013
This is called whenever \&2 is a control sequence other than \scan_stop: in meaning. We special-case \wd, \ht, \dp (see later) and otherwise assume that it is a register, but carefully unpack it with \tex_the:D within braces. First, we find the exponent following \&2. Then we unpack \&2 with \tex_the:D, and the auxii auxiliary distinguishes integer registers from dimensions/skips from mskips, according to the presence of a period and/or of pt. For integers, simply convert \langle value \rangle e\langle exponent \rangle to a floating point number with \__fp_parse:n (this is somewhat wasteful). For other registers, the decimal rounding provided by T\TeX{} does not accurately represent the binary value that it manipulates, so we extract this binary value as a number of scaled points with \int_value:w \dim_to_decimal_in_sp:n \langle decimal value \rangle pt, and use an auxiliary of \dim_to_fp:n, which performs the multiplication by $2^{-16}$, correctly rounded.
\cs_new:Npn \__fp_parse_one_register_dim:ww #1; #2;
{\exp_after:wN \__fp_from_dim_test:ww
  \int_value:w #2 \exp_after:wN ,
  \int_value:w \dim_to_decimal_in_sp:n { #1 pt } ;}
\__fp_parse_one_register_special:N
\__fp_parse_one_register_math:NNw
\__fp_parse_one_register_wd:w
\__fp_parse_one_register_wd:Nw

The \wd, \dp, \ht primitives expect an integer argument. We abuse the exponent parser to find the integer argument: simply include the exponent marker e. Once that “exponent” is found, use \tex_the:D to find the box dimension and then copy what we did for dimensions.

\cs_new:Npn \__fp_parse_one_register_special:N #1
{\if_meaning:w \box_wd:N #1 \__fp_parse_one_register_wd:w \fi:
 \if_meaning:w \box_ht:N #1 \__fp_parse_one_register_wd:w \fi:
 \if_meaning:w \box_dp:N #1 \__fp_parse_one_register_wd:w \fi:
 \if_meaning:w \infty #1 \__fp_parse_one_register_math:NNw \infty #1 \fi:
 \if_meaning:w \pi #1 \__fp_parse_one_register_math:NNw \pi #1 \fi:
}
\cs_new:Npn \__fp_parse_one_register_math:NNw #1#2#3#4 \__fp_parse_expand:w
{#3\str_if_eq:nnTF {#1} {#2}
 \msg-expandable-error:nnn
 \{ fp \} \infty-pi \{#1\}
 \c_nan_fp
 \}
 {#4 \__fp_parse_expand:w }
}
\cs_new:Npn \__fp_parse_one_register_wd:w #1#2 \exp_after:wN #3#4 \__fp_parse_expand:w
{\if_meaning:w \__fp_parse_one_register_wd:w \fi:
 \exp_after:wN \__fp_from_dim_test:ww
 \exp_after:wN 0 \exp_after:wN ,
 \int_value:w \dim_to_decimal_in_sp:n { #1 #2 } ;}
\__fp_parse_one_digit:NN

A digit marks the beginning of an explicit floating point number. Once the number is found, we catch the case of overflow and underflow with \__fp-sanitize:WN.
then \_\_fp_parse_infix_after_operand:NwN expands \_\_fp_parse_infix:NN after the number we find, to wrap the following infix operator as required. Finding the number itself begins by removing leading zeros: further steps are described later.

\cs_new:Npn \_\_fp_parse_one_digit:NN #1 \begin{verbatim}
\exp_after:wN \_\_fp_parse_infix_after_operand:NwN \exp_after:wN #1 \exp:w \exp_end_continue_f:w \exp_after:wN \_\_fp_sanitize:wN \int_value:w \__fp_int_eval:w 0 \_\_fp_parse_trim_zeros:N \end{verbatim}
\begin{verbatim}(End definition for \_\_fp_parse_one_digit:NN.)\end{verbatim}

\_\_fp_parse_one_other:NN For this function, #2 is a character token which is not a digit. If it is an ASCII letter, \_\_fp_parse_letters:N beyond this one and give the result to \_\_fp_parse_word:Nw. Otherwise, the character is assumed to be a prefix operator, and we build \_\_fp_parse_prefix_:Nw.
\cs_new:Npn \_\_fp_parse_one_other:NN #1 #2 \begin{verbatim}
\if_int_compare:w \__fp_int_eval:w (#2 \if_int_compare:w #2 > 'Z - 32 \fi:w ) / 26 = 3 \exp_stop_f:
\exp_after:wN \_\_fp_parse_word:Nw \exp_after:wN #1 \exp_after:wN #2 \cs:w \_\_fp_parse_prefix_ \token_to_str:N #2 \__fp_parse_letters:N \exp:w \else:w \exp_after:wN \_\_fp_parse_prefix:NNNN \exp_after:wN #1 \exp_after:wN #2 \cs:w \_\_fp_parse_prefix_ \token_to_str:N \__fp_parse_letters:N \cs_end:w \exp:w \fi:w \_\_fp_parse_expand:w \end{verbatim}
\begin{verbatim}(End definition for \_\_fp_parse_one_other:NN.)\end{verbatim}

\_\_fp_parse_word:Nw \_\_fp_parse_letters:N Finding letters is a simple recursion. Once \_\_fp_parse_letters:N has done its job, we try to build a control sequence from the word #2. If it is a known word, then the corresponding action is taken, and otherwise, we complain about an unknown word, yield \c_nan_fp, and look for the following infix operator. Note that the unknown word could be a mistyped function as well as a mistyped constant, so there is no way to tell whether to look for arguments; we do not. The standard requires “inf” and “infinity” and “nan” to be recognized regardless of case, but we probably don’t want to allow every \l3fp word to have an arbitrary mixture of lower and upper case, so we test and use a differently-named control sequence.
For this function, #1 is the previous \langle precedence \rangle, #2 is the operator just seen, and #3 is a control sequence which implements the operator if it is a known operator. If this control sequence is \scan_stop:, then the operator is in fact unknown. Either the expression is missing a number there (if the operator is valid as an infix operator), and we put \nan, wrapping the infix operator in a csname as appropriate, or the character is simply invalid in floating point expressions, and we continue looking for a number, starting again from \__fp_parse_one:Nw.

\__fp_parse_prefix:NNN

\__fp_parse_prefix_unknown:NNN

\exp_after:wN \c_nan_fp \exp:w \exp_end_continue_f:w
\__fp_parse_infix:NN
\__fp_parse_letter:NN
\__fp_parse_letter:N
\exp_after:wN \__fp_parse_letters:N
\__fp_parse_letters:NN

\__fp_parse_prefix:NNN
\__fp_parse_prefix_unknown:NNN
69.4.1 Numbers: trimming leading zeros

Numbers are parsed as follows: first we trim leading zeros, then if the next character is a digit, start reading a significand $\geq 1$ with the set of functions $\__fp_parse_large...$; if it is a period, the significand is $< 1$; and otherwise it is zero. In the second case, trim additional zeros after the period, counting them for an exponent shift $(exp_i) < 0$, then read the significand with the set of functions $\__fp_parse_small...$. Once the significand is read, read the exponent if $e$ is present.

This function expects an already expanded token. It removes any leading zero, then distinguishes three cases: if the first non-zero token is a digit, then call $\__fp_parse_large:N$ (the significand is $\geq 1$); if it is $.$, then continue trimming zeros with $\__fp_parse_strim_zeros:N$; otherwise, our number is exactly zero, and we call $\__fp_parse_zero:$ to take care of that case.
If we have removed all digits until a period (or if the body started with a period), then enter the “small trim” loop which outputs −1 for each removed 0. Those −1 are added to an integer expression waiting for the exponent. If the first non-zero token is a digit, call \_fp_parse_small:N (our significand is smaller than 1), and otherwise, the number is an exact zero. The name strim stands for “small trim”.

\cs_new:Npn \_fp_parse_strim_zeros:N \#1
\if:w 0 \exp_not:N #1 - 1 \exp_after:wN \_fp_parse_strim_zeros:N \exp:w \else:\_fp_parse_strim_end:w #1 \fi:\_fp_parse_expand:w
\cs_new:Npn \_fp_parse_strim_end:w #1 \fi: \_fp_parse_expand:w
\cs_new:Npn \_fp_parse_zero:
\exp_after:wN ; \exp_after:wN 1 \int_value:w \_fp_parse_exponent:N

\_fp_parse_zero: After reading a significand of 0, find any exponent, then put a sign of 1 for \_fp_\-sanitize:w\N, which removes everything and leaves an exact zero.

\_fp_parse_small:N This function is called after we have passed the decimal separator and removed all leading zeros from the significand. It is followed by a non-zero digit (with any catcode). The goal is to read up to 16 digits. But we can’t do that all at once, because \int_value:w (which allows us to collect digits and continue expanding) can only go up to 9 digits. Hence we grab digits in two steps of 8 digits. Since \#1 is a digit, read seven more digits using \_fp_parse_digits_vii:N. The small_leading auxiliary leaves those digits in the \int_value:w, and grabs some more, or stops if there are no more digits. Then the
pack_leading auxiliary puts the various parts in the appropriate order for the processing further up.

```latex
\texttt{\cs_new:Npn \_\_fp_parse_small:N \ #1}
\texttt{\{}
\texttt{\exp_after:wN \_\_fp_parse_pack_leading:N\#1 \token_to_str:N \#1}
\texttt{\exp_after:wN \_\_fp_int_eval:w 1 \token_to_str:N \#1}
\texttt{\exp_after:wN \_\_fp_parse_small_leading:ww\#1}
\texttt{\int_value:w 1}
\texttt{\exp_after:wN \_\_fp_parse_digits_vii:N}
\texttt{\exp:w \_\_fp_parse_expand:w}
\texttt{\}}
\texttt{(End definition for \_\_fp_parse_small:N.)}
\texttt{\_\_fp_parse_small_leading:ww\#1}
\texttt{(digits) ; (zeros) ; (number of zeros)}

We leave \texttt{(digits)} \texttt{(zeros)} in the input stream: the functions used to grab digits are such that this constitutes digits 1 through 8 of the significand. Then prepare to pack 8 more digits, with an exponent shift of zero (this shift is used in the case of a large significand). If \#4 is a digit, leave it behind for the packing function, and read 6 more digits to reach a total of 15 digits: further digits are involved in the rounding. Otherwise put 8 zeros in to complete the significand, then look for an exponent.

```latex
\texttt{\cs_new:Npn \_\_fp_parse_small_leading:ww\#1 \#1 \#2 \#3 \#4}
\texttt{\{}
\texttt{#1 \#2}
\texttt{\exp_after:wN \_\_fp_parse_pack_trailing:N\#1 \#2 \#3 \#4}
\texttt{\exp_after:wN \_\_fp_int_eval:w 1}
\texttt{\if_int_compare:w 9 \textless \texttt{1} \exp_stop_f:}
\texttt{\token_to_str:N \#4 \exp_stop_f:}
\texttt{\exp_after:wN \_\_fp_parse_small_trailing:ww\#1 \#2 \#3 \#4}
\texttt{\int_value:w 1}
\texttt{\exp_after:wN \_\_fp_parse_digits_vi:N}
\texttt{\exp:w}
\texttt{\}}
\texttt{(End definition for \_\_fp_parse_small_leading:ww\#1.)}
\texttt{\_\_fp_parse_small_trailing:ww\#1}
\texttt{(digits) ; (zeros) ; (number of zeros)}

Leave digits 10 to 15 (arguments \#1 and \#2) in the input stream. If the \texttt{(next token)} is a digit, it is the 16th digit, we keep it, then the small_round auxiliary considers this digit and all further digits to perform the rounding: the function expands to nothing, to \texttt{+0} or to \texttt{+1}. Otherwise, there is no 16-th digit, so we put a 0, and look for an exponent.

```latex
\texttt{\cs_new:Npn \_\_fp_parse_small_trailing:ww\#1 \#1 \#2 \#3 \#4}
\texttt{\{}
\texttt{\#1 \#2}
\texttt{\if_int_compare:w 9 \textless \texttt{1} \exp_stop_f:}
\texttt{\token_to_str:N \#4 \exp_stop_f:}
\texttt{\exp_after:wN \_\_fp_parse_small_round:NN}
```
Those functions are expanded after all the digits are found, we took care of the rounding, as well as the exponent. The last argument is the exponent. The previous five arguments are 8 digits which we pack in groups of 4, and the argument before that is 1, except in the rare case where rounding lead to a carry, in which case the argument is 2. The trailing function has an exponent shift as its first argument, which we add to the exponent found in the e... syntax. If the trailing digits cause a carry, the integer expression for the leading digits is incremented (+1 in the code below). If the leading digits propagate this carry all the way up, the function \_fp_parse_pack_carry:w increments the exponent, and changes the significand from 0000... to 1000.... this is simple because such a carry can only occur to give rise to a power of 10.

\[ \text{End definition for } \_\text{fp}_\text{parse}_\text{small}_\text{trailing:wwNN}. \]

69.4.3 Number: large significand

Parsing a significand larger than 1 is a little bit more difficult than parsing small significands. We need to count the number of digits before the decimal separator, and add that to the final exponent. We also need to test for the presence of a dot each time we run out of digits, and branch to the appropriate parse_small function in those cases.

\[ \text{End definition for } \_\text{fp}_\text{parse}_\text{pack}_\text{trailing:NNNNNww}, \_\text{fp}_\text{parse}_\text{pack}_\text{leading:NNNNNww}, \text{and } \_\text{fp}_\text{parse}_\text{pack}_\text{carry:w}. \]

69.4.3 Number: large significand

Parsing a significand larger than 1 is a little bit more difficult than parsing small significands. We need to count the number of digits before the decimal separator, and add that to the final exponent. We also need to test for the presence of a dot each time we run out of digits, and branch to the appropriate parse_small function in those cases.

This function is followed by the first non-zero digit of a “large” significand (≥ 1). It is called within an integer expression for the exponent. Grab up to 7 more digits, for a total of 8 digits.

\[ \text{End definition for } \_\text{fp}_\text{parse}_\text{large:NN}. \]
We shift the exponent by the number of digits in \#1, namely the target number, 8, minus the \( \text{number of zeros} \) (number of digits missing). Then prepare to pack the 8 first digits. If the \( \text{next token} \) is a digit, read up to 6 more digits (digits 10 to 15). If it is a period, try to grab the end of our 8 first digits, branching to the small functions since the number of digit does not affect the exponent anymore. Finally, if this is the end of the significand, insert the \( \text{zeros} \) to complete the 8 first digits, insert 8 more, and look for an exponent.

We have just read 15 digits. If the \( \text{next token} \) is a digit, then the exponent shift caused by this block of 8 digits is 8, first argument to the \text{pack_trailing} function. We keep the \( \text{digits} \) and this 16-th digit, and find how this should be rounded using \text{__fp_parse_large_round:NN}. Otherwise, the exponent shift is the number of \( \text{digits} \), 7 minus the \( \text{number of zeros} \), and we test for a decimal point. This case happens in 123451234512345.67 with exactly 15 digits before the decimal separator. Then branch to the appropriate small auxiliary, grabbing a few more digits to complement the digits.
we already grabbed. Finally, if this is truly the end of the significand, look for an exponent after using the \( \langle \text{zeros} \rangle \) and providing a 16-th digit of 0.

\[
\begin{align*}
\text{\texttt{\_fp_parse_large_trailing:wwNN}} & \text{ 1 \#1 ; \#2; \#3 \#4} \\
\text{\texttt{\_fp_parse_large_trailing:NNNNNNww}} & \\
\text{\texttt{\_fp_int_eval:w 1 \#1 \token_to_str:N \#4}} \\
\text{\texttt{\_fp_parse_half_prec_int}} \\
\text{\texttt{\_fp_parse_large_round:NN}} \\
\text{\texttt{\_fp_parse_small_trailing:wwNN}} \\
\text{\texttt{\_fp_parse_digits_}} \\
\text{\texttt{\_fp_int_to_roman:w \#3}} \\
\text{\texttt{\exp_after:wN}} \\
\text{\texttt{\_fp_parse_round_loop:N}} \\
\text{\texttt{\_fp_parse_round_up:N}} \\
\text{\texttt{\exp:w}}
\end{align*}
\]

(End definition for \_fp_parse_large_trailing:wwNN.)

69.4.4 Number: beyond 16 digits, rounding

This loop is called when rounding a number (whether the mantissa is small or large). It should appear in an integer expression. This function reads digits one by one, until reaching a non-digit, and adds 1 to the integer expression for each digit. If all digits found are 0, the function ends the expression by ;0, otherwise by ;1. This is done by switching the loop to round up at the first non-zero digit, thus we avoid to test whether digits are 0 or not once we see a first non-zero digit.

\[
\begin{align*}
\text{\texttt{\_fp_parse_round_loop:N \#1}} \\
\text{\texttt{\exp_stop_f:}} \\
\text{\texttt{\_fp_parse_round_loop:N \#1}} \\
\text{\texttt{\_fp_parse_round_up:N}} \\
\text{\texttt{\_fp_parse_round_loop:N}} \\
\text{\texttt{\_fp_parse_round_up:N}} \\
\text{\texttt{\exp:w}}
\end{align*}
\]
\_fp_parse_round_after:wN\n\__fp_parse_round_after:wN\n\_fp_parse_round_loop:N\n\_fp_parse_round_after:wN
\_fp_parse_small_round:NN\n\_fp_parse_round_after:wN

After the loop \_fp_parse_round_loop:N, this function fetches an exponent with \_fp_parse_exponent:N, and combines it with the number of digits counted by \_fp_parse_round_loop:N. At the same time, the result 0 or 1 is added to the surrounding integer expression.

\_fp_parse_small_round:NN\n\_fp_parse_round_after:wN\n
Here, #1 is the digit that we are currently rounding (we only care whether it is even or odd). If #2 is not a digit, then fetch an exponent and expand to \{exponent\} only. Otherwise, we expand to +0 or +1, then \{exponent\}. To decide which, call \_fp_round_s:NNw to know whether to round up, giving it as arguments a sign 0 (all explicit numbers are positive), the digit #1 to round, the first following digit #2, and either +0 or +1 depending on whether the following digits are all zero or not. This last argument is obtained by \_fp_parse_round_loop:N, whose number of digits we discard by multiplying it by 0. The exponent which follows the number is also fetched by \_fp_parse_round_after:wN.
Large numbers are harder to round, as there may be a period in the way. Again, \#1 is the digit that we are currently rounding (we only care whether it is even or odd). If there are no more digits (\#2 is not a digit), then we must test for a period: if there is one, then switch to the rounding function for small significands, otherwise fetch an exponent. If there are more digits (\#2 is a digit), then round, checking with \_\_fp_parse_round_\_loop:N if all further digits vanish, or some are non-zero. This loop is not enough, as it is stopped by a period. After the loop, the aux function tests for a period: if it is present, then we must continue looking for digits, this time discarding the number of digits we find.

\begin{verbatim}
\cs_new:Npn \_fp_parse_large_round:NN #1#2
{ \if_int_compare:w 9 < 1 \token_to_str:N #2 \exp_stop_f: +
  \exp_after:wN \_fp_round_s:NNNw \exp_after:wN 0 \exp_after:wN #1 \exp_after:wN #2 \int_value:w \__fp_int_eval:w
  \exp_after:wN \_fp_parse_large_round_aux:wNN \int_value:w \__fp_int_eval:w 1 \exp_after:wN \_fp_parse_round_loop:N \exp:w \exp_after:wN \_fp_parse_expand:w
\else: \A could be dot, or e, or other
  \exp_after:wN \_fp_parse_exponent:Nw #2 \fi:
  \_fp_parse_expand:w
\}
\cs_new:Npn \_fp_parse_large_round_test:NN #1#2
{ \if:w . \exp_not:N #2 \exp_after:wN \_fp_parse_small_round:NN \exp_after:wN #1 \exp:w
  \_fp_parse_exponent:Nw #2 \fi:
  \_fp_parse_expand:w
}\cs_new:Npn \_fp_parse_large_round_aux:wNN #1 ; #2 #3
{ \if:w + \_fp_int_eval:w 1 \fi:
  \_fp_parse_expand:w
\}
\cs_new:Npn \_fp_parse_large_round:NN and \_fp_parse_round_after:wN
\end{verbatim}
69.4.5 Number: finding the exponent

Expansion is a little bit tricky here, in part because we accept input where multiplication
is implicit.

\_fp_parse:n { 3.2 erf(0.1) }
\_fp_parse:n { 3.2 e\_i_my_int }
\_fp_parse:n { 3.2 \c\_pi_fp }

The first case indicates that just looking one character ahead for an “e” is not enough,
since we would mistake the function erf for an exponent of “rf”. An alternative would
be to look two tokens ahead and check if what follows is a sign or a digit, considering
in that case that we must be finding an exponent. But taking care of the second case
requires that we unpack registers after \texttt{e}. However, blindly expanding the two tokens
ahead completely would break the third example (unpacking is even worse). Indeed, in
the course of reading 3.2, \texttt{c\_pi\_fp} is expanded to \texttt{s\_fp \_fp\_chk:w 1 0 {-1} {3141}}
\ldots; and \texttt{s\_fp} stops the expansion. Expanding two tokens ahead would then force
the expansion of \texttt{\_fp\_chk:w} (despite it being protected), and that function tries to
produce an error.

What can we do? Really, the reason why this last case breaks is that just as \TeX
does, we should read ahead as little as possible. Here, the only case where there may be
an exponent is if the first token ahead is \texttt{e}. Then we expand (and possibly unpack) the
second token.

\texttt{\_fp\_parse\_exponent:Nw}

This auxiliary is convenient to smuggle some material through \texttt{fi}: ending conditional
processing. We place those \texttt{fi:} (argument \#2) at a very odd place because this allows
us to insert \texttt{\_fp\_int\_eval:w} \ldots there if needed.

\texttt{cs\_new:Np \_fp\_parse\_exponent:Nw \#1 \#2 \_fp\_parse\_expand:w}
\{\exp\_after:wN ;
\int\_value:w \#2 \_fp\_parse\_exponent:N \#1 \}

(End definition for \texttt{\_fp\_parse\_exponent:Nw})

\texttt{\_fp\_parse\_exponent:N \_fp\_parse\_exponent\_aux:NN}

This function should be called within an \texttt{\int\_value:w} expansion (or within an integer
expression). It leaves digits of the exponent behind it in the input stream, and terminates
the expansion with a semicolon. If there is no \texttt{e} (or \texttt{E}), leave an exponent of \texttt{0}. If there
is an \texttt{e} or \texttt{E}, expand the next token to run some tests on it. The first rough test is that
if the character code of \#1 is greater than that of \texttt{9} (largest code valid for an exponent,
less than any code valid for an identifier), there was in fact no exponent; otherwise, we
search for the sign of the exponent.

\texttt{cs\_new:Np \_fp\_parse\_exponent:N \#1}
\{
\__fp_parse_exponent\__

Read signs one by one (if there is any).
\__fp_parse_exponent\__

An exponent can be an explicit integer (most common case), or various other things (most of which are invalid).
\__fp_parse_exponent\__
\_fp\_parse\_exponent\_digits:N

Read digits one by one, and leave them behind in the input stream. When finding a non-digit, stop, and insert a semicolon. Note that we do not check for overflow of the exponent, hence there can be a TeX error. It is mostly harmless, except when parsing 0e9876543210, which should be a valid representation of 0, but is not.

\begin{verbatim}
__fp_parse_exponent_digits:N { \if_int_compare:w 9 < 1 \token_to_str:N #1 \exp_stop_f: \exp:w \else: \__fp_parse_return_semicolon:w #1 \fi: \__fp_parse_expand:w }
\end{verbatim}

\_fp\_parse\_exponent\_keep:N

This is the last building block for parsing exponents. The argument \#1 is already fully expanded, and neither + nor - nor a digit. It can be:

- \s__fp, marking the start of an internal floating point, invalid here;
- another control sequence equal to \relax, probably a bad variable;
- a register: in this case we make sure that it is an integer register, not a dimension;
- a character other than +, - or digits, again, an error.

\begin{verbatim}
__fp_parse_exponent_digits:N \cs_new:Npn \__fp_parse_exponent_digits:N \if_catcode:w \scan_stop: \exp_not:N #1 \if_meaning:w \scan_stop: #1 \if:w 0 \__fp_str_if_eq:nn { \s__fp } { \exp_not:N #1 } 0 \msg_expandable_error:nnn { \relax } { after-e } { floating-point~ } \prg_return_true: \else: \if:w 0 \__fp_str_if_eq:nn { \int_value:w #1 } { \tex_the:D #1 } \int_value:w #1 \else: \msg_expandable_error:nnn { \relax } { bad-variable } \prg_return_false: \fi: \else: \if:w 0 \__fp_str_if_eq:nn { \int_value:w #1 } { \tex_the:D #1 } \int_value:w #1 \else: \msg_expandable_error:nnn { \relax } { dimension~#1 } \prg_return_false: \fi: \prg_return_false: \end{verbatim}
69.5 Constants, functions and prefix operators

69.5.1 Prefix operators

A unary + does nothing: we should continue looking for a number.

\__fp_parse_prefix_+:Nw

A unary + does nothing: we should continue looking for a number.

\cs_new_eq:cN { __fp_parse_prefix_+:Nw } __fp_parse_one:Nw

(End definition for \__fp_parse_prefix_+:Nw.)

\__fp_parse_apply_function:NNNw

Here, #1 is a precedence, #2 is some extra data used by some functions, #3 is e.g., \__fp_sin_o:w, and expands once after the calculation, #4 is the operand, and #5 is a \__fp_parse_infix_...:N function. We feed the data #2, and the argument #4, to the function #3, which expands \exp:w thus the infix function #5.

\cs_new:Npn __fp_parse_apply_function:NNNw #1#2#3#4@#5

\__fp_parse_apply_unary:NNNw

In contrast to \__fp_parse_apply_function:NNNw, this checks that the operand #4 is a single argument (namely there is a single ;). We use the fact that any floating point starts with a "safe" token like \s__fp. If there is no argument produce the fp-no-arg error; if there are at least two produce fp-multi-arg. For the error message extract the mathematical function name (such as sin) from the expl3 function that computes it, such as \__fp_sin_o:w.

\cs_new:Npn __fp_parse_apply_unary:NNNw #1#2#3#4@#5

In addition, since there is a single argument we can dispatch on type and check that the resulting function exists. This catches things like sin((1,2)) where it does not make sense to take the sine of a tuple.

\cs_new:Npn __fp_parse_apply_unary:NNNw #1#2#3#4@#5

\__fp_parse_apply_unary_type:NNN

\__fp_parse_apply_unary_error:NNw

In contrast to \__fp_parse_apply_function:NNNw, this checks that the operand #4 is a single argument (namely there is a single ;). We use the fact that any floating point starts with a "safe" token like \s__fp. If there is no argument produce the fp-no-arg error; if there are at least two produce fp-multi-arg. For the error message extract the mathematical function name (such as sin) from the expl3 function that computes it, such as \__fp_sin_o:w.

In addition, since there is a single argument we can dispatch on type and check that the resulting function exists. This catches things like sin((1,2)) where it does not make sense to take the sine of a tuple.
\_fp\_parse\_prefix\_,\_::\_Nw
\_fp\_parse\_prefix\_,\!:\_Nw

The unary \(-\) and boolean not are harder: we parse the operand using a precedence equal to the maximum of the previous precedence \#1 and the precedence \_fp\_prec\_not\_\_-\_int\_ of the unary operator, then call the appropriate \_fp\_⟨\_operation\_⟩\_o:w function, where the \(⟨\_operation\_⟩\) is set\_\_\_sign or not.

\_fp\_parse\_prefix\_,\_::\_Nw

Numbers which start with a decimal separator (a period) end up here. Of course, we do not look for an operand, but for the rest of the number. This function is very similar to \_fp\_parse\_one\_digit:\_Nw but calls \_fp\_parse\_strim\_zeros:\_N to trim zeros after the decimal point, rather than the trim\_zeros function for zeros before the decimal point.
The left parenthesis is treated as a unary prefix operator because it appears in exactly the same settings. If the previous precedence is \c\__fp_prec_func_int we are parsing arguments of a function and commas should not build tuples; otherwise commas should build tuples. We distinguish these cases by precedence: \c\__fp_prec_comma_int for the case of arguments, \c\__fp_prec_tuple_int for the case of tuples. Once the operand is found, the \texttt{lparen_after} auxiliary makes sure that there was a closing parenthesis (otherwise it complains), and leaves in the input stream an operand, fetching the following infix operator.

The right parenthesis can appear as a prefix in two similar cases: in an empty tuple or tuple ending with a comma, or in an empty argument list or argument list ending with a comma, such as in \texttt{max(1,2,) or in \texttt{rand()}}.
69.5.2 Constants

\_\_fp_parse_word_inf:N Some words correspond to constant floating points. The floating point constant is left as a result of \_\_fp_parse_one:Nw after expanding \_\_fp_parse_infix:NN.
\_\_fp_parse_word_nan:N \_\_fp_parse_word_pi:N \_\_fp_parse_word_deci:N \_\_fp_parse_word_true:N \_\_fp_parse_word_false:N
\_\_fp_parse_word_pt:N \_\_fp_parse_word_in:N \_\_fp_parse_word_cm:N \_\_fp_parse_word_mm:N \_\_fp_parse_word_dd:N \_\_fp_parse_word_cc:N \_\_fp_parse_word_nd:N \_\_fp_parse_word_nc:N \_\_fp_parse_word_bp:N \_\_fp_parse_word_sp:N

Dimension units are also floating point constants but their value is not stored as a floating point constant. We give the values explicitly here.
\_\_fp_parse_word_pt:N \_\_fp_parse_word_in:N \_\_fp_parse_word_cm:N \_\_fp_parse_word_mm:N \_\_fp_parse_word_deci:N \_\_fp_parse_word_true:N \_\_fp_parse_word_false:N \_\_fp_parse_word_pt:N \_\_fp_parse_word_in:N \_\_fp_parse_word_cm:N \_\_fp_parse_word_mm:N \_\_fp_parse_word_deci:N \_\_fp_parse_word_true:N \_\_fp_parse_word_false:N

(End definition for \_\_fp_parse_prefix:Nw.)
The font-dependent units \texttt{em} and \texttt{ex} must be evaluated on the fly. We reuse an auxiliary of \texttt{dim_to_fp:n}.

\begin{verbatim}
\tl_map_inline:nn { {em} {ex} }
  {\cs_new:cpn { __fp_parse_word_#1:N } {\exp_after:wN \__fp_from_dim_test:ww \exp_after:wN 0 \exp_after:wN , \int_value:w \dim_to_decimal_in_sp:n { 1 #1 } \exp_after:wN ; \exp:w \exp_end_continue_f:w \__fp_parse_infix:NN}
}\end{verbatim}

(End definition for \texttt{\_fp_parse_word_em:N} and \texttt{\_fp_parse_word_ex:N}.)

69.5.3 Functions

\begin{verbatim}
\cs_new:Npn \__fp_parse_unary_function:NNN #1#2#3
  {\exp_after:wN \__fp_parse_apply_unary:NNNwN \exp_after:wN #3 \exp_after:wN #2 \exp_after:wN #1 \exp:w \__fp_parse_operand:Nw \c__fp_prec_func_int \__fp_parse_expand:w}
\cs_new:Npn \__fp_parse_function:NNN #1#2#3
  {\exp_after:wN \__fp_parse_apply_function:NNNwN \exp_after:wN #3 \exp_after:wN #2 \exp_after:wN #1 \exp:w \__fp_parse_operand:Nw \c__fp_prec_func_int \__fp_parse_expand:w}
\end{verbatim}

(End definition for \texttt{\_fp_parse_unary_function:NNN} and \texttt{\_fp_parse_function:NNN}.)
69.6 Main functions

Start an \texttt{\_fp\_parse:n} expansion so that \texttt{\_fp\_parse:n} expands in two steps. The \texttt{\_fp\_parse_operand:Nw} function performs computations until reaching an operation with precedence \texttt{c\_fp\_prec\_end\_int} or less, namely, the end of the expression. The marker \texttt{\_fp\_expr\_mark} indicates that the next token is an already parsed version of an infix operator, and \texttt{\_fp\_parse\_infix\_end:N} has infinitely negative precedence. Finally, clean up a (well-defined) set of extra tokens and stop the initial expansion with \texttt{\exp-_end:}.

\begin{verbatim}
\cs_new:Npn \_fp_parse:n #1
{\exp:w \exp_after:wN \_fp_parse_after:ww \exp:w \_fp_parse_operand:Nw \c\_fp_prec\_end\_int \_fp_parse_expand:w \_fp_expr_stop \_fp_expr_mark \_fp_parse_infix\_end:N \_fp_parse_stop #2 \exp_end:}
\cs_new:Npn \_fp_parse_after:ww #1@ \_fp_parse_infix\_end:N \_fp_expr_stop #2 { #2 #1 }
\cs_new:Npn \_fp_parse_o:n #1
{\exp:w \exp_after:wN \_fp_parse_after:ww \exp:w \_fp_parse_operand:Nw \c\_fp_prec\_end\_int \_fp_parse_expand:w \_fp_expr_stop \_fp_expr_mark \_fp_parse_infix\_end:N \_fp_parse_stop}
\exp_end:
\end{verbatim}

(End definition for \texttt{\_fp_parse:n}, \texttt{\_fp_parse_o:n}, and \texttt{\_fp_parse_after:ww}.)

\texttt{\_fp_parse_operand:Nw} and \texttt{\_fp_parse_continue:NwN} This is just a shorthand which sets up both \texttt{\_fp_parse_continue:NwN} and \texttt{\_fp_parse_one:N} with the same precedence. Note the trailing \texttt{\exp:w}.

\begin{verbatim}
\cs_new:Npn \_fp_parse_operand:Nw #1
{\exp:w \exp_after:wN \_fp_parse_after:ww \exp:w \_fp_parse_operand:Nw \c\_fp_prec\_end\_int \_fp_parse_expand:w \_fp_expr_stop \_fp_expr_mark \_fp_parse_infix\_end:N \_fp_parse_stop}
\end{verbatim}

(End definition for \texttt{\_fp_parse_operand:Nw} and \texttt{\_fp_parse_continue:NwN}.)
\_fp\_parse\_apply\_binary:NwNwN
\_fp\_parse\_apply\_binary\_chk:NN
\_fp\_parse\_apply\_binary\_error:NNN

Receives \langle precedence \rangle \langle operand1 \rangle \circ \langle operation \rangle \langle operand2 \rangle @ \langle infix command \rangle. Builds the appropriate call to the \langle operation \rangle \#3, dispatching on both types. If the resulting control sequence does not exist, the operation is not allowed.

This is redefined in 13fp-extras.

24191 \cs\_new:Npn \_fp\_parse\_apply\_binary:NwNwN \#1 \#2\#3 \#4 \#5\#6 \#7
24192 { \exp\_after:wN \_fp\_parse\_continue:NwN
24193 \exp\_after:wN \#1
24194 \exp:w \exp\_end\_continue\_f:w
24195 \exp\_after:wN \_fp\_parse\_apply\_binary\_chk:NN
24196 \cs:w
24197 \_fp
24198 \_fp\_type\_from\_scan:N \#2
24199 \#4
24200 \_fp\_type\_from\_scan:N \#5
24201 \#o:ww
24202 \cs\_end:
24203 \#4
24204 \#2\#3 \#5\#6
24205 \exp:w \exp\_end\_continue\_f:w \#7 \#1
24206 \exp:w \exp\_end\_continue\_f:w \#7 \#1
24207 }
24208 \cs\_new:Npn \_fp\_parse\_apply\_binary\_chk:NN \#1\#2
24209 { \if\_meaning:w \scan\_stop: \#1
24210 \_fp\_parse\_apply\_binary\_error:NNN \#2
24211 \fi:
24212 \#1
24213 }
24214 \cs\_new:Npn \_fp\_parse\_apply\_binary\_error:NNN \#1\#2\#3
24215 { \#2
24216 \_fp\_invalid\_operation\_o:NNw \#1
24217 }
24218 (End definition for \_fp\_parse\_apply\_binary:NwNwN, \_fp\_parse\_apply\_binary\_chk:NN, and \_fp\_parse\_apply\_binary\_error:NNN.)

\_fp\_binary\_type\_o:Nww
\_fp\_binary\_rev\_type\_o:Nww

Applies the operator \#1 to its two arguments, dispatching according to their types, and expands once after the result. The rev version swaps its arguments before doing this.

24220 \cs\_new:Npn \_fp\_binary\_type\_o:Nww \#1 \#2\#3 \#4
24221 { \exp\_after:wN \_fp\_binary\_type\_o:Nww \#1 \#2\#3 \#4
24222 \cs:w
24223 \_fp
24224 \_fp\_type\_from\_scan:N \#2
24225 \#1
24226 \_fp\_type\_from\_scan:N \#4
24227 \#o:ww
24228 \cs\_end:
24229 \#1
24230 \#2 \#3 \#4
24231 }
24232 \cs\_new:Npn \_fp\_binary\_rev\_type\_o:Nww \#1 \#2\#3 \#4\#5
24233 {
(End definition for \__fp_binary_type_o:Nww and \__fp_binary_rev_type_o:Nww.)

69.7 Infix operators

\cs_new:Npn \__fp_parse_infix_after_operand:NwN #1 #2;{
  \__fp_exp_after_f:nw { \__fp_parse_infix:NN #1 } #2;
}\cs_new:Npn \__fp_parse_infix:NN #1 #2{
  \if_catcode:w \scan_stop: \exp_not:N #2
    \if:w 0 \__fp_str_if_eq:nn { \s__fp_expr_mark } { \exp_not:N #2 }
    \exp_after:wN \exp_after:wN \exp_after:wN \__fp_parse_infix_juxt:N
    \else:
      \exp_after:wN \__fp_parse_infix_check:NNN \cs:w __fp_parse_infix_ \token_to_str:N #2 :N
      \exp_after:wN \exp_after:wN \exp_after:wN \cs_end:
    \fi:
  \else:
    \if_int_compare:w \__fp_int_eval:w ( '#2 > 'Z - 32 \fi: ) / 26 = 3 \exp_stop_f:
      \exp_after:wN \exp_after:wN \exp_after:wN \__fp_parse_infix_juxt:N
    \else:
      \exp_after:wN \__fp_parse_infix_check:NNN \cs:w __fp_parse_infix_ \token_to_str:N #2 :N
      \exp_after:wN \exp_after:wN \exp_after:wN \cs_end:
    \fi:
    \if_int_compare:w 1036
69.7.1 Closing parentheses and commas

As an infix operator, \s__fp_expr_mark means that the next token (#3) has already gone through \__fp_parse_infix:NN and should be provided the precedence #1. The scan mark #2 is discarded.

\cs_new:Npn \__fp_parse_infix_mark:NNN #1#2#3 { #3 #1 }

(End definition for \__fp_parse_infix_after_paren:NN.)
This one is a little bit odd: force every previous operator to end, regardless of the precedence.

\cs_new:Npn \_fp_parse_infix_end:N #1
{ @ \use_none:n \_fp_parse_infix_end:N }

This is very similar to \_fp_parse_infix_end:N, complaining about an extra closing parenthesis if the previous operator was the beginning of the expression, with precedence \c__fp_prec_end_int.

\cs_set_protected:Npn \_fp_tmp:w #1
\cs_new:Npn #1 ##1
{ \if_int_compare:w ##1 > \c__fp_prec_end_int
\exp_after:wN @
\exp_after:wN \use_none:n
\exp_after:wN #1
\else:
\msg_expandable_error:nnn { fp } { extra } { ) }
\exp_after:wN \_fp_parse_infix:NN
\exp_after:wN ##1
\exp:w \exp_after:wN \_fp_parse_expand:w
\fi:
}
\exp_args:Nc \_fp_tmp:w { \_fp_parse_infix_):N }

As for other infix operations, if the previous operations has higher precedence the comma waits. Otherwise we call \_fp_parse_operand:Nw to read more comma-delimited arguments that \_fp_parse_infix_comma:w simply concatenates into a $@$-delimited array. The first comma in a tuple that is not a function argument is distinguished: in that case call \_fp_parse_apply_comma:NwWwW whose job is to convert the first item of the tuple and an array of the remaining items into a tuple. In contrast to \_fp_parse_apply.binary:NwWwW this function’s operands are not single-object arrays.
\exp_after:wN \__fp_parse_infix_comma:w
\exp:w
\fi:
\__fp_parse_operand:Nw \c__fp_prec_comma_int
\exp_after:wN \__fp_parse_expand:w
\fi:
}
\exp_args:Nc \__fp_tmp:w \{ \__fp_parse_infix_,:N \}
\cs_new:Npn \__fp_parse_infix_comma:w \#1 @
{ \#1 @ \use_none:n }
\cs_new:Npn \__fp_parse_apply_comma:NwNwN \#1 #2@ #3 #4@ #5
{ \exp_after:wN \__fp_parse_continue:NwN
\exp_after:wN \exp_end_continue_f:w
\__fp_exp_after_tuple_f:nw { }\s__fp_tuple \__fp_tuple_chk:w \{ #2 #4 \}; #5 #1 }
(End definition for \__fp_parse_infix_,:N, \__fp_parse_infix_comma:w, and \__fp_parse_apply_comma:NwNwN)

69.7.2 Usual infix operators

As described in the “work plan”, each infix operator has an associated \..._infix... function, a computing function, and precedence, given as arguments to \__fp_tmp:w. Using the general mechanism for arithmetic operations. The power operation must be associative in the opposite order from all others. For this, we use two distinct precedences.

\cs_set_protected:Npn \__fp_tmp:w #1#2#3#4
{ \if_int_compare:w #1 < #3
 \exp_after:wN \@ \exp_after:wN \__fp_parse_apply_binary:NwNwN
 \exp_after:wN #2
 \exp:w
 \__fp_parse_operand:Nw \#4
 \exp_after:wN \__fp_parse_expand:w
 \else:
 \exp_after:wN \@ \exp_after:wN \use_none:n
 \exp_after:wN \#1
 \fi:
}
\exp_args:Nc \__fp_tmp:w \{ \__fp_parse_infix_+:N \}
\c__fp_prec_hatii_int \c__fp_prec_hat_int
\exp_args:Nc \__fp_tmp:w \{ \__fp_parse_infix_-:N \}
\c__fp_prec_juxt_int \c__fp_prec_juxt_int
\exp_args:Nc \__fp_tmp:w \{ \__fp_parse_infix_/:N \}
\c__fp_prec_times_int \c__fp_prec_times_int

1039
69.7.3 Juxtaposition

\_\_fp\_parse\_infix\_(:N

When an opening parenthesis appears where we expect an infix operator, we compute the product of the previous operand and the contents of the parentheses using \_\_fp\_parse\_infix\_mul\:N.

69.7.4 Multi-character cases

\_\_fp\_parse\_infix\_+:N

\_\_fp\_parse\_infix\_|:N

\_\_fp\_parse\_infix\&_N

\_\_fp\_parse\_infix\_or:W

(End definition for \_\_fp\_parse\_infix\_+:N and others.)
69.7.5 Ternary operator

\__fp_parse_infix_?:N
\__fp_parse_infix_:N

\cs_set_protected:Npn \__fp_tmp:w #1#2#3#4
{ \cs_new:Npn #1 ##1
{ \if_int_compare:w ##1 < \c__fp_prec_quest_int
  #4
  \exp_after:wN @
  \exp_after:wN #2
  \exp:w \__fp_parse_operand:Nw #3
  \exp_after:wN \__fp_parse_expand:w
  \else:
  \exp_after:wN @
  \exp_after:wN \use_none:n
  \exp_after:wN #1
  \fi:
}
\exp_args:Nc \__fp_tmp:w { \__fp_parse_infix_?:Nw } | \__fp_parse_infix_or:N
\exp_args:Nc \__fp_tmp:w { \__fp_parse_infix &:Nw } & \__fp_parse_infix_and:N

(End definition for \__fp_parse_infix_?:Nw and \__fp_parse_infix &:Nw.)

69.7.6 Comparisons

\__fp_parse_infix_<:N
\__fp_parse_infix_=:N
\__fp_parse_infix>:N
\__fp_parse_excl_error:
\__fp_parse_compare:NNNNNNN
\__fp_parse_compare_auxi:NNNNNN
\__fp_parse_compare_end:NNNN
\__fp_compare:wNNNNw

\cs_new:cpn { \__fp_parse_infix_<:N } #1
{ \__fp_parse_compare:NNNNNNN #1 1 0 0 0 0 < }
\cs_new:cpn { \__fp_parse_infix_=:N } #1
{ \__fp_parse_compare:NNNNNNN #1 1 0 0 0 0 = }
\cs_new:cpn { \__fp_parse_infix>:N } #1
{ \__fp_parse_compare:NNNNNNN #1 1 0 0 0 0 > }
\cs_new:cpn { \__fp_parse_infix_!:N } #1
{ \__fp_parse_compare:NNNNNNN #1 1 0 0 0 0 ! }

(End definition for \__fp_parse_infix_?:N and \__fp_parse_infix ::N.)

1041
\__fp_parse_infix_<:N

(End definition for \__fp_parse_infix_<:N and others.)

69.8 Tools for functions

\__fp_parse_function_all_fp:o:fw

Followed by \{\langle function name\rangle\} \{\langle code\rangle\} \{\langle float array\rangle\} @ this checks all floats are floating point numbers (no tuples).
This is followed by {⟨function name⟩} {⟨code⟩} ⟨float array⟩ @. It checks that the ⟨float array⟩ consists of one or two floating point numbers (not tuples), then leaves the ⟨code⟩ (if there is one float) or its tail (if there are two floats) followed by the ⟨float array⟩. The ⟨code⟩ should start with a single token such as \__fp_atan_default:w that deals with the single-float case.

The first \__fp_if_type_fp:NTwFw test catches the case of no argument and the case of a tuple argument. The next one distinguishes the case of a single argument (no error, just add \c_one_fp) from a tuple second argument. Finally check there is no further argument.
(End definition for \_fp_parse_function_one_two:nnw and others.)

Apply \#1 to all items in the following tuple and expand once afterwards. The code \#1 should itself expand once after its result.

\cs_new:Npn \_fp_tuple_map_o:nw \ #1 \ #2 ; #3 ; #4 ; #5
\cs_new:Npn \_fp_tuple_map_loop_o:nw \ #1 \ #2 \ #3 ; #4 ; #5
\cs_new:Npn \_fp_tuple_mapthread_o:nww \ #1 \ #2 ; #3
\cs_new:Npn \_fp_tuple_mapthread_loop_o:nw \ #1 \ #2 ; #3 ; #4 ; #5

(End definition for \_fp_tuple_map_o:nw and \_fp_tuple_map_loop_o:nw.)
69.9 Messages

\msg_new:nnn { fp } { deprecated }
\msg_new:nnn { '#1'-deprecated;\-use\-\#2 } { unknown-fp-word }
\msg_new:nnn { fp } { missing }
\msg_new:nnn { fp } { extra }
\msg_new:nnn { fp } { early-end }
\msg_new:nnn { fp } { after-e }
\msg_new:nnn { fp } { missing-number }
\msg_new:nnn { fp } { unknown-symbol }
\msg_new:nnn { fp } { extra-comma }
\msg_new:nnn { fp } { no-arg }
\msg_new:nnn { fp } { multi-arg }
\msg_new:nnn { fp } { num-args }
\msg_new:nnn { fp } { bad-args }
\msg_new:nnn { fp } { infty-pi }
\cs_if_exist:cT { @unexpandable@protect }
\msg_new:nnn { fp } { robust-cmd }
\msg_new:nnn { fp } { robust-command\-\#1 invalid-in-fp-expression! }
{/package}
Chapter 70

l3fp-assign implementation

70.1 Assigning values

Floating point variables are initialized to be +0.

\fp_new:N\cs_new_protected:Npn \fp_new:N #1\cs_new_eq:NN #1 \c_zero_fp \cs_generate_variant:Nn \fp_new:N {c}

(End definition for \fp_new:N. This function is documented on page 240.)

\fp_set:Nn\fp_set:cn\fp_gset:Nn\fp_gset:cn\fp_const:Nn\fp_const:cn\cs_new_protected:Npn \fp_set:Nn #1#2\cs_new_protected:Npn \fp_gset:Nn #1#2\cs_new_protected:Npn \fp_const:Nn #1#2\cs_generate_variant:Nn \fp_set:Nn {c}\cs_generate_variant:Nn \fp_gset:Nn {c}\cs_generate_variant:Nn \fp_const:Nn {c}

Copying a floating point is the same as copying the underlying token list.

\fp_set_eq:NN\fp_set_eq:cN\fp_set_eq:cc\fp_gset_eq:NN\fp_gset_eq:cN\fp_gset_eq:cc\cs_new_eq:NN \fp_set_eq:NN \tl_set_eq:NN\cs_new_eq:NN \fp_gset_eq:NN \tl_gset_eq:NN\cs_generate_variant:Nn \fp_set_eq:NN { c , Nc , cc}\cs_generate_variant:Nn \fp_gset_eq:NN { c , Nc , cc}

(End definition for \fp_set_eq:NN and \fp_gset_eq:NN. These functions are documented on page 240.)

Setting a floating point to zero: copy \c_zero_fp.

\fp_zero:N\fp_zero:c\fp_gzero:N\fp_gzero:c\cs_new_protected:Npn \fp_zero:N #1\cs_new_protected:Npn \fp_gzero:N #1\cs_generate_variant:Nn \fp_zero:N { c }\cs_generate_variant:Nn \fp_gzero:N { c }

Set the floating point to zero, or define it if needed.

\[ \text{\texttt{\textbackslash fp_zero:new:N}} \quad \text{\texttt{\textbackslash fp_zero:new:c}} \quad \text{\texttt{\textbackslash fp_gzero:new:N}} \quad \text{\texttt{\textbackslash fp_gzero:new:c}} \]

These match the equivalent functions in l3int and l3skip.

\[ \text{\texttt{\textbackslash fp_add:Nn}} \quad \text{\texttt{\textbackslash fp_add:cn}} \quad \text{\texttt{\textbackslash fp_gadd:Nn}} \quad \text{\texttt{\textbackslash fp_gadd:cn}} \quad \text{\texttt{\textbackslash fp_sub:Nn}} \quad \text{\texttt{\textbackslash fp_sub:cn}} \quad \text{\texttt{\textbackslash fp_gsub:Nn}} \quad \text{\texttt{\textbackslash fp_gsub:cn}} \]

For the sake of error recovery we should not simply set \#1 to \#1±\#2: for instance, if \#2 is 0)+2, the parsing error would be raised at the last closing parenthesis rather than at the closing parenthesis in the user argument. Thus we evaluate \#2 instead of just putting parentheses. As an optimization we use \texttt{\textbackslash __fp_parse:n} rather than \texttt{\textbackslash fp_eval:n}, which would convert the result away from the internal representation and back.

\[ \text{\texttt{\textbackslash cs_new_protected:Npn \textbackslash fp_zero:new:N \#1}} \]
\[ \text{\texttt{\textbackslash cs_new_protected:Npn \textbackslash fp_gzero:new:N \#1}} \]
\[ \text{\texttt{\textbackslash cs_generate_variant:Nn \textbackslash fp_zero:new:N \{ c \}}} \]
\[ \text{\texttt{\textbackslash cs_generate_variant:Nn \textbackslash fp_gzero:new:N \{ c \}}} \]

(End definition for \texttt{\textbackslash fp_zero:new:N} and \texttt{\textbackslash fp_gzero:new:N}. These functions are documented on page 240.)

\[ \text{\texttt{\textbackslash fp_show:N}} \quad \text{\texttt{\textbackslash fp_show:c}} \quad \text{\texttt{\textbackslash fp_log:N}} \quad \text{\texttt{\textbackslash fp_log:c}} \quad \text{\texttt{\textbackslash __fp_show:NN}} \quad \text{\texttt{\textbackslash __fp_show_validate:w}} \]

This shows the result of computing its argument by passing the right data to \texttt{\textbackslash tl_show:n} or \texttt{\textbackslash tl_log:n}.

\[ \text{\texttt{\textbackslash cs_new_protected:Npn \textbackslash fp_show:N \{ \__fp_show:NN \textbackslash tl_show:n \} \}}} \]
\[ \text{\texttt{\textbackslash cs_new_protected:Npn \textbackslash fp_log:N \{ \__fp_show:NN \textbackslash tl_log:n \} \}}} \]
\[ \text{\texttt{\textbackslash cs_generate_variant:Nn \textbackslash fp_show:N \{ c \}}} \]
\[ \text{\texttt{\textbackslash cs_generate_variant:Nn \textbackslash fp_log:N \{ c \}}} \]
\[ \text{\texttt{\textbackslash cs_new_protected:Npn \textbackslash __fp_show:NN \textbackslash __fp_show_validate:w \#2 \{ fp \}}} \]

(End definition for \texttt{\textbackslash fp_show:N} and others. These functions are documented on page 240.)

## 70.2 Updating values

These match the equivalent functions in l3int and l3skip.

\[ \text{\texttt{\textbackslash cs_new_protected:Npn \textbackslash fp_zero:new:N \#1}} \]
\[ \text{\texttt{\textbackslash cs_new_protected:Npn \textbackslash fp_gzero:new:N \#1}} \]
\[ \text{\texttt{\textbackslash cs_generate_variant:Nn \textbackslash fp_zero:new:N \{ c \}}} \]
\[ \text{\texttt{\textbackslash cs_generate_variant:Nn \textbackslash fp_gzero:new:N \{ c \}}} \]

(End definition for \texttt{\textbackslash fp_zero:new:N} and \texttt{\textbackslash fp_gzero:new:N}. These functions are documented on page 240.)

## 70.3 Showing values

This shows the result of computing its argument by passing the right data to \texttt{\textbackslash tl_show:n} or \texttt{\textbackslash tl_log:n}.

\[ \text{\texttt{\textbackslash cs_new_protected:Npn \textbackslash fp_show:N \{ \__fp_show:NN \textbackslash tl_show:n \} \}}} \]
\[ \text{\texttt{\textbackslash cs_new_protected:Npn \textbackslash fp_log:N \{ \__fp_show:NN \textbackslash tl_log:n \} \}}} \]
\[ \text{\texttt{\textbackslash cs_generate_variant:Nn \textbackslash fp_show:N \{ c \}}} \]
\[ \text{\texttt{\textbackslash cs_generate_variant:Nn \textbackslash fp_log:N \{ c \}}} \]
\[ \text{\texttt{\textbackslash cs_new_protected:Npn \textbackslash __fp_show:NN \textbackslash __fp_show_validate:w \#2 \{ fp \}}} \]

\[ \text{\texttt{\textbackslash __kernel_chk_tl_type:NnnT \#2 \{ fp \}}} \]

(End definition for \texttt{\textbackslash fp_show:N} and others. These functions are documented on page 240.)
Some useful constants and scratch variables

70.4 Some useful constants and scratch variables

\c_one_fp
\c_e_fp
\c_pi_fp
\c_one_degree_fp
\l_tmpa_fp
\l_tmpb_fp
\g_tmpa_fp
\g_tmpb_fp

(End definition for \l_tmpa_fp and others. These variables are documented on page 247.)

⟨/package⟩
Chapter 71

\texttt{l3fp-cl\text{-}logic} Implementation

Those functions may receive a variable number of arguments.

\begin{verbatim}
\__fp_parse_word_max:N \__fp_parse_word_min:N
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__fp_parse_word_max:N { \__fp_parse_function:NNN \__fp_minmax_o:Nw 2 }
\cs_new:Npn \__fp_parse_word_min:N { \__fp_parse_function:NNN \__fp_minmax_o:Nw 0 }
\end{verbatim}

(End definition for \_\_\_fp\_parse\_word\_max:N and \_\_\_fp\_parse\_word\_min:N.)

### 71.1 Syntax of internal functions

- \_\_\_fp\_compare\_npos:nww \{(\text{expo}_1)\} \{\text{body}_1\} ; \{(\text{expo}_2)\} \{\text{body}_2\} ;
- \_\_\_fp\_minmax_o:Nw \langle \text{sign} \rangle \langle \text{floating point array} \rangle
- \_\_\_fp\_not_o:w \langle \text{floating point array} \rangle \langle \text{floating point array} \rangle \langle \text{floating point number only} \rangle
- \_\_\_fp\_\_o:ww \langle \text{floating point} \rangle \langle \text{floating point} \rangle
- \_\_\_fp\_\_|_o:ww \langle \text{floating point} \rangle \langle \text{floating point} \rangle
- \_\_\_fp\_ternary:NwwN, \_\_\_fp\_ternary\_auxi:NwwN, \_\_\_fp\_ternary\_auxii:NwwN
  have to be understood.

### 71.2 Tests

\begin{verbatim}
\fp_if_exist_p:N \fp_if_exist_p:c \fp_if_exist:N \fp_if_exist:c 
\prg_new_eq_conditional:NNn \fp_if_exist:N \cs_if_exist:N \{ TF , T , F , p \}
\prg_new_eq_conditional:NNn \fp_if_exist:c \cs_if_exist:c \{ TF , T , F , p \}
\end{verbatim}

(End definition for \_\_\_fp\_if\_exist:NTF. This function is documented on page 242.)
\texttt{\textbackslash fp\_if\_nan\_p:n}\n\texttt{\textbackslash fp\_if\_nan\_nTF}

Evaluate and check if the result is a floating point of the same kind as \texttt{nan}.

\begin{verbatim}
\prg_new_conditional:Npnn \fp_if_nan:n #1 { TF , T , F , p }
\{ \if:w 3 \exp_last_unbraced:Nf \__fp_kind:w \{ \__fp_parse:n \#1 \}
\prg_return_true:
\else:
\prg_return_false:
\fi:
\}
\end{verbatim}

(End definition for \texttt{\fp\_if\_nan\_nTF}. This function is documented on page 301.)

71.3 Comparison

Within floating point expressions, comparison operators are treated as operations, so we evaluate \#1, then compare with ±0. Tuples are \texttt{true}.

\begin{verbatim}
\prg_new_conditional:Npnn \fp_compare:n #1 { p , T , F , TF }
\{ \exp_after:wN \__fp_compare_return:w
\exp:w \exp_end_continue_f:w \__fp_parse:n \#1
\}
\cs_new:Npn \__fp_compare_return:w #1#2#3;
\{ \if_charcode:w 0 \__fp_if_type_fp:NTwFw
#1 { \__fp_use_i_delimit_by_s_stop:nw #3 \s__fp_stop }
\s__fp 1 \s__fp_stop
\prg_return_false:
\else:
\prg_return_true:
\fi:
\}
\end{verbatim}

(End definition for \texttt{\fp\_compare:nTF} and \texttt{\__fp\_compare\_return:w}. This function is documented on page 244.)

\texttt{\textbackslash fp\_compare\_p:n}\n\texttt{\textbackslash fp\_compare\_nTF}\n\texttt{\__fp\_compare\_return:w}\n\texttt{\textbackslash _fp\_compare\_aux:wn}

Evaluate \#1 and \#3, using an auxiliary to expand both, and feed the two floating point numbers swapped to \texttt{\__fp\_compare\_back\_any:ww}, defined below. Compare the result with ‘\#2’=, which is \texttt{-1} for \texttt{<}, \texttt{0} for \texttt{=}\texttt{,} \texttt{1} for \texttt{>} and \texttt{2} for \texttt{?}.

\begin{verbatim}
\prg_new_conditional:Npnn \fp_compare:nNn #1#2#3 { p , T , F , TF }
\{ \exp_after:wN \__fp_compare_aux:wn
\exp:w \exp_end_continue_f:w \__fp_parse:n \#1 \#3 \}
\__fp_int_eval:w \#2 - '=' \__fp_int_eval_end:
\prg_return_true:
\else:
\prg_return_false:
\fi:
\}
\cs_new:Npn \__fp_compare_aux:wn #1#2; #3
\{ \exp_after:wN \__fp_compare_back_any:ww \#3 \}
\end{verbatim}

1051
\__fp_compare_back_any:ww \__fp_compare_back:ww \__fp_compare_nan:w

\cs_new:Npn \__fp_compare_back_any:ww #1#2; #3
\cs_new:Npn \__fp_compare_back:ww
{
\__fp_if_type_fp:NTwFw #1 { \__fp_if_type_fp:NTwFw #3 \use_i:nn \s__fp \use_ii:nn \s__fp_stop }
\s__fp \use_ii:nn \s__fp_stop
\__fp_compare_back:ww
{
\cs:w
\__fp
\__fp_type_from_scan:N #1
\__compare_back
\__fp_type_from_scan:N #3
:ww
\cs_end:
}
\cs_new:Npn \__fp_compare_back:ww
{\s__fp \__fp_chk:w #1 #2 #3;
\s__fp \__fp_chk:w #4 #5 #6;
{\int_value:w
\if_meaning:w 3 #1 \exp_after:wN \__fp_compare_nan:w \fi:
\if_meaning:w 3 #3 \exp_after:wN \__fp_compare_nan:w \fi:
\if_meaning:w 2 #5 - \fi:
\if_meaning:w #2 #5
\if_meaning:w #1 #4
\if_meaning:w 1 #1
\__fp_compare_npos:nwnw #6; #3;
\else:
0
\fi:
\else:
\if_int_compare:w #4 < #1 - \fi: 1
\fi:
\else:
\if_int_compare:w #1#4 = \c_zero_int
0
\else:
1

(End definition for \fp_compare:nNnTF and \__fp_compare_aux:wn. This function is documented on page 243.)

\__fp_compare_back_any:ww \__fp_compare_back:ww \__fp_compare_nan:w

\exp:w \exp_end_continue_f:w \__fp_parse:n {#2} #1;
}

Expands (in the same way as \int_eval:n) to −1 if \textit{x} < \textit{y}, 0 if \textit{x} = \textit{y}, 1 if \textit{x} > \textit{y}, and 2 otherwise (denoted as \textit{x}?\textit{y}). If either operand is nan, stop the comparison with \__fp_compare_nan:w returning 2. If \textit{x} is negative, swap the outputs 1 and −1 (i.e., > and <); we can henceforth assume that \textit{x} ≥ 0. If \textit{y} ≥ 0, and they have the same type, either they are normal and we compare them with \__fp_compare_npos:nwnw, or they are equal. If \textit{y} ≥ 0, but of a different type, the highest type is a larger number. Finally, if \textit{y} ≤ 0, then \textit{x} > \textit{y}, unless both are zero.
Tuples and floating point numbers are not comparable so return 2 in mixed cases or when tuples have a different number of items. Otherwise compare pairs of items with \_\_fp\_compare\_back\_tuple:ww and any don't match return 2 (as int_value:w 02 \exp\_stop\_f:).

\cs_new:Npn \_\_fp\_compare\_back\_tuple:ww #1; #2; { 2 }
\cs_new:Npn \_\_fp\_tuple\_compare\_back:ww #1; #2; { 2 }
\cs_new:Npn \_\_fp\_tuple\_compare\_back\_tuple:ww
\s__fp\_tuple\_chk:w #1;
\s__fp\_tuple\_chk:w #2;
\{ 
\int\_compare:nNnTF { \_\_fp\_array\_count:n {#1} } = 
{ \_\_fp\_array\_count:n {#2} }
{ 
\int\_value:w 0
\_\_fp\_tuple\_compare\_back\_loop:w #1 { \s\_fp \prg\_break: } ; #2 { \s\_fp \prg\_break: } ; 
\prg\_break\_point:
\exp\_stop\_f:
}
\}
\cs_new:Npn \_\_fp\_tuple\_compare\_back\_loop:w #1#2 ; #3#4 ;
{ 
\if\int\_compare:w
\_\_fp\_compare\_back\_any:ww #1 #2 ; #4 #5 ; = \c\_zero\_int
\else:
2 \exp\_after:w \prg\_break:
\fi:
\_\_fp\_tuple\_compare\_back\_loop:w #3 #5 ;
\}
(End definition for \_\_fp\_compare\_back\_tuple:ww and others.)

\_\_fp\_compare\_npos:nwv
\_fp\_compare\_significant:nmmnnn
\_\_fp\_compare\_npos:nwv \{\langle exp01\rangle \} \{\langle body1\rangle \} ; \{\langle exp02\rangle \} \{\langle body2\rangle \} ;
Within an int_value:w ... \exp\_stop\_f: construction, this expands to 0 if the two numbers are equal, –1 if the first is smaller, and 1 if the first is bigger. First compare the exponents: the larger one denotes the larger number. If they are equal, we must compare significands. If both the first 8 digits and the next 8 digits coincide, the numbers are equal. If only the first 8 digits coincide, the next 8 decide. Otherwise, the first 8 digits are compared.
\cs_new:Npn \_\_fp\_compare\_npos:nwv #1#2 ; #3#4 ;
{ 
\if\int\_compare:w #1 = #3 \exp\_stop\_f:

1053
\input{sections/71-fp-do-loop.tex}

\section{Floating point expression loops}

These are quite easy given the above functions. The \dountil and \dowhile versions execute the body, then test. The \untildo and \whiledo do it the other way round.

\begin{verbatim}
\cs_new:Npn \do_until:nn #1#2 { #2 \fp_compare:nF {#1} { \do_until:nn {#1} {#2} } }
\cs_new:Npn \dowhile:nn #1#2 { #2 \fp_compare:nT {#1} { \dowhile:nn {#1} {#2} } }
\cs_new:Npn \untildo:nn #1#2 { \fp_compare:nF {#1} { \untildo:nn {#1} {#2} } }
\cs_new:Npn \whiledo:nn #1#2 { \fp_compare:nT {#1} { \whiledo:nn {#1} {#2} } }
\end{verbatim}

(End definition for \dountil and others. These functions are documented on page 245.)
\texttt{\textbackslash fp\_do\textbackslash until\:nNnn} As above but not using the \texttt{nNn} syntax.
\begin{verbatim}
\cs_new:Npn \fp\_do\textbackslash until\:nNnn \#1\#2\#3\#4
{ \fp\_compare:nNnF \{\#1\} \#2 \{\#3\} \{\#4\} }
\end{verbatim}
\begin{verbatim}
\cs_new:Npn \fp\_do\textbackslash while\:nNnn \#1\#2\#3\#4
{ \fp\_compare:nNnT \{\#1\} \#2 \{\#3\} \{\#4\} }
\end{verbatim}
\begin{verbatim}
\cs_new:Npn \fp\_until\_do\:nNnn \#1\#2\#3\#4
{ \fp\_compare:nNnF \{\#1\} \#2 \{\#3\} \fp\_until\_do\:nNnn \{\#1\} \#2 \{\#3\} \{\#4\} }
\end{verbatim}
\begin{verbatim}
\cs_new:Npn \fp\_while\_do\:nNnn \#1\#2\#3\#4
{ \fp\_compare:nNnT \{\#1\} \#2 \{\#3\} \fp\_while\_do\:nNnn \{\#1\} \#2 \{\#3\} \{\#4\} }
\end{verbatim}

(End definition for \texttt{\fp\_do\textbackslash until\:nNnn} and others. These functions are documented on page \texttt{244}.)

\texttt{\textbackslash fp\_step\_function:nNN}\texttt{N}
\texttt{\textbackslash fp\_step\_function:nNNc}
\texttt{\_fp\_step\_function:wNN}
\texttt{\_fp\_step\_function:wwNN}
\texttt{\_fp\_step\_function:NnnnN}
\texttt{\_fp\_step\_function:NnNNN}

The approach here is somewhat similar to \texttt{\textbackslash int\_step\_function:nNN}. There are two subtleties: we use the internal parser \texttt{\_fp\_parse:n} to avoid converting back and forth from the internal representation; and (due to rounding) even a non-zero step does not guarantee that the loop counter increases.
\begin{verbatim}
\cs_new:Npn \fp\_step\_function:nNN \#1\#2\#3
{ \exp_after:wN \_fp\_step:wwNN
 \exp:w \exp_end_continue_f:w \_fp\_parse:o:n \{\#1\} 
 \exp:w \exp_end_continue_f:w \_fp\_parse:o:n \{\#2\} 
 \exp:w \exp_end_continue_f:w \_fp\_parse:n \{\#3\} }
\end{verbatim}
\begin{verbatim}
\cs_generate_variant:Nn \fp\_step\_function:nNN { nNNc }
\end{verbatim}
\begin{verbatim}
\% \begin{macrocode}
\% Only floating point numbers (not tuples) are allowed arguments.
\% Only \texttt{\textquote{normal}} floating points (not \texttt{\textbackslash pm 0},
\% \textbackslash pm\texttt{\textbackslash inf}}, \texttt{\textbackslash nan}) can be used as step; if positive,
\% call \cs{\_fp\_step:NnnnN} with argument \texttt{|>}; otherwise \texttt{-<}. This
\% function has one more argument than its integer counterpart, namely
\% the previous value, to catch the case where the loop has made no
\% progress. Conversion to decimal is done just before calling the
\% user's function.
\% \end{macrocode}

1055
\cs_new:Npn \_\_fp_step:wwwN #1#2; #3#4; #5#6; #7
{
  \_\_fp_if_type_fp:NTwFw #1 { } \s__fp \prg_break: \s__fp_stop
  \_\_fp_if_type_fp:NTwFw #3 { } \s__fp \prg_break: \s__fp_stop
  \_\_fp_if_type_fp:NTwFw #5 { } \s__fp \prg_break: \s__fp_stop
  \use_i:nnnn { \_\_fp_step_fp:wwwN #1#2; #3#4; #5#6; #7 }
\prg_break_point:
\use:n

{} \_\_fp_error:nfff { step-tuple } { \fp_to_tl:n { #1#2 ; } }
{ \fp_to_tl:n { #3#4 ; } } { \fp_to_tl:n { #5#6 ; } }
\use:n
{} \_\_fp_error:nnfn { bad-step } { #6}

\cs_new:Npn \_\_fp_step_fp:wwwN #1 ; \s__fp \_\_fp_chk:w #2#3#4 ; #5; #6
{
  \token_if_eq_meaning:NNTF #2 1
  \token_if_eq_meaning:NNTF #3 0
  \_\_fp_step:NnnnnN >
  \_\_fp_step:NnnnnN <
}\use_none:nnnnn

\token_if_eq_meaning:NNTF #2 0
\msg_expandable_error:nnn { kernel }
{ zero-step } {#6}

{} \_\_fp_error:nnfn { bad-step } { }
{ \fp_to_tl:n { \s__fp \_\_fp_chk:w #2#3#4 ; } } {#6}
\use_none:nnnnn

{} \_\_fp_step:NnnnnN #1#2#3#4#5#6

\exp_args:Nf #6 { \_\_fp_to_decimal_dispatch:w #2 }
\_\_fp_step:NfnnnN #1 { \_\_fp_parse:n { #2 + #4 } } {#2} {#4} {#5} #6
\cs_generate_variant:Nn \_\_fp_step:NnnnnN { Nf }

(End definition for \fp_step_function:nnnN and others. This function is documented on page 246.)
As for \texttt{\int_step_inline:nnnn}, create a global function and apply it, following up with a break point.

\begin{verbatim}
\cs_new_protected:Npn \fp_step_inline:nnnn
{ \int_gincr:N \g__kernel_prg_map_int
\exp_args:NNc \__fp_step:NNnnnn
\cs_gset_protected:Npn
\{ \_fp_map_ \int_use:N \g__kernel_prg_map_int :w \}
} \cs_new_protected:Npn \fp_step_variable:nnnNn #1#2#3#4#5
{ \int_gincr:N \g__kernel_prg_map_int
\exp_args:NNc \__fp_step:NNnnnn
\cs_gset_protected:Npx
\{ \_fp_map_ \int_use:N \g__kernel_prg_map_int :w \}
{#1} {#2} {#3}
\tl_set:Nn \exp_not:N #4 {##1}
\exp_not:n {#5}
} \cs_new_protected:Npn \__fp_step:NNnnnn #1#2#3#4#5#6
{ #1 #2 ##1 {#6}
\fp_step_function:nnnN {#3} {#4} {#5} #2
prg_break_point:Nn \scan_stop: \{ \int_gdecr:N \g__kernel_prg_map_int \}
} \end{verbatim}

(End definition for \texttt{\fp_step_inline:nnnn}, \texttt{\fp_step_variable:nnnNn}, and \texttt{\_fp_step:NNnnnn}. These functions are documented on page 246.)

\begin{verbatim}
\cs_new:Npn \_fp_minmax_o:Nw #1
\{ \_fp_parse_function_all_fp_o:fnw
\{ \token_if_eq_meaning:NNTF 0 #1 { min } { max } \}
\_fp_minmax_aux_o:Nw \}
\end{verbatim}

\section*{71.5 Extrema}

First check all operands are floating point numbers. The argument \#1 is 2 to find the maximum of an array \#2 of floating point numbers, and 0 to find the minimum. We read numbers sequentially, keeping track of the largest (smallest) number found so far. If numbers are equal (for instance ±0), the first is kept. We append −\infty (\infty), for the case of an empty array. Since no number is smaller (larger) than that, this additional item only affects the maximum (minimum) in the case of \texttt{\texttt{max()}} and \texttt{\texttt{min()}} with no argument. The weird \texttt{fp-like trailing marker breaks the loop correctly: see the precise definition of \_fp_minmax_loop:Nw.}

\begin{verbatim}
\cs_new:Npn \_fp_minmax_o:Nw \_fp_minmax_aux_o:Nw
\{ \_fp_parse_function_all_fp_o:fnw
\{ \token_if_eq_meaning:NNTF 0 \#1 \{ min \} \{ max \}\}
\_fp_minmax_aux_o:Nw \}
\end{verbatim}

\section*{71.5 Extrema}

First check all operands are floating point numbers. The argument \#1 is 2 to find the maximum of an array \#2 of floating point numbers, and 0 to find the minimum. We read numbers sequentially, keeping track of the largest (smallest) number found so far. If numbers are equal (for instance ±0), the first is kept. We append −\infty (\infty), for the case of an empty array. Since no number is smaller (larger) than that, this additional item only affects the maximum (minimum) in the case of \texttt{\texttt{max()}} and \texttt{\texttt{min()}} with no argument. The weird \texttt{fp-like trailing marker breaks the loop correctly: see the precise definition of \_fp_minmax_loop:Nw.}

\begin{verbatim}
\cs_new:Npn \_fp_minmax_o:Nw \_fp_minmax_aux_o:Nw
\{ \_fp_parse_function_all_fp_o:fnw
\{ \token_if_eq_meaning:NNTF 0 \#1 \{ min \} \{ max \}\}
\_fp_minmax_aux_o:Nw \}
\end{verbatim}
\texttt{\_fp_minmax\_loop:Nww} The first argument is \(-\) or \(+\) to denote the case where the currently largest (smallest) number found (first floating point argument) should be replaced by the new number (second floating point argument). If the new number is \texttt{nan}, keep that as the extremum, unless that extremum is already a \texttt{nan}. Otherwise, compare the two numbers. If the new number is larger (in the case of \texttt{max}) or smaller (in the case of \texttt{min}), the test yields \texttt{true}, and we keep the second number as a new maximum; otherwise we keep the first number. Then loop.

\texttt{\_fp_minmax\_auxi:ww} \texttt{\_fp_minmax\_auxii:ww} Keep the first/second number, and remove the other.

(End definition for \texttt{\_fp_minmax\_loop:Nww}.)
This function is called from within an \texttt{\if_meaning:w} test. Skip to the end of the tests, close the current test with \texttt{\fi:}, clean up, and return the appropriate number with one post-expansion.

\begin{verbatim}
cs_new:Npn \_fp_minmax_break_o:w #1 \fi: \fi: #2 \s__fp #3; #4; {
  \fi: \_fp_exp_after_o:w \s__fp #3; }
\end{verbatim}

71.6 Boolean operations

Return \texttt{true} or \texttt{false}, with two expansions, one to exit the conditional, and one to please \texttt{l3fp-parse}. The first argument is provided by \texttt{l3fp-parse} and is ignored.

\begin{verbatim}
cs_new:Npn \_fp_not_o:w #1 \s__fp \_fp_chk:w #2#3; @ {
  \if_meaning:w 0 #2
    \exp_after:wN \exp_after:wN \exp_after:wN \c_one_fp
  \else:
    \exp_after:wN \exp_after:wN \exp_after:wN \c_zero_fp
  \fi:
  \_fp_exp_after_o:w
} \cs_new:Npn \_fp_tuple_not_o:w #1 @ { \exp_after:wN \c_zero_fp }
\end{verbatim}

For \texttt{and}, if the first number is zero, return it (with the same sign). Otherwise, return the second one. For \texttt{or}, the logic is reversed: if the first number is non-zero, return it, otherwise return the second number: we achieve that by hi-jacking \texttt{\_fp\_\&\_o:ww}, inserting an extra argument, \texttt{\else:}, before \texttt{\s__fp}. In all cases, expand after the floating point number.

\begin{verbatim}
\group_begin: \char_set_catcode_letter:N \& \char_set_catcode_letter:N \|
\cs_new:Npn \_fp\&_o:ww #1 \s__fp \_fp_chk:w #2#3; {
  \if_meaning:w 0 #2 #1
    \_fp_and_return:wNw \s__fp \_fp_chk:w #2#3;
  \fi:
  \_fp_exp_after_o:w \s__fp \_fp_chk:w #2#3;}
\cs_new:Npn \_fp\&_tuple_o:ww #1; { \_fp_exp_after_o:w }
\cs_new:Npn \_fp\&_tuple_o:ww #1; { \_fp_exp_after_tuple_o:w }
\cs_new:Npn \_fp\_\&\_tuple_o:ww #1; #2; { \_fp_exp_after_tuple_o:w #1; }
\group_end:
\end{verbatim}

1059
71.7 Ternary operator

The first function receives the test and the true branch of the `?:` ternary operator. It calls `\__fp_ternary_auxii:NwwN` if the test branch is a floating point number ±0, and otherwise calls `\__fp_ternary_auxi:NwwN`. These functions select one of their two arguments.

```
\cs_new:Npn \__fp_ternary:NwwN #1 #2#3@ #4@ #5
  {\if_meaning:w \__fp_parse_infix_::N #5
    \if_charcode:w 0
      \__fp_if_type_fp:NTwFw #2 { \use_i:nn \__fp_use_i_delimit_by_s_stop:nw #3 \s__fp_stop }
      \s__fp 1 \s__fp_stop
    \exp_after:wN \exp_after:wN \exp_after:wN \__fp_ternary_auxii:NwwN
    \else:
      \exp_after:wN \exp_after:wN \exp_after:wN \__fp_ternary_auxi:NwwN
      \fi:
    \exp_after:wN #1
    \exp:w \exp_end_continue_f:w
    \__fp_exp_after_array_f:w #4 \s__fp_expr_stop
    \else:
      \msg_expandable_error:nnnn { fp } { missing } { : } { ~for~?: }
      \exp_after:wN \__fp_parse_continue:NwN
      \exp_after:wN #1
      \exp:w \exp_end_continue_f:w
      \__fp_exp_after_array_f:w #4 \s__fp_expr_stop
      \exp_after:wN #5
      \exp_after:wN #1
      \fi:
  }
\cs_new:Npn \__fp_ternary_auxi:NwwN #1#2@#3@#4
  { \exp_after:wN \__fp_parse_continue:NwN \exp_after:wN #1
    \exp:w \exp_end_continue_f:w
    \__fp_exp_after_array_f:w #2 \s__fp_expr_stop
    \else:
      \msg_expandable_error:nnnn { fp } { missing } { : } { ~for~?: }
      \exp_after:wN \__fp_parse_continue:NwN
      \exp_after:wN #1
      \exp:w \exp_end_continue_f:w
      \__fp_exp_after_array_f:w #4 \s__fp_expr_stop
      \exp_after:wN #5
      \exp_after:wN #1
      \fi:
  }
\cs_new:Npn \__fp_ternary_auxii:NwwN #1#2@#3@#4
  { \exp_after:wN \__fp_parse_continue:NwN \exp_after:wN #1
    \exp:w \exp_end_continue_f:w
    \__fp_exp_after_array_f:w #2 \s__fp_expr_stop
    \else:
      \msg_expandable_error:nnnn { fp } { missing } { : } { ~for~?: }
      \exp_after:wN \__fp_parse_continue:NwN
      \exp_after:wN #1
      \exp:w \exp_end_continue_f:w
      \__fp_exp_after_array_f:w #4 \s__fp_expr_stop
      \exp_after:wN #5
      \exp_after:wN #1
      \fi:
  }
```

(End definition for `\__fp_&_o:ww` and others.)
\exp_after:wN \__fp_parse_continue:NwN
\exp_after:wN \#1
\exp:w \exp_end_continue_f:w
\__fp_exp_after_array_f:w \#3 \s__fp_expr_stop
\#4 \#1
\}

(End definition for \__fp_ternary:NwwN, \__fp_ternary_auxi:NwwN, and \__fp_ternary_auxii:NwwN.)

(/package)
Chapter 72

\textbf{l3fp-basics Implementation}

The \texttt{l3fp-basics} module implements addition, subtraction, multiplication, and division of two floating points, and the absolute value and sign-changing operations on one floating point. All operations implemented in this module yield the outcome of rounding the infinitely precise result of the operation to the nearest floating point.

Some algorithms used below end up being quite similar to some described in “What Every Computer Scientist Should Know About Floating Point Arithmetic”, by David Goldberg, which can be found at \url{http://cr.yp.to/2005-590/goldberg.pdf}.

\begin{Verbatim}
\__fp_parse_word_abs:N
\__fp_parse_word_logb:N
\__fp_parse_word_sign:N
\__fp_parse_word_sqrt:N
\end{Verbatim}

\textbf{Unary functions.}

\begin{Verbatim}
\cs_new:Npn \__fp_parse_word_abs:N
{ \__fp_parse_unary_function:NNN \__fp_set_sign_o:w 0 }
\cs_new:Npn \__fp_parse_word_logb:N
{ \__fp_parse_unary_function:NNN \__fp_logb_o:w ? }
\cs_new:Npn \__fp_parse_word_sign:N
{ \__fp_parse_unary_function:NNN \__fp_sign_o:w ? }
\cs_new:Npn \__fp_parse_word_sqrt:N
{ \__fp_parse_unary_function:NNN \__fp_sqrt_o:w ? }
\end{Verbatim}

(End definition for \__fp_parse_word_abs:N and others.)

72.1 Addition and subtraction

We define here two functions, \texttt{\_\_fp\_\_o:w} and \texttt{\_\_fp\_+o:w}, which perform the subtraction and addition of their two floating point operands, and expand the tokens following the result once.

A more obscure function, \texttt{\_\_fp\_add\_big\_i\_o:wNww}, is used in \texttt{l3fp-expo}.

The logic goes as follows:

- \texttt{\_\_fp\_\_o:w} calls \texttt{\_\_fp\_+o:w} to do the work, with the sign of the second operand flipped;

- \texttt{\_\_fp\_+o:w} dispatches depending on the type of floating point, calling specialized auxiliaries;
• in all cases except summing two normal floating point numbers, we return one or the other operands depending on the signs, or detect an invalid operation in the case of \( \infty - \infty \);

• for normal floating point numbers, compare the signs;

• to add two floating point numbers of the same sign or of opposite signs, shift the significand of the smaller one to match the bigger one, perform the addition or subtraction of significands, check for a carry, round, and pack using the \_\_fp_-\_basics_pack\_\_... functions.

The trickiest part is to round correctly when adding or subtracting normal floating point numbers.

72.1.1 Sign, exponent, and special numbers

The \_\_fp_-\_o:ww auxiliary has a hook: it takes one argument between the first \s-_fp and \_\_fp_chk:ww, which is applied to the sign of the second operand. Positioning the hook there means that \_\_fp_-\_o:ww can still perform the sanity check that it was followed by \s-_fp.

\cs_new:cpx { \_\_fp_-\_o:ww } \s-_fp
\{\exp_not:c { \_\_fp_-\_o:ww } \s-_fp
\exp_not:n { \s-_fp \_\_fp_neg_sign:N }
\}

(End definition for \_\_fp_-\_o:ww.)

This function is either called directly with an empty \#1 to compute an addition, or it is called by \_\_fp_-\_o:ww with \_\_fp_neg_sign:N as \#1 to compute a subtraction, in which case the second operand's sign should be changed. If the \{\textit{types}\} \#2 and \#4 are the same, dispatch to case \#2 (0, 1, 2, or 3), where we call specialized functions: thanks to \int_value:w, those receive the tweaked \{\textit{sign2}\} (expansion of \#1\#5) as an argument. If the \{\textit{types}\} are distinct, the result is simply the floating point number with the highest \{\textit{type}\}. Since case 3 (used for two \textit{nan}) also picks the first operand, we can also use it when \{\textit{type1}\} is greater than \{\textit{type2}\}. Also note that we don't need to worry about \{\textit{sign2}\} in that case since the second operand is discarded.

\cs_new:cpn { \_\_fp_+\_o:ww } \s-_fp \{ \if_case:w \exp_stop_f:
\if_meaning:w \#2 \#4 \else:
\if_int_compare:w \#2 > \#4 \exp_stop_f:
\else:
\else:
\fi:
\ti:
\exp_stop_f:
\exp_after:wN \_\_fp_add_zeros_o:Nww \int_value:w
\exp_after:wN \_\_fp_add_normal_o:Nww \int_value:w
1063
\or: \exp_after:wN \__fp_add_inf_o:Nww \int_value:w
\or: \__fp_case_return_i_o:ww
\else: \exp_after:wN \__fp_add_return_i_o:o:ww \int_value:w
\fi:
\fi: #1 #5
\if: \__fp \__fp_chk:w #2 #3 ;
\s__fp \__fp_chk:w #4 #5
\}

(End definition for \__fp_add_return_i_o:ww.)

\__fp_add_return_i_o:ww Ignore the first operand, and return the second, but using the sign #1 rather than #4. As usual, expand after the floating point.
\cs_new:Npn \__fp_add_return_i_o:ww #1 #2 ; \s__fp \__fp_chk:w #1 #3
{ \__fp_exp_after_o:w \s__fp \__fp_chk:w #1 #3 #1 }

(End definition for \__fp_add_return_i_o:ww.)

\__fp_add_zeros_o:ww Adding two zeros yields \c_zero_fp, except if both zeros were −0.
\cs_new:Npn \__fp_add_zeros_o:ww #1 \s__fp \__fp_chk:w 0 #2
{ \if_int_compare:w #2 #1 = 20 \exp_stop_f:
  \exp_after:wN \__fp_add_return_i_o:o:ww \__fp_add_return_i_o:o:ww
\else:
  \__fp_case_return_i_o:ww
\fi: #1
\s__fp \__fp_chk:w 0 #2
}

(End definition for \__fp_add_zeros_o:ww.)

\__fp_add_inf_o:ww If both infinities have the same sign, just return that infinity, otherwise, it is an invalid operation. We find out if that invalid operation is an addition or a subtraction by testing whether the tweaked \langle sign2 \rangle (#1) and the \langle sign2 \rangle (#4) are identical.
\cs_new:Npn \__fp_add_inf_o:ww #1 \s__fp \__fp_chk:w 2 #2 #3; \s__fp \__fp_chk:w 2 #4
{ \if_meaning:w #1 #2
  \__fp_case_return_i_o:ww
\else:
  \__fp_case_use:nw
  \ifexpr_last_unbraced:Nf \__fp_invalid_operation_o:Nww \__fp_invalid_operation_o:Nww
  \token_if_eq_meaning:NNTF #1 #4 + - }
\fi: #1
\s__fp \__fp_chk:w 2 #2 #3;
\s__fp \__fp_chk:w 2 #4
}

(End definition for \__fp_add_inf_o:ww.)
We now have two normal numbers to add, and we have to check signs and exponents more carefully before performing the addition.

(End definition for \_\_fp_add_normal_o:Nww.)

72.1.2 Absolute addition

In this subsection, we perform the addition of two positive normal numbers.

Since we are doing an addition, the final sign is \langle sign1 \rangle. Start an \__fp_int_eval:w, responsible for computing the exponent: the result, and the \langle final sign \rangle are then given to \__fp_sanitise:Nw which checks for overflow. The exponent is computed as the largest exponent #2 or #5, incremented if there is a carry. To add the significands, we decimate the smaller number by the difference between the exponents. This is done by \__fp_minus_add_big_i:o:NNww or \__fp_minus_add_big_ii:o:NNww. We need to bring the final sign with us in the midst of the calculation to round properly at the end.

(End definition for \_\_fp_add_npos_o:NwwNww.)

Used in l3fp-expo. Shift the significand of the small number, then add with \__fp_minus_add_significand_o:NnnwnnnN.
To round properly, we must know at which digit the rounding should occur. This requires to know whether the addition produces an overall carry or not. Thus, we do the computation now and check for a carry, then go back and do the rounding. The rounding may cause a carry in very rare cases such as $0.99\cdots95 \rightarrow 1.00\cdots0$, but this situation always gives an exact power of 10, for which it is easy to correct the result at the end.

If there's no carry, grab all the digits again and round. The packing function \texttt{\_fp_add\_significand\_pack:NNNNNN} takes care of the case where rounding brings a carry.
The case where there is a carry is very similar. Rounding can even raise the first
digit from 1 to 2, but we don’t care.

Rounding properly in some modes requires to know what the sign of the result will
be. Thus, we start by comparing the exponents and significands. If the numbers coincide,
return zero. If the second number is larger, swap the numbers and call \_fp\_sub\_npos\_ii_o:NNw
with the opposite of \(\text{sign}_1\).

Rounding properly in some modes requires to know what the sign of the result will
be. Thus, we start by comparing the exponents and significands. If the numbers coincide,
return zero. If the second number is larger, swap the numbers and call \_fp\_sub\_npos\_ii_o:NNw
with the opposite of \(\text{sign}_1\).
\__fp_sub_npos_i_o:Nnwnw

After the computation is done, \__fp_sanitize:Nw checks for overflow/underflow. It expects the \langle final sign \rangle and the \langle exponent \rangle (delimited by ;). Start an integer expression for the exponent, which starts with the exponent of the largest number, and may be decreased if the two numbers are very close. If the two numbers have the same exponent, call the near auxiliary. Otherwise, decimate \( y \), then call the far auxiliary to evaluate the difference between the two significands. Note that we decimate by 1 less than one could expect.

\cs_new:Npn \__fp_sub_npos_i_o:Nnwnw \#1 \#2 \#3; \#4 \#5; {
  \exp_after:wN \__fp_sanitize:Nw
  \exp_after:wN \#1
  \int_value:w \__fp_int_eval:w \#2
  \if_int_compare:w \#2 = \#4 \exp_stop_f:
    \exp_after:wN \__fp_sub_back_near_o:nnnnnnnnN
  \else:
    \exp_after:wN \__fp_decimate:nNnnnn \exp_after:wN
    \{ \int_value:w \__fp_int_eval:w \#2 - \#4 - 1 \exp_after:wN \}
    \exp_after:wN \__fp_sub_back_far_o:Nnnwnnnn
  \fi:
  \#5
  \#3
  \#1
}

(End definition for \__fp_sub_npos_i_o:Nnwnw.)

\__fp_sub_back_near_o:nnnnnnnnN \__fp_sub_back_near_pack:NNNNN
\__fp_sub_back_after:wNNNN

In this case, the subtraction is exact, so we discard the \langle final sign \rangle \#9. The very large shifts of \( 10^9 \) and \( 1.1 \cdot 10^9 \) are unnecessary here, but allow the auxiliaries to be reused later. Each integer expression produces a 10 digit result. If the resulting 16 digits start with a 0, then we need to shift the group, padding with trailing zeros.

\cs_new:Npn \__fp_sub_back_near_o:nnnnnnnnN \{ \#1 \} \{ \#2 \} \{ \#3 \} \{ \#4 \} \{ \#5 \} \{ \#6 \} \langle final sign \rangle

\cs_new:Npn \__fp_sub_back_near_pack:NNNNN \#1 \#2 \#3 \#4 \#5 \#6
\cs_new:Npn \__fp_sub_back_after:wNNNN \#1 \#2 \#3 \#4 \#5

(End definition for \__fp_sub_back_near_o:NNNNN.)
\fi:
; \ifnum\cntcnt=\#5 \fi
}

(End definition for \_\_fp_sub_back_near_o:nnnnnnnnN, \_\_fp_sub_back_near_pack:NNNNNNw, and \_\_fp_sub_back_near_after:WWWWww.)

\_\_fp_sub_back_shift:wnnnn
\_\_fp_sub_back_shift_ii:ww
\_\_fp_sub_back_shift_iii:NNNNNNww
\_\_fp_sub_back_shift_iv:nnnnw
\cs_new:Npn \__fp_sub_back_shift:wnnnn ; #1#2#3#4; {#5}
  \ifnum\cntcnt=\#5
    \exp_after:wN \__fp_sub_back_shift_ii:ww
    \int_value:w #1 #2 0 ;
  \else:
    \__fp_sub_back_shift_iii:NNNNNNww #1 123456789;
    \fi:
\exp_after:wN \__fp_pack_twice_four:wNNNNNNww
\exp_after:wN \__fp_pack_twice_four:wNNNNNNww
\exp_after:wN \__fp_sub_back_shift_iv:nnnnw
\exp_after:wN ;
\int_value:w
#1 ~ #2#3 0 ~ 0000 0000 0000 000 ;
\fi:
\cs_new:Npn \__fp_sub_back_shift_ii:ww #1 0 ; #2#3; { ; #1 ; }
\begin{Verbatim}
\end{Verbatim}
\cs_new:Npn \__fp_sub_back_far_o:NnnwnnnN #1 #2#3 #4; #5#6#7#8
\cs_new:Npn \__fp_sub_back_far_o:NnnwnnnN #1 #2#3#4#5#6#7#8#9; {#10}
\cs_new:Npn \__fp_sub_back_far_o:NnnwnnnN #1 #2#3#4; {#5#6#7#8}

(End definition for \_\_fp_sub_back_shift:wnnnn and others.)

\_\_fp_sub_back_far_o:NNnnnn
\_\_fp_sub_back_far_o:NNnnnnN (rounding) \{ \(Y_1\)\} \{ \(Y_2\)\}
\_\_fp_sub_back_far_o:NNnnnnN (extra-digits) \{ \(X_1\)\} \{ \(X_2\)\} \{ \(X_3\)\} \{ \(X_4\)\}
\_\_fp_sub_back_far_o:NNnnnnN (final sign)

If the difference is greater than \texttt{10^{exp_{pos}}}, call the very_far auxiliary. If the result is less than \texttt{10^{exp_{neg}}}, call the not_far auxiliary. If it is too close a call to know yet, namely if \texttt{1\langle Y_1\rangle \langle Y_2\rangle} = \texttt{\langle X_1\rangle \langle X_2\rangle \langle X_3\rangle \langle X_4\rangle 0}, then call the quite_far auxiliary. We use the odd combination of space and semi-colon delimiters to allow the not_far auxiliary to grab each piece individually, the very_far auxiliary to use \_\_fp_pack_eight:wwNNNNNNNN, and the quite_far to ignore the significands easily (using the ; delimiter).
\if_case:w
  \if_int_compare:w #2 = #5#6 \use_i:nnnn #7 \exp_stop_f:
    \if_int_compare:w #3 = \use_none:n #7#8 0 \exp_stop_f:
      0
    \else:
      \if_int_compare:w #3 > \use_none:n #7#8 0 - \fi: 1
    \fi:
  \else:
    \if_int_compare:w 1 #2 > #5#6 \use_i:nnnn #7 - \fi: 1
  \fi:
  \exp_stop_f:
  \exp_after:wN \__fp_sub_back_quite_far_o:wwNN
  \or: \exp_after:wN \__fp_sub_back_very_far_o:wNNNN
  \else: \exp_after:wN \__fp_sub_back_not_far_o:wNNNN
  \fi:
  #2 ~ #3 ; #5 #6 ~ #7 #8 ; #1
\}
(End definition for \__fp_sub_back_far_o:NnnwnnnnN.)
\__fp_sub_back_quite_far_o:wwNN
\__fp_sub_back_quite_far_ii:NN
The easiest case is when \(x - y\) is extremely close to a power of 10, namely the first digit of \(x\) is 1, and all others vanish when subtracting \(y\). Then the \(\langle\text{rounding}\rangle\) #3 and the \(\langle\text{final sign}\rangle\) #4 control whether we get 1 or 0.9999999999999999. In the usual round-to-nearest mode, we get 1 whenever the \(\langle\text{rounding}\rangle\) digit is less than or equal to 5 (remember that the \(\langle\text{rounding}\rangle\) digit is only equal to 5 if there was no further non-zero digit).
\cs_new:Npn \__fp_sub_back_quite_far_o:wwNN #1; #2; #3#4
  { \exp_after:wN \__fp_sub_back_quite_far_ii:NN #3#4 }\cs_new:Npn \__fp_sub_back_quite_far_ii:NN #1#2
  { \if_case:w \__fp_round_neg:NNN #2 0 #1 \exp_after:wN \use_i:nn
    \else:
      \exp_after:wN \use_ii:nn
    \fi:
      { ; {1000} {0000} {0000} {0000} ; }
      { - 1 ; {9999} {9999} {9999} {9999} ; }
  }\exp_after:wN \__fp_round_neg:NNN #2 0 #1
  \exp_after:wN \use_i:nn
  \else:
    \exp_after:wN \use_ii:nn
  \fi:
  \__fp_sub_back_quite_far_o:wwNN #1; #2; #3#4
\__fp_sub_back_not_far_o:wNNNN
In the present case, \(x\) and \(y\) have different exponents, but \(y\) is large enough that \(x - y\) has a smaller exponent than \(x\). Decrement the exponent (with -1). Then proceed in a way similar to the near auxiliaries seen earlier, but multiplying \(x\) by 10 (#30 and #40 below), and with the added quirk that the \(\langle\text{rounding}\rangle\) digit has to be taken into account. Namely, we may have to decrease the result by one unit if \__fp_round_neg:NNN returns 1. This function expects the \(\langle\text{final sign}\rangle\) #6, the last digit of 1100000000+40-2, and the \(\langle\text{rounding}\rangle\) digit. Instead of redoing the computation for the second argument, we note that \__fp_round_neg:NNN only cares about its parity, which is identical to that of the last digit of #2.
The case where $x - y$ and $x$ have the same exponent is a bit more tricky, mostly because it cannot reuse the same auxiliaries. Shift the $y$ significand by adding a leading 0. Then the logic is similar to the not_far functions above. Rounding is a bit more complicated: we have two (rounding) digits #3 and #6 (from the decimation, and from the new shift) to take into account, and getting the parity of the main result requires a computation. The first \int_value:w triggers the second one because the number is unfinished; we can thus not use 0 in place of 2 there.

\begin{verbatim}
\cs_new:Npn \__fp_sub_back_not_far_o:wwwwNN #1 - #2; #3 - #4; #5#6
\{ - 1 \exp_after:wN \__fp_sub_back_near_after:wNNNw \int_value:w \__fp_int_eval:w 1 #30 - #1 - 11 \exp_after:wN \__fp_sub_back_near_pack:NNNNNNw \int_value:w \__fp_int_eval:w 11 0000 0000 + #40 - #2 - \exp_after:wN \__fp_round_neg:NNN \exp_after:wN #6 \use_none:nnnnnnn #2 #5 \exp_after:wN ; \}
\end{verbatim}

(End definition for \__fp_sub_back_not_far_o:wwwwNN.)

72.2 Multiplication

72.2.1 Signs, and special numbers

\_fp_*_o:ww We go through an auxiliary, which is common with \_fp_/o:ww. The first argument is the operation, used for the invalid operation exception. The second is inserted in a formula to dispatch cases slightly differently between multiplication and division. The
third is the operation for normal floating points. The fourth is there for extra cases
needed in \_\_fp/_\_o:ww.

\cs_new:c { \_\_fp_*_o:ww }
\__fp_mul_cases_o:NnNnww
\*\{ - 2 + \}
\__fp_mul_npos_o:Nww\{ \}

(End definition for \_\_fp_*_o:ww.)

\_\_fp_mul_cases_o:nNnnww Split into 10 cases (12 for division). If both numbers are normal, go to case 0 (same sign)
or case 1 (opposite signs): in both cases, call \_\_fp_mul_npos_o:Nww to do the work. If
the first operand is nan, go to case 2, in which the second operand is discarded; if the
second operand is nan, go to case 3, in which the first operand is discarded (note the
weird interaction with the final test on signs). Then we separate the case where the first
number is normal and the second is zero: this goes to cases 4 and 5 for multiplication,
10 and 11 for division. Otherwise, we do a computation which dispatches the products
0\times0 = 0\times1 = 1\times0 = 0 to case 4 or 5 depending on the combined sign, the products 0\times\infty
and \infty\times0 to case 6 or 7 (invalid operation), and the products \infty\times\infty = \infty\times\infty = \infty
to cases 8 and 9. Note that the code for these two cases (which return ±\infty) is inserted
as argument #4, because it differs in the case of divisions.

\cs_new:Npn \_\_fp_mul_cases_o:NnNnww
\#1\#2\#3\#4 \s__fp \__fp_chk:w \#5\#6\#7; \s__fp \__fp_chk:w \#8\#9
\__fp_int_eval:w \_\_fp_int_eval:w
\__fp_case_return_i_o:ww \c_zero_fp
\_\_fp_case_return_ii_o:ww
72.2.2 Absolute multiplication

In this subsection, we perform the multiplication of two positive normal numbers.

\Reporter{\_fp.mpos.np:Nw} \Reporter{\_fp chk:w} 1 \Reporter{\_fp sanitize:Nw} \Reporter{\_fp int eval:w} \Reporter{\_fp basics pack low:NWNNW} \Reporter{\_fp basics pack high:NWNNW}

Note the three semicolons at the end of the definition. One is for the last \Reporter{\_fp_mul_significant drop:NNNNNw}; one is for \Reporter{\_fp round digit:NW later on}; and one, preceded by \Reporter{\_fp after:WN}, which is correctly expanded (within an \Reporter{\_fp int eval:w}), is used by \Reporter{\_fp basics pack low:NWNNW}.

The product of two 16 digit integers has 31 or 32 digits, but it is impossible to know which one before computing. The place where we round depends on that number of digits, and may depend on all digits until the last in some rare cases. The approach is thus to compute the 5 first blocks of 4 digits (the first one is between 100 and 9999 inclusive), and a compact version of the remaining 3 blocks. Afterwards, the number of digits is known, and we can do the rounding within yet another set of \Reporter{\_fp int eval:w}. 

\Reporter{\_fp int eval:w} 99990000 + \Reporter{\_fp int eval:w} #1*#6 +
\exp_after:wN \__fp_mul_significand_keep:NNNNNw
\int_value:w \__fp_int_eval:w 99990000 + #1*#7 + #2*#6 +
\exp_after:wN \__fp_mul_significand_drop:NNNNNw
\int_value:w \__fp_int_eval:w 99990000 + #1*#8 + #2*#7 + #3*#6 +
\exp_after:wN \__fp_mul_significand_drop:NNNNNw
\int_value:w \__fp_int_eval:w 99990000 + #1*#9 + #2*#8 +
#3*#7 + #4*#6 +
\exp_after:wN \__fp_mul_significand_drop:NNNNNw
\int_value:w \__fp_int_eval:w 99990000 + #2*#9 + #3*#8 +
#4*#7 +
\exp_after:wN \__fp_mul_significand_drop:NNNNNw
\int_value:w \__fp_int_eval:w 99990000 + #3*#9 + #4*#8 +
\exp_after:wN \__fp_mul_significand_drop:NNNNNw
\int_value:w \__fp_int_eval:w 100000000 + #4*#9 ;
; \exp_after:wN ;
}
\cs_new:Npn \__fp_mul_significand_drop:NNNNNw #1#2#3#4#5 #6;
{ #1#2#3#4#5 ; + #6 }
\cs_new:Npn \__fp_mul_significand_keep:NNNNNw #1#2#3#4#5 #6;
{ #1#2#3#4#5 ; #6 ; }
\__fp_mul_significand_test_f:NNN \__fp_mul_significand_test_f:NNN
\langle \text{sign} \rangle 1 \langle \text{digits 1–8} \rangle ; \langle \text{digits 9–12} \rangle ;
\langle \text{digits 13–16} \rangle ; + \langle \text{digits 17–20} \rangle + \langle \text{digits 21–24} \rangle + \langle \text{digits 25–28} \rangle + \langle \text{digits 29–32} \rangle ; \exp_after:wN ;
If the \langle \text{digit 1} \rangle is non-zero, then for rounding we only care about the digits 16 and
17, and whether further digits are zero or not (check for exact ties). On the other hand,
if \langle \text{digit 1} \rangle is zero, we care about digits 17 and 18, and whether further digits are zero.
\cs_new:Npn \__fp_mul_significand_test_f:NNN #1 #2 #3
{ \if_meaning:w 0 #3 \exp_after:wN \__fp_mul_significand_small_f:NNNNNw
\else: \exp_after:wN \__fp_mul_significand_large_f:NwwNNNw \fi:
#1 #3 }
\__fp_mul_significand_large_f:NwwNNNw
\__fp_mul_significand_test_f:NNN (\text{sign}) 1 (\text{digits 1–8}) ; (\text{digits 9–12}) ;
(\text{digits 13–16}) ; + (\text{digits 17–20}) + (\text{digits 21–24}) + (\text{digits 25–28}) + (\text{digits 29–32}) ; \exp_after:wN ;
\exp_after:wN \__fp_mul_significand_large_f:NwwNNNw
\__fp_mul_significand_large_f:NwwNNNw
\__fp_mul_significand_large_f:NwwNNNw
\__fp_mul_significand_large_f:NwwNNNw
In this branch, \langle \text{digit 1} \rangle is non-zero. The result is thus \langle \text{digits 1–16} \rangle, plus some rounding
which depends on the digits 16, 17, and whether all subsequent digits are zero or not.
Here, \__fp_round_digit:Nw takes digits 17 and further (as an integer expression), and
replaces it by a \langle \text{rounding digit} \rangle, suitable for \__fp_round:Nw.
\cs_new:Npn \__fp_mul_significand_large_f:NwwNNNw #1 #2 #3 #4 #5 #6 #7 +
{ \exp_after:wN \__fp_basics_pack_high:NNNNNw \int_value:w \__fp_int_eval:w 1#2
\exp_after:wN \__fp_basics_pack_low:NNNNNw \int_value:w \__fp_int_eval:w 1#3#4#5#6#7
+ \exp_after:wN \__fp_round:Nw
\exp_after:wN \__fp_round:Nw #1
}
In this branch, \(\text{(digit 1)}\) is zero. Our result is thus \(\text{(digits 2–17)}\), plus some rounding which depends on the digits \(17, 18\), and whether all subsequent digits are zero or not. The 8 digits \(1#3\) are followed, after expansion of the \text{small_pack} auxiliary, by the next digit, to form a 9 digit number.

72.3 Division

72.3.1 Signs, and special numbers

Time is now ripe to tackle the hardest of the four elementary operations: division.

Filtering special floating point is very similar to what we did for multiplications, with a few variations. Invalid operation exceptions display / rather than *. In the formula for dispatch, we replace \(- 2 + \) by \(-\). The case of normal numbers is treated using \text{\_fp\_div\_npos\_o:Nww} rather than \text{\_fp\_mul\_npos\_o:Nww}. There are two additional cases: if the first operand is normal and the second is a zero, then the division by zero exception is raised; cases 10 and 11 of the \text{\_if\_case:w} construction in \text{\_fp\_mul\_cases\_o:NnNww} are provided as the fourth argument here.
We want to compute \( A / Z \). As for multiplication, \( \__fp_sanitize:Nw \) checks for overflow or underflow; we provide it with the \( \langle \text{final sign} \rangle \), and an integer expression in which we compute the exponent. We set up the arguments of \( \__fp_div_significand_i_o:wnnw \), namely an integer \( \langle y \rangle \) obtained by adding 1 to the first 5 digits of \( Z \) (explanation given soon below), then the four \( \{ A_i \} \), then the four \( \{ Z_i \} \), a semi-colon, and the \( \langle \text{final sign} \rangle \), used for rounding at the end.

\[
\__fp_div_npos_o:Nww \{final sign\} \s__fp\__fp_chk:w1 \{exp A\} \{A1\} \{A2\} \{A3\} \{A4\} \{exp Z\} \{Z1\} \{Z2\} \{Z3\} \{Z4\} ;
\]

We proceed as follows.

- Find an integer \( Q_A \approx 10^4 A / Z \).
- Replace \( A \) by \( B = 10^4 A - Q_A Z \).
- Find an integer \( Q_B \approx 10^4 B / Z \).
- Replace \( B \) by \( C = 10^4 B - Q_B Z \).
- Find an integer \( Q_C \approx 10^4 C / Z \).
- Replace \( C \) by \( D = 10^4 C - Q_C Z \).
- Find an integer \( Q_D \approx 10^4 D / Z \).
- Consider \( E = 10^4 D - Q_D Z \), and ensure correct rounding.

72.3.2 Work plan

In this subsection, we explain how to avoid overflowing \( \text{TeX} \)'s integers when performing the division of two positive normal numbers.

We are given two numbers, \( A = 0.A_1A_2A_3A_4 \) and \( Z = 0.Z_1Z_2Z_3Z_4 \), in blocks of 4 digits, and we know that the first digits of \( A_1 \) and of \( Z_1 \) are non-zero. To compute \( A / Z \), we proceed as follows.

- Find an integer \( Q_A \approx 10^4 A / Z \).
- Replace \( A \) by \( B = 10^4 A - Q_A Z \).
- Find an integer \( Q_B \approx 10^4 B / Z \).
- Replace \( B \) by \( C = 10^4 B - Q_B Z \).
- Find an integer \( Q_C \approx 10^4 C / Z \).
- Replace \( C \) by \( D = 10^4 C - Q_C Z \).
- Find an integer \( Q_D \approx 10^4 D / Z \).
- Consider \( E = 10^4 D - Q_D Z \), and ensure correct rounding.
The result is then \( Q = 10^{-4}Q_A + 10^{-8}Q_B + 10^{-12}Q_C + 10^{-16}Q_D + \text{rounding} \). Since the \( Q_i \) are integers, \( B, C, D, \) and \( E \) are all exact multiples of \( 10^{-16} \), in other words, computing with 16 digits after the decimal separator yields exact results. The problem is the risk of overflow: in general \( B, C, D, \) and \( E \) may be greater than 1.

Unfortunately, things are not as easy as they seem. In particular, we want all intermediate steps to be positive, since negative results would require extra calculations at the end. This requires that \( Q_A \leq 10^4 A / Z \) etc. A reasonable attempt would be to define \( Q_A \) as

\[
\inteval{n} \left\{ \frac{A_1 A_2}{Z_1+1} - 1 \right\} \leq 10^4 A / Z
\]

Subtracting 1 at the end takes care of the fact that \( \varepsilon \text{-\TeX} \)’s \texttt{\_fp_int_eval:w} rounds divisions instead of truncating (really, \( 1/2 \) would be sufficient, but we work with integers). We add 1 to \( Z \) because \( Z_1 \leq 10^4 Z < Z_1 + 1 \) and we need \( Q_A \) to be an underestimate. However, we are now underestimating \( Q_A \) too much: it can be wrong by up to 100, for instance when \( Z = 0.1 \) and \( A \approx 1 \). Then \( B \) could take values up to 10 (maybe more), and a few steps down the line, we would run into arithmetic overflow, since \TeX{} can only handle integers less than roughly \( 2 \cdot 10^9 \).

A better formula is to take

\[
Q_A = \inteval{n} \left\{ \frac{10 \cdot A_1 A_2}{10^{-3} \cdot Z_1 Z_2} + 1 \right\}.
\]

This is always less than \( 10^9 A / (10^5 Z) \), as we wanted. In words, we take the 5 first digits of \( Z \) into account, and the 8 first digits of \( A \), using 0 as a 9-th digit rather than the true digit for efficiency reasons. We shall prove that using this formula to define all the \( Q_i \) avoids any overflow. For convenience, let us denote

\[
y = \left\lfloor 10^{-3} \cdot Z_1 Z_2 \right\rfloor + 1,
\]

so that, taking into account the fact that \( \varepsilon \text{-\TeX} \) rounds ties away from zero,

\[
Q_A = \left\lfloor \frac{A_1 A_2 0 \cdot 1}{y} \right\rfloor \leq \frac{A_1 A_2 0 \cdot 1}{y} - \frac{1}{2} \leq A_1 A_2 0 \cdot y - \frac{3}{2}.
\]

Note that \( 10^4 < y \leq 10^5 \), and 999 \( \leq Q_A \leq 99989 \). Also note that this formula does not cause an overflow as long as \( A < (2^{31} - 1) / 10^9 \approx 2.147 \ldots \), since the numerator involves an integer slightly smaller than \( 10^9 A \).

Let us bound \( B \):

\[
10^5 B = A_1 A_2 0 + 10 \cdot 0 A_3 A_4 - 10 \cdot Z_1 Z_2 Z_3 Z_4 \cdot Q_A < A_1 A_2 0 \cdot \left( 1 - 10 \cdot \frac{Z_1 Z_2 Z_3 Z_4}{y} \right) + \frac{3}{2} \cdot 10 \cdot Z_1 Z_2 Z_3 Z_4 + 10 \leq A_1 A_2 0 \cdot \left( y - 10 \cdot \frac{Z_1 Z_2 Z_3 Z_4}{y} \right) + \frac{3}{2} y + 10 \leq A_1 A_2 0 \cdot \frac{1}{y} \cdot 3 y + 10 \leq \frac{10^9 A}{y} + 1.6 \cdot y.
\]

1077
At the last step, we hide 10 into the second term for later convenience. The same reasoning yields

\[
\begin{align*}
10^5 B &< 10^9 A/y + 1.6y, \\
10^5 C &< 10^9 B/y + 1.6y, \\
10^5 D &< 10^9 C/y + 1.6y, \\
10^5 E &< 10^9 D/y + 1.6y.
\end{align*}
\]

The goal is now to prove that none of \( B, C, D, \) and \( E \) can go beyond \( (2^{31} - 1)/10^9 = 2.147\ldots \).

Combining the various inequalities together with \( A < 1 \), we get

\[
\begin{align*}
10^5 B &< 10^9/y + 1.6y, \\
10^5 C &< 10^13/y^2 + 1.6(y + 10^4), \\
10^5 D &< 10^17/y^3 + 1.6(y + 10^4 + 10^8/y), \\
10^5 E &< 10^{21}/y^4 + 1.6(y + 10^4 + 10^8/y + 10^{12}/y^2).
\end{align*}
\]

All of those bounds are convex functions of \( y \) (since every power of \( y \) involved is convex, and the coefficients are positive), and thus maximal at one of the end-points of the allowed range \( 10^4 < y \leq 10^5 \). Thus,

\[
\begin{align*}
10^5 B &< \max(1.16 \cdot 10^5, 1.7 \cdot 10^5), \\
10^5 C &< \max(1.32 \cdot 10^5, 1.77 \cdot 10^5), \\
10^5 D &< \max(1.48 \cdot 10^5, 1.777 \cdot 10^5), \\
10^5 E &< \max(1.64 \cdot 10^5, 1.7777 \cdot 10^5).
\end{align*}
\]

All of those bounds are less than \( 2.147 \cdot 10^5 \), and we are thus within \TeX\’s bounds in all cases!

We later need to have a bound on the \( Q_i \). Their definitions imply that \( Q_A < 10^9 A/y - 1/2 < 10^5 A \) and similarly for the other \( Q_i \). Thus, all of them are less than 177770.

The last step is to ensure correct rounding. We have

\[
A/Z = \sum_{i=1}^{4} (10^{-4i}Q_i) + 10^{-16}E/Z
\]

exactly. Furthermore, we know that the result is in \([0.1, 10)\), hence will be rounded to a multiple of \( 10^{-16} \) or of \( 10^{-15} \), so we only need to know the integer part of \( E/Z \), and a “rounding” digit encoding the rest. Equivalently, we need to find the integer part of \( 2E/Z \), and determine whether it was an exact integer or not (this serves to detect ties). Since

\[
\frac{2E}{Z} = 2 \frac{10^5 E}{10^6 Z} \leq 2 \frac{10^5 E}{10^9} < 36,
\]

1078
this integer part is between 0 and 35 inclusive. We let \( \varepsilon \text{-TeX} \) round

\[
P = \intevalm \left\{ \frac{2 \cdot E_1 E_2}{Z_1 Z_2} \right\},
\]

which differs from \( 2E/Z \) by at most

\[
\frac{1}{2} + 2 \left| \frac{E}{Z} - \frac{E}{10^{-8} Z_1 Z_2} \right| + 2 \left| \frac{10^8 E - E_1 E_2}{Z_1 Z_2} \right| < 1,
\]

(1/2 comes from \( \varepsilon \text{-TeX} \)'s rounding) because each absolute value is less than \( 10^{-7} \). Thus \( P \) is either the correct integer part, or is off by 1; furthermore, if \( 2E/Z \) is an integer, \( P = 2E/Z \). We will check the sign of \( 2E - PZ \). If it is negative, then \( E/Z \in ((P-1)/2, P/2) \). If it is zero, then \( E/Z \in (P/2, (P-1)/2) \). In each case, we know how to round to an integer, depending on the parity of \( P \), and the rounding mode.

### 72.3.3 Implementing the significand division

\[
\_\_fp_div_significand_i:o:wnn\{y\} \{A_1\} \{A_2\} \{A_3\} \{A_4\}
\{\langle Z_1\rangle\} \{\langle Z_2\rangle\} \{\langle Z_3\rangle\} \{\langle Z_4\rangle\} \{\langle sign\rangle\}
\]

Compute \( 10^6 + Q_A \) (a 7 digit number thanks to the shift), unbrace \( \langle A_1 \rangle \) and \( \langle A_2 \rangle \), and prepare the \( \langle \text{continuation} \rangle \) arguments for 4 consecutive calls to \_\_fp_div_-_significand_calc:wnnnnnnn. Each of these calls needs \( \langle y \rangle \#1 \), and it turns out that we need post-expansion there, hence the \( \text{int_value:w} \). Here, \#4 is six brace groups, which give the six first \( n \)-type arguments of the \text{calc} function.

\[
\text{cs_new:Nn} \_\_fp_div_significand_i:o:wnn \#1 \#2\#3 \#4
\{
\exp_after:w \_\_fp_div_significand_test_o:w
\inteval:w \_\_fp_int_eval:w
\exp_after:w \_\_fp_div_significand_calc:wnnnnnnn
\inteval:w \_\_fp_int_eval:w 999999 + \#2 \#3 / \#1
\#2 \#3
\#4
\exp_after:w \_\_fp_div_significand_i:i:w:wnnnnnnn\{A_1\} \{A_2\} \{A_3\} \{A_4\}
\{\langle Z_1\rangle\} \{\langle Z_2\rangle\} \{\langle Z_3\rangle\} \{\langle Z_4\rangle\} \{\langle continuation\rangle\}
\]

expands to

\[
\langle 10^6 + Q_A \rangle \langle continuation \rangle \{B_1\} \{B_2\} \{B_3\} \{B_4\} \{\langle Z_1\rangle\} \{\langle Z_2\rangle\} \{\langle Z_3\rangle\} \{\langle Z_4\rangle\}
\]

where \( B = 10^4 A - Q_A \cdot Z \). This function is also used to compute \( C \), \( D \), \( E \) (with the input shifted accordingly), and is used in \text{3fp-expo}.

We know that \( 0 < Q_A < 1.8 \cdot 10^5 \), so the product of \( Q_A \) with each \( Z_i \) is within \text{TeX}'s bounds. However, it is a little bit too large for our purposes: we would not be able to
use the usual trick of adding a large power of 10 to ensure that the number of digits is fixed.

The bound on $Q_A$, implies that $10^6 + Q_A$ starts with the digit 1, followed by 0 or 1. We test, and call different auxiliaries for the two cases. An earlier implementation did the tests within the computation, but since we added a \textit{continuation}, this is not possible because the macro has 9 parameters.

The result we want is then (the overall power of 10 is arbitrary):

$$10^{-4}(#2 - #1 \cdot #5 - 10 \cdot (i) \cdot #6) + 10^{-8}(#3 - #1 \cdot #6 - 10 \cdot (i) \cdot #7)$$

$$+ 10^{-12}(#4 - #1 \cdot #7 - 10 \cdot (i) \cdot #8) + 10^{-16}(-#1 \cdot #8),$$

where $(i)$ stands for the $10^5$ digit of $Q_A$, which is 0 or 1, and #1, #2, etc. are the parameters of either auxiliary. The factors of 10 come from the fact that $Q_A = 10 \cdot 10^4 \cdot (i) + #1$. As usual, to combine all the terms, we need to choose some shifts which must ensure that the number of digits of the second, third, and fourth terms are each fixed. Here, the positive contributions are at most $10^8$ and the negative contributions can go up to $10^9$. Indeed, for the auxiliary with $(i) = 1$, #1 is at most 80000, leading to contributions of at worse $-8 \cdot 10^4$, while the other negative term is very small $< 10^6$ (except in the first expression, where we don’t care about the number of digits); for the auxiliary with $(i) = 0$, #1 can go up to 99999, but there is no other negative term. Hence, a good choice is $2 \cdot 10^9$, which produces totals in the range $[10^9, 2 \cdot 10^9]$. We are flirting with \TeX{}’s limits once more.

\begin{verbatim}
\cs_new:Npn \__fp_div_significand_calc:wwnnnnnnn 1#1
{ \if_meaning:w 1 #1 \exp_after:wN \__fp_div_significand_calc_i:wwnnnnnnn \else: \exp_after:wN \__fp_div_significand_calc_ii:wwnnnnnnn \fi: }
\cs_new:Npn \__fp_div_significand_calc_i:wwnnnnnnn  #1; #2;#3#4 #5#6#7#8 #9
{ \int_value:w \__fp_int_eval:w \c__fp_Bigg_leading_shift_int + #2 - #1 * #5 - #5#60 \exp_after:wN \__fp_pack_Bigg:NNNNNNw \int_value:w \__fp_int_eval:w \c__fp_Bigg_middle_shift_int + #3 - #1 * #6 - #70 \exp_after:wN \__fp_pack_Bigg:NNNNNNw \int_value:w \__fp_int_eval:w \c__fp_Bigg_middle_shift_int + #4 - #1 * #7 - #80 \exp_after:wN \__fp_pack_Bigg:NNNNNNw \int_value:w \__fp_int_eval:w \c__fp_Bigg_trailing_shift_int - #1 * #8 ; (#5){#6}{#7}{#8} }
\cs_new:Npn \__fp_div_significand_calc_ii:wwnnnnnnn  #1; #2;#3#4 #5#6#7#8 #9
{ \int_value:w \__fp_int_eval:w \c__fp_Bigg_trailing_shift_int - #1 * #8 ; (#5){#6}{#7}{#8} }
\cs_new:Npn \__fp_div_significand_calc:wwnnnnnnn 1#1
{ \if_meaning:w 1 #1 \exp_after:wN \__fp_div_significand_calc_i:wwnnnnnnn \else: \exp_after:wN \__fp_div_significand_calc_ii:wwnnnnnnn \fi: }
\end{verbatim}
\Copyright

\__fp_div_significand_ii:wwn
\__fp_div_significand_ii:wwn
⟨y⟩; ⟨E1⟩; {⟨E2⟩} {⟨E3⟩} {⟨E4⟩} {⟨Z1⟩} {⟨Z2⟩} {⟨Z3⟩} {⟨Z4⟩} ⟨continuations⟩ ⟨sign⟩

Compute \( Q_B \) by evaluating \( (B_1)(B_2)/y - 1 \). The result is output to the left, in an
\__fp_int_eval:w which we start now. Once that is evaluated (and the other \( Q_i \) also,
since later expansions are triggered by this one), a packing auxiliary takes care of placing
the digits of \( Q_B \) in an appropriate way for the final addition to obtain \( Q \). This auxiliary
is also used to compute \( Q_C \) and \( Q_D \) with the inputs \( C \) and \( D \) instead of \( B \).

\__fp_div_significand_ii:wwn
\__fp_div_significand_iv:wwnnnnnnn
\__fp_div_significand_v:NNw
\__fp_div_significand_vi:Nw
\__fp_div_significand_iii:wwnnnnn (y) ; ⟨E1⟩ ; {⟨E2⟩} {⟨E3⟩} {⟨E4⟩}
{⟨Z1⟩} {⟨Z2⟩} {⟨Z3⟩} {⟨Z4⟩} ⟨sign⟩

We compute \( P \approx 2E/Z \) by rounding \( 2E_1E_2/Z_1Z_2 \). Note the first 0, which multiplies
\( Q_D \) by 10: we later add (roughly) \( 5 \cdot P \), which amounts to adding \( P/2 \approx E/Z \) to \( Q_D \),
the appropriate correction from a hypothetical \( Q_E \).

\__fp_div_significand_iii:wwnnnn
\__fp_div_significand_iv:wwnnnnnnn
\__fp_div_significand_v:NNw
\__fp_div_significand_vi:Nw
\__fp_div_significand_iii:wwnnnn (y) ; ⟨E1⟩ ; {⟨E2⟩} {⟨E3⟩} {⟨E4⟩}
{⟨Z1⟩} {⟨Z2⟩} {⟨Z3⟩} {⟨Z4⟩} ⟨sign⟩

\Copyright
This adds to the current expression \((10^7 + 10 \cdot Q_D)\) a contribution of \(5 \cdot P + \text{sign}(T)\) with \(T = 2E - PZ\). This amounts to adding \(P/2\) to \(Q_D\), with an extra \((\text{rounding})\) digit. This \((\text{rounding})\) digit is 0 or 5 if \(T\) does not contribute, i.e., if \(0 = T = 2E - PZ\), in other words if \(10^{16} A/Z\) is an integer or half-integer. Otherwise it is in the appropriate range, [1, 4] or [6, 9]. This is precise enough for rounding purposes (in any mode).

It seems an overkill to compute \(T\) exactly as I do here, but I see no faster way right now.

Once more, we need to be careful and show that the calculation \(\#1 \cdot \#6\#7\) below does not cause an overflow: naively, \(P\) can be up to 35, and \(\#6\#7\) up to \(10^8\), but both cannot happen simultaneously. To show that things are fine, we split in two (non-disjoint) cases.

- For \(P < 10\), the product obeys \(P \cdot \#6\#7 < 10^8 \cdot P < 10^9\).
- For large \(P \geq 3\), the rounding error on \(P\), which is at most 1, is less than a factor of 2, hence \(P \leq 4E/Z\). Also, \#6\#7 \(\leq 10^8 \cdot Z\), hence \(P \cdot \#6\#7 \leq 4E \cdot 10^8 < 10^9\).

Both inequalities could be made tighter if needed.

Note however that \(P \cdot \#8\#9\) may overflow, since the two factors are now independent, and the result may reach \(3.5 \cdot 10^9\). Thus we compute the two lower levels separately. The rest is standard, except that we use \(+\) as a separator (ending integer expressions explicitly). \(T\) is negative if the first character is \(-\), it is positive if the first character is neither \(0\) nor \(-\). It is also positive if the first character is \(0\) and second argument of \(\_\_\_\_\_\_\_\text{fp} \_\_\text{div} \_\_\_\_\_\_\_\text{significand} \_\_\_\_\_\_\_\text{vi}:Nw\), a sum of several terms, is also zero. Otherwise, there was an exact agreement: \(T = 0\).

\[
\text{At this stage, we are in the following situation: } \TeX \text{ is in the process of expanding several integer expressions, thus functions at the bottom expand before those above.}
\]
Here, \( \varepsilon = \text{sign}(T) \) is 0 in case \( 2E = PZ \), 1 in case \( 2E > PZ \), which means that \( P \) was the correct value, but not with an exact quotient, and -1 if \( 2E < PZ \), i.e., \( P \) was an overestimate. The packing function we define now does nothing special: it removes the \( 10^6 \) and carries two digits (for the \( 10^5 \)'s and the \( 10^4 \)'s).

\[
\begin{align*}
\text{\texttt{\_fp_div_significand_test_o:w\ 10^6 + QA \_fp_div_significand_pack:\ NNN 10^6 + QB \_fp_div_significand_pack:\ NNN 10^6 + QC \_fp_div_significand_pack:\ NNN 10^7 + 10 \cdot QD + 5 \cdot P + \varepsilon ; (sign)}}
\end{align*}
\]

We know that the final result cannot reach 10, hence \texttt{1#1#2}, together with contributions from the level below, cannot reach \( 2 \cdot 10^9 \). For rounding, we build the \textit{(rounding digit)} from the last two of our 18 digits.
\[ \exp_after:wN \__fp_basics_pack_weird_high:NNNNNNNN \]
\[ \int_value:w \__fp_int_eval:w 1 \#1 \#2 \]
\[ \exp_after:wN \__fp_basics_pack_weird_low:NNNN \]
\[ \int_value:w \__fp_int_eval:w 1 \#3 \#4 \#5 \#6 + \]
\[ \exp_after:wN \__fp_round:NNN \]
\[ \exp_after:wN \#9 \]
\[ \exp_after:wN \#6 \]
\[ \int_value:w \__fp_round_digit:Nw \#7 \#8 ; \]
\[ \exp_after:wN ; \]

\[ \quad \]

(End definition for \_\_fp_div_significand_large_o:wwwNNNNw.)

### 72.4 Square root

The square root function \_\_fp_sqrt_o:w\] Zeros are unchanged: \( \sqrt{-0} = -0 \) and \( \sqrt{+0} = +0 \). Negative numbers (other than \(-0\)) have no real square root. Positive infinity, and \texttt{nan}, are unchanged. Finally, for normal positive numbers, there is some work to do.

\begin{verbatim}
\cs_new:Npn \_\_fp_sqrt_o:w \s__fp \_\_fp_chk:w \#2\#3\#4; @
\end{verbatim}

\begin{verbatim}
\if_meaning:w 0 \#2 \_\_fp_case_return_same_o:w \fi:
\if_meaning:w 2 \#3
\_\_fp_case_use:nw \{ \_\_fp_invalid_operation_o:nw \{ sqrt \} \}
\fi:
\if_meaning:w 1 \#2 \else: \_\_fp_case_return_same_o:w \fi:
\_\_fp_sqrt_npos_o:w
\s__fp \_\_fp_chk:w \#2 \#3 \#4;
\end{verbatim}

(End definition for \_\_fp_sqrt_o:w.)

\begin{verbatim}
\_\_fp_sqrt_npos_o:w
\_\_fp_sqrt_npos_auxi_o:wwnnN
\_\_fp_sqrt_npos_auxii_o:wNNNNNNNN
\end{verbatim}

Prepare \_\_fp_sanitize:Nw to receive the final sign 0 (the result is always positive) and the exponent, equal to half of the exponent \#1 of the argument. If the exponent \#1 is even, find a first approximation of the square root of the significand \( 10^6a_1 + a_2 = 10^6\#2\#3 + \#4\#5 \) through Newton's method, starting at \( x = 57234133 \approx 10^{7.75} \). Otherwise, first shift the significand by one digit, getting \( 1 \leq a_1' < 10^6, 10^5 \) instead of \( 10^7, 10^5 \), then use Newton's method starting at \( 17782794 \approx 10^{7.25} \).

\begin{verbatim}
\cs_new:Npn \_\_fp_sqrt_npos_o:w \s__fp \_\_fp_chk:w \#2 \#3\#4\#5;
\end{verbatim}

\begin{verbatim}
\exp_after:wN \_\_fp_sqrt_npos_o:w \s__fp \_\_fp_chk:w \#2 \#3\#4\#5;
\exp_after:wN \_\_fp_int_eval:w
\if_int_odd:w \#1 \exp_stop_f:
\exp_after:wN \_\_fp_sqrt_npos_auxi_o:wwnnN
\else:
\exp_after:wN \_\_fp_int_eval:w
\if_int_odd:w \#1 \exp_stop_f:
\exp_after:wN \_\_fp_sqrt_npos_auxii_o:wwnnN
\else:
\_\_fp_sqrt_Newton_o:wwn 56234133; 0; \{\#2\#3\} \{\#4\#5\} 0
\else:
\_\_fp_sqrt_Newton_o:wwn 17782794; 0; \{\#2\#3\} \{\#4\#5\} 0
\end{verbatim}

(End definition for \_\_fp_sqrt_o:w.)
Newton’s method maps \( x \mapsto \left( \frac{x + \left[ 10^8 a_1 / x \right]}{2} \right) \) in each iteration, where \([b/c]\) denotes \(\varepsilon\text{-\TeX}'s division. This division rounds the real number \(b/c\) to the closest integer, rounding ties away from zero, hence when \(c\) is even, \(b/c - 1/2 + 1/c \leq \left[ b/c \right] \leq b/c + 1/2\) and when \(c\) is odd, \(b/c - 1/2 + 1/(2c) \leq \left[ b/c \right] \leq b/c + 1/2 - 1/(2c)\). For all \(c\), \(b/c - 1/2 + 1/(2c) \leq \left[ b/c \right] \leq b/c + 1/2\).

Let us prove that the method converges when implemented with \(\varepsilon\text{-\TeX}'s division, for any \(10^6 \leq a_1 < 10^8\) and starting value \(10^6 \leq x < 10^8\). Using the inequalities above and the arithmetic-geometric inequality \((x + t) / 2 \geq \sqrt{xt}\) for \(t = 10^8 a_1 / x\), we find

\[
x' = \frac{x + \left[ 10^8 a_1 / x \right]}{2} \geq \frac{x + 10^8 a_1 / x - 1/2 + 1/(2x)}{2} \geq \sqrt{10^8 a_1} - \frac{1}{4} + \frac{1}{4x}.
\]

After any step of iteration, we thus have \(\delta = x - \sqrt{10^8 a_1} \geq -0.25 + 0.25 \cdot 10^{-8}\). The new difference \(\delta' = x' - \sqrt{10^8 a_1}\) after one step is bounded above as

\[
x' - \sqrt{10^8 a_1} \leq \frac{x + 10^8 a_1 / x + 1/2}{2} - \frac{\delta}{2 \sqrt{10^8 a_1} + \delta} + \frac{3}{4}.
\]

For \(\delta > 3/2\), this last expression is \(\leq \delta / 2 + 3/4 < \delta\), hence \(\delta\) decreases at each step: since all \(x\) are integers, \(\delta\) must reach a value \(-1/4 < \delta \leq 3/2\). In this range of values, we get \(\delta' \leq \frac{3}{2 \sqrt{10^8 a_1}} + \frac{3}{4} \leq 0.75 + 1.125 \cdot 10^{-7}\). We deduce that the difference \(\delta = x - \sqrt{10^8 a_1}\) eventually reaches a value in the interval \([-0.25 + 0.25 \cdot 10^{-8}, 0.75 + 11.25 \cdot 10^{-8}]\), whose width is \(1 + 11 \cdot 10^{-8}\). The corresponding interval for \(x\) may contain two integers, hence \(x\) might oscillate between those two values.

However, the fact that \(x \mapsto x - 1\) and \(x - 1 \mapsto x\) puts stronger constraints, which are not compatible: the first implies

\[
x + \left[ 10^8 a_1 / x \right] \leq 2x - 2
\]

hence \(10^8 a_1 / x \leq x - 3/2\), while the second implies

\[
x - 1 + \left[ 10^8 a_1 / (x - 1) \right] \geq 2x - 1
\]

hence \(10^8 a_1 / (x - 1) \geq x - 1/2\). Combining the two inequalities yields \(x^2 - 3x/2 \geq 10^8 a_1 \geq x - 3x/2 + 1/2\), which cannot hold. Therefore, the iteration always converges to a single integer \(x\). To stop the iteration when two consecutive results are equal, the function \(\text{\_fp_sqrt\_Newton\_o:wn}\) receives the newly computed result as \#1, the previous result as \#2, and \(a_1\) as \#3. Note that \(\varepsilon\text{-\TeX}\) combines the computation of a multiplication and a following division, thus avoiding overflow in \(\#3 \times 100000000 / \#1\). In any case, the result is within \([10^7, 10^8]\).
This receives a continuation function \#1, then five blocks of 4 digits for \( y \), then two 8-digit blocks and a single digit for \( a \). A common estimate of \( \sqrt{a} - y = (a - y^2)/(\sqrt{a} + y) \) is \( a - y^2 \), which leads to alternating overestimates and underestimates. We tweak this, to only work with underestimates (no need then to worry about signs in the computation). Each step finds the largest integer \( j \) such that \( 10^{16}(a - y^2) < 2 \cdot 10^8 \), then computes the integer (with \( \varepsilon \)-FPX’s rounding division)

\[
10^{16} z = \left( \frac{10^{16} (a - y^2)}{257} \cdot (0.5 \cdot 10^6) \right) / (10^8 y + 1)
\]

The choice of \( j \) ensures that \( 10^{16} z < 2 \cdot 10^8 \cdot 0.5 \cdot 10^6 / 10^7 = 10^9 \), thus \( 10^9 + 10^{16} z \) has exactly 10 digits, does not overflow FPX’s integer range, and starts with 1. Incidentally, since all \( a - y^2 \leq 3.2 \cdot 10^{-8} \), we know that \( j \geq 3 \).

Let us show that \( z \) is an underestimate of \( \sqrt{a} - y \). On the one hand, \( \sqrt{a} - y \leq 16 \cdot 10^{-8} \) because this holds for the initial \( y \) and values of \( y \) can only increase. On the other hand, the choice of \( j \) implies that \( \sqrt{a} - y \leq 5(\sqrt{a} + y)(\sqrt{a} - y) = 5(a - y^2) < 10^9 - 10^j \). For \( j = 3 \),
the first bound is better, while for larger \( j \), the second bound is better. For all \( j \in [3, 6] \), we find \( \sqrt{a} - y < 16 \cdot 10^{-2j} \). From this, we deduce that

\[
10^{4j}(\sqrt{a} - y) = \frac{10^{4j}(a - y^2 - (\sqrt{a} - y)^2)}{2y} \geq \frac{10^{4j}(a - y^2)}{2 \cdot 10^{-8}(10^8y + 1)} + \frac{1}{2}
\]

where we have replaced the bound \( 10^{4j}(16 \cdot 10^{-2j}) = 256 \) by 257 and extracted the corresponding term \( 1/(2 \cdot 10^{-8}(10^8y + 1)) \geq 1/2 \). Given that \( \epsilon \)-FPX's integer division obeys \( [b/c] \leq b/c + 1/2 \), we deduce that \( 10^{4j}z \leq 10^{4j}(\sqrt{a} - y) \), hence \( y + z \leq \sqrt{a} \) is an underestimate of \( \sqrt{a} \), as claimed. One implementation detail: because the computation involves \(-#4*#4 - 2*#3*#5 - 2*#2*#6\) which may be as low as \(-5 \cdot 10^8\), we need to use the \texttt{pack\_big} functions, and the \texttt{big shifts}.

We receive here the difference \( a - y^2 = d = \sum d_i \cdot 10^{-4i} \), as \( (d_2) \); \( (d_3) \) ... \( (d_{10}) \), where each block has 4 digits, except \( (d_2) \). This function finds the largest \( j \leq 6 \) such that \( 10^{4j}(a - y^2) < 2 \cdot 10^8 \), then leaves an open parenthesis and the integer \( \lfloor 10^{4j}(a - y^2) \rfloor \)
in an integer expression. The closing parenthesis is provided by the caller \texttt{\_\_fp_sqrt\_auxii\_o:NnnnnnnN}, which completes the expression
\[
10^{4j}z = \left(\left[\left(10^{4j}(a - y^2)\right] - 257 \right) \cdot (0.5 \cdot 10^8) \right) \left\lceil \frac{10^8y + 1}{10^8y + 1} \right\rceil
\]
for an estimate of $10^{4j}(\sqrt{a} - y)$. If $d_2 \geq 2$, $j = 3$ and the \texttt{auxiv} auxiliary receives $10^{12}z$. If $d_2 \leq 1$ but $10^{4j}d_2 + 10^{4j}d_3 \geq 2$, $j = 4$ and the \texttt{auxv} auxiliary is called, and receives $10^{16}z$, and so on. In all those cases, the \texttt{auxviii} auxiliary is set up to add $z$ to $y$, then go back to the \texttt{auxi} step with continuation \texttt{auxiii} (the function we are currently describing). The maximum value of $j$ is 6, regardless of whether $10^{12}d_2 + 10^{12}d_3 + 10^{12}d_4 + d_5 \geq 1$. In this last case, we detect when $10^{24}z < 10^5$, which essentially means $\sqrt{a} - y \lesssim 10^{-17}$: once this threshold is reached, there is enough information to find the correctly rounded $\sqrt{a}$ with only one more call to \texttt{\_\_fp_sqrt\_auxii\_o:NnnnnnnN}. Note that the iteration cannot be stuck before reaching $j = 6$, because for $j < 6$, one has $2 \cdot 10^8 \leq 10^{2(j+1)}(a - y^2)$, hence
\[
10^{4j}z \geq \frac{(20000 - 257) \cdot (0.5 \cdot 10^8)}{10^8y + 1} \geq (20000 - 257) \cdot 0.5 > 0.
\]
Simply add the two 8-digit blocks of \( z \), aligned to the last four of the five 4-digit blocks of \( y \), then call the auxii auxiliary to evaluate \( y^2 = (y + z)^2 \).
The difference $0 \leq a + 10^{-16} - m^2 \leq 10^{-16} + (\sqrt{a} - m)(\sqrt{a} + m) \leq 2 \cdot 10^{-16}$ was just computed: its first 8 digits vanish, as do the next four, #1, and most of the following four, #2. The guess $m$ is an overestimate if $a + 10^{-16} - m^2 < 10^{-16}$, that is, #1#2 vanishes. Otherwise it is an underestimate, unless $a + 10^{-16} - m^2 = 10^{-16}$ exactly. For an underestimate, call the auxxiv function with argument 9998. For an exact result call it with 9999, and for an overestimate call it with 10000.

```latex
\cs_new:Npn \__fp_sqrt_auxxiv_o:wnnnnnnnN #1; #2#3#4#5#6 #7#8#9
\{
\exp_after:wN \__fp_basics_pack_high:NNNNNw
\int_value:w \__fp_int_eval:w 1 0000 0000 + #2#3
\}
```

This receives 9998, 9999 or 10000 as #1 when $m$ is an underestimate, exact, or an overestimate, respectively. Then comes $m$ as five blocks of 4 digits, but where the last block #6 may be 0, 5000, or 10000. In the latter case, we need to add a carry, unless $m$ is an overestimate (#1 is then 10000). Then comes $a$ as three arguments. Rounding is done by \_\_\_fp_round:NNN, whose first argument is the final sign (square roots are positive). We fake its second argument. It should be the last digit kept, but this is only used when ties are “rounded to even”, and only when the result is exactly half-way between two representable numbers rational square roots of numbers with 16 significant digits have: this situation never arises for the square root, as any exact square root of a 16 digit number has at most 8 significant digits. Finally, the last argument is the next digit, possibly shifted by 1 when there are further nonzero digits. This is achieved by \_\_\_fp_\-round_digit:Nw, which receives (after removal of the 10000’s digit) one of 0000, 0001, 4999, 5000, 5001, or 9999, which it converts to 0, 1, 4, 5, 6, and 9, respectively.

```latex
\cs_new:Npn \__fp_sqrt_auxxiv_o:wnnnnnnnN \{
\exp_after:wN \__fp_basics_pack_high:NNNNNw
\int_value:w \__fp_int_eval:w 1 0000 0000 + #2\#3
\}
```

1090
25935 \exp_after:wN \__fp_basics_pack_low:NNNNNw
25936 \int_value:w \__fp_int_eval:w 1 0000 0000
25937 \if_int_compare:w #6 > #1 \exp_stop_f: + 1 \fi:
25938 + \exp_after:wN \__fp_round:NNN
25939 \exp_after:wN 0
25940 \exp_after:wN 0
25941 \int_value:w
25942 \exp_after:wN \use_i:nn
25943 \exp_after:wN \__fp_round_digit:Nw
25944 \int_value:w \__fp_int_eval:w #6 + 19999 - #1 ;
25945 \exp_after:wN ;}
25946 (End definition for \__fp_sqrt_auxxiv_o:wnnnnnnnN.)

72.5 About the sign and exponent

The exponent of a normal number is its \{exponent\} minus one.

\cs_new:Npn \__fp_logb_o:w ? \s__fp \__fp_chk:w #1#2; @
{ \if_case:w #1 \exp_stop_f:
  \__fp_case_use:nw
  \__fp_division_by_zero_o:Nnw \c_minus_inf_fp \{ logb \}
  \or: \exp_after:wN \__fp_logb_aux:o:w
  \or: \__fp_case_return_o:Nw \c_inf_fp
  \else: \__fp_case_return_same_o:w
  \fi:
  \s__fp \__fp_chk:w #1 #2;
}

(End definition for \__fp_logb_o:w and \__fp_logb_aux:o:w.)

Find the sign of the floating point: nan, +0, -0, +1 or -1.

\cs_new:Npn \__fp_sign_o:w ? \s__fp \__fp_chk:w #1#2; @
{ \if_case:w #1 \exp_stop_f:
  \__fp_case_return_same_o:w
  \or: \exp_after:wN \__fp_sign_aux:o:w
  \or: \__fp_case_return_same_o:w
  \else: \__fp_case_return_same_o:w
  \fi:
  \s__fp \__fp_chk:w #1 #2;
}

(End definition for \__fp_sign_o:w and \__fp_sign_aux:o:w.)
This function is used for the unary minus and for \texttt{abs}. It leaves the sign of \texttt{nan} invariant, turns negative numbers (sign 2) to positive numbers (sign 0) and positive numbers (sign 0) to positive or negative numbers depending on \#1. It also expands after itself in the input stream, just like \_\_fp_+\_o:ww.

\begin{verbatim}
25975  \cs_new:Npn \_\_fp_set_sign_o:w #1 \s__fp \_\_fp_chk:w #2#3#4; \@ 
25976  { \exp_after:wN \_\_fp_exp_after_o:w 
25977    \exp_after:wN \s__fp 
25978    \exp_after:wN \_\_fp_chk:w 
25979    \exp_after:wN \int_value:w 
25980    \if_case:w #3 \exp_stop_f: #1 \or: 1 \or: 0 \fi: \exp_stop_f: 
25981    #4; 
25982  }

\end{verbatim}

(End definition for \_\_fp_set_sign_o:w.)

\subsection{Operations on tuples}

Two cases: \texttt{abs(⟨tuple⟩)} for which \#1 is 0 (invalid for tuples) and -(⟨tuple⟩) for which \#1 is 2. In that case, map over all items in the tuple an auxiliary that dispatches to the type-appropriate sign-flipping function.

\begin{verbatim}
25985  \cs_new:Npn \_\_fp_tuple_set_sign_o:w #1 
25986  { \if_meaning:w 2 #1 \exp_after:wN \_\_fp_tuple_set_sign_aux_o:Nnw \fi: 
25987    \__fp_invalid_operation_o:nw { abs } 
25988  }

25989  \cs_new:Npn \__fp_tuple_set_sign_aux_o:Nnw #1#2#3 @ 
25990  { \__fp_tuple_map_o:nw \__fp_tuple_set_sign_aux_o:w #3 }

25991  \cs_new:Npn \__fp_tuple_set_sign_aux_o:w #1#2 ; 
25992  { \__fp_change_func_type:NNN #1 \__fp_set_sign_o:w 
25993    \__fp_parse_apply_unary_error:NNw #2 #1 #2; @ 
25994  }

\end{verbatim}

(End definition for \_\_fp_tuple_set_sign_o:w, \_\_fp_tuple_set_sign_aux_o:Nnw, and \_\_fp_tuple_set_sign_aux_o:w.)

\subsection*{For ⟨number⟩*(⟨tuple⟩) and ⟨tuple⟩*(⟨number⟩) and ⟨tuple⟩/(⟨number⟩), loop through the \langle tuple⟩ some code that multiplies or divides by the appropriate ⟨number⟩. Importantly we need to dispatch according to the type, and we make sure to apply the operator in the correct order.

\begin{verbatim}
26001  \cs_new:cpn { \__fp_\_\_fp_tuple_o:ww } #1 ; 
26002  { \__fp_tuple_map_o:nw { \__fp_binary_type_o:Nww * #1 ; } }
26003  \cs_new:cpn { \__fp_\_fp_tuple_o:ww } #1 ; #2 
26004  { \__fp_tuple_map_o:nw { \__fp_binary_rev_type_o:Nww * #2 ; } #1 ; }
26005  \cs_new:cpn { \__fp_binary_rev_type_o:Nww } #1 ; #2 
26006  { \__fp_tuple_map_o:nw { \__fp_binary_rev_type_o:Nww / #2 ; } #1 ; }

\end{verbatim}

(End definition for \_\_fp_\_\_fp_tuple_o:ww, \_\_fp_\_fp_tuple_o:ww, \_\_fp_tuple_o:ww, and \_\_fp_\_fp_tuple_/_o:ww.)
Check the two tuples have the same number of items and map through these a helper that dispatches appropriately depending on the types. This means \((1,2)+((1,1),2)\) gives \((\text{nan},4)\).

\begin{verbatim}
\cs_set_protected:Npn \__fp_tmp:w #1
\cs_new:cpn { \__fp_tuple_#1_tuple_o:ww }
\s__fp_tuple \__fp_tuple_chk:w ##1 ;
\s__fp_tuple \__fp_tuple_chk:w ##2 ;
{
\int_compare:nNnTF { \__fp_array_count:n {##1} } = { \__fp_array_count:n {##2} }
{ \__fp_tuple_mapthread_o:nww { \__fp_binary_type_o:Nww #1 } }
{ \__fp_invalid_operation_o:nww #1 }
\s__fp_tuple \__fp_tuple_chk:w {##1} ;
\s__fp_tuple \__fp_tuple_chk:w {##2} ;
}
\__fp_tmp:w +
\__fp_tmp:w -
\end{verbatim}

(End definition for \__fp_tuple_+_tuple_o:ww and \__fp_tuple_-_tuple_o:ww.)
Chapter 73

l3fp-extended implementation

73.1 Description of fixed point numbers

This module provides a few functions to manipulate positive floating point numbers with extended precision (24 digits), but mostly provides functions for fixed-point numbers with this precision (24 digits). Those are used in the computation of Taylor series for the logarithm, exponential, and trigonometric functions. Since we eventually only care about the 16 first digits of the final result, some of the calculations are not performed with the full 24-digit precision. In other words, the last two blocks of each fixed point number may be wrong as long as the error is small enough to be rounded away when converting back to a floating point number. The fixed point numbers are expressed as

\[ \langle a_1 \rangle \{ \langle a_2 \rangle \{ \langle a_3 \rangle \{ \langle a_4 \rangle \{ \langle a_5 \rangle \{ \langle a_6 \rangle ; \right. \]

where each \( a_i \) is exactly 4 digits (ranging from 0000 to 9999), except \( a_1 \), which may be any “not-too-large” non-negative integer, with or without leading zeros. Here, “not-too-large” depends on the specific function (see the corresponding comments for details). Checking for overflow is the responsibility of the code calling those functions. The fixed point number \( a \) corresponding to the representation above is \( a = \sum_{i=1}^{6} \langle a_i \rangle \cdot 10^{-4i} \).

Most functions we define here have the form

\[ \text{\_fp_fixed_(calculation):wwn} \{ \langle \text{operand}_1 \rangle \}; \{ \langle \text{operand}_2 \rangle \}; \{ \langle \text{continuation} \rangle \} \]

They perform the \( \langle \text{calculation} \rangle \) on the two \( \langle \text{operands} \rangle \), then feed the result (6 brace groups followed by a semicolon) to the \( \langle \text{continuation} \rangle \), responsible for the next step of the calculation. Some functions only accept an \( \text{N-type} \) \( \langle \text{continuation} \rangle \). This allows constructions such as

\[ \text{\_fp_fixed_add:wwn} \{ \langle X_1 \rangle \}; \{ X_2 \}; \]
\[ \text{\_fp_fixed_mul:wwn} \{ \langle X_3 \rangle \}; \]
\[ \text{\_fp_fixed_add:wwn} \{ \langle X_4 \rangle \}; \]

1094
to compute \((X_1 + X_2) \cdot X_3 + X_4\). This turns out to be very appropriate for computing continued fractions and Taylor series.

At the end of the calculation, the result is turned back to a floating point number using \texttt{fp_fixed_to_float:o:W}. This function has to change the exponent of the floating point number: it must be used after starting an integer expression for the overall exponent of the result.

### 73.2 Helpers for numbers with extended precision

#### \_fp_one_fixed_tl

The fixed-point number 1, used in \texttt{l3fp-expo}.

\begin{verbatim}
\tl_const:Nn \_fp_one_fixed_tl { {10000} {0000} {0000} {0000} {0000} {0000} ; }
\end{verbatim}

(End definition for \_fp_one_fixed_tl.)

#### \_fp_fixed_continue:wn

This function simply calls the next function.

\begin{verbatim}
\cs_new:Npn \_fp_fixed_continue:wn #1; #2 { #2 #1; }
\end{verbatim}

(End definition for \_fp_fixed_continue:wn.)

#### \_fp_fixed_add_one:wN

\begin{verbatim}
\_fp_fixed_add_one:wN \_fp_fixed_add_one:wN \langle a \rangle ; \langle continuation \rangle
\end{verbatim}

This function adds 1 to the fixed point \(\langle a \rangle\), by changing \(a_1\) to \(10000 + a_1\), then calls the \langle continuation \rangle. This requires \(a_1 + 10000 \lt 2^{31}\).

\begin{verbatim}
\cs_new:Npn \_fp_fixed_add_one:wN #1#2; #3
\exp_after:wN \_fp_fixed_mul_after:wwn \int_value:w \_fp_int_eval:w \c__fp_myriad_int + #1 ; {#2}{#3}{#4}{#5};
\end{verbatim}

(End definition for \_fp_fixed_add_one:wN.)

#### \_fp_fixed_div_myriad:wn

Divide a fixed point number by 10000. This is a little bit more subtle than just removing the last group and adding a leading group of zeros: the first group \(\#1\) may have any number of digits, and we must split \#1 into the new first group and a second group of exactly 4 digits. The choice of shifts allows \#1 to be in the range \([0, 5 \cdot 10^8 - 1]\).

\begin{verbatim}
\cs_new:Npn \_fp_fixed_div_myriad:wn #1#2#3#4#5#6;
\exp_after:wN \_fp_fixed_mul_after:wwn \int_value:w \_fp_int_eval:w \c__fp_leading_shift_int
\exp_after:wN \_fp_pack:NNNNNw \int_value:w \_fp_trailing_shift_int + #1 ; {#2}{#3}{#4}{#5};
\end{verbatim}

(End definition for \_fp_fixed_div_myriad:wn.)

#### \_fp_fixed_mul_after:wwn

The fixed point operations which involve multiplication end by calling this auxiliary. It braces the last block of digits, and places the \langle continuation \rangle \#3 in front.

\begin{verbatim}
\cs_new:Npn \_fp_fixed_mul_after:wwn #1; #2; #3 { #3 \langle #1 \rangle \#2; }
\end{verbatim}

(End definition for \_fp_fixed_mul_after:wwn.)
73.3 Multiplying a fixed point number by a short one

\_\_fp\_fixed\_mul\_short:wwn

\{⟨a1⟩\} \{⟨b1⟩\} \{⟨a2⟩\} \{⟨a3⟩\} \{⟨a4⟩\} ;
\{⟨b0⟩\} \{⟨b⟩\} ; \{⟨continuation⟩\}

Computes the product \(c = ab\) of \(a = \sum_i ⟨a_i⟩ \cdot 10^{-4i}\) and \(b = \sum_i ⟨b_i⟩ \cdot 10^{-4i}\), rounds it to the closest multiple of \(10^{-24}\), and leaves \(⟨continuation⟩\) \{⟨c1⟩\} ... \{⟨c6⟩\} ; in the input stream, where each of the \(⟨c_i⟩\) are blocks of 4 digits, except \(⟨c_1⟩\), which is any \(\text{TeX}\) integer. Note that indices for \(⟨b⟩\) start at 0: for instance a second operand of \{0001\}{0000}{0000} leaves the first operand unchanged (rather than dividing it by \(10^4\), as \_\_fp\_fixed\_mul:wwn would).

26043 \cs_new:Npn \_\_fp\_fixed\_mul\_short:wwn #1#2#3#4#5#6; #7#8#9;
26044 \begin{align*}
\exp_after:wN \_\_fp\_fixed\_mul\_after:wwn
\int_value:w \_\_fp\_int\_eval:w \c__fp\_leading\_shift\_int + #1*#7
\exp_after:wN \_\_fp\_pack:NNNNNw
\int_value:w \_\_fp\_int\_eval:w \c__fp\_middle\_shift\_int + #1*#8 + #2*#7
\exp_after:wN \_\_fp\_pack:NNNNNw
\int_value:w \_\_fp\_int\_eval:w \c__fp\_middle\_shift\_int + #1*#9 + #2*#8 + #3*#7
\exp_after:wN \_\_fp\_pack:NNNNNw
\int_value:w \_\_fp\_int\_eval:w \c__fp\_middle\_shift\_int + #2*#9 + #3*#8 + #4*#7
\exp_after:wN \_\_fp\_pack:NNNNNw
\int_value:w \_\_fp\_int\_eval:w \c__fp\_middle\_shift\_int + #3*#9 + #4*#8 + #5*#7
\exp_after:wN \_\_fp\_pack:NNNNNw
\int_value:w \_\_fp\_int\_eval:w \c__fp\_trailing\_shift\_int + #4*#9 + #5*#8 + #6*#7
+ ( #5*#9 + #6*#8 + #6*#9 / \c__fp\_myriad\_int ) / \c__fp\_myriad\_int ; ;
\end{align*}

(End definition for \_\_fp\_fixed\_mul\_short:wwn.)

\_\_fp\_fixed\_div\_int:wwN
\_\_fp\_fixed\_div\_int:wwN
\_\_fp\_fixed\_div\_int\_auxi:wnn
\_\_fp\_fixed\_div\_int\_auxii:wnn
\_\_fp\_fixed\_div\_int\_pack:Nw
\_\_fp\_fixed\_div\_int\_after:Nw

73.4 Dividing a fixed point number by a small integer

\_\_fp\_fixed\_div\_int:wwN \{⟨a⟩\}; \{⟨n⟩\}; \{⟨continuation⟩\}

Divides the fixed point number \(⟨a⟩\) by the (small) integer \(0 < ⟨n⟩ < 10^4\) and feeds the result to the \{⟨continuation⟩\}. There is no bound on \(a_1\).

The arguments of the \(r\) auxiliary are 1: one of the \(a_i\), 2: \(n\), 3: the \(ii\) or the \(iii\) auxiliary. It computes a (somewhat tight) lower bound \(Q_i\) for the ratio \(a_i/n\). The \(ii\) auxiliary receives \(Q_i\), \(n\), and \(a_i\) as arguments. It adds \(Q_i\) to a surrounding integer expression, and starts a new one with the initial value 9999, which ensures that the result of this expression has 5 digits. The auxiliary also computes \(a_i - n \cdot Q_i\), placing the result in front of the 4 digits of \(a_{i+1}\). The resulting \(a_{i+1}' = 10^4(a_i - n \cdot Q_i) + a_{i+1}\) serves as the first argument for a new call to the \(r\) auxiliary.

When the \(iii\) auxiliary is called, the situation looks like this:
where expansion is happening from the last line up. The \texttt{iii} auxiliary adds $Q_6 + 2 \simeq a_6/n + 1$ to the last 9999, giving the integer closest to $10000 + a_6/n$.

Each \texttt{pack} auxiliary receives 5 digits followed by a semicolon. The first digit is added as a carry to the integer expression above, and the 4 other digits are braced. Each call to the \texttt{pack} auxiliary thus produces one brace group. The last brace group is produced by the \texttt{after} auxiliary, which places the (continuation) as appropriate.

```
\cs_new:Npn \__fp_fixed_div_int:wwN #1 #2 #3 #4 #5 #6 ; #7 ; #8
\exp_after:wN \__fp_fixed_div_int_after:Nw
\exp_after:wN #8
\int_value:w \__fp_int_eval:w - 1
\__fp_fixed_div_int:wnN #1; {#7} \__fp_fixed_div_int_auxi:wnn #2; {#7} \__fp_fixed_div_int_auxi:wnn #3; {#7} \__fp_fixed_div_int_auxi:wnn #4; {#7} \__fp_fixed_div_int_auxi:wnn #5; {#7} \__fp_fixed_div_int_auxi:wnn #6; {#7} \__fp_fixed_div_int_auxii:wnn ;
\}
\cs_new:Npn \__fp_fixed_div_int:wnN #1; #2 #3
\exp_after:wN #3
\int_value:w \__fp_int_eval:w #1 / #2 - 1;
\{#2\}
\{#1\}
\}
\cs_new:Npn \__fp_fixed_div_int_auxi:wnn #1; #2 #3
\{ + #1 \exp_after:wN \__fp_fixed_div_int_pack:Nw
\int_value:w \__fp_int_eval:w 9999
\exp_after:wN \__fp_fixed_div_int:wnN
\int_value:w \__fp_int_eval:w #3 - #1*#2 \__fp_int_eval_end:
\}
\cs_new:Npn \__fp_fixed_div_int_auxii:wnn #1; #2 #3 \{ + #1 + 2 ; \}
\cs_new:Npn \__fp_fixed_div_int_pack:Nw #1 #2 #3 #4 #5 #6
\__fp_fixed_div_int:wwN #1#2#3#4#5#6 ; #7 ; #8
\cs_new:Npn \__fp_fixed_div_int_after:Nw #1 #2; {#1} {#2}
```

(End definition for \__fp_fixed_div_int:wwN and others.)

### 73.5 Adding and subtracting fixed points

```
\__fp_fixed_add:wwn \langle \textit{a} \rangle ; \langle \textit{b} \rangle ; \{ \langle \textit{continuation} \rangle \}
```

1097
Computes $a + b$ (resp. $a - b$) and feeds the result to the \langle continuation \rangle. This function requires $0 \leq a_1, b_1 \leq 114748$, its result must be positive (this happens automatically for addition) and its first group must have at most 5 digits: $(a \pm b)_1 < 100000$. The two functions only differ by a sign, hence use a common auxiliary. It would be nice to grab the 12 brace groups in one go; only 9 parameters are allowed. Start by grabbing the sign, $a_1, \ldots, a_4$, the rest of $a$, and $b_1$ and $b_2$. The second auxiliary receives the rest of $a$, the sign multiplying $b$, the rest of $b$, and the \langle continuation \rangle as arguments. After going down through the various level, we go back up, packing digits and bringing the \langle continuation \rangle ($#8$, then $#7$) from the end of the argument list to its start.

\begin{verbatim}
\cs_new:Npn \__fp_fixed_add:wwn { \__fp_fixed_add:Nnnnnwnn + }
\cs_new:Npn \__fp_fixed_sub:wwn { \__fp_fixed_add:Nnnnnwnn - }
\cs_new:Npn \__fp_fixed_add:Nnnnnwnn #1 #2#3#4#5 #6; #7#8 { \exp_after:wN \__fp_fixed_add_after:NNNNNwn 1 #1 #2#3#4#5 #6; #7#8 }
\cs_new:Npn \__fp_fixed_add:nnNnnnwn #1#2 #3 #4#5 #6#7 ; #8 { \exp_after:wN \__fp_fixed_add_after:NNNNNwn 1#2#3#4#5 #6 #1#2#3#4#5 #6 #7#8
\int_value:w \__fp_int_eval:w 9 9999 9998 + #2#3 #1 #7#8
\exp_after:wN \__fp_fixed_add_pack:NNNNNwn #6 #1
\int_value:w \__fp_int_eval:w 2 0000 0000 #3 #6#7 + #1#2 ; {#8} ;
\exp_after:wN \__fp_fixed_add_pack:NNNNNwn #3 #4#5 #6#7 ; #8
\exp_after:wN \__fp_fixed_add_pack:NNNNNwn
\int_value:w \__fp_int_eval:w 2 0000 0000 #3 #6#7 + #1#2 ; (#8) ;
\exp_after:wN \__fp_fixed_add_pack:NNNNNwn #1 #2#3#4#5 #6; #7
{ + #1 ; {#7} {#2#3#4#5} {#6} } \cs_new:Npn \__fp_fixed_add_after:NNNNNwn 1 #1 #2#3#4#5 #6 #7 #8 { {#7} {#1#2#3#4#5} {#6} }
\end{verbatim}

(End definition for \__fp_fixed_add:wwn and others.)

### 73.6 Multiplying fixed points

\begin{verbatim}
\__fp_fixed_mul:wwn (a) ; (b) ; \langle continuation \rangle
\end{verbatim}

Computes $a \times b$ and feeds the result to \langle continuation \rangle. This function requires $0 \leq a_1, b_1 < 10000$. Once more, we need to play around the limit of 9 arguments for \TeX macros. Note that we don’t need to obtain an exact rounding, contrarily to the $* \operatorname{operator}$, so things could be harder. We wish to perform carries in

\[
a \times b = a_1 \cdot b_1 \cdot 10^{-8} \\
+ (a_1 \cdot b_2 + a_2 \cdot b_1) \cdot 10^{-12} \\
+ (a_1 \cdot b_3 + a_2 \cdot b_2 + a_3 \cdot b_1) \cdot 10^{-16} \\
+ (a_1 \cdot b_4 + a_2 \cdot b_3 + a_3 \cdot b_2 + a_4 \cdot b_1) \cdot 10^{-20} \\
+ (a_2 \cdot b_4 + a_3 \cdot b_3 + a_4 \cdot b_2 \\
+ a_3 \cdot b_4 + a_4 \cdot b_3 + a_5 \cdot b_2 + a_6 \cdot b_1) \cdot 10^4 \\
+ a_1 \cdot b_5 + a_5 \cdot b_1) \cdot 10^{-24} + O(10^{-24}),
\]

1098
where the $O(10^{-24})$ stands for terms which are at most $5 \cdot 10^{-24}$; ignoring those leads to an error of at most 5 ulp. Note how the first 15 terms only depend on $a_1, \ldots, a_4$ and $b_1, \ldots, b_4$, while the last 6 terms only depend on $a_1, a_2, a_5, a_6$, and the corresponding parts of $b$. Hence, the first function grabs $a_1, \ldots, a_4$, the rest of $a$, and $b_1, \ldots, b_4$, and writes the 15 first terms of the expression, including a left parenthesis for the fraction. The auxiliary receives $a_5, a_6, b_1, b_2, a_1, a_2, b_5, b_6$ and finally the (continuation) as arguments. It writes the end of the expression, including the right parenthesis and the denominator of the fraction. The (continuation) is finally placed in front of the 6 brace groups by \_\_fp_fixed_mul_after:wwn.

\begin{verbatim}
\cs_new:Npn \__fp_fixed_mul:wwn #1#2#3#4 #5; #6#7#8#9
\exp_after:wN \__fp_fixed_mul_after:wwn
\int_value:w \__fp_int_eval:w \c__fp_leading_shift_int
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
+ #1*#6
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
+ #1*#7 + #2*#6
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
+ #1*#8 + #2*#7 + #3*#6
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
+ #1*#9 + #2*#8 + #3*#7 + #4*#6
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
+ #1*#9 + #2*#8 + #3*#7 + #4*#6
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
+ #1*#9 + #2*#8 + #3*#7 + #4*#6
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_trailing_shift_int
+ #2*#9 + #3*#8 + #4*#7
+ ( #3*#9 + #4*#8
+ \__fp_fixed_mul:nnnnnnnw #5 {#6}{#7} {#1}{#2}
\cs_new:Npn \__fp_fixed_mul:nnnnnnnw #1#2 #3#4 #5#6 #7#8 ;
\exp_after:wN \__fp_fixed_mul_after:wwn
\int_value:w \__fp_int_eval:w \c__fp_leading_shift_int
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
+ #1*#9 + #2*#8 + #3*#7 + #4*#6
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
+ #1*#9 + #2*#8 + #3*#7 + #4*#6
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
+ #1*#9 + #2*#8 + #3*#7 + #4*#6
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
+ #1*#9 + #2*#8 + #3*#7 + #4*#6
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_trailing_shift_int
+ #2*#9 + #3*#8 + #4*#7
+ ( #3*#9 + #4*#8
+ \__fp_fixed_mul:nnnnnnnw #5 {#6}{#7} {#1}{#2}
\end{verbatim}

(End definition for \_\_fp_fixed_mul:wwn and \_\_fp_fixed_mul:nnnnnnnw.)

73.7 Combining product and sum of fixed points

\_\_fp_fixed_mul_add:wwwn
\_\_fp_fixed_mul_sub:wwwn
\_\_fp_fixed_one_minus_mul:wwn
\_\_fp_fixed_mul_add:wwwn
\_\_fp_fixed_mul_sub:wwwn
\_\_fp_fixed_one_minus_mul:wwn
\_\_fp_fixed_mul_add:wwwn
\_\_fp_fixed_mul_sub:wwwn
\_\_fp_fixed_one_minus_mul:wwn
\_\_fp_fixed_mul_add:wwwn
\_\_fp_fixed_mul_sub:wwwn
\_\_fp_fixed_one_minus_mul:wwn

\begin{verbatim}
\_\_fp_fixed_mul_add:wwwn (a); (b); (c); (continuation)
\_\_fp_fixed_mul_sub:wwwn (a); (b); (c); (continuation)
\_\_fp_fixed_one_minus_mul:wwn (a); (b); (continuation)
\end{verbatim}

Sometimes called FMA (fused multiply-add), these functions compute $a \times b + c, c - a \times b$, and $1 - a \times b$ and feed the result to the (continuation). Those functions require $0 \leq a_1, b_1, c_1 \leq 10000$. Since those functions are at the heart of the computation of Taylor expansions, we over-optimize them a bit, and in particular we do not factor out the common parts of the three functions.
For definiteness, consider the task of computing $a \times b + c$. We perform carries in

$$a \times b + c = (a_1 \cdot b_1 + c_1 c_2) \cdot 10^{-8}$$

$$+ (a_1 \cdot b_2 + a_2 \cdot b_1) \cdot 10^{-12}$$

$$+ (a_1 \cdot b_3 + a_2 \cdot b_2 + a_3 \cdot b_1 + c_3 c_4) \cdot 10^{-16}$$

$$+ (a_1 \cdot b_4 + a_2 \cdot b_3 + a_3 \cdot b_2 + a_4 \cdot b_1) \cdot 10^{-20}$$

$$+ \left( \frac{a_1 \cdot b_4 + a_4 \cdot b_1 + a_5 \cdot b_2 + a_5 \cdot b_2}{10^4} \right)$$

$$+ a_1 \cdot b_5 + a_5 \cdot b_1 + c_5 c_6 \right) \cdot 10^{-24} + O(10^{-24}),$$

where $c_1 c_2, c_3 c_4, c_5 c_6$ denote the 8-digit number obtained by juxtaposing the two blocks of digits of $c$, and \cdot denotes multiplication. The task is obviously tough because we have 18 brace groups in front of us.

Each of the three functions starts the first two levels (the first, corresponding to $10^{-4}$, is empty), with $c_1 c_2$ in the first level, calls the i auxiliary with arguments described later, and adds a trailing $+ c_5 c_6$; (continuation) ; The $+ c_5 c_6$ piece, which is omitted for __fp_fixed_one_minus_mul:wwn, is taken in the integer expression for the $10^{-24}$ level.

\begin{verbatim}
\cs_new:Npn \__fp_fixed_mul_add:wwwn #1; #2; #3#4#5#6#7#8;
\begin{verbatim}
\exp_after:wN \__fp_fixed_mul_after:wwn
\int_value:w \__fp_int_eval:w \c__fp_big_leading_shift_int
\exp_after:wN \__fp_pack_big:NNNNNNw
\int_value:w \__fp_int_eval:w \c__fp_big_middle_shift_int + #3 #4
\__fp_fixed_mul_add:Nwnnnwnnn +
+ #5 #6 ; #2 ; #1 ; #2 ; +
+ #7 #8 ; ;
\end{verbatim}
\end{verbatim}
\cs_new:Npn \__fp_fixed_mul_sub_back:wwwn #1; #2; #3#4#5#6#7#8;
\begin{verbatim}
\exp_after:wN \__fp_fixed_mul_after:wwn
\int_value:w \__fp_int_eval:w \c__fp_big_leading_shift_int
\exp_after:wN \__fp_pack_big:NNNNNNw
\int_value:w \__fp_int_eval:w \c__fp_big_middle_shift_int + #3 #4
\__fp_fixed_mul_add:Nwnnnwnnn -
+ #5 #6 ; #2 ; #1 ; #2 ; -
+ #7 #8 ; ;
\end{verbatim}
\cs_new:Npn \__fp_fixed_one_minus_mul:wwn #1; #2;
\begin{verbatim}
\exp_after:wN \__fp_fixed_mul_after:wwn
\int_value:w \__fp_int_eval:w \c__fp_big_leading_shift_int
\exp_after:wN \__fp_pack_big:NNNNNNw
\int_value:w \__fp_int_eval:w \c__fp_big_middle_shift_int + #3 #4
\__fp_fixed_mul_add:Nwnnnwnnn -
; #2 ; #1 ; #2 ; -
; ;
\end{verbatim}
\end{verbatim}

1100
(End definition for \_fp_fixed_mul_add:wwnnnnnnnn, \_fp_fixed_mul_sub_back:wwnn, and \_fp_fixed_one_minus_mul:wwnn.)

\_fp_fixed_mul_add:nnnnwnnnn \text{ (op)} + \langle c_3 \rangle \langle c_4 \rangle ;
\langle b \rangle ; \langle a \rangle ; \langle b \rangle ; \langle \text{op} \rangle
+ \langle c_5 \rangle \langle c_6 \rangle ;

Here, \langle \text{op} \rangle is either + or −. Arguments #3, #4, #5 are \langle b_1 \rangle, \langle b_2 \rangle, \langle b_3 \rangle; arguments #7, #8, #9 are \langle a_1 \rangle, \langle a_2 \rangle, \langle a_3 \rangle. We can build three levels: \( a_1 \cdot b_1 \) for \( 10^{-20} \), \( a_1 \cdot b_2 + a_2 \cdot b_1 \) for \( 10^{-12} \), and \( a_1 \cdot b_3 + a_2 \cdot b_2 + a_3 \cdot b_1 + c_3c_4 \) for \( 10^{-10} \). The \( a \cdot b \) products use the sign #1. Note that #2 is empty for \_fp_fixed_one_minus_mul:wwnn. We call the \i auxiliary for levels \( 10^{-20} \) and \( 10^{-24} \), keeping the pieces of \langle a \rangle we’ve read, but not \langle b \rangle, since there is another copy later in the input stream.

\cs_new:Npn \_fp_fixed_mul_add:nnnnwnnnn \#1 #2; \#3\#4\#5; \#7\#8\#9
  {
    \#1 \#7\#3
    \exp_after:wN \_fp_pack_big:NNNNNw
    \int_value:w \_fp_int_eval:w \c__fp_big_middle_shift_int
    \#1 \#7\#4 \#1 \#8\#3
    \exp_after:wN \_fp_pack_big:NNNNNw
    \int_value:w \_fp_int_eval:w \c__fp_big_middle_shift_int
    \#1 \#7\#5 \#1 \#8\#4 \#1 \#9\#3 \#2
    \exp_after:wN \_fp_pack_big:NNNNNw
    \int_value:w \_fp_int_eval:w \c__fp_big_middle_shift_int
    \#1 \_fp_fixed_mul_add:nnnnwnnnn \#7 \#8 \#9
  }

(End definition for \_fp_fixed_mul_add:nnnnwnnnnn.)

\_fp_fixed_mul_add:nnnnwnnnn \langle a \rangle ; \langle b \rangle ; \langle \text{op} \rangle
+ \langle c_5 \rangle \langle c_6 \rangle ;

Level \( 10^{-20} \) is \( (a_1 \cdot b_4 + a_2 \cdot b_3 + a_3 \cdot b_2 + a_4 \cdot b_1) \), multiplied by the sign, which was inserted by the \i auxiliary. Then we prepare level \( 10^{-24} \). We don’t have access to all parts of \langle a \rangle and \langle b \rangle needed to make all products. Instead, we prepare the partial expressions

\begin{align*}
  b_1 + a_1 \cdot b_2 + a_3 \cdot b_3 + a_2 \cdot b_4 + a_1 \\
  b_2 + a_1 \cdot b_3 + a_3 \cdot b_4 + a_2.
\end{align*}

Obviously, those expressions make no mathematical sense: we complete them with \( a_5 \cdot b_5 \), and \( a b_5 \cdot a_5 \cdot b_5 \), and \( a c_5c_6 \), and of course with the trailing \( c_5c_6 \).

To do all this, we keep \( a_1, a_5, a_6 \), and the corresponding pieces of \langle b \rangle.

\cs_new:Npn \_fp_fixed_mul_add:nnnnwnnnn \#1\#2\#3\#4\#5; \#6\#7\#8\#9
  {
    ( \#1\#9 + \#2\#8 + \#3\#7 + \#4\#6 )
    \exp_after:wN \_fp_pack_big:NNNNNw
    \int_value:w \_fp_int_eval:w \c__fp_big_trailing_shift_int
    \_fp_fixed_mul_add:nnnnwnnnn
      \{ \#6 + \#4\#7 + \#3\#8 + \#2\#9 + \#1 \}
      \{ \#7 + \#4\#8 + \#3\#9 + \#2 \}
      \{ \#1 \} \#5;
    \{ \#6 \}
  }

1101
73.8 Extended-precision floating point numbers

In this section we manipulate floating point numbers with roughly 24 significant figures ("extended-precision" numbers, in short, "ep"), which take the form of an integer exponent, followed by a comma, then six groups of digits, ending with a semicolon. The first group of digits may be any non-negative integer, while other groups of digits have 4 digits. In other words, an extended-precision number is an exponent ending in a comma, then a fixed point number. The corresponding value is $0.\langle\text{digits}\rangle \cdot 10^{\langle\text{exponent}\rangle}$. This convention differs from floating points.

Converting an extended-precision number with an exponent at most 4 and a first block less than $10^8$ to a fixed point number whose first block has 12 digits, hopefully starting with many zeros.

\begin{verbatim}
\cs_new:Npn \__fp_ep_to_fixed:wwn #1,#2
{ \exp_after:wN \__fp_ep_to_fixed_auxi:www \int_value:w \__fp_int_eval:w 1 0000 0000 + #2 \exp_after:wN ; \prg_replicate:nn { 4 - \int_max:nn {#1} { -32 } } { 0 } ; }
\cs_new:Npn \__fp_ep_to_fixed_auxi:www \prg_replicate:nnn {\prg_replicate:nnn } {#1#2}{#3#4#5#6#7} 0000 !
\end{verbatim}

(End definition for \__fp_fixed_mul_add:nnnnwnnnn.)
Normalize an extended-precision number. More precisely, leading zeros are removed from the mantissa of the argument, decreasing its exponent as appropriate. Then the digits are packed into 6 groups of 4 (discarding any remaining digit, not rounding). Finally, the continuation #8 is placed before the resulting exponent–mantissa pair. The input exponent may in fact be given as an integer expression. The loop auxiliary grabs a digit: if it is 0, decrement the exponent and continue looping, and otherwise call the end auxiliary, which places all digits in the right order (the digit that was not 0, and any remaining digits), followed by some 0, then packs them up neatly in \(3 \times 2 = 6\) blocks of four. At the end of the day, remove with \_fp_use_i:ww any digit that did not make it in the final mantissa (typically only zeros, unless the original first block has more than 4 digits).

\begin{verbatim}
\cs_new:Npn \_fp_ep_to_ep:wwN #1,#2#3#4#5#6#7; #8
\{ \exp_after:wN #8 \int_value:w \__fp_int_eval:w #1 + 4 \exp_after:wN \use_i:nn \exp_after:wN \_fp_ep_to_ep_loop:N \int_value:w \__fp_int_eval:w 1 0000 00000 + #2 \_fp_int_eval_end: #3#4#5#6#7; ; ! \}
\cs_new:Npn \_fp_ep_to_ep_loop:N #1
{ \if_meaning:w 0 #1 - 1 \else: \_fp_ep_to_ep_end:www #1 \fi: \_fp_ep_to_ep_loop:N \}
\cs_new:Npn \_fp_ep_to_ep_end:www #1 \fi: \_fp_ep_to_ep_loop:N #2; #3!
\{ \fi: \_fp_ep_to_ep_zero:ww \fi: , ; #1 #2 0000 0000 0000 0000 0000 0000 ; \}
\cs_new:Npn \_fp_ep_to_ep_zero:ww \fi: #1; #2; #3;
{ \fi: , {1000}{0000}{0000}{0000}{0000}{0000} ; }
\end{verbatim}

In l3fp-trig we need to compare two extended-precision numbers. This is based on the same function for positive floating point numbers, with an extra test if comparing only 16 decimals is not enough to distinguish the numbers. Note that this function only works if the numbers are normalized so that their first block is in \([1000, 9999]\).

\begin{verbatim}
\cs_new:Npn \_fp_ep_compare:wwww \_fp_ep_compare_aux:wwww
\{ \_fp_ep_compare:wwww #1,#2#3#4#5#6#7; \}
\end{verbatim}
Multiply two extended-precision numbers: first normalize them to avoid losing too much precision, then multiply the mantissas \#2 and \#4 as fixed point numbers, and sum the exponents \#1 and \#3. The result’s first block is in \([100, 9999]\).

73.9 Dividing extended-precision numbers

Divisions of extended-precision numbers are difficult to perform with exact rounding: the technique used in \l3fp\-basics for 16-digit floating point numbers does not generalize easily to 24-digit numbers. Thankfully, there is no need for exact rounding.

Let us call \(\langle n \rangle\) the numerator and \(\langle d \rangle\) the denominator. After a simple normalization step, we can assume that \(\langle n \rangle \in [0.1, 1)\) and \(\langle d \rangle \in [0.1, 1)\), and compute \(\langle n \rangle/(10\langle d \rangle) \in (0.01, 1)\). In terms of the 6 blocks of digits \(\langle n_1 \rangle \cdots \langle n_6 \rangle\) and the 6 blocks \(\langle d_1 \rangle \cdots \langle d_6 \rangle\), the condition translates to \(\langle n_1 \rangle, \langle d_1 \rangle \in [1000, 9999]\).
We first find an integer estimate \( a \simeq 10^8/\langle d \rangle \) by computing

\[
\begin{align*}
\alpha &= \left\lfloor \frac{10^9}{\langle d_1 \rangle + 1} \right\rfloor \\
\beta &= \left\lfloor \frac{10^9}{\langle d_1 \rangle} \right\rfloor \\
a &= 10^3\alpha + (\beta - \alpha) \cdot \left( 10^3 - \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor \right) - 1250,
\end{align*}
\]

where \( \left\lfloor \frac{d}{2} \right\rfloor \) denotes \( \varepsilon \)-\TeX’s rounding division, which rounds ties away from zero. The idea is to interpolate between \( 10^3\alpha \) and \( 10^3\beta \) with a parameter \( \langle d_2 \rangle/10^4 \), so that when \( \langle d_2 \rangle = 0 \) one gets \( a = 10^3\beta - 1250 \simeq 10^{12}/\langle d_1 \rangle \simeq 10^8/\langle d \rangle \), while when \( \langle d_2 \rangle = 9999 \) one gets \( a = 10^3\alpha - 1250 \simeq 10^{12}/(\langle d_1 \rangle + 1) \simeq 10^8/\langle d \rangle \). The shift by 1250 helps to ensure that \( a \) is an underestimate of the correct value. We shall prove that

\[
1 - 1.755 \cdot 10^{-5} < \frac{\langle d \rangle a}{10^8} < 1.
\]

We can then compute the inverse of \( \langle d \rangle a/10^8 = 1 - \epsilon \) using the relation \( 1/(1 - \epsilon) \simeq (1 + \epsilon)(1 + \epsilon^2) + \epsilon^4 \), which is correct up to a relative error of \( \epsilon^6 < 1.6 \cdot 10^{-24} \). This allows us to find the desired ratio as

\[
\langle n \rangle/\langle d \rangle = \frac{\langle n \rangle a}{10^8} ((1 + \epsilon)(1 + \epsilon^2) + \epsilon^4).
\]

Let us prove the upper bound first (multiplied by \( 10^{15} \)). Note that \( 10^7 \langle d \rangle < 10^3\langle d_1 \rangle + 10^{-1}\langle d_2 \rangle + 1 \), and that \( \varepsilon \)-\TeX’s division \( \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor \) underestimates \( 10^{-1}\langle d_2 \rangle + 1 \) by 0.5 at most, as can be checked for each possible last digit of \( \langle d_2 \rangle \). Then,

\[
10^7 \langle d \rangle a < \left( 10^3\langle d_1 \rangle + \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor + \frac{1}{2} \right) \left( 10^3 - \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor \right) \beta + \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor \alpha - 1250
\]

\[
< 10^3\langle d_1 \rangle + \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor + \frac{1}{2} \left( 10^3 - \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor \right) \left( \frac{\langle d_1 \rangle + 1}{2} \right) \left( \frac{\langle d_2 \rangle}{10} \right) \beta + \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor \alpha - 1250
\]

\[
< \left( 10^3\langle d_1 \rangle + \left\lfloor \frac{\langle d_2 \rangle}{10} \right\rfloor + \frac{1}{2} \right) \left( \frac{10^9}{\langle d_1 \rangle} + \frac{1}{2} \right) \left( \frac{\langle d_2 \rangle}{10} \right) \alpha - 750
\]

We recognize a quadratic polynomial in \( \left\lfloor \langle d_2 \rangle/10 \right\rfloor \) with a negative leading coefficient: this polynomial is bounded above, according to \( (\langle d_2 \rangle/10) \cdot (b - c(\langle d_2 \rangle/10)) \leq (b-c)^2/(4c) \). Hence,

\[
10^7 \langle d \rangle a < \frac{10^{15}}{\langle d_1 \rangle(\langle d_1 \rangle + 1)} \left( \frac{1}{2} + \frac{1}{4} 10^{-3} - \frac{3}{8} \cdot 10^{-9} \langle d_1 \rangle(\langle d_1 \rangle + 1) \right)^2
\]

Since \( \langle d_1 \rangle \) takes integer values within \([1000, 9999]\), it is a simple programming exercise to check that the squared expression is always less than \( \langle d_1 \rangle(\langle d_1 \rangle + 1) \), hence \( 10^7 \langle d \rangle a < 10^{15} \). The upper bound is proven. We also find that \( \frac{7}{8} \) can be replaced by slightly smaller numbers, but nothing less than 0.374563..., and going back through the derivation of
the upper bound, we find that 1250 is as small a shift as we can obtain without breaking the bound.

Now, the lower bound. The same computation as for the upper bound implies

\[
10^7 \langle d \rangle a > \left( 10^3 \langle d_1 \rangle + \frac{\langle d_2 \rangle}{10} - \frac{1}{2} \right) \left( 10^{12} - \frac{\langle d_2 \rangle}{10} \right) - 1750
\]

This time, we want to find the minimum of this quadratic polynomial. Since the leading coefficient is still negative, the minimum is reached for one of the extreme values \( y/10 = 0 \) or \( y/10 = 100 \), and we easily check the bound for those values.

We have proven that the algorithm gives us a precise enough answer. Incidentally, the upper bound that we derived tells us that \( a < \frac{10^8}{\langle d \rangle} \leq 10^9 \), hence we can compute \( a \) safely as a \( \text{TeX} \) integer, and even add \( 10^9 \) to it to ease grabbing of all the digits. The lower bound implies \( 10^8 - 1755 < a \), which we do not care about.

\[
\text{\_fp\_ep\_div:wwwwn}
\]

Compute the ratio of two extended-precision numbers. The result is an extended-precision number whose first block lies in the range \([100,9999]\), and is placed after the \( \langle \text{continuation} \rangle \) once we are done. First normalize the inputs so that both first block lie in \([1000,9999]\), then call \( \text{\_fp\_ep\_div\_esti:wwwwn} \) (denominator) \( \langle \text{numerator} \rangle \), responsible for estimating the inverse of the denominator.

\[
\text{\_fp\_ep\_div\_esti:wwwwn}
\]

The \( \text{esti} \) function evaluates \( \alpha = \frac{10^9}{\langle d_1 \rangle + 1} \), which is used twice in the expression for \( a \), and combines the exponents \#1 and \#4 (with a shift by 1 because we later compute \( \langle n \rangle/(10\langle d \rangle) \)). Then the \( \text{estii} \) function evaluates \( 10^9 + a \), and puts the exponent \#2 after the continuation \#7: from there on we can forget exponents and focus on the mantissa. The \( \text{estiii} \) function multiplies the denominator \#7 by \( 10^{-8}a \) (obtained as a split into the single digit \#1 and two blocks of 4 digits, \#2#3#4#5 and \#6). The result \( 10^{-8}a(d) = (1 - \epsilon) \), and a partially packed \( 10^{-9}a \) (as a block of four digits, and five individual digits, not packed by lack of available macro parameters here) are passed to \( \text{\_fp\_ep\_div\_epsi:wnNNNNn} \), which computes \( 10^{-9}a/(1 - \epsilon) \), that is, \( 1/(10\langle d \rangle) \) and we finally multiply this by the numerator \#8.

\[
\text{\_fp\_ep\_div\_epsi:wnNNNNn}
\]

\[
\text{\_fp\_ep\_div\_epsi:wwNNWWWn}
\]

\[
\text{\_fp\_ep\_div\_epsi:wwNNnnWWnn}
\]

\[
\text{\_fp\_ep\_div\_epsi:wwNNNNn}
\]

\[
\text{\_fp\_ep\_div\_epsi:wwNNNNn}
\]

\[
\text{\_fp\_ep\_div\_epsi:wwNNNNn}
\]

\[
\text{\_fp\_ep\_div\_epsi:wwNNNNn}
\]

\[
\text{\_fp\_ep\_div\_epsi:wwNNNNn}
\]
The bounds shown above imply that the \texttt{epsi} function's first operand is 
$(1 - \epsilon)$ with \(\epsilon \in [0, 1.755 \cdot 10^{-5}]\). The \texttt{epsi} function computes \(\epsilon\) as 
$1 - (1 - \epsilon)$. Since \(\epsilon < 10^{-4}\), its first block vanishes and there is no need to explicitly use \#1 (which is 9999). Then \texttt{epsii} evaluates 
$10^{-9}a/(1 - \epsilon)$ as $(1 + \epsilon^2)(1 + \epsilon)(10^{-9}a\epsilon) + 10^{-9}a$. Importantly, we compute 
$10^{-9}a\epsilon$ before multiplying it with the rest, rather than multiplying by \(\epsilon\) and then 
$10^{-9}a$, as this second option loses more precision. Also, the combination of \texttt{short_mul} and 
\texttt{div_myriad} is both faster and more precise than a simple \texttt{mul}.

\begin{verbatim}
\exp_after:wN \__fp_ep_div_estiii:NNNNNwwwn
\int_value:w \__fp_int_eval:w 10 0000 0000 - 1750
+ #1 000 + (10 0000 0000 / #3 - #1) * (1000 - #4 / 10);
\exp_after:wN \__fp_fixed_mul_short:wwn #7; {#1#2#3#4#5}{#6}; #6; { #7 #2, }
\cs_new:Npn \__fp_ep_div_estiii:NNNNNwwwn #1#2#3#4#5#6; #7;
{ \__fp_fixed_mul_short:wwn #7; {#1}{#2#3#4#5}{#6};
  \__fp_ep_div_epsii:wnNNNNN #1#2#3#4#5#6
  \__fp_fixed_mul:wwn
} (End definition for \__fp_ep_div_estiii:NNNNNwwwn, \__fp_ep_div_estii:wwnnwwn, and \__fp_ep_div_estiii:NNNNNwwwn.)
\__fp_ep_div_epsii:wwnNNNNN
\__fp_ep_div_eps_pack:NNNNNw
\__fp_ep_div_epsii:wwnNNNNN
\_fp_ep_div_epsi:wnNNNNN
\_fp_ep_div_eps_pack:NNNNNw
\_fp_ep_div_epsi:wwnNNNNN
\end{verbatim}
73.10 Inverse square root of extended precision numbers

The idea here is similar to division. Normalize the input, multiplying by powers of 100 until we have \( x \in [0.01, 1) \). Then find an integer approximation \( r \in [101, 1003] \) of \( 10^2/\sqrt{x} \), as the fixed point of iterations of the Newton method: essentially \( r \mapsto (r + 10^8/(rx))/2 \), starting from a guess that optimizes the number of steps before convergence. In fact, just as there is a slight shift when computing divisions to ensure that some inequalities hold, we replace \( 10^8 \) by a slightly larger number which ensures that \( r^2x \geq 10^4 \). This also causes \( r \in [101, 1003] \). Another correction to the above is that the input is actually normalized to \([0.1, 1)\), and we use either \( 10^8 \) or \( 10^9 \) in the Newton method, depending on the parity of the exponent. Skipping those technical hurdles, once we have the approximation \( r \), we set \( y = 10^{-4}r^2x \) (or rather, the correct power of 10 to get \( y \approx 1 \)) and compute \( y^{-1/2} \) through another application of Newton’s method. This time, the starting value is \( z = 1 \), each step maps \( z \mapsto (1.5 - 0.5yz^2) \), and we perform a fixed number of steps. Our final result combines \( r \) with \( y^{-1/2} \) as \( x^{-1/2} = 10^{−2}ry^{-1/2} \).

First normalize the input, then check the parity of the exponent #1. If it is even, the result’s exponent will be \(-#1/2\), otherwise it will be \((#1 - 1)/2\) (except in the case where the input was an exact power of 100). The auxii function receives as #1 the result’s exponent just computed, as #2 the starting value for the iteration giving \( r \) (the values 168 and 535 lead to the least number of iterations before convergence, on average), as #3 and #4 one empty argument and one 0, depending on the parity of the original exponent, as #5 and #6 the normalized mantissa (\( #5 \in [1000, 9999] \)), and as #7 the continuation. It sets up the iteration giving \( r \): the esti function thus receives the initial two guesses \( #2 \) and 0, an approximation \( #5 \) of \( 10^4x \) (its first block of digits), and the empty/zero arguments #3 and #4, followed by the mantissa and an altered continuation where we have stored the result’s exponent.

If the last two approximations gave the same result, we are done: call the estii function to clean up. Otherwise, evaluate \(((\text{prev}) + 1.005 \cdot 10^8)/(\text{prev} \cdot x))/2\), as the next approximation: omitting the 1.005 factor, this would be Newton’s method. We can
check by brute force that if #4 is empty (the original exponent was even), the process computes an integer slightly larger than $100/\sqrt{x}$, while if #4 is 0 (the original exponent was odd), the result is an integer slightly larger than $100/\sqrt{x}/10$. Once we are done, we evaluate $100r^2/2$ or $10r^2/2$ (when the exponent is even or odd, respectively) and feed that to estiii. This third auxiliary finds $y_{even}/2 = 10^{-4}r^2x/2$ or $y_{odd}/2 = 10^{-5}r^2x/2$ (again, depending on earlier parity). A simple program shows that $y \in [1, 1.0201]$. The number $y/2$ is fed to \_\_fp_ep_isqrt_epsi:wN, which computes $1/\sqrt{y}$, and we finally multiply the result by $r$.

\begin{verbatim}
\cs_new:Npn \_\_fp_ep_isqrt_esti:wwwnn #1, #2, #3, #4
\{\if_int_compare:w #1 = #2 \exp_stop_f:\exp_after:wN \_\_fp_ep_isqrt_estii:wwwnn #1, #3, {#4}\fi: \exp_after:wN \_\_fp_ep_isqrt_esti:wwwnn #1, #3, {#4}\int_value:w \_\_fp_int_eval:w \(#1 + 1 0050 0000 #4 / (#1 * #3)) / 2 , #1, #3, {#4}\}
\cs_new:Npn \_\_fp_ep_isqrt_estii:wwwnn #1, #2, #3, #4#5
\{\exp_after:wN \_\_fp_ep_isqrt_estiii:NNNNNwwwn #1, \int_value:w \_\_fp_int_eval:w 1000 0000 + #2 * #2 #5 * 5
\exp_after:wN , \int_value:w \_\_fp_int_eval:w 10000 + #2 ; #1, #3, {#4}\}
\cs_new:Npn \_\_fp_ep_isqrt_estiii:NNNNNwwwn 1#1#2#3#4#5#6, 1#7#8; #9;
\{\_\_fp_fixed_mul_short:wwn #9; {#1} \{#2#3#4#5} \{#600} ; \_\_fp_ep_isqrt_epsii:wN \_\_fp_fixed_mul_short:wwn \{#7\} \{#80\} {0000} ;
\}
\end{verbatim}

Here, we receive a fixed point number $y/2$ with $y \in [1, 1.0201]$. Starting from $z = 1$ we iterate $z \mapsto z(3/2 - z^2y/2)$. In fact, we start from the first iteration $z = 3/2 - y/2$ to avoid useless multiplications. The epsii auxiliary receives $z$ as #1 and $y$ as #2.

\begin{verbatim}
\cs_new:Npn \_\_fp_ep_isqrt_epsii:wN #1; \_\_fp_ep_isqrt_epsii:wN \{15000}{0000}{0000}{0000}; #1;
\cs_new:Npn \_\_fp_ep_isqrt_epsii:wN #1; \_\_fp_ep_isqrt_epsii:wN \{15000}{0000}{0000}{0000}; #1;
\cs_new:Npn \_\_fp_ep_isqrt_epsii:wN #1; \_\_fp_ep_isqrt_epsii:wN \{15000}{0000}{0000}{0000}; #1;
\cs_new:Npn \_\_fp_ep_isqrt_epsii:wN #1; \_\_fp_ep_isqrt_epsii:wN \{15000}{0000}{0000}{0000}; #1;
\end{verbatim}

(End definition for \_\_fp_ep_isqrt_esti:wwwnn, \_\_fp_ep_isqrt_estii:wwwnn, and \_\_fp_ep_isqrt_estiii:NNNNNwwwn.)
73.11 Converting from fixed point to floating point

After computing Taylor series, we wish to convert the result from extended precision (with or without an exponent) to the public floating point format. The functions here should be called within an integer expression for the overall exponent of the floating point.

An extended-precision number is simply a comma-delimited exponent followed by a fixed point number. Leave the exponent in the current integer expression then convert the fixed point number.

\begin{verbatim}
\_fp_ep_to_float_o:wwN \_fp_ep_inv_to_float_o:wwN
\end{verbatim}

Another function which reduces to converting an extended precision number to a float.

\begin{verbatim}
\_fp_fixed_to_float_rad_o:wN
\end{verbatim}

And the \texttt{to_fixed} version gives six brace groups instead of 4, ensuring that 1000 ≤ ⟨a₁⟩ ≤ 9999. At this stage, we know that ⟨a₁⟩ is positive (otherwise, it is sign of an error before), and we assume that it is less than 10^8.¹¹

¹¹ Bruno: I must double check this assumption.
Chapter 74

l3fp-expo implementation

Unary functions.

\__fp_parse_word_exp:N
\cs_new:Npn \__fp_parse_word_exp:N { \__fp_parse_unary_function:NNN \__fp_exp_o:w ? }
\__fp_parse_word_ln:N
\cs_new:Npn \__fp_parse_word_ln:N { \__fp_parse_unary_function:NNN \__fp_ln_o:w ? }
\__fp_parse_word_fact:N
\cs_new:Npn \__fp_parse_word_fact:N { \__fp_parse_unary_function:NNN \__fp_fact_o:w ? }

(End definition for \__fp_parse_word_exp:N, \__fp_parse_word_ln:N, and \__fp_parse_word_fact:N)

74.1 Logarithm

74.1.1 Work plan

As for many other functions, we filter out special cases in \__fp_ln_o:w. Then \__fp_-ln_npos_o:w receives a positive normal number, which we write in the form \( a \cdot 10^b \) with \( a \in [0,1) \).

The rest of this section is actually not in sync with the code. Or is the code not in sync with the section? In the current code, \( c \in [1,10] \) is such that \( 0.7 \leq ac < 1.4 \).

We are given a positive normal number, of the form \( a \cdot 10^b \) with \( a \in [0,1) \). To compute its logarithm, we find a small integer \( 5 \leq c < 50 \) such that \( 0.91 \leq \frac{ac}{5} < 1.1 \), and use the relation

\[
\ln(a \cdot 10^b) = b \cdot \ln(10) - \ln(c/5) + \ln(ac/5).
\]

The logarithms \( \ln(10) \) and \( \ln(c/5) \) are looked up in a table. The last term is computed using the following Taylor series of \( \ln \) near 1:

\[
\ln \left( \frac{ac}{5} \right) = \ln \left( \frac{1 + t}{1 - t} \right) = 2t \left( 1 + t^2 \left( \frac{1}{3} + t^2 \left( \frac{1}{5} + t^2 \left( \frac{1}{7} + t^2 \left( \frac{1}{9} + \ldots \right) \right) \right) \right) \right)
\]

where \( t = 1 - 10/(ac + 5) \). We can now see one reason for the choice of \( ac \sim 5 \): then \( ac + 5 = 10(1 - \epsilon) \) with \( -0.05 < \epsilon \leq 0.045 \), hence

\[
t = \frac{\epsilon}{1 - \epsilon} = \epsilon(1 + \epsilon)(1 + \epsilon^2)(1 + \epsilon^4) \ldots,
\]
is not too difficult to compute.

### 74.1.2 Some constants

A few values of the logarithm as extended fixed point numbers. Those are needed in the implementation. It turns out that we don’t need the value of \( \ln(5) \).  

\begin{verbatim}
\tl_const:Nn \c__fp_ln_i_fixed_tl { {0000}{0000}{0000}{0000}{0000}{0000};}
\tl_const:Nn \c__fp_ln_ii_fixed_tl { {6931}{4718}{0559}{9453}{0941}{7232};}
\tl_const:Nn \c__fp_ln_iii_fixed_tl { {10986}{1228}{8668}{1096}{9139}{5245};}
\tl_const:Nn \c__fp_ln_iv_fixed_tl { {13862}{9436}{1119}{8906}{1883}{4464};}
\tl_const:Nn \c__fp_ln_vi_fixed_tl { {17917}{5946}{9228}{0550}{0081}{2477};}
\tl_const:Nn \c__fp_ln_vii_fixed_tl { {19459}{1014}{9055}{3133}{0510}{5353};}
\tl_const:Nn \c__fp_ln_viii_fixed_tl { {20794}{4154}{1679}{8359}{2825}{1896};}
\tl_const:Nn \c__fp_ln_ix_fixed_tl { {21972}{2457}{7336}{2193}{8279}{0490};}
\tl_const:Nn \c__fp_ln_x_fixed_tl { {23025}{8509}{2994}{0456}{8401}{7991};}
\end{verbatim}

(End definition for \c__fp_ln_i_fixed_tl and others.)

### 74.1.3 Sign, exponent, and special numbers

\__fp_ln_o:w

The logarithm of negative numbers (including \( -\infty \) and \( -0 \)) raises the “invalid” exception. The logarithm of \(+0\) is \(-\infty\), raising a division by zero exception. The logarithm of \(+\infty\) or a \( \text{nan} \) is itself. Positive normal numbers call \__fp_ln_npos_o:w.

\begin{verbatim}
\cs_new:Npn \__fp_ln_o:w #1 \s__fp \__fp_chk:w #2#3#4; @
\if_meaning:w 2 #3
\__fp_case_use:nw { \__fp_invalid_operation_o:nw { ln } }
\fi:
\if_case:w #2 \exp_stop_f:
\__fp_case_use:nw{ \__fp_division_by_zero_o:Nnw \c_minus_inf_fp { ln } }
\or:
\else:
\__fp_case_return_same_o:w
\__fp_ln_npos_o:w \s__fp \__fp_chk:w #2#3#4;
\end{verbatim}

(End definition for \__fp_ln_o:w.)

### 74.1.4 Absolute \( \ln \)

\__fp_ln_npos_o:w

We catch the case of a significand very close to 0.1 or to 1. In all other cases, the final result is at least \( 10^{-4} \), and then an error of \( 0.5 \cdot 10^{-20} \) is acceptable.

\begin{verbatim}
\cs_new:Npn \__fp_ln_npos_o:w \s__fp \__fp_chk:w \l0\#1\#2\#3;
{ \%"A todo: \ln(1) should be "exact zero", not "underflow"
\exp_after:wN \__fp_sanitize:Nw
\int_value:w % for the overall sign
\if_int_compare:w #1 < \c_one_int
\_2
\else:
0
\fi:
\end{verbatim}

1113
This function expands to
\begin{align*}
\langle X_1 \rangle \{ \langle X_2 \rangle \} \{ \langle X_3 \rangle \} \{ \langle X_4 \rangle \};
\end{align*}
where $Y = -\ln(X)$ as an extended fixed point.

\begin{verbatim}
\cs_new:Npn \__fp_ln_significand:NNNNnnnN #1#2#3#4 
{ \exp_after:wN \__fp_ln_x_ii:wnnn \int_value:w #1 \exp_stop_f: 
\or: \if_int_compare:w #2 < 4 \exp_stop_f: 
\__fp_int_eval:w 10 - #2 
\else: 4 \or: 3 \or: 2 \or: 2 \or: 2 \else: 1 \fi: 
\fi: \or: 4 \or: 2 \or: 2 \else: 1 \fi: 
\fi: \or: 6 \fi: 
\or: 4 \or: 2 \or: 2 \else: 1 \fi: 
\fi: \or: 6 \fi: 
\or: 4 \or: 2 \else: 1 \fi: 
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__fp_ln_x_ii:wnnn \int_value:w #1 \exp_stop_f: 
\exp_after:wN \__fp_ln_div_after:Nw \cs:w c__fp_ln_ \__fp_int_to_roman:w #1 _fixed_tl \exp_after:wN \cs_end:
\int_value:w \__fp_int_eval:w \__fp_ln_x_iii_var:NNNNNw \int_value:w \__fp_int_eval:w #1 10 0000 0000 + #1*#2#3 + #4 * #5 ;
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__fp_x_ii:wnnn \int_value:w #1 \exp_stop_f: 
\exp_after:wN \__fp_ln_div_after:Nw \cs:w c__fp_ln_ \__fp_int_to_roman:w #1 _fixed_tl \exp_after:wN \cs_end:
\int_value:w \__fp_int_eval:w \__fp_x_iii_var:NNNNNw \int_value:w \__fp_int_eval:w 9999 9990 + #1*#2#3 + #4*#5 +;
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__fp_x_iii_var:NNNNNw #1#2#3#4#5 #6; 
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__fp_ln_x_ii:wnnn \int_value:w #1 \exp_stop_f: 
\exp_after:wN \__fp_ln_div_after:Nw \cs:w c__fp_ln_ \__fp_int_to_roman:w #1 _fixed_tl \exp_after:wN \cs_end:
\int_value:w \__fp_int_eval:w \__fp_x_iii_var:NNNNNw \int_value:w \__fp_int_eval:w #1 10 0000 0000 + #1*#2#3 + #4*#5 +;
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__fp_x_iii_var:NNNNNw #1#2#3#4#5 #6; 
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__fp_ln_x_iii_var:NNNNNw #1 #2 #3 #4 #5 #6; 
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__fp_x_iii_var:NNNNNw #1#2#3#4#5 #6; 
\end{verbatim}

\begin{verbatim}
\cs_new:Npn \__fp_x_iii_var:NNNNNw #1#2#3#4#5 #6; 
\end{verbatim}
The Taylor series to be used is expressed in terms of \( t = (x - 1)/(x + 1) = 1 - 2/(x + 1) \). We now compute the quotient with extended precision, reusing some code from \texttt{\_fp_\_o:ww}. Note that 1 + \( x \) is known exactly.

To reuse notations from \texttt{l3fp-basics}, we want to compute \( A/Z \) with \( A = 2 \) and \( Z = x + 1 \). In \texttt{l3fp-basics}, we considered the case where both \( A \) and \( Z \) are arbitrary, in the range \([0.1, 1]\), and we had to monitor the growth of the sequence of remainders \( A, B, C, \) etc. to ensure that no overflow occurred during the computation of the next quotient. The main source of risk was our choice to define the quotient as roughly \( 10^9 \cdot A/10^5 \cdot Z \): then \( A \) was bound to be below \( 2 \cdot 1.47 \cdots \), and this limit was never far.

In our case, we can simply work with \( 10^8 \cdot A \) and \( 10^4 \cdot Z \), because our reason to work with higher powers has gone: we needed the integer \( y \approx 10^5 \cdot Z \) to be at least \( 10^4 \), and now, the definition \( y \approx 10^4 \cdot Z \) suffices.

Let us thus define \( y = \lfloor 10^4 \cdot Z \rfloor + 1 \in (1.7 \cdot 10^4, 2.4 \cdot 10^4] \), and

\[
Q_1 = \left\lfloor \frac{10^8 \cdot A}{y} - \frac{1}{2} \right\rfloor.
\]

(The 1/2 comes from how \( \varepsilon \)-\LaTeX{} rounds.) As for division, it is easy to see that \( Q_1 \leq 10^4 A/Z \), i.e., \( Q_1 \) is an underestimate.

Exactly as we did for division, we set \( B = 10^4 A - Q_1 Z \). Then

\[
10^4 B \leq A_1 A_2 A_3 A_4 - \left( \frac{A_1 A_2}{y} - \frac{3}{2} \right) 10^4 Z
\]

\[
\leq A_1 A_2 \left( 1 - \frac{10^4 Z}{y} \right) + 1 + \frac{3}{2} y
\]

\[
\leq 10^8 \frac{A}{y} + 1 + \frac{3}{2} y
\]

In the same way, and using \( 1.7 \cdot 10^4 \leq y \leq 2.4 \cdot 10^4 \), and convexity, we get

\[
10^4 A = 2 \cdot 10^4
\]

\[
10^4 B \leq 10^8 \frac{A}{y} + 1.6 \leq 4.7 \cdot 10^4
\]

\[
10^4 C \leq 10^8 \frac{B}{y} + 1.6 \leq 5.8 \cdot 10^4
\]

\[
10^4 D \leq 10^8 \frac{C}{y} + 1.6 \leq 6.3 \cdot 10^4
\]

\[
10^4 E \leq 10^8 \frac{D}{y} + 1.6 \leq 6.5 \cdot 10^4
\]

\[
10^4 F \leq 10^8 \frac{E}{y} + 1.6 \leq 6.6 \cdot 10^4
\]

Note that we compute more steps than for division: since \( t \) is not the end result, we need to know it with more accuracy (on the other hand, the ending is much simpler, as we don’t need an exact rounding for transcendental functions, but just a faithful rounding).
The number is $x$. Compute $y$ by adding 1 to the five first digits.

We now have essentially

\begin{verbatim}
\_fp_ln_x_iv:wnnnnnnn (1 or 2) \{8d\} ; \{4d\} \{(4d)\} \{\fixed-tl\}
\end{verbatim}

where \{\fixed tl\} holds the logarithm of a number in $[1, 10]$, and \{\exponent\} is the exponent. Also, the expansion is done backwards. Then \_fp_div_significand_pack:NNN puts things in the correct order to add the $Q_i$ together and put semicolons between each piece. Once those have been expanded, we get

\begin{verbatim}
\_fp_ln_div_after:Nw \{\fixed tl\} \_fp_div_significand_pack:NNN 10^0 + Q_1;
\_fp_div_significand_pack:NNN 10^0 + Q_2;
\_fp_div_significand_pack:NNN 10^0 + Q_3;
\_fp_div_significand_pack:NNN 10^0 + Q_4;
\_fp_div_significand_pack:NNN 10^0 + Q_5;
\_fp_div_significand_pack:NNN 10^0 + Q_6;
\{\exponent\}; \{\continuation\}
\end{verbatim}

1116
Just as with division, we know that the first two digits are 1 and 0 because of bounds on the final result of the division $2/(x + 1)$, which is between roughly 0.8 and 1.2. We then compute $1 - 2/(x + 1)$, after testing whether $2/(x + 1)$ is greater than or smaller than 1.

\[\text{Compute the square } t_2, \text{ and keep } t \text{ at the end with its sign. We know that } t < 0.1765, \text{ so every piece has at most 4 digits. However, since we were not careful in } \text{this part}, \text{ they can have less than 4 digits.}\]
\int_value:w \__fp_int_eval:w 9999 + 2*#5
\exp_after:wN \__fp_ln_twice_t_pack:Nw
\int_value:w \__fp_int_eval:w 9999 + 2*#6
\exp_after:wN \__fp_ln_twice_t_pack:Nw
\int_value:w \__fp_int_eval:w 9999 + 2*#7
\exp_after:wN \__fp_ln_twice_t_pack:Nw
\int_value:w \__fp_int_eval:w 10000 + 2*#8 ; ;
{ \__fp_ln_c:NwNw #1 }
#2
}
\cs_new:Npn \__fp_ln_twice_t_pack:Nw #1 #2; { + #1 ; {#2} }
\cs_new:Npn \__fp_ln_twice_t_after:w #1; { ;;; {#1} }
\cs_new:Npn \__fp_ln_square_t_pack:NNNNNw #1 #2#3#4#5 #6; { + #1#2#3#4#5 ; {#6} }
\cs_new:Npn \__fp_ln_square_t_after:w 1 0 #1#2#3 #4; { \__fp_ln_Taylor:wwNw {0#1#2#3} {#4} }
(End definition for \__fp_ln_x_ii:wnnnn.)
\__fp_ln_Taylor:wwNw Denoting $T = t^2$, we get
\__fp_ln_Taylor:wwNw
{{(T_1)}} {{(T_2)}} {{(T_3)}} {{(T_4)}} {{(T_5)}} ; ;
{{(2t_1)}} {{(2t_2)}} {{(2t_3)}} {{(2t_4)}} {{(2t_5)}} ; ;
\__fp_ln_c:NwNw (sign)
\{ fixed tl \} (exponent); (continuation)

And we want to compute

$$\ln \left( \frac{1+t}{1-t} \right) = 2t \left( 1 + T \left( \frac{1}{3} + T \left( \frac{1}{5} + T \left( \frac{1}{7} + T \left( \frac{1}{9} + \cdots \right) \right) \right) \right) \right)$$

The process looks as follows
\loop 5; A;
\div_int 5; 1.0; \add A; \mul T; {\loop \eval 5-2;}
\add 0.2; A; \mul T; {\loop \eval 5-2;}
\mul B; T; {\loop 3;}
\loop 3; C;

This uses the routine for dividing a number by a small integer ($< 10^4$).
We are now reduced to finding \( \ln(c) \) and \( \langle \text{exponent} \rangle \ln(10) \) in a table, and adding it to the mixture. The first step is to get \( \ln(c) - \ln(x) = -\ln(a) \), then we get \( b \ln(10) \) and add or subtract.

For now, \( \ln(x) \) is given as \( \cdot 10^0 \). Unless both the exponent is 1 and \( c = 1 \), we shift to working in units of \( \cdot 10^4 \), since the final result is at least \( \ln(10/7) \approx 0.35 \).

Compute \( \langle \text{exponent} \rangle \times \ln(10) \). Apart from the cases where \( \langle \text{exponent} \rangle \) is 0 or 1, the result is necessarily at least \( \ln(10) \approx 2.3 \) in magnitude. We can thus drop the least significant 4 digits. In the case of a very large (positive or negative) exponent, we can (and we need to) drop 4 additional digits, since the result is of order \( 10^4 \). Naively, one would think that in both cases we can drop 4 more digits than we do, but that would be slightly too tight for rounding to happen correctly. Besides, we already have addition and subtraction for 24 digits fixed point numbers.
Now we painfully write all the cases.\(^{12}\) No overflow nor underflow can happen, except when computing \(\ln(1)\).

For small exponents, we just drop one block of digits, and set the exponent of the log to 4 (minus any shift coming from leading zeros in the conversion from fixed point to floating point). Note that here the exponent has been made positive.

\(\text{End definition for } \_\_\text{fp\_ln\_exponent:wn.}\)

### 74.2 Exponential

#### 74.2.1 Sign, exponent, and special numbers

\(\text{End definition for } \_\_\text{fp\_exp:ow.}\)
\or:
\__fp_case_return_same_o:w
\fi:
\s__fp \__fp_chk:w #2#3#4;
}

(End definition for \__fp_exp_o:w)
\__fp_exp_normal_o:w
\__fp_exp_pos_o:Nnwnw
\__fp_exp_overflow:NN
\__fp_exp_normal_o:w \__fp_exp_pos_o:Nnwnw \__fp_chk:w #2#3#4;
{\if_meaning:w 0 #1
\__fp_exp_pos_o:NNwnw + \__fp_fixed_to_float_o:wN
\else:
\__fp_exp_pos_o:NNwnw - \__fp_fixed_inv_to_float_o:wN
\fi: }
\cs_new:Npn \__fp_exp_pos_o:NNwnw #1#2#3 \fi: #4#5;
{\fi: \if_int_compare:w #4 > \c__fp_max_exp_exponent_int
\token_if_eq_charcode:NNTF + #1
{ \__fp_exp_overflow:NN \__fp_overflow:o:w \c_zero_fp }
{ \__fp_exp_overflow:NN \__fp_underflow:o:w \c_zero_fp }
\exp:w
\else:
\exp_after:wN \__fp_sanitze:Nw
\exp_after:wN 0
\int_value:w #1 \__fp_int_eval:w
\if_int_compare:w #4 < \c_zero_int
\exp_after:wN \use_i:nn
\else:
\exp_after:wN \use_ii:nn
\fi: 
{ 0
\__fp_decimate:nNnnnn { - #4 }
\__fp_exp_Taylor:nNnNn
}
{ \__fp_decimate:nNnnnn { \c__fp_prec_int - #4 }
\__fp_exp_pos_large:NnnNn
}
#5
{#4}
#1 #2 0
\exp:w
\fi: 
\exp_after:wN \exp_end:
}
\cs_new:Npn \__fp_exp_overflow:NN #1#2#3 #4#5
{\exp:w
\exp_after:wN \exp_after:wN
\exp_after:wN #1
1121
This function is called for numbers in the range \([10^{-9}, 10^{-1})\). We compute 10 terms of the Taylor series. The first argument is irrelevant (rounding digit used by some other functions). The next three arguments, at least 16 digits, delimited by a semicolon, form a fixed point number, so we pack it in blocks of 4 digits.

\begin{verbatim}
\cs_new:Npn \__fp_exp_Taylor:Nnnwn #1#2#3 #4; #5 #6 
{ #6 \__fp_pack_twice_four:wNNNNNNNN \__fp_pack_twice_four:wNNNNNNNN \__fp_pack_twice_four:wNNNNNNNN \__fp_exp_Taylor_ii:ww ; #2#3#4 0000 0000 ;
}
\cs_new:Npn \__fp_exp_Taylor_ii:ww #1; #2; 
{ \__fp_exp_Taylor_loop:www 10 ; #1 ; #1 ; \s__fp_stop }
\cs_new:Npn \__fp_exp_Taylor_loop:www #1; #2; #3; 
{ \if_int_compare:w #1 = \c_one_int \exp_after:wN \__fp_exp_Taylor_break:Nww 
\fi: \__fp_fixed_div_int:ww \#3 ; #1 ; \__fp_fixed_add_one:wN 
\__fp_fixed_mul:wwn \#2 ;
\exp_after:wN \__fp_exp_Taylor_loop:www \int_value:w \__fp_int_eval:w #1 - 1 ; \#2 ; 
}
\cs_new:Npn \__fp_exp_Taylor_break:Nww #1 #2; #3 \s__fp_stop 
{ \__fp_fixed_add_one:wN \#2 ; }
\end{verbatim}

The integer array has \(6 \times 9 \times 4 = 216\) items encoding the values of \(\exp(j \times 10^i)\) for \(j = 1, \ldots, 9\) and \(i = -1, \ldots, 4\). Each value is expressed as \(\approx 10^p \times 0.m_1 m_2 m_3\) with three 8-digit blocks \(m_1, m_2, m_3\) and an integer exponent \(p\) (one more than the scientific exponent), and these are stored in the integer array as four items: \(p, 10^8 + m_1, 10^8 + m_2, 10^8 + m_3\). The various exponentials are stored in increasing order of \(j \times 10^i\).

Storing this data in an integer array makes it slightly harder to access (slower, too), but uses 16 bytes of memory per exponential stored, while storing as tokens used around 40 tokens; tokens have an especially large footprint in Unicode-aware engines.

\begin{verbatim}
\intarray_const_from_clist:Nn \c__fp_exp_intarray

\end{verbatim}
The first two arguments are irrelevant (a rounding digit, and a brace group with 8 zeros).
The third argument is the integer part of our number, then we have the decimal part delimited by a semicolon, and finally the exponent, in the range $[0, 5]$. Remove leading zeros from the integer part: putting \texttt{#4} in there too ensures that an integer part of 0 is also removed. Then read digits one by one, looking up \( \exp((\text{digit}) \cdot 10^{(\text{exponent})}) \) in a table, and multiplying that to the current total. The loop is done by \texttt{__fp_exp_large:NwN}, whose \texttt{#1} is the \( (\text{exponent}) \), \texttt{#2} is the current mantissa, and \texttt{#3} is the \( (\text{digit}) \). At the end, \texttt{__fp_exp_large_after:wwn} moves on to the Taylor series, eventually multiplied with the mantissa that we have just computed.

\begin{verbatim}
\cs_new:Npn \__fp_exp_pos_large:NnnNwn #1#2#3 #4#5; #6
\exp_after:wN \exp_after:wN \exp_after:wN \__fp_exp_large:NwN
\exp_after:wN \c__fp_one_fixed_tl \int_value:w #3 #4 \exp_stop_f:
\if_meaning:w 0 #1 \exp_after:wN \__fp_exp_large_after:wwn \else:
\exp_after:wN \__fp_exp_large:NwN \int_value:w \__fp_int_eval:w #1 - 1 \exp_after:wN \scan_stop:
\fi:
\if_case:w #2 ~
\exp_after:wN \__fp_fixed_continue:wn \else:
\exp_after:wN \__fp_exp_intarray:w \int_value:w \__fp_int_eval:w 36 * #1 + 4 * #3 \exp_after:wN ;
\fi:
\exp_after:wN \__fp_fixed_mul:wwn \int_value:w 0 \exp_after:wN \__fp_exp_intarray_aux:w
\int_value:w \__kernel_intarray_item:Nn \c__fp_exp_intarray { \__fp_int_eval:w #1 - 2 }
\exp_after:wN \use_i:nnn \__fp_int_eval:w #1 - 1 \exp_after:wN \scan_stop:
\if: \exp_after:wN \__fp_exp_intarray:wn #1 \; #2 \; #3
\exp_after:wN \__fp_exp_intarray_aux:w
\int_value:w \__kernel_intarray_item:Nn \c__fp_exp_intarray { \__fp_int_eval:w #1 - 3 } \; \scan_stop: \}
\exp_after:wN \__fp_exp_intarray:wn #1 \; \c__fp_one_fixed_tl \int_value:w #3 #4 \exp_stop_f:
\exp_after:wN \__fp_int_eval:w 36 * #1 + 4 * #3 \exp_after:wN ;
\__fp_fixed_continue:wn
\end{verbatim}
74.3 Power

Raising a number \( a \) to a power \( b \) leads to many distinct situations.

<table>
<thead>
<tr>
<th>( a^b )</th>
<th>(-\infty)</th>
<th>(-\infty, -0)</th>
<th>-integer</th>
<th>±0</th>
<th>+integer</th>
<th>((0,\infty))</th>
<th>(+\infty)</th>
<th>nan</th>
</tr>
</thead>
<tbody>
<tr>
<td>+∞</td>
<td>+0</td>
<td>+0</td>
<td>+1</td>
<td>+∞</td>
<td>+∞</td>
<td>nan</td>
<td>nan</td>
<td>nan</td>
</tr>
<tr>
<td>(1, ∞)</td>
<td>+0</td>
<td>(+</td>
<td>a</td>
<td>^b)</td>
<td>+1</td>
<td>+</td>
<td></td>
<td>+∞</td>
</tr>
<tr>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>+1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(0, 1)</td>
<td>+∞</td>
<td>(+</td>
<td>a</td>
<td>^b)</td>
<td>+1</td>
<td>+</td>
<td></td>
<td>+∞</td>
</tr>
<tr>
<td>+0</td>
<td>+∞</td>
<td>+∞</td>
<td>+1</td>
<td>+0</td>
<td>+0</td>
<td>nan</td>
<td>nan</td>
<td>nan</td>
</tr>
<tr>
<td>-0</td>
<td>+∞</td>
<td>nan</td>
<td>((-1)^b\infty)</td>
<td>+1</td>
<td>((-1)^b0)</td>
<td>+0</td>
<td>+0</td>
<td>nan</td>
</tr>
<tr>
<td>(-1, 0)</td>
<td>+∞</td>
<td>nan</td>
<td>((-1)^b</td>
<td>a</td>
<td>^b)</td>
<td>+1</td>
<td>((-1)^b</td>
<td>a</td>
</tr>
<tr>
<td>-1</td>
<td>+1</td>
<td>nan</td>
<td>((-1)^b)</td>
<td>+1</td>
<td>((-1)^b)</td>
<td>nan</td>
<td>+1</td>
<td>nan</td>
</tr>
<tr>
<td>(-∞, -1)</td>
<td>+0</td>
<td>nan</td>
<td>((-1)^b</td>
<td>a</td>
<td>^b)</td>
<td>+1</td>
<td>((-1)^b</td>
<td>a</td>
</tr>
<tr>
<td>-∞</td>
<td>+0</td>
<td>+0</td>
<td>((-1)^b0)</td>
<td>+1</td>
<td>((-1)^b\infty)</td>
<td>nan</td>
<td>+∞</td>
<td>nan</td>
</tr>
</tbody>
</table>

We distinguished in this table the cases of finite (positive or negative) integer exponents, as \((-1)^b\) is defined in that case. One peculiarity of this operation is that \( nan^0 = 1 \), because this relation is obeyed for any number, even \( ±\infty \).

\( _\text{fp}^\wedge_o:ww \)

We cram most of the tests into a single function to save csnames. First treat the case \( b = 0: a^0 = 1 \) for any \( a \), even \( nan \). Then test the sign of \( a \).

- If it is positive, and \( a \) is a normal number, call \( _\text{fp}^\text{pow_normal_o:ww} \) followed by the two \( fp \) \( a \) and \( b \). For \( a = +0 \) or \( +\inf \), call \( _\text{fp}^\text{pow_zero_or_inf:ww} \) instead, to return either \( +0 \) or \( +\infty \) as appropriate.

- If \( a \) is a \( nan \), then skip to the next semicolon (which happens to be conveniently the end of \( b \)) and return \( nan \).

- Finally, if \( a \) is negative, compute \( a^b \) \( (_\text{fp}^\text{pow_normal_o:ww} \) which ignores the sign of its first operand), and keep an extra copy of \( a \) and \( b \) (the second brace group, containing \{ \( b \ a \} \), is inserted between \( a \) and \( b \)). Then do some tests to find the final sign of the result if it exists.
\ else:
\exp_after:wN \__fp_pow_neg:ww
\exp:w \exp_end_continue_f:w \exp_after:wN \use:nn
\fi:
\{ 
\if_meaning:w 1 \#1
\exp_after:wN \__fp_pow_normal_o:ww
\else:
\exp_after:wN \__fp_pow_zero_or_inf:ww
\fi:
\s__fp \__fp_chk:w \#1\#2\#3;
\{ \s__fp \__fp_chk:w \#4\#5\#6; \s__fp \__fp_chk:w \#1\#2\#3; \}
\s__fp \__fp_chk:w \#4\#5\#6;
\}
(End definition for \__fp_^_o:ww.)

\__fp_pow_zero_or_inf:ww
Raising \(-0\) or \(-\infty\) to \textit{nan} yields \textit{nan}. For other powers, the result is \(+0\) if \(0\) is raised to a positive power or \(+\infty\) otherwise. Thus, if the type of \(a\) and the sign of \(b\) coincide, the result is \(0\), since those conveniently take the same possible values, \(0\) and \(2\). Otherwise, either \(a = \pm \infty\) and \(b > 0\) and the result is \(+\infty\), or \(a = \pm 0\) with \(b < 0\) and we have a division by zero unless \(b = -\infty\).

\cs_new:Npn \__fp_pow_zero_or_inf:ww
\s__fp \__fp_chk:w \#1\#2; \s__fp \__fp_chk:w \#3\#4
\{ 
\if_meaning:w 1 \#4
\__fp_case_return_same_o:Nw \c_zero_fp
\fi:
\if_meaning:w \#1 \#4
\__fp_case_return_o:Nw \c_zero_fp
\fi:
\if_meaning:w 2 \#1
\__fp_case_return_o:Nw \c_inf_fp
\fi:
\if_meaning:w 2 \#3
\__fp_case_return_o:Nw \c_inf_fp
\else:
\__fp_case_use:nw
\{ 
\__fp_division_by_zero_o:NNww \c_inf_fp \wedge
\s__fp \__fp_chk:w \#1 \#2 ;
\}
\fi:
\s__fp \__fp_chk:w \#3\#4
\}
(End definition for \__fp_^_o:ww.)

\__fp_pow_normal_o:ww
We have in front of us \(a\), and \(b \neq 0\), we know that \(a\) is a normal number, and we wish to compute \(|a|^b\). If \(|a| = 1\), we return \(1\), unless \(a = -1\) and \(b\) is \textit{nan}. Indeed, returning \(1\) at this point would wrongly raise “invalid” when the sign is considered. If \(|a| \neq 1\), test the type of \(b\):

0 Impossible, we already filtered \(b = \pm 0\).
1. Call \texttt{\_fp\_pow\_npos\_o:Nww}.

2. Return $+\infty$ or $+0$ depending on the sign of $b$ and whether the exponent of $a$ is positive or not.

3. Return $b$.

\begin{verbatim}
\cs_new:Npn \__fp_pow_normal_o:ww
\__fp_chk:w 1 \#1\#2\#3; \s__fp \__fp_chk:w #4#5
{
  \if:w 0 \__fp_str_if_eq:nn { #2 #3 } { 1 {1000} {0000} {0000} } {1}
    \exp_after:wN \__fp_case_return_ii:o:ww
  \fi:
  \__fp_case_return_o:Nww \c_one_fp
  \fi:
  \if_case:w #4 \exp_stop_f:
   \or:
  \exp_after:wN \__fp_pow_npos_o:Nww
\exp_after:wN \c_zero_fp
  \fi:
  \s__fp \__fp_chk:w 1 #1 {#2} #3 ;
}\end{definition}

\__fp\_pow\_npos\_o:Nww
\end{verbatim}

We now know that $a \neq \pm1$ is a normal number, and $b$ is a normal number too. We want to compute $|a|^b = (|x| \cdot 10^n)^y \cdot 10^p = \exp((\ln|x| + n \ln(10)) \cdot y \cdot 10^p) = \exp(z)$. To compute the exponential accurately, we need to know the digits of $z$ up to the 16-th position. Since the exponential of $10^5$ is infinite, we only need at most 21 digits, hence the fixed point result of \texttt{\_fp\_ln\_o:w} is precise enough for our needs. Start an integer expression for the decimal exponent of $e^z$. If $z$ is negative, negate that decimal exponent, and prepare to take the inverse when converting from the fixed point to the floating point result.
\else:
\exp_after:wN \__fp_pow_npos_aux:NNnww
\exp_after:wN -
\exp_after:wN \__fp_fixed_inv_to_float_o:wN
\fi:
{#3}
}

(End definition for \__fp_pow_npos_o:Nww.)

The first argument is the conversion function from fixed point to float. Then comes an
exponent and the 4 brace groups of \(x\), followed by \(b\). Compute \(-\ln(x)\).

\cs_new:Npn \__fp_pow_npos_aux:NNnww \#1\#2\#3\#4\#5; \s__fp \__fp_chk:w 1\#6\#7\#8;
{
#1
\__fp_int_eval:w
\__fp_ln_significand:NNNnNn \#4\#5
\__fp_pow_exponent:wnN {#3}
\__fp_fixed_mul:wn \#6 {0000}{0000};
\__fp_pow_B:wwN \#7;
\#1 \#2 0 \% fixed_to_float_o:wN
}
\cs_new:Npn \__fp_pow_exponent:wnN \#1; \#2
{
\if_int_compare:w \#2 > \c_zero_int
\exp_after:wN \__fp_pow_exponent:Nwnnnnnw % n\ln(10) - (-\ln(x))
\exp_after:wN +
\else:
\exp_after:wN \__fp_pow_exponent:Nwnnnnnw % -(|n|\ln(10) + (-\ln(x)))
\exp_after:wN -
\fi:
{#2; \#1;
}
}\cs_new:Npn \__fp_pow_exponent:Nwnnnnnw \#1\#2; \#3\#4\#5\#6\#7\#8;
{ %"A todo: use that in ln.
\exp_after:wN \__fp_fixed_mul_after:wN
\int_value:w \__fp_int_eval:w \c__fp_leading_shift_int
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
\#1\#2*23025 - \#1 \#3
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
\#1 \#2*8401 - \#1 \#4
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
\#1 \#2*8509 - \#1 \#5
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
\#1 \#2*0456 - \#1 \#6
\exp_after:wN \__fp_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_trailing_shift_int
\#1 \#2*7991 - \#8 / 1 0000 ; ;
}
This function is followed by three floating point numbers: \( a^b \), \( a \in [\infty, -0] \), and \( b \). If \( b \) is an even integer (case \(-1\)), \( a^b = a^b \). If \( b \) is an odd integer (case \(0\)), \( a^b = -a^b \), obtained by a call to \( \__fp_pow_neg:www \). Otherwise, the sign is undefined. This is invalid, unless \( a^b \) turns out to be +0 or \text{nан}, in which case we return that as \( a^b \). In particular, since the
underflow detection occurs before \_\_fp\_pow\_neg:www is called, \((-0.1)**(12345.67)\) gives +0 rather than complaining that the sign is not defined.

\cs_new:Npn \_\_fp\_pow\_neg:www \s\_fp \_\_fp\_chk:w #1; #3; #4;
{\if_case:w \_\_fp\_pow\_neg\_case:w #4 ; \exp_after:wN \_\_fp\_pow\_neg\_aux:wNN
\or: \_\_fp\_int\_compare:w \_\_fp\_int\_eval:w #1 / 2 = \c\_one\_int
\_\_fp\_invalid\_operation\_o:NNw ^ #3; #4; \exp:w \exp\_end\_continue\_f:w
\exp\_after:wN \exp\_after:wN \_\_fp\_use\_none\_until\_s:w
\fi:
\_\_fp\_exp\_after\_o:w \s\_fp \_\_fp\_chk:w #1#2;
}
\cs_new:Npn \_\_fp\_pow\_neg\_case:w \_\_fp\_chk:w #1#2#3;
{\if_case:w #1 \exp_stop_f:
\_\_fp\_pow\_neg\_case\_aux:nnnnn \_\_fp\_pow\_neg\_case\_aux:Nnnw
\or: \_\_fp\_int\_compare\_aux:nnnnn \_\_fp\_pow\_neg\_case\_aux:Nnnw
\or: \_\_fp\_int\_eval\_o:w \_\_fp\_int\_eval:w #2 - #3 \_\_fp\_int\_eval\_end:
\fi:
\_\_fp\_exp\_after\_o:w \s\_fp \_\_fp\_chk:w #1#2#3;
}
(End definition for \_\_fp\_pow\_neg:www and \_\_fp\_pow\_neg\_aux:NNw.)

This function expects a floating point number, and determines its “parity”. It should be used after \_\_fp\_pow\_neg:www or in an integer expression. It gives \(-1\) if the number is an even integer, 0 if the number is an odd integer, and 1 otherwise. Zeros and \pm\infty are even (because very large finite floating points are even), while nan is a non-integer. The sign of normal numbers is irrelevant to parity. After \_\_fp\_decimate:nNNnnn the argument \#1 of \_\_fp\_pow\_neg\_case\_aux:Nnnw is a rounding digit, 0 if and only if the number was an integer, and \#3 is the 8 least significant digits of that integer.
The maximum integer whose factorial fits in the exponent range is 3248, as \(3249! \approx 10^{10000.8}\). 

First detect \(\pm 0\) and \(+\infty\) and \(\text{nan}\). Then note that factorial of anything with a negative sign (except \(-0\)) is undefined. Then call \(\texttt{__fp_small_int:wTF}\) to get an integer as the argument, and start a loop. This is not the most efficient way of computing the factorial, but it works all right. Of course we work with 24 digits instead of 16. It is easy to check that computing factorials with this precision is enough.
Then check the input is an integer, and call \_\_fp_facorial_int\_o:n with that int as an argument. If it’s too big the factorial overflows. Otherwise call \_\_fp_sanitize:Nw with a positive sign marker 0 and an integer expression that will mop up any exponent in the calculation.

\begin{verbatim}
\cs_new:Npn \_\_fp_fact_pos_o:w #1; 
{ \_\_fp_small_int:wTF #1; 
  \{ \_\_fp_fact_int_o:n \} 
  \{ \_\_fp_invalid_operation_o:fw \{ fact \} #1; \} 
} 
\cs_new:Npn \_\_fp_fact_int_o:n #1 
{ \if_int_compare:w #1 > \c__fp_fact_max_arg_int 
    \_\_fp_case_return:nw 
    \{ \exp_after:wN \exp_after:wN \exp_after:wN \_\_fp_overflow:w 
      \exp_after:wN \exp_after:wN \_\_fp_invalid_operation_o:fw \{ fact \} #1; \} 
  \fi: 
  \exp_after:wN \_\_fp_sanitize:Nw 
  \exp_after:wN \_\_fp_case_return:nw 
  \{ \exp_after:wN \exp_after:wN \_\_fp_overflow:w 
    \_\_fp_case_return:nw 
  \} 
} 
\cs_new:Npn \_\_fp_fact_loop_o:w #1 . #2 ; 
{ \if_int_compare:w #1 < 12 \exp_stop_f: 
    \_\_fp_fact_small_o:w #1 
  \fi: 
    \exp_after:wN \_\_fp_ep_mul:wwwwn 
    \exp_after:wN \_\_fp_case_return:nw 
    \int_value:w \_\_fp_int_eval:w \_\_fp_fact_loop_o:w #1 . 4 , \{ 1 \} \} \} \} \} \} \} 
\end{verbatim}

The loop receives an integer \#1 whose factorial we want to compute, which we progressively decrement, and the result so far as an extended-precision number \#2 in the form \langle exponent\rangle, \langle mantissa\rangle. The loop goes in steps of two because we compute \#1\#1-1 as an integer expression (it must fit since \#1 is at most 3248), then multiply with the result so far. We don’t need to fill in most of the mantissa with zeros because \_\_fp_ep_mul:wwwwn first normalizes the extended precision number to avoid loss of precision. When reaching a small enough number simply use a table of factorials less than 10^8. This limit is chosen because the normalization step cannot deal with larger integers.

\begin{verbatim}
\cs_new:Npn \_\_fp_fact_loop_o:w #1 . #2 ; 
{ \_\_fp_small_int:wTF \#1; 
  \{ \_\_fp_fact_int_o:n \} 
  \{ \_\_fp_invalid_operation_o:fw \{ fact \} #1; \} 
} 
\cs_new:Npn \_\_fp_fact_small_o:w #1 \fi: #2 ; #3 ; #4 
{ \_\_fp_fact_small_int:wTF \#1; 
  \{ \_\_fp_invalid_operation_o:fw \{ fact \} #1; \} 
} 
\end{verbatim}
\exp_after:wN \__fp_fact_loop_o:w
\exp_after:wN \__fp_fact_loop_o:w
\exp_after:wN \__fp_fact_loop_o:w
\exp_after:wN \__fp_fact_loop_o:w
{ \int_value:w
  \if_case:w #1 \exp_stop_f:
    1 \or: 1 \or: 2 \or: 6 \or: 24 \or: 120 \or: 720 \or: 5040 \or: 40320 \or: 362880 \or: 3628800 \or: 39916800
  \fi:
  } \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} \{ \} 

(End definition for \__fp_fact_loop_o:w.)

//package
Chapter 75

13fp-trig Implementation

Unary functions.

\__fp_parse_word_acos:N
\__fp_parse_word_acosd:N
\__fp_parse_word_acsc:N
\__fp_parse_word_acscd:N
\__fp_parse_word_asec:N
\__fp_parse_word_asecd:N
\__fp_parse_word_asin:N
\__fp_parse_word_asind:N
\__fp_parse_word_cos:N
\__fp_parse_word_cosd:N
\__fp_parse_word_cot:N
\__fp_parse_word_cotd:N
\__fp_parse_word_csc:N
\__fp_parse_word_cscd:N
\__fp_parse_word_sec:N
\__fp_parse_word_secd:N
\__fp_parse_word_sin:N
\__fp_parse_word_sind:N
\__fp_parse_word_tan:N
\__fp_parse_word_tand:N

Those functions may receive a variable number of arguments.

\cs_new:Npn \__fp_parse_word_acot:N
\cs_new:Npn \__fp_parse_word_atan:N
\cs_new:Npn \__fp_parse_word_atand:N

(End definition for \__fp_parse_word_acos:N and others.)
75.1 Direct trigonometric functions

The approach for all trigonometric functions (sine, cosine, tangent, cotangent, cosecant, and secant), with arguments given in radians or in degrees, is the same.

- Filter out special cases (±0, ±inf and nan).
- Keep the sign for later, and work with the absolute value |x| of the argument.
- Small numbers (|x| < 1 in radians, |x| < 10 in degrees) are converted to fixed point numbers (and to radians if |x| is in degrees).
- For larger numbers, we need argument reduction. Subtract a multiple of π/2 (in degrees, 90) to bring the number to the range to [0, π/2) (in degrees, [0, 90)).
- Reduce further to [0, π/4] (in degrees, [0, 45]) using \( \sin x = \cos(\pi/2 - x) \), and when working in degrees, convert to radians.
- Use the appropriate power series depending on the octant \( \lfloor \frac{x}{\pi/4} \rfloor \mod 8 \) (in degrees, the same formula with \( \pi/4 \to 45 \)), the sign, and the function to compute.

75.1.1 Filtering special cases

This function, and its analogs for cos, csc, sec, tan, and cot instead of sin, are followed either by \use_i:nn and a float in radians or by \use_ii:nn and a float in degrees. The sine of ±0 or nan is the same float. The sine of ±∞ raises an invalid operation exception with the appropriate function name. Otherwise, call the \trig function to perform argument reduction and if necessary convert the reduced argument to radians. Then, \_fp_sin_series_o:NNwwww is called to compute the Taylor series: this function receives a sign #3, an initial octant of 0, and the function \_fp_ep_to_float_o:wwN which converts the result of the series to a floating point directly rather than taking its inverse, since \( \sin(x) = #3 \sin|x| \).

\[ \_\_fp\_sin\_o:w \]

This function, and its analogs for cos, csc, sec, tan, and cot instead of sin, are followed either by \use_i:nn and a float in radians or by \use_ii:nn and a float in degrees. The sine of ±0 or nan is the same float. The sine of ±∞ raises an invalid operation exception. The cosine of nan is itself. Otherwise, the \trig function reduces the argument to at most half a right-angle and converts if necessary to radians. We then call the same series as
for sine, but using a positive sign 0 regardless of the sign of $x$, and with an initial octant of 2, because $\cos(x) = \sin(\pi/2 + |x|)$.

The cosecant of $\pm 0$ is $\pm \infty$ with the same sign, with a division by zero exception (see \_fp_cot_zero_o:Nfw defined below), which requires the function name. The cosecant of $\pm \infty$ raises an invalid operation exception. The cosecant of nan is itself. Otherwise, the trig function performs the argument reduction, and converts if necessary to radians before calling the same series as for sine, using the sign #3, a starting octant of 0, and inverting during the conversion from the fixed point sine to the floating point result, because $\csc(x) = \pm 1 / \sin(\pi/2 + |x|)$.

The secant of $\pm 0$ is 1. The secant of $\pm \infty$ raises an invalid operation exception. The secant of nan is itself. Otherwise, the trig function reduces the argument and turns it to radians before calling the same series as for sine, using a positive sign 0, a starting octant of 2, and inverting upon conversion, because $\sec(x) = 1 / \sin(\pi/2 + |x|)$. 

(End definition for \_fp_cos_o:w.)

(End definition for \_fp_csc_o:w.)

(End definition for \_fp_sec_o:w.)
\_fp\_tan\_o:w

The tangent of ±0 or \texttt{nan} is the same floating point number. The tangent of ±∞ raises an invalid operation exception. Once more, the \texttt{trig} function does the argument reduction step and conversion to radians before calling \_fp\_tan\_series\_o:NNwww, with a sign \#3 and an initial octant of 1 (this shift is somewhat arbitrary). See \_fp\_\-cot\_o:w for an explanation of the 0 argument.

\_fp\_cot\_o:w

\_fp\_cot\_zero\_o:Nfw

The cotangent of ±0 is ±∞ with the same sign, with a division by zero exception (see \_fp\_cot\_zero\_o:Nfw). The cotangent of ±∞ raises an invalid operation exception. The cotangent of \texttt{nan} is itself. We use \( \cot x = -\tan(\pi/2 + x) \), and the initial octant for the tangent was chosen to be 1, so the octant here starts at 3. The change in sign is obtained by feeding \_fp\_tan\_series\_o:NNwww two signs rather than just the sign of the argument: the first of those indicates whether we compute tangent or cotangent. Those signs are eventually combined.
75.1.2 Distinguishing small and large arguments

The first argument is \use_i:nn if the operand is in radians and \use_ii:nn if it is in degrees. Arguments #2 to #5 control what trigonometric function we compute, and #6 to #8 are pieces of a normal floating point number. Call the _series function #2, with arguments #3, either a conversion function (\_fp_to_float:o:w or \_fp_inv_to_float:o:w) or a sign 0 or 2 when computing tangent or cotangent; #4, a sign 0 or 2; the octant, computed in an integer expression starting with #5 and stopped by a period; and a fixed point number obtained from the floating point number by argument reduction (if necessary) and conversion to radians (if necessary). Any argument reduction adjusts the octant accordingly by leaving a (positive) shift into its integer expression. Let us explain the integer comparison. Two of the four \exp_after:wN are expanded, the expansion hits the test, which is true if the float is at least 1 when working in radians, and at least 10 when working in degrees. Then one of the remaining \exp_after:wN hits #1, which picks the trig or trigd function in whichever branch of the conditional was taken. The final \exp_after:wN closes the conditional. At the end of the day, a number is large if it is ≥ 1 in radians or ≥ 10 in degrees, and small otherwise. All four trig/trigd auxiliaries receive the operand as an extended-precision number.
75.1.3 Small arguments

\_\_fp\_trig\_small:ww

This receives a small extended-precision number in radians and converts it to a fixed
point number. Some trailing digits may be lost in the conversion, so we keep the original
floating point number around: when computing sine or tangent (or their inverses), the
last step is to multiply by the floating point number (as an extended-precision number)
rather than the fixed point number. The period serves to end the integer expression for
the octant.

\cs_new:Npn \__fp_trig_small:ww #1,#2;
\{ \__fp_ep_to_fixed:wwn #1,#2; . #1,#2; \}

(End definition for \_\_fp\_trig\_small:ww.)

\_\_fp\_trigd\_small:ww

Convert the extended-precision number to radians, then call \_\_fp\_trig\_small:ww to
massage it in the form appropriate for the _series auxiliary.

\cs_new:Npn \__fp_trigd_small:ww #1,#2;
\{ \__fp_ep_mul_raw:wwwwN -1,{1745}{3292}{5199}{4329}{5769}{2369}; #1,#2; \__fp_trig_small:ww \}

(End definition for \_\_fp\_trig\_dsmall:ww.)

75.1.4 Argument reduction in degrees

\_\_fp\_trig\_d\_large:ww

\_\_fp\_trig\_d\_large\_auxi:nnnnwNNNN
\_\_fp\_trig\_d\_large\_auxii:wNw
\_\_fp\_trig\_d\_large\_auxiii:www

Note that $25 \times 360 = 9000$, so $10^{k+1} \equiv 10^k$ (mod 360) for $k \geq 3$. When the exponent #1
is very large, we can thus safely replace it by 22 (or even 19). We turn the floating point
number into a fixed point number with two blocks of 8 digits followed by five blocks of
4 digits. The original float is $100 \times \langle block_1 \rangle \cdots \langle block_3 \rangle . \langle block_4 \rangle \cdots \langle block_7 \rangle$, or is equal to
it modulo 360 if the exponent #1 is very large. The first auxiliary finds $\langle block_1 \rangle + \langle block_2 \rangle$
(mod 9), a single digit, and prepends it to the 4 digits of $\langle block_3 \rangle$. It also unpacks
$\langle block_4 \rangle$ and grabs the 4 digits of $\langle block_7 \rangle$. The second auxiliary grabs the $\langle block_3 \rangle$ plus
any contribution from the first two blocks as #1, the first digit of $\langle block_4 \rangle$ (just after the
decimal point in hundreds of degrees) as #2, and the third other digits as #3. It finds the
quotient and remainder of #1#2 modulo 9, adds twice the quotient to the integer
expression for the octant, and places the remainder (between 0 and 8) before #3 to form
a new $\langle block_4 \rangle$. The resulting fixed point number is $x \in [0,0.9]$. If $x \geq 0.45$, we add 1 to
the octant and feed $0.9 - x$ with an exponent of 2 (to compensate the fact that we are
working in units of hundreds of degrees rather than degrees) to \_\_fp\_trig\_small:ww.
Otherwise, we feed it $x$ with an exponent of 2. The third auxiliary also discards digits
which were not packed into the various $\langle blocks \rangle$. Since the original exponent #1 is at
least 2, those are all 0 and no precision is lost (#6 and #7 are four 0 each).

```
\cs_new:Npn \__fp_trigd_large:ww #1, #2#3#4#5#6#7;
\{ \exp_after:wN \__fp_pack_eight:wNNNNNNNN \exp_after:wN \__fp_pack_eight:wNNNNNNNN \exp_after:wN \__fp_pack_twice_four:wNNNNNNNN \exp_after:wN \__fp_pack_twice_four:wNNNNNNNN \exp_after:wN \__fp_trigd_large_auxi:nnnnwNNNN \exp_after:wN \__fp_trigd_large_auxii:wNw \exp_after:wN \__fp_trigd_large_auxiii:www \exp:w \exp_end_continue_f:w
```

1139
\begin{verbatim}
\prg_replicate:nn { \int_max:nn { 22 - #1 } { 0 } } { 0 }
#2\#3\#4\#5\#6\#7 0000 0000 0000
\}
cs_new:Npn \__fp_trigd_large_auxi:nnnnwNNNN #1#2#3#4#5; #6#7#8#9
{\exp_after:wN \__fp_trigd_large_auxii:wNw
\int_value:w \__fp_int_eval:w #1 + #2
- (#1 + #2 - 4) / 9 * 9 \__fp_int_eval_end:
#3;
#4; #5{#6#7#8#9};}
cs_new:Npn \__fp_trigd_large_auxii:wNw #1; #2#3;
{+(#1#2 - 4) / 9 * 2
\exp_after:wN \__fp_trigd_large_auxiii:ww
\int_value:w \__fp_int_eval:w #1#2
- (#1#2 - 4) / 9 * 9 \__fp_int_eval_end: #3 ;}
cs_new:Npn \__fp_trigd_large_auxiii:www #1; #2; #3!
{\if_int_compare:w #1 < 4500 \exp_stop_f:
\exp_after:wN \__fp_use_i_until_s:nw
\exp_after:wN \__fp_fixed_continue:wn
\else:
  + 1
\fi:
\__fp_fixed_sub:wwn {9000}{0000}{0000}{0000}{0000}{0000};
\#1\#2(0000){0000};
{ \__fp_trigd_small:ww 2, }
}
\end{verbatim}

(End definition for \_\_fp_trigd_large:ww and others.)

75.1.5 Argument reduction in radians

Arguments greater or equal to 1 need to be reduced to a range where we only need a few terms of the Taylor series. We reduce to the range \([0, 2\pi]\) by subtracting multiples of \(2\pi\), then to the smaller range \([0, \pi/2]\) by subtracting multiples of \(\pi/2\) (keeping track of how many times \(\pi/2\) is subtracted), then to \([0, \pi/4]\) by mapping \(x \rightarrow \pi/2 - x\) if appropriate. When the argument is very large, say, \(10^{100}\), an equally large multiple of \(2\pi\) must be subtracted, hence we must work with a very good approximation of \(2\pi\) in order to get a sensible remainder modulo \(2\pi\).

Specifically, we multiply the argument by an approximation of \(1/(2\pi)\) with 10048 digits, then discard the integer part of the result, keeping 52 digits of the fractional part. From the fractional part of \(x/(2\pi)\) we deduce the octant (quotient of the first three digits by 125). We then multiply by 8 or \(-8\) (the latter when the octant is odd), ignore any integer part (related to the octant), and convert the fractional part to an extended precision number, before multiplying by \(\pi/4\) to convert back to a value in radians in \([0, \pi/4]\).

It is possible to prove that given the precision of floating points and their range of exponents, the 52 digits may start at most with 24 zeros. The 5 last digits are affected by carries from computations which are not done, hence we are left with at least 52 - 24 - 5 = 23 significant digits, enough to round correctly up to 0.6 \cdot ulp in all cases.

1140
This integer array stores blocks of 8 decimals of $10^{-16}/(2\pi)$. Each entry is $10^8$ plus an 8 digit number storing 8 decimals. In total we store 10112 decimals of $10^{-16}/(2\pi)$. The number of decimals we really need is the maximum exponent plus the number of digits we later need, 52, plus 12 ($4 - 1$ groups of 4 digits). The memory footprint (1/2 byte per digit) is the same as an earlier method of storing the data as a control sequence name, but the major advantage is that we can unpack specific subsets of the digits without unpacking the 10112 decimals.
The exponent #1 is between 1 and 10000. We wish to look up decimals \(10^{#1-16}/(2\pi)\) starting from the digit \(\#1+1\). Since they are stored in batches of 8, compute \([\#1/8]\) and fetch blocks of 8 digits starting there. The numbering of items in \c__fp_trig_intarray\ starts at 1, so the block \([\#1/8]+1\) contains the digit we want, at one of the eight positions. Each call to \int_value:w \__kernel_intarray_item:NNn expands the next, until being stopped by \__fp_trig_large_auxiii:w\ using \exp_stop_f:. Once all these blocks are unpacked, the \exp_stop_f: and 0 to 7 digits are removed by \use_\ldots:nn. Finally, \__fp_trig_large_auxii:w\ packs 64 digits (there are between 65 and 72 at this point) into groups of 4 and the auxv auxiliary is called.

\begin{verbatim}
\cs_new:Npn \__fp_trig_large:w \#1, \#2\#3\#4\#5\#6;
\{
\exp_after:wN \__fp_trig_large_auxi:w
\int_value:w \__fp_int_eval:w (#1 - 4) / 8 \exp_after:wN ,
\int_value:w \#1 , ;
\{#2\#3\#4\#5\};
\}
\cs_new:Npn \__fp_trig_large_auxi:w \#1, \#2,
\{
\exp_after:wN \exp_after:wN
\exp_after:wN \__fp_trig_large_auxii:w
\cs:w
use_none:n \prg_replicate:nn { #2 - \#1 * 8 } { n }
\exp_after:wN \exp_after:wN
\cs_end:
\int_value:w \__kernel_intarray_item:NNn \c__fp_trig_intarray
{ \__fp_int_eval:w \#1 + 1 \scan_stop: }
\exp_after:wN \__fp_trig_large_auxii:w \int_value:w
\__kernel_intarray_item:NNn \c__fp_trig_intarray
{ \__fp_int_eval:w \#1 + 2 \scan_stop: }
\exp_after:wN \__fp_trig_large_auxii:w \int_value:w
\__kernel_intarray_item:NNn \c__fp_trig_intarray
{ \__fp_int_eval:w \#1 + 3 \scan_stop: }
\exp_after:wN \__fp_trig_large_auxii:w \int_value:w
\__kernel_intarray_item:NNn \c__fp_trig_intarray
{ \__fp_int_eval:w \#1 + 4 \scan_stop: }
\exp_after:wN \__fp_trig_large_auxii:w \int_value:w
\__kernel_intarray_item:NNn \c__fp_trig_intarray
{ \__fp_int_eval:w \#1 + 5 \scan_stop: }
\exp_after:wN \__fp_trig_large_auxii:w \int_value:w
\__kernel_intarray_item:NNn \c__fp_trig_intarray
{ \__fp_int_eval:w \#1 + 6 \scan_stop: }
\exp_after:wN \__fp_trig_large_auxii:w \int_value:w
\__kernel_intarray_item:NNn \c__fp_trig_intarray
{ \__fp_int_eval:w \#1 + 7 \scan_stop: }
\end{verbatim}
First come the first 64 digits of the fractional part of $10^{#1 - 16}/(2\pi)$, arranged in 16 blocks of 4, and ending with a semicolon. Then a few more digits of the same fractional part, ending with a semicolon, then 4 blocks of 4 digits holding the significand of the original argument. Multiply the 16-digit significand with the 64-digit fractional part: the auxvi auxiliary receives the significand as #2#3#4#5 and 16 digits of the fractional part as #6#7#8#9, and computes one step of the usual ladder of pack functions we use for multiplication (see e.g., \_fp_fixed_mul:wwn), then discards one block of the fractional part to set things up for the next step of the ladder. We perform 13 such steps, replacing the last middle shift by the appropriate trailing shift, then discard the significand and remaining 3 blocks from the fractional part, as there are not enough digits to compute any more step in the ladder. The last semicolon closes the ladder, and we return control to the auxvii auxiliary.

\"fp_trig_large_auxv:www \fppack_large_auxvi:wnnnnnn \fppack_large_pack:NNNNWw

(End definition for \_fp_trig_large:ww and others.)
The auxiliary is followed by 52 digits and a semicolon. We find the octant as the integer part of 8 times what follows, or equivalently as the integer part of \( \frac{\#1\#2\#3}{125} \), and add it to the surrounding integer expression for the octant. We then compute 8 times the 52-digit number, with a minus sign if the octant is odd. Again, the last \textit{middle} shift is converted to a \textit{trailing} shift. Any integer part (including negative values which come up when the octant is odd) is discarded by \texttt{\_fp_use_i_until_s:nw}. The resulting fractional part should then be converted to radians by multiplying by \( 2\pi/8 \), but first, build an extended precision number by abusing \texttt{\_fp_ep_to_ep_loop:N} with the appropriate trailing markers. Finally, \texttt{\_fp_trig_small:ww} sets up the argument for the functions which compute the Taylor series.

\begin{verbatim}
\cs_new:Npn \__fp_trig_large_auxvii:w \#1\#2\#3
\exp_after:wN \__fp_trig_large_auxviii:ww
\int_value:w \__fp_int_eval:w \#1\#2\#3 - 62 / 125 ; \#1\#2\#3

\cs_new:Npn \__fp_trig_large_auxviii:ww \#1;
\+ \#1
\if_int_odd:w \#1 \exp_stop_f:
\exp_after:wN \__fp_trig_large_auxix:Nw
\exp_after:wN -
\else:
\exp_after:wN \__fp_trig_large_auxix:Nw
\exp_after:wN +
\fi:
\fi:
\cs_new:Npn \__fp_trig_large_auxix:Nw
\exp_after:wN \__fp_use_i_until_s:nw
\exp_after:wN \__fp_trig_large_auxxi:w
\int_value:w \__fp_int_eval:w \c__fp_leading_shift_int
\prg_replicate:nn { 13 }
\{ \__fp_trig_large_auxx:wNNNNN \} + \c__fp_trailing_shift_int - \c__fp_middle_shift_int
;
\cs_new:Npn \__fp_trig_large_auxx:wNNNNN \#1; \#2 \#3\#4\#5\#6
\exp_after:wN \__fp_trig_large_pack:NNNNNw
\int_value:w \__fp_int_eval:w \c__fp_middle_shift_int
\#2 \#3\#4\#5\#6 \#1; \#2
\cs_new:Npn \__fp_trig_large_auxxi:w \#1;
\{ \exp_after:wN \__fp_ep_mul_raw:wwwwN
\int_value:w \__fp_int_eval:w 0 \c__fp_trig_large_pack:NNNNNw \#1 ; ; !
0, \{7853\}\{9816\}\{3397\}\{4483\}\{0961\}\{5661\};
\__fp_trig_small:ww
\end{verbatim}
75.1.6 Computing the power series

Here we receive a conversion function \_fp_ep_to_float_o:wwN or \_fp_ep_inv_to_float_o:wwN, a \langle sign \rangle (0 or 2), a (non-negative) \langle octant \rangle delimited by a dot, a (fixed point) number delimited by a semicolon, and an extended-precision number. The auxiliary receives:

- the conversion function \#1;
- the final sign, which depends on the octant \#3 and the sign \#2;
- the octant \#3, which controls the series we use;
- the square \#4 * \#4 of the argument as a fixed point number, computed with \_fp_fixed_mul:wwn;
- the number itself as an extended-precision number.

If the octant is in \{1, 2, 5, 6,...\}, we are near an extremum of the function and we use the series

$$\cos(x) = 1 - x^2 \left( \frac{1}{2!} - x^2 \left( \frac{1}{4!} - x^2 \left( \cdots \right) \right) \right).$$

Otherwise, the series

$$\sin(x) = x \left( 1 - x^2 \left( \frac{1}{3!} - x^2 \left( \frac{1}{5!} - x^2 \left( \cdots \right) \right) \right) \right)$$

is used. Finally, the extended-precision number is converted to a floating point number with the given sign, and \_fp_sanitize:Nw checks for overflow and underflow.
Contrarily to \_\_fp_sin_series_o:NNwwww which received a conversion auxiliary as \#1, here, \#1 is 0 for tangent and 2 for cotangent. Consider first the case of the tangent. The octant \#3 starts at 1, which means that it is 1 or 2 for $|x| \in [0, \pi/2]$, it is 3 or 4 for $|x| \in [\pi/2, \pi]$, and so on: the intervals on which $\tan |x| \geq 0$ coincide with those for which $\lfloor (\#3 + 1)/2 \rfloor$ is odd. We also have to take into account the original sign of $x$ to get the sign of the final result; it is straightforward to check that the first \int_value:w expansion produces 0 for a positive final result, and 2 otherwise. A similar story holds for $\cot(x)$.

The auxiliary receives the sign, the octant, the square of the (reduced) input, and the (reduced) input (an extended-precision number) as arguments. It then computes the numerator and denominator of

$$\tan(x) \approx \frac{x(1 - x^2(a_1 - x^2(a_2 - x^2(a_3 - x^2(a_4 - x^2a_5)))))}{1 - x^2(b_1 - x^2(b_2 - x^2(b_3 - x^2(b_4 - x^2b_5))))}.$$

The ratio is computed by \_\_fp_ep_div:wwww, then converted to a floating point number. For octants \#3 (really, quadrants) next to a pole of the functions, the fixed point
numerator and denominator are exchanged before computing the ratio. Note that this
\if_int_odd:w test relies on the fact that the octant is at least 1.

(End definition for \__fp_tan_series_o:NNwwww and \__fp_tan_series_aux_o:Nwww.)

75.2 Inverse trigonometric functions

All inverse trigonometric functions (arcsine, arccosine, arctangent, arccotangent, arcosecant, and arcsecant) are based on a function often denoted \texttt{atan2}. This func-
The arctangent function \( \text{atan} \) is accessed directly by feeding two arguments to arctangent, and is defined by \( \text{atan}(y, x) = \text{atan}(y/x) \) for generic \( y \) and \( x \). Its advantages over the conventional arctangent is that it takes values in \([-\pi, \pi]\) rather than \([-\pi/2, \pi/2]\), and that it is better behaved in boundary cases. Other inverse trigonometric functions are expressed in terms of arctan as

\[
\text{acos} x = \text{atan}(\sqrt{1-x^2}, x) \\
\text{asin} x = \text{atan}(x, \sqrt{1-x^2}) \\
\text{asec} x = \text{atan}(\sqrt{x^2-1}, 1) \\
\text{acsc} x = \text{atan}(1, \sqrt{x^2-1}) \\
\text{atan} x = \text{atan}(x, 1) \\
\text{acot} x = \text{atan}(1, x).
\]

Rather than introducing a new function, \( \text{atan} \), the arctangent function \( \text{atan} \) is overloaded: it can take one or two arguments. In the comments below, following many texts, we call the first argument \( y \) and the second \( x \), because \( \text{atan}(y, x) = \text{atan}(y/x) \) is the angular coordinate of the point \((x, y)\).

As for direct trigonometric functions, the first step in computing \( \text{atan}(y, x) \) is argument reduction. The sign of \( y \) gives that of the result. We distinguish eight regions where the point \((x, |y|)\) can lie, of angular size roughly \( \pi/8 \), characterized by their “octant”, between 0 and 7 included. In each region, we compute an arctangent as a Taylor series, then shift this arctangent by the appropriate multiple of \( \pi/4 \) and sign to get the result. Here is a list of octants, and how we compute the arctangent (we assume \( y > 0 \): otherwise replace \( y \) by \( -y \) below):

- \( 0 \) 0 < \( |y| \) < 0.41421\( x \), then \( \text{atan} \frac{|y|}{x} \) is given by a nicely convergent Taylor series;
- \( 1 \) 0 < 0.41421\( x \) < \( |y| \) < \( x \), then \( \text{atan} \frac{|y|}{x} = \frac{\pi}{4} - \text{atan} \frac{x-|y|}{x+|y|} \);
- \( 2 \) 0 < 0.41421\( |y| \) < \( x \) < \( |y| \), then \( \text{atan} \frac{|y|}{x} = \frac{\pi}{4} + \text{atan} \frac{x+|y|}{x-|y|} \);
- \( 3 \) 0 < \( x \) < 0.41421\( |y| \), then \( \text{atan} \frac{|y|}{x} = \frac{\pi}{2} - \text{atan} \frac{x}{|y|} \);
- \( 4 \) 0 < \( -x \) < 0.41421\( |y| \), then \( \text{atan} \frac{|y|}{x} = \frac{\pi}{2} + \text{atan} \frac{x}{|y|} \);
- \( 5 \) 0 < 0.41421\( |y| \) < \( -x \) < \( |y| \), then \( \text{atan} \frac{|y|}{x} = \frac{3\pi}{4} - \text{atan} \frac{x+|y|}{x-|y|} \);
- \( 6 \) 0 < -0.41421\( x \) < \( |y| \) < \( -x \), then \( \text{atan} \frac{|y|}{x} = \frac{3\pi}{4} + \text{atan} \frac{x-|y|}{x+|y|} \);
- \( 7 \) 0 < \( |y| \) < -0.41421\( x \), then \( \text{atan} \frac{|y|}{x} = \pi - \text{atan} \frac{x}{|y|} \).

In the following, we denote by \( z \) the ratio among \( |\frac{x}{y}| \), \( |\frac{y}{x}| \), \( |\frac{x+|y|}{x-|y|}| \), \( |\frac{x-|y|}{x+|y|}| \) which appears in the right-hand side above.

### 75.2.1 Arctangent and arccotangent

The parsing step manipulates \( \text{atan} \) and \( \text{acot} \) like \( \text{min} \) and \( \text{max} \), reading in an array of operands, but also leaves \( \text{use_i:nn} \) or \( \text{use_ii:nn} \) depending on whether the result
should be given in radians or in degrees. The helper \_\_\_fp_parse_function_one_two:nnw checks that the operand is one or two floating point numbers (not tuples) and leaves its second argument or its tail accordingly (its first argument is used for error messages). More precisely if we are given a single floating point number \_\_\_fp_atan_default:w places \c_one_fp (expanded) after it; otherwise \_\_\_fp_parse_function_one_two:nnw is omitted by \_\_\_fp_parse_function_one_two:nnw.

\cs_new:Npn \_\_\_fp_atan_o:Nw #1
{ \_\_\_fp_parse_function_one_two:nnw
\begin{itemize}
\item\_\_\_fp_atan_default:w \_\_\_fp_atanii_o:Nww #1
\end{itemize}
}
\cs_new:Npn \_\_\_fp_acot_o:Nw #1
{ \_\_\_fp_parse_function_one_two:nnw
\begin{itemize}
\item\_\_\_fp_acot_default:w \_\_\_fp_acotii_o:Nww #1
\end{itemize}
}
\cs_new:Npx \_\_\_fp_atan_default:w #1#2#3 @ { #1 #2 #3 \c_one_fp @ } (End definition for \_\_\_fp_atan_o:Nw, \_\_\_fp_acot_o:Nw, and \_\_\_fp_atan_default:w.)

\_\_\_fp_atanii_o:Nww \_\_\_fp_acotii_o:Nww
If either operand is \textit{nan}, we return it. If both are normal, we call \_\_\_fp_atan_normal_o:NNNw. If both are zero or both infinity, we call \_\_\_fp_atan_inf_o:NNNw with argument 2, leading to a result among \{±π/4, ±3π/4\} (in degrees, \{±45, ±135\}). Otherwise, one is much bigger than the other, and we call \_\_\_fp_atan_inf_o:NNNw with either an argument of 4, leading to the values ±π/2 (in degrees, ±90), or 0, leading to \{±0, ±π\} (in degrees, \{±0, ±180\}). Since \textit{acot}(x, y) = \textit{atan}(y, x), \_\_\_fp_acotii_o:ww simply reverses its two arguments.

\cs_new:Npn \_\_\_fp_atanii_o:Nww #1 \s__fp \_\_\_fp_chk:w #2#3#4; \s__fp \_\_\_fp_chk:w #5 #6 @
{ \if_meaning:w 3 #2 \_\_\_fp_case_return_i_o:ww \fi:
\if_meaning:w 3 #5 \_\_\_fp_case_return_ii_o:ww \fi:
\if_case:w
\if_meaning:w #2 #5
\if_meaning:w 1 #2 10 \else: 0 \fi:
\else:
\if_int_compare:w #2 > #5 \exp_stop_f: 1 \else: 2 \fi:
\fi:
\exp_stop_f:
\_\_\_fp_case_return:nw { \_\_\_fp_atan_inf_o:NNNw #1 #3 #2 }
\or: \_\_\_fp_case_return:nw { \_\_\_fp_atan_inf_o:NNNw #1 #3 #4 }
\or: \_\_\_fp_case_return:nw { \_\_\_fp_atan_inf_o:NNNw #1 #3 #0 }
\fi:
\_\_\_fp_atan_normal_o:NNNw\_\_\_fp_atan_normal_o:NNNw\_\_\_fp_atan_normal_o:NNNw #1
\_\_\_fp_acotii_o:ww
(End definition for \_\_\_fp_atanii_o:Nww and \_\_\_fp_acotii_o:Nww.)
This auxiliary is called whenever one number is ±0 or ±∞ (and neither is nan). Then
the result only depends on the signs, and its value is a multiple of π/4. We use the same
auxiliary as for normal numbers, \__fp_atan_combine_o:NwwwwN, with arguments the
final sign #2; the octant #3; atan z/z = 1 as a fixed point number; z = 0 as a fixed
point number; and z = 0 as an extended-precision number. Given the values we provide,
atan z is computed to be 0, and the result is [#3/2] · π/4 if the sign #5 of x is positive,
and (7 − #3)/2 · π/4 for negative x, where the divisions are rounded up.

\cs_new:Npn \__fp_atan_inf_o:NNNw #1#2#3 \s__fp \__fp_chk:w #4#5#6;
\exp_after:wN \__fp_atan_combine_o:NwwwwwN
\exp_after:wN #2
\int_value:w \__fp_int_eval:w
\if_meaning:w 2 #4 7 - \fi: #3 \exp_after:wN ;
\c__fp_one_fixed_tl
{0000}{0000}{0000}{0000}{0000}{0000}; #1
}

(End definition for \__fp_atan_inf_o:NNNw.)

\__fp_atan_normal_o:NNnwNnw
Here we simply reorder the floating point data into a pair of signed extended-precision
numbers, that is, a sign, an exponent ending with a comma, and a six-block mantissa
ending with a semi-colon. This extended precision is required by other inverse trigono-
metric functions, to compute things like atan(x, √1 − x²) without intermediate rounding
errors.

\cs_new_protected:Npn \__fp_atan_normal_o:NNnwNnw #1 \s__fp \__fp_chk:w 1#2#3#4; \s__fp \__fp_chk:w 1#5#6#7;
\__fp_atan_test_o:NwwNwwN
\exp_after:wN \__fp_atan_combine_o:NwwwwwN
\exp_after:wN #2
\int_value:w \__fp_int_eval:w
\if_meaning:w 2 #4
\__fp_int_eval:w
\fi: #3 \exp_after:wN ;
\c__fp_one_fixed_tl
{0000}{0000}{0000}{0000}{0000}{0000}; #1
}

(End definition for \__fp_atan_normal_o:NNnwNnw.)

\__fp_atan_test_o:NwwNwwN
This receives: the sign #1 of y, its exponent #2, its 24 digits #3 in groups of 4, and
similarly for x. We prepare to call \__fp_atan_combine_o:NwwwwN which expects the
sign #1, the octant, the ratio (atan z)/z = 1 − ⋯, and the value of z, both as a fixed
point number and as an extended-precision floating point number with a mantissa in
[0.01,1). For now, we place #1 as a first argument, and start an integer expression for
the octant. The sign of x does not affect z, so we simply leave a contribution to the
octant: (octant) → 7 − (octant) for negative x. Then we order |y| and |x| in a non-
decreasing order: if |y| > |x|, insert 3− in the expression for the octant, and swap the
two numbers. The finer test with 0.41421 is done by \__fp_atan_div:wnwwnw after the
operands have been ordered.

\cs_new:Npn \__fp_atan_test_o:NwwNwwN #1#2,#3; #4#5,#6;
\exp_after:wN \__fp_atan_combine_o:NwwwwwN
\exp_after:wN #1
\int_value:w \__fp_int_eval:w
\if_meaning:w 2 #4
\__fp_int_eval:w
\fi: 7 - \__fp_int_eval:w
\fi:
This receives two positive numbers \( a \) and \( b \) (equal to \(|x|\) and \(|y|\) in some order), each as an exponent and 6 blocks of 4 digits, such that \( 0 < a < b \). If \( 0.41421b < a \), the two numbers are “near”, hence the point \((y,x)\) that we started with is closer to the diagonals \( \{ |y| = |x| \} \) than to the axes \( \{ xy = 0 \} \). In that case, the octant is 1 (possibly combined with the \( -7 \) and \( -3 \) inserted earlier) and we wish to compute \( \arctan \frac{b}{a} \). Otherwise, the octant is 0 (again, combined with earlier terms) and we wish to compute \( \arctan \frac{x}{y} \). In any case, call \( \fp_atan_div:wwwn \) followed by \( z \), as a comma-delimited exponent and a fixed point number.
{\
\_fp_atan_Taylor_loop:www 39 ;
\}\
! #1;
}

(End definition for \_fp_atan_auxi:ww and \_fp_atan_auxii:w)

\_fp_atan_Taylor_loop:www \_fp_atan_Taylor_break:w

We compute the series of \((\text{atan} z)/z\). A typical intermediate stage has 
\#1 = 2k − 1, 
\#2 = \frac{1}{2k+1} − z^2(\frac{1}{2k+3} − z^2(\cdots − z^2(\frac{-1}{2g})))\), and \#3 = \(z^2\). To go to the next step \(k \rightarrow k - 1\), we compute \(\frac{1}{2k+1}\), then subtract from it \(z^2\) times \#2. The loop stops when \(k = 0\): then \#2 is \((\text{atan} z)/z\), and there is a need to clean up all the unnecessary data, end the integer expression computing the octant with a semicolon, and leave the result \#2 afterwards.

\cs_new:Npn \_fp_atan_Taylor_loop:www #1; #2; #3;\{
\if_int_compare:w #1 = - \c_one_int
\_fp_atan_Taylor_break:w
\fi: 
\exp_after:wN \_fp_fixed_div_int:wwN \_fp_one_fixed_tl #1; 
\_fp_rrot:www \_fp_fixed_mul_sub_back:wwwn #2; #3;\{
\exp_after:wN \_fp_atan_Taylor_loop:www
\int_value:w \_fp_int_eval:w #1 - 2 ;
#3;
\}
\cs_new:Npn \_fp_atan_Taylor_break:w
\fi: \_fp_fixed_mul_sub_back:wwwn #2; #3 !
{ \fi: ; #2 ; }
\}

(End definition for \_fp_atan_Taylor_loop:www and \_fp_atan_Taylor_break:w)

\_fp_atan_combine_o:NwwwwwN \_fp_atan_combine_aux:ww

This receives a \((\text{sign})\), an \((\text{octant})\), a fixed point value of \((\text{atan} z)/z\), a fixed point number \(z\), and another representation of \(z\), as an \((\text{exponent})\) and the fixed point number \(10^{\text{exponent}}z\), followed by either \use_i:nn (when working in radians) or \use_ii:nn (when working in degrees). The function computes the floating point result \((\text{sign})\left(\left\lceil \frac{\text{octant}}{2} \right\rceil \frac{\pi}{4} + (-1)^{\text{octant}}\frac{\text{atan} z \cdot z}{z} \right)\), (11) multiplied by \(180/\pi\) if working in degrees, and using in any case the most appropriate representation of \(z\). The floating point result is passed to \_fpsanitize:Nw, which checks for overflow or underflow. If the octant is 0, leave the exponent \#5 for \_fp-sanitize:Nw, and multiply \#3 = \frac{\text{atan} z}{z} with \#6, the adjusted \(z\). Otherwise, multiply \#3 = \frac{\text{atan} z}{z} with \#4 = \(z\), then compute the appropriate multiple of \(\frac{\pi}{4}\) and add or subtract the product \#3 \cdot \#4. In both cases, convert to a floating point with \_fp_fixed_to_float:o:wN.

\cs_new:Npn \_fp_atan_combine_o:NwwwwwN #1 #2; #3; #4; #5,#6; #7\{
\exp_after:wN \_fpsanitize:Nw
\exp_after:wN \_fp_fixed_to_float:o:wN
\int_value:w \_fp_int_eval:w

1155
75.2.2 Arcsine and arccosine

\__fp_asin_o:w

Again, the first argument provided by l3fp-parse is \use_i:nn if we are to work in radians and \use_i:ii:nn for degrees. Then comes a floating point number. The arcsine of ±0 or \texttt{nan} is the same floating point number. The arcsine of ±\infty raises an invalid operation exception. Otherwise, call an auxiliary common with \__fp_acos_o:w, feeding it information about what function is being performed (for “invalid operation” exceptions).
The arcsine of ±0 is π/2 (in degrees, 90). The arcsine of ±∞ raises an invalid operation exception. The arcsine of nan is itself. Otherwise, call an auxiliary common with \_\_fp_sin_o:w, informing it that it was called by acs or acosd, and preparing to swap some arguments down the line.

\_\_fp_acos_o:w

\cs_new:Npn \_\_fp_acos_o:w #1 \s__fp \_\_fp_chk:w #2#3; @
{\if_case:w #2 \exp_stop_f:
  \_\_fp_case_use:nw { \_\_fp_atan_inf_o:NNNw #1 0 4 }
  \or:
  \_\_fp_case_use:nw
  \{ \_\_fp_acos_normal_o:NfwNnnnnw #1 #1 { acos } { acosd } \}
  \_\_fp_reverse_args:Nww
\or:
  \_\_fp_case_use:nw
  \{ \_\_fp_invalid_operation_o:fw {#2} \}
\else:
  \_\_fp_case_return_same_o:w
  \s__fp \_\_fp_chk:w #2 #3;
\fi:}

\_\_fp_acos_normal_o:NfwNnnnnw

If the exponent #5 is at most 0, the operand lies within (−1,1) and the operation is permitted: call \_\_fp_sin_normal_o:NfwNnnnnw with the appropriate arguments. If the number is exactly ±1 (the test works because we know that #5 ≥ 1, #6#7 ≥ 10000000, #8#9 ≥ 0, with equality only for ±1), we also call \_\_fp_sin_normal_o:NfwNnnnnw. Otherwise, \_\_fp_use_i:ww gets rid of the sin auxiliary, and raises instead an invalid operation, because the operand is outside the domain of arcsine or arcsine.

\_\_fp_sin_normal_o:NfwNnnnnw

We compute $x/\sqrt{1-x^2}$. This function is used by asin and acos, but also by acsc and asec after inverting the operand, thus it must manipulate extended-precision numbers. First evaluate $1-x^2$ as \((1+x)(1-x)\): this behaves better near $x=1$. We do the
addition/subtraction with fixed point numbers (they are not implemented for extended-
precision floats), but go back to extended-precision floats to multiply and compute the
inverse square root $1/\sqrt{1-x^2}$. Finally, multiply by the (positive) extended-precision float $|x|$, and feed the (signed) result, and the number $+1$, as arguments to the arctangent function. When computing the arccosine, the arguments $x/\sqrt{1-x^2}$ and $+1$ are swapped by \texttt{\_\_fp_reverse_args:Nww} in that case before \texttt{\_\_fp_atan_test_o:NwwNwwN} is evaluated. Note that the arctangent function requires normalized arguments, hence the need for \texttt{ep_to_ep} and \texttt{continue} after \texttt{ep_mul}.

\begin{verbatim}
28026 \cs_new:Npn \__fp_asin_auxi_o:NnNww #1#2#3#4,#5; 28027 { 28028 \__fp_ep_to_fixed:wwn #4,#5; 28029 \__fp_asin_isqrt:wn 28030 \__fp_ep_mul:wwwwn #4,#5; 28031 \__fp_ep_to_ep:wwN 28032 \__fp_fixed_continue:wn 28033 { #2 \__fp_atan_test_o:NwwNwwN #3 } 28034 0 1,{1000}{0000}{0000}{0000}{0000}{0000}; #1 28035 } 28036 \cs_new:Npn \__fp_asin_isqrt:wn #1; 28037 { 28038 \exp_after:wN \__fp_fixed_sub:wwn \c__fp_one_fixed_tl #1; 28039 \__fp_fixed_add_one:wN #1; 28040 \__fp_fixed_continue:wn { \__fp_ep_mul:wwwwn 0, } 0, 28041 \__fp_ep_isqrt:wwn 28042 } 28043 \) 28044 \)
\end{verbatim}

(End definition for \texttt{\_\_fp_asin_auxi_o:Nww} and \texttt{\_\_fp_asin_isqrt:wn}.)

### 75.2.3 Arccosecant and arcsecant

\begin{verbatim}
\_fp_acsc_o:w
Cases are mostly labelled by #2, except when #2 is 2: then we use #3#2, which is \texttt{02 = 2}
when the number is $+\infty$ and $22$ when the number is $-\infty$. The arccosecant of $0$ raises
an invalid operation exception. The arccosecant of $\pm\infty$ is $0$ with the same sign. The
arcosecant of \texttt{nan} is itself. Otherwise, \texttt{\_\_fp_acsc_normal_o:NfwNw} does some more
tests, keeping the function name (\texttt{acsc} or \texttt{acscd}) as an argument for invalid operation
exceptions.

\begin{verbatim}
28044 \cs_new:Npn \__fp_acsc_o:w #1 \s__fp \__fp_chk:w #2#3#4; @ 28045 { 28046 \if_case:w \if_meaning:w 2 #2 #3 \fi: #2 \exp_stop_f: 28047 \_\_fp_case_use:nw 28048 \_\_fp_invalid_operation_o:fw \{ #1 \acsc \} \acscd \} } 28049 \or: \_\_fp_case_use:nw 28050 \_\_fp_acsc_normal_o:NfwNw \#1 \{ #1 \acsc \} \acscd \} } 28051 \or: \_\_fp_case_return_o:Nw \c_zero_fp 28052 \or: \_\_fp_case_return_same_o:w 28053 \else: \_\_fp_case_return_o:Nw \c_minus_zero_fp 28054 \fi: 28055 \s__fp \_\_fp_chk:w #2 #3 #4; 28056 } 28057 \)
\end{verbatim}

(End definition for \texttt{\_\_fp_acsc_o:w}.)

1158
The arcsecant of $\pm0$ raises an invalid operation exception. The arcsecant of $\pm\infty$ is $\pi/2$ (in degrees, 90). The arcsecant of \texttt{nan} is itself. Otherwise, do some more tests, keeping
the function name \texttt{asec} (or \texttt{asecd}) as an argument for invalid operation exceptions, and
a \texttt{\_\_fp_reverse_args:Nww} following precisely that appearing in \texttt{\_\_fp_acos:o:w}.

\begin{verbatim}
\cs_new:Npn \_\_fp_asec_o:w #1 \s__fp \_\_fp_chk:w #2#3; @
\{
\if_case:w #2 \exp_stop_f:
\__fp_case_use:nw
\__fp_invalid_operation_o:fw { #1 { asec } { asecd } }
\or:
\__fp_case_use:nw
\__fp_acsc_normal_o:NfwNnw #1 { #1 { asec } { asecd } }
\__fp_reverse_args:Nww
\} \or: \__fp_case_use:nw \__fp_atan_inf_o:NNNw #1 0 4 \fi:
\s__fp \_\_fp_chk:w #2 #3;
\}
\end{verbatim}

(End definition for \texttt{\_\_fp_asec_o:w}.)

\begin{verbatim}
\cs_new:Npn \_\_fp_acsc_normal_o:NfwNnw #1#2#3 \s__fp \_\_fp_chk:w 1#4#5#6;
\{
\int_compare:nNnTF {#5} < 1
\{
\__fp_invalid_operation_o:fw {#2}
\s__fp \_\_fp_chk:w 1#4#5#6;
\}
\{
\__fp_ep_div:wwwwn
1,(1000){0000}{0000}{0000}{0000}{0000};
#5,#6{0000}{0000};
\__fp_asin_auxi_o:NnNww #1 {#3} #4 }
\}
\end{verbatim}

(End definition for \texttt{\_\_fp_acsc_normal_o:NfwNnw}.)

⟨/package⟩
Chapter 76

13fp-convert implementation

76.1 Dealing with tuples

The first argument is for instance \_fp_to_tl_dispatch:w, which converts any floating point object to the appropriate representation. We loop through all items, putting \,\, between all of them and making sure to remove the leading \,\,.

(End definition for \_fp_tuple_convert:Nw, \_fp_tuple_convert_loop:nW, \_fp_tuple_convert_end:w.)

76.2 Trimming trailing zeros

If \#1 ends with a 0, the loop auxiliary takes that zero as an end-delimiter for its first argument, and the second argument is the same loop auxiliary. Once the last trailing
zero is reached, the second argument is the dot auxiliary, which removes a trailing dot if any. We then clean-up with the end auxiliary, keeping only the number.

\cs_new:Npn \__fp_trim_zeros:w #1 ;
\__fp_trim_zeros_loop:w #1 ; \__fp_trim_zeros_loop:w 0; \__fp_trim_zeros_dot:w .; \s__fp_stop
\cs_new:Npn \__fp_trim_zeros_loop:w #1 0; \__fp_trim_zeros_end:w #1 ; 
\cs_new:Npn \__fp_trim_zeros_end:w #1 ; #2 \s__fp_stop \{ \}

(End definition for \__fp_trim_zeros:w and others.)

76.3 Scientific notation

The three public functions evaluate their argument, then pass it to \__fp_to_scientific_dispatch:w.

\cs_new:Npn \fp_to_scientific:N #1 
\exp_after:wN \__fp_to_scientific_dispatch:w #1
\cs_generate_variant:Nn \fp_to_scientific:N { c }
\cs_new:Npn \fp_to_scientific:n 
\exp_after:wN \__fp_to_scientific_dispatch:w \exp:w \exp_end_continue_f:w \__fp_parse:n

(End definition for \fp_to_scientific:N and \fp_to_scientific:n. These functions are documented on page 242.)

\__fp_to_scientific_dispatch:w
\__fp_to_scientific_recover:w
\__fp_tuple_to_scientific:w

We allow tuples.

\cs_new:Npn \__fp_to_scientific_dispatch:w #1 
\{ \__fp_change_func_type:NNN \__fp_to_scientific:w \__fp_to_scientific_recover:w \__fp_to_scientific_normal:wnnnnn
\__fp_to_scientific_normal:wNw

(End definition for \__fp_to_scientific_dispatch:w, \__fp_to_scientific_recover:w, \__fp_tuple_to_scientific:w, \__fp_to_scientific_normal:wnnnn, \__fp_to_scientific_normal:wNw)

Expressing an internal floating point number in scientific notation is quite easy: no rounding, and the format is very well defined. First cater for the sign: negative numbers (\#2 = 2) start with \texttt{-}; we then only need to care about positive numbers and \texttt{nan}. Then filter the special cases: \pm 0 are represented as 0; infinities are converted to a number slightly larger than the largest after an “invalid_operation” exception; \texttt{nan} is represented as 0 after an “invalid_operation” exception. In the normal case, decrement the exponent.
and unbrace the 4 brace groups, then in a second step grab the first digit (previously hidden in braces) to order the various parts correctly.

```latex
\cs_new:Npn \__fp_to_scientific:w \s__fp \__fp_chk:w \#1\#2
\{ 28141  \if_meaning:w 2 \#2 \exp_after:wN - \exp:w \exp_end_continue_f:w \fi:
\if_case:w \#1 \exp_stop_f:
\__fp_case_return:nw { 0.000000000000000e0 } \or:
\exp_after:wN \__fp_to_scientific_normal:wnnnn
\or:
\__fp_case_use:nw
\{ 28149 \__fp_invalid_operation:nnw
\{ \fp_to_scientific:N \c__fp_overflowing_fp \}
\{ \fp_to_scientific \}
\}
\or:
\__fp_case_use:nw
\{ 28158 \__fp_invalid_operation:nnw
\{ \fp_to_scientific:N \c_zero_fp \}
\{ \fp_to_scientific \}
\}
\fi:
\s__fp \__fp_chk:w \#1 \#2
\}
\cs_new:Npn \__fp_to_scientific_normal:wnnnn
\s__fp \__fp_chk:w 1 \#1 \#2 \#3\#4\#5\#6 ;
\{ 28169 \exp_after:wN \__fp_to_scientific_normal:wNw
\exp_after:wN e \int_value:w \__fp_int_eval:w \#2 - 1
\; \#3 \#4 \#5 \#6 ;
\}
\cs_new:Npn \__fp_to_scientific_normal:wNw \#1 ; \#2\#3;
\{ \#2.\#3 \#1 \}
(End definition for \__fp_to_scientific:w, \__fp_to_scientific_normal:wnnnn, and \__fp_to_scientific_normal:wNw.)
```

### 76.4 Decimal representation

\verb!\fp_to_decimal:N! \verb!\fp_to_decimal:c! \verb!\fp_to_decimal:n! All three public variants are based on the same \__fp_to_decimal_dispatch:w after evaluating their argument to an internal floating point.

```latex
\cs_new:Npn \fp_to_decimal:N \#1
\{ \exp_after:wN \__fp_to_decimal_dispatch:w \#1 \}
\cs_generate_variant:Nn \fp_to_decimal:N { c }
\cs_new:Npn \fp_to_decimal:n
\{ \exp_after:wN \__fp_to_decimal_dispatch:w
\exp:w \exp_end_continue_f:w \__fp_parse:n
\}
(End definition for \fp_to_decimal:N and \fp_to_decimal:n. These functions are documented on page 241.)
```
We allow tuples.

The structure is similar to \texttt{\_fp_to_scientific:w}. Insert - for negative numbers. Zero gives 0, \pm \infty and \texttt{nan} yield an “invalid operation” exception; note that \pm \infty produces a very large output, which we don’t expand now since it most likely won’t be needed. Normal numbers with an exponent in the range [1, 15] have that number of digits before the decimal separator: “decimate” them, and remove leading zeros with \texttt{\_int_value:w}, then trim trailing zeros and dot. Normal numbers with an exponent 16 or larger have no decimal separator, we only need to add trailing zeros. When the exponent is non-positive, the result should be 0,\langle\texttt{zeros}\rangle\langle\texttt{digits}\rangle, trimmed.
\int_compare:nNnTF {#2} < \c__fp_prec_int
\{ 
    \__fp_decimate:nNnnn { \c__fp_prec_int - #2 }
    \__fp_to_decimal_large:Nnnw
\}
\{ \exp_after:wN \exp_after:wN 
    \exp_after:wN \__fp_to_decimal_huge:wnnnn 
    \prg_replicate:nn { #2 - \c__fp_prec_int } { 0 } ;
\}
\{ #3 \} \{ #4 \} \{ #5 \} \{ #6 \} 
\} 
\{ \exp_after:wN \__fp_trim_zeros:w \int_value:w 
    \if_int_compare:w #2 > \c_zero_int 
    \exp_stop_f: 
    \exp:w \exp_end_continue_f:w #3#4#5#6 ; 
\}
\cs_new:Npn \__fp_to_decimal_large:Nnnw #1#2#3#4 ;
\cs_new:Npn \__fp_to_decimal_huge:wnnnn #1 ; #2#3#4#5 { #2#3#4#5 #1 }
(End definition for \__fp_to_decimal:w and others.)

76.5 Token list representation

\fp_to_tl:N \fp_to_tl:c \fp_to_tl:n
These three public functions evaluate their argument, then pass it to \__fp_to_tl_dispatch:w.
\cs_new:Npn \fp_to_tl:N #1 { \exp_after:wN \__fp_to_tl_dispatch:w #1 }
\cs_new:Npn \fp_to_tl:n \fp_to_tl:N \cs_generate_variant:Nn \fp_to_tl:N \c }
\cs_new:Npn \fp_to_tl:n \fp_to_tl:n \fp_to_tl:n
\exp:w \exp_end_continue_f:w \__fp_tuple_to_tl:w
(End definition for \fp_to_tl:N and \fp_to_tl:n. These functions are documented on page 242.)

\__fp_to_tl_dispatch:w \__fp_to_tl_recover:w \__fp_tuple_to_tl:w
We allow tuples.
\cs_new:Npn \__fp_to_tl_dispatch:w #1
\cs_new:Npn \__fp_to_tl_recover:w #1 \__fp_to_tl:w \__fp_to_tl_recover:w #1
\cs_new:Npn \__fp_to_tl_recover:w #1 \__fp_to_tl:w \__fp_to_tl_recover:w #1
\{ \exp_after:wN \__fp_error:nffn \unknown-type \} \tl_to_str:n \{ #2 ; \} \} \} \}
A structure similar to \__fp_to_scientific_dispatch:w and \__fp_to_decimal_dispatch:w, but without the “invalid operation” exception. First filter special cases. We express normal numbers in decimal notation if the exponent is in the range \([-2, 16]\), and otherwise use scientific notation.

\cs_new:Npn \__fp_to_tl:w \s__fp \__fp_chk:w #1#2,
  \if_meaning:w 2 #2 \exp_after:wN \exp:w \exp_end_continue_f:w \fi:
  \if_case:w #1 \exp_stop_f:
    \__fp_case_return:nw \{ 0 \}
  \or: \exp_after:wN \__fp_to_tl_normal:nnnnn
  \or: \__fp_case_return:nw \{ inf \}
  \else: \__fp_case_return:nw \{ nan \}
  \fi:
\}
\cs_new:Npn \__fp_to_tl_normal:nnnnn #1,
  \int_compare:nTF
    \{-2 <= \#1 <= \c__fp_prec_int \}
  \{ \__fp_to_decimal_normal:wnnnnn \}
  \{ \__fp_to_tl_scientific:wnnnn \}
  \s__fp \__fp_chk:w 1 0 {#1}
\cs_new:Npn \__fp_to_tl_scientific:wnnnnn
  \s__fp \__fp_chk:w 1 #1 #2 #3 #4 #5 #6 ;
  \exp_after:wN \__fp_to_tl_scientific:wNw
  \exp_after:wN e
  \int_value:w \__fp_int_eval:w \#2 - 1
  \; \#3 \#4 \#5 \#6 ;
\}
\cs_new:Npn \__fp_to_tl_scientific:wNw
  \s__fp \__fp_chk:w 1 0 {1}\}
\}
\cs_new:Npn \__fp_to_tl_normal:nnnnn #1
  \s__fp \__fp_chk:w 1 #1 #2 #3 #4 #5 #6 ;
  \{ \exp_after:wN \__fp_to_tl_scientific:wNw
  \exp_after:wN e
  \int_value:w \__fp_int_eval:w \#2 - 1
  \; \#3 \#4 \#5 \#6 ;
  \}
\cs_new:Npn \__fp_to_tl_scientific:wNw
  \s__fp \__fp_chk:w 1 #1 #2 #3\#2 #3\; #1 \}
(End definition for \__fp_to_tl:w and others.)

76.6 Formatting

This is not implemented yet, as it is not yet clear what a correct interface would be, for this kind of structured conversion from a floating point (or other types of variables) to a string. Ideas welcome.

76.7 Convert to dimension or integer

\fp_to_dim:N
\fp_to_dim:c
\fp_to_dim:n
\__fp_to_dim_dispatch:w
\__fp_to_dim_recover:w
\__fp_to_dim:w

All three public variants are based on the same \__fp_to_dim_dispatch:w after evaluating their argument to an internal floating point. We only allow floating point numbers,
not tuples.

\cs_new:Npn \fp_to_dim:N #1
\cs_generate_variant:Nn \fp_to_dim:N { c }
\cs_new:Npn \fp_to_dim:n
\exp_after:wN \__fp_to_dim_dispatch:w #1
\exp:w \exp_end_continue_f:w \__fp_parse:n
\cs_new:Npn \__fp_to_dim_dispatch:w #1#2 ;
\__fp_change_func_type:NNN #1 \__fp_to_dim:w \__fp_to_dim_recover:w #1 #2 ;
\cs_new:Npn \__fp_to_int_recover:w #1 { \__fp_invalid_operation:nnw { 0 } { fp_to_int } }
\cs_new:Npn \__fp_to_int:w #1; 
\exp_after:wN \__fp_to_decimal:w \exp:w \exp_end_continue_f:w \__fp_round:Nwn \__fp_round_to_nearest:NNN #1; { 0 }
(End definition for \fp_to_dim:N and others. These functions are documented on page \textit{241}.)

\fp_to_int:N
\fp_to_int:c
\fp_to_int:N
\__fp_to_int_dispatch:w
\__fp_to_int_recover:w
For the most part identical to \fp_to_dim:N but without pt, and where \__fp_to_int:w does more work. To convert to an integer, first round to 0 places (to the nearest integer), then express the result as a decimal number: the definition of \__fp_to_decimal_dispatch:w is such that there are no trailing dot nor zero.

\cs_new:Npn \fp_to_int:N #1 { \exp_after:wN \__fp_to_int_dispatch:w #1 }
\cs_generate_variant:Nn \fp_to_int:N { c }
\cs_new:Npn \fp_to_int:n
\exp_after:wN \__fp_to_int_dispatch:w
\exp:w \exp_end_continue_f:w \__fp_parse:n
\cs_new:Npn \__fp_to_int_dispatch:w #1#2 ;
\__fp_change_func_type:NNN #1 \__fp_to_int:w \__fp_to_int_recover:w #1 #2 ;
\cs_new:Npn \__fp_to_int_recover:w #1 { \__fp_invalid_operation:nnw { 0 } { fp_to_int } }
\cs_new:Npn \__fp_to_int:w #1; 
\exp_after:wN \__fp_to_decimal:w \exp:w \exp_end_continue_f:w \__fp_round:Nwn \__fp_round_to_nearest:NNN #1; { 0 }
(End definition for \fp_to_int:N and others. These functions are documented on page \textit{241}.)

76.8 Convert from a dimension

\dim_to_fp:n
\__fp_from_dim_test:ww
\__fp_from_dim:ww
\__fp_from_dim:WW
\__fp_from_dim:WNNnnnn
\__fp_from_dim:WNNnnW
The dimension expression (which can in fact be a glue expression) is evaluated, converted to a number (i.e., expressed in scaled points), then multiplied by $2^{-16} = 0.0000152587890625$ to give a value expressed in points. The auxiliary \__fp_mul_-npos_o:Nww expects the desired (final sign) and two floating point operands (of the form
\s__fp... \) as arguments. This set of functions is also used to convert dimension registers to floating points while parsing expressions: in this context there is an additional exponent, which is the first argument of \_fp_from_dim_test:ww, and is combined with the exponent $-4$ of $2^{-16}$. There is also a need to expand afterwards: this is performed by \_fp_mul_npos_o:Nww, and cancelled by \prg_do_nothing: here.

\begin{verbatim}
\cs_new:Npn \dim_to_fp:n #1
\exp_after:wN \__fp_from_dim_test:ww
\exp_after:wN 0
\exp_after:wN ,
\int_value:w \tex_glueexpr:D #1 ;
\}
\cs_new:Npn \__fp_from_dim_test:ww #1, #2
\if_meaning:w 0 #2 \__fp_case_return:nw { \exp_after:wN \c_zero_fp }
\else:
\exp_after:wN \__fp_from_dim:wNw
\int_value:w \__fp_int_eval:w #1 - 4
\if_meaning:w - #2 \exp_after:wN , \exp_after:wN 2 \int_value:w #2
\else:
\exp_after:wN , \exp_after:wN 0 \int_value:w #2
\fi:
\fi:
\}
\cs_new:Npn \__fp_from_dim:wNw #1,#2#3;
\__fp_pack_twice_four:wNNNNNNNN \__fp_from_dim:wNNnnnnnn ;
#3 000 0000 00 \{10\}987654321; #2 \{#1\}
\}
\cs_new:Npn \__fp_from_dim:wNNnnnnnn #1; #2#3#4#5#6#7#8#9
{ \__fp_from_dim:wnnnnwNn #1 {#2#300} {0000} ; }
\cs_new:Npn \__fp_from_dim:wnnnnwNn #1; #2#3#4#5; #7#8
{ \__fp_mul_npos_o:Nww #7
\s__fp \__fp_chk:w 1 #7 \{#5\} \{#1\};
\s__fp \__fp_chk:w 1 0 \{#8\} \{1525\} \{8789\} \{0625\} \{0000\} ;
\prg_do_nothing:
\}
\end{verbatim}

(End definition for \dim_to_fp:n and others. This function is documented on page 213.)

76.9 Use and eval

\begin{verbatim}
\fp_use:N
\fp_use:c
\fp_eval:n
\end{verbatim}

Those public functions are simple copies of the decimal conversions.

\begin{verbatim}
\cs_new_eq:NN \fp_use:N \fp_to_decimal:N
\cs_generate_variant:Nn \fp_use:N { c }
\cs_new_eq:NN \fp_eval:n \fp_to_decimal:N
\end{verbatim}

(End definition for \fp_use:N and \fp_eval:n. These functions are documented on page 242.)

1167
\texttt{fp_sign:n} Trivial but useful. See the implementation of \texttt{fp_add:Nn} for an explanation of why to use \texttt{__fp_parse:n}, namely, for better error reporting.

\begin{verbatim}
28370 \cs_new:Npn \fp_sign:n #1
28371 { \fp_to_decimal:n { sign \__fp_parse:n {#1} } }
\end{verbatim}

(End definition for \texttt{fp_sign:n}. This function is documented on page 241.)

\texttt{fp_abs:n} Trivial but useful. See the implementation of \texttt{fp_add:Nn} for an explanation of why to use \texttt{__fp_parse:n}, namely, for better error reporting.

\begin{verbatim}
28372 \cs_new:Npn \fp_abs:n #1
28373 { \fp_to_decimal:n { abs \__fp_parse:n {#1} } }
\end{verbatim}

(End definition for \texttt{fp_abs:n}. This function is documented on page 257.)

\texttt{fp_max:nn} and \texttt{fp_min:nn} Similar to \texttt{fp_abs:n}, for consistency with \texttt{int_max:nn}, etc.

\begin{verbatim}
28374 \cs_new:Npn \fp_max:nn #1#2
28375 { \fp_to_decimal:n { max ( \__fp_parse:n {#1} , \__fp_parse:n {#2} ) } }
28376 \cs_new:Npn \fp_min:nn #1#2
28377 { \fp_to_decimal:n { min ( \__fp_parse:n {#1} , \__fp_parse:n {#2} ) } }
\end{verbatim}

(End definition for \texttt{fp_max:nn} and \texttt{fp_min:nn}. These functions are documented on page 257.)

### 76.10 Convert an array of floating points to a comma list

\texttt{__fp_array_to_clist:n} \texttt{__fp_array_to_clist_loop:Nw}

Converts an array of floating point numbers to a comma-list. If speed here ends up irrelevant, we can simplify the code for the auxiliary to become

\begin{verbatim}
28380 \cs_new:Npn \__fp_array_to_clist_loop:Nw #1#2;
28381 { \tl_if_empty:nF {#1}
28382 { \exp_last_unbraced:Ne \use_ii:nn
28383 { \__fp_array_to_clist_loop:Nw { ? \prg_break: } ;
28384 \__fp_array_to_clist_loop:Nw
28385 } }
28386 \use_none:n #1
28387 , ~
\end{verbatim}

The \texttt{use_ii:nn} function is expanded after \texttt{__fp_expand:n} is done, and it removes ,~ from the start of the representation.

\begin{verbatim}
28395 \cs_new:Npn \__fp_array_to_clist:n #1
28396 { \tl_if_empty:nF {#1}
28397 { \exp_last_unbraced:Ne \use_ii:nn
28398 { \__fp_array_to_clist_loop:Nw { ? \prg_break: } ;
28399 \__fp_array_to_clist_loop:Nw
28400 } }
28401 \use_none:n #1
28402 , ~
\end{verbatim}
\exp_not:f \__fp_to_tl_dispatch:w \#1 \#2 ;
\__fp_array_to_clist_loop:Nw
}

(End definition for \__fp_array_to_clist:n and \__fp_array_to_clist_loop:Nw.)

⟨/package⟩
Chapter 77

l3fp-random Implementation

Those functions may receive a variable number of arguments. We won’t use the argument ?.

(End definition for \_\_fp_parse_word_rand:N and \_\_fp_parse_word_randint:N.)

77.1 Engine support

Most engines provide random numbers, but not all. We write the test twice simply in order to write the false branch first.
Quite problematic, it is uncomfortably close to being so, and it becomes worse as \( k \) takes at the very least 16 seconds on a 2 giga-hertz processor). While this bias is not directly \( N \), \( N \) to \( K \) and \( K \) of \( N \) does not divide \( N \). This is detectable after about \( 2^{56}/2^{28} \) random numbers. For \( k = 15 \), \( N = 98304 \), this means roughly \( 2^{20} \) calls to the RNG (experimentally this takes at the very least 16 seconds on a 2 giga-hertz processor). While this bias is not quite problematic, it is uncomfortably close to being so, and it becomes worse as \( N \) is increased. In our code, we shall thus combine several results from the RNG.

The RNG has three types of unexpected correlations. First, everything is linear modulo \( 2^{28} \), hence the lowest \( k \) bits of the random numbers only depend on the lowest \( k \) bits of the seed (and of course the number of times the RNG was called since setting the seed). The recommended way to get a number from 0 to \( N - 1 \) is thus to scale the raw \( 28 \)-bit integer, as the engine's RNG does. We will go further and in fact typically we discard some of the lowest bits.

Second, suppose that we call the RNG with the same argument \( N \) to get a set of \( K \) integers in \([0, N - 1]\) (throwing away repeats), and suppose that \( N > K^3 \) and \( K > 55 \). The recursion used to construct more \( 28 \)-bit numbers from previous ones is linear: \( x_n = x_{n-55} - x_{n-24} \) or \( x_n = x_{n-55} - x_{n-24} + 2^{28} \). After rescaling and rounding we find that the result \( N_n \in [0, N - 1] \) is among \( N_{n-55} - N_{n-24} + \{-1, 0, 1\} \) modulo \( N \) (a more detailed analysis shows that 0 appears with frequency close to 3/4). The resulting set thus has more triplets \((a, b, c)\) than expected obeying \( a = b + c \) modulo \( N \). Namely it will have of order \((K - 55) \times 3/4\) such triplets, when one would expect \( K^3/(6N) \). This starts to be detectable around \( N = 2^{18} > 55^3 \) (earlier if one keeps track of positions too, but this is more subtle than it looks because the array of \( 28 \)-bit integers is read backwards by the engine). Hopefully the correlation is subtle enough to not affect realistic documents so we do not specifically mitigate against this. Since we typically use two calls to the RNG per \( \texttt{int_rand:nn} \) we would need to investigate linear relations between the \( x_{2n} \) on the one hand and between the \( x_{2n+1} \) on the other hand. Such relations will have more complicated coefficients than \( \pm 1 \), which alleviates the issue.

Third, consider successive batches of 165 calls to the RNG (with argument \( 2^{28} \) or with argument 2 for instance), then most batches have more odd than even numbers. Note

\[
\begin{align*}
\{ & 1 \\
\} \\
\} \\
\{ \texttt{sys_if_rand_exist:T} \\
\{ \\
\}
\end{align*}
\]

Obviously, every word “random” below means “pseudo-random”, as we have no access to entropy (except a very unreliable source of entropy: the time it takes to run some code).

The primitive random number generator (RNG) is provided as \( \texttt{int_rand:nn} \). Under the hood, it maintains an array of \( 55 \) \( 28 \)-bit numbers, updated with a linear recursion relation (similar to Fibonacci numbers) modulo \( 2^{28} \). When \( \texttt{int_uniformdeviate:D} \langle \texttt{integer} \rangle \) is called (for brevity denote by \( N \) the \( \langle \texttt{integer} \rangle \)), the next \( 28 \)-bit number is read from the array, scaled by \( N/2^{28} \), and rounded. To prevent 0 and \( N \) from appearing half as often as other numbers, they are both mapped to the result 0.

This process means that \( \texttt{int_uniformdeviate:D} \) only gives a uniform distribution from 0 to \( N - 1 \) if \( N \) is a divisor of \( 2^{28} \), so we will mostly call the RNG with such power of 2 arguments. If \( N \) does not divide \( 2^{28} \), then the relative non-uniformity (difference between probabilities of getting different numbers) is about \( N/2^{28} \). This implies that detecting deviation from \( 1/N \) of the probability of a fixed value \( X \) requires about \( 2^{56}/N \) random trials. But collective patterns can reduce this to about \( 3\times2^k \); the modulo 3 partition of such random numbers is biased with a non-uniformity about \( 2^k/2^{28} \) (which is much worse than the circa \( 3/2^{28} \) non-uniformity from taking directly \( N = 3 \)). This is detectable after about \( 2^{56}/2^{28} = 9 \cdot 2^{38}/N^2 \) random numbers.

When \( k = 15 \), \( N = 98304 \), this means roughly \( 2^{20} \) calls to the RNG. In our code, we shall thus combine several results from the RNG.

The recommended way to get a number from 0 to \( N - 1 \) is thus to scale the raw \( 28 \)-bit integer, as the engine’s RNG does. We will go further and in fact typically we discard some of the lowest bits.

Second, suppose that we call the RNG with the same argument \( N \) to get a set of \( K \) integers in \([0, N - 1]\) (throwing away repeats), and suppose that \( N > K^3 \) and \( K > 55 \). The recursion used to construct more \( 28 \)-bit numbers from previous ones is linear: \( x_n = x_{n-55} - x_{n-24} \) or \( x_n = x_{n-55} - x_{n-24} + 2^{28} \). After rescaling and rounding we find that the result \( N_n \in [0, N - 1] \) is among \( N_{n-55} - N_{n-24} + \{-1, 0, 1\} \) modulo \( N \) (a more detailed analysis shows that 0 appears with frequency close to 3/4). The resulting set thus has more triplets \((a, b, c)\) than expected obeying \( a = b + c \) modulo \( N \). Namely it will have of order \((K - 55) \times 3/4\) such triplets, when one would expect \( K^3/(6N) \). This starts to be detectable around \( N = 2^{18} > 55^3 \) (earlier if one keeps track of positions too, but this is more subtle than it looks because the array of \( 28 \)-bit integers is read backwards by the engine). Hopefully the correlation is subtle enough to not affect realistic documents so we do not specifically mitigate against this. Since we typically use two calls to the RNG per \( \texttt{int_rand:nn} \) we would need to investigate linear relations between the \( x_{2n} \) on the one hand and between the \( x_{2n+1} \) on the other hand. Such relations will have more complicated coefficients than \( \pm 1 \), which alleviates the issue.

Third, consider successive batches of 165 calls to the RNG (with argument \( 2^{28} \) or with argument 2 for instance), then most batches have more odd than even numbers. Note
that this does not mean that there are more odd than even numbers overall. Similar issues are discussed in Knuth’s TAOCP volume 2 near exercise 3.3.2-31. We do not have any mitigation strategy for this.

Ideally, our algorithm should be:

- Uniform. The result should be as uniform as possible assuming that the RNG’s underlying 28-bit integers are uniform.

- Uncorrelated. The result should not have detectable correlations between different seeds, similar to the lowest-bit ones mentioned earlier.

- Quick. The algorithm should be fast in \TeX, so no “bit twiddling”, but “digit twiddling” is ok.

- Simple. The behaviour must be documentable precisely.

- Predictable. The number of calls to the RNG should be the same for any \texttt{int-nn rand:nn}, because then the algorithm can be modified later without changing the result of other uses of the RNG.

- Robust. It should work even for \texttt{int_rand:nn \{ - \texttt{c_max_int} \} \{ \texttt{c_max_int} \}} where the range is not representable as an integer. In fact, we also provide later a floating-point \texttt{randint} whose range can go all the way up to $2 \times 10^{16} - 1$ possible values.

Some of these requirements conflict. For instance, uniformity cannot be achieved with a fixed number of calls to the RNG.

Denote by \texttt{random(N)} one call to \texttt{tex_uniformdeviate:D} with argument \texttt{N}, and by \texttt{ediv(p, q)} the $\varepsilon$-\TeX rounding division giving $[p/q + 1/2]$. Denote by $\langle \text{min} \rangle$, $\langle \text{max} \rangle$ and $R = \langle \text{max} \rangle - \langle \text{min} \rangle + 1$ the arguments of \texttt{int_min:nn} and the number of possible outcomes. Note that $R \in [1, 2^{12} - 1]$ cannot necessarily be represented as an integer (however, $R - 2^{31}$ can). Our strategy is to get two 28-bit integers $X$ and $Y$ from the RNG, split each into 14-bit integers, as $X = X_1 \times 2^{14} + X_0$ and $Y = Y_1 \times 2^{14} + Y_0$ then return essentially $\langle \text{min} \rangle + \lfloor R(X_1 \times 2^{-14} + Y_1 \times 2^{-28} + Y_0 \times 2^{-42} + X_0 \times 2^{-56}) \rfloor$. For small $R$ the $X_0$ term has a tiny effect so we ignore it and we can compute $R \times Y/2^{28}$ much more directly by \texttt{random(R)}.

- If $R \leq 2^{17} - 1$ then return $\text{ediv}(R\text{random}(2^{14}) + \text{random}(R) + 2^{13}, 2^{14}) - 1 + \langle \text{min} \rangle$. The shifts by $2^{13}$ and $-1$ convert $\varepsilon$-\TeX division to truncated division. The bound on $R$ ensures that the number obtained after the shift is less than $\texttt{c_max_int}$. The non-uniformity is at most of order $2^{17}/2^{42} = 2^{-25}$.

- Split $R = R_2 \times 2^{28} + R_1 \times 2^{14} + R_0$, where $R_2 \in [0, 15]$. Compute $\langle \text{min} \rangle + R_2 X_1 \times 2^{14} + (R_2 Y_1 + R_1 X_1) + \text{ediv}(R_2 Y_0 + R_1 Y_1 + R_0 X_1 + \text{ediv}(R_2 X_0 + R_0 Y_1 + \text{ediv}((2^{14} R_1 + R_0)(2^{14} Y_0 + X_0), 2^{28}), 2^{14}), 2^{14})$ then map a result of $\langle \text{max} \rangle + 1$ to $\langle \text{min} \rangle$. Writing each \texttt{ediv} in terms of truncated division with a shift, and using $[(p + [r/s])/q] = [(ps + r)/(sq)]$, what we compute is equal to $\lfloor (\text{exact}) + 2^{-28} + 2^{-15} + 2^{-1} \rfloor$ with $\langle \text{exact} \rangle = \langle \text{min} \rangle + R \times 0.X_1 Y_1 Y_0 X_0$. Given we map $\langle \text{max} \rangle + 1$ to $\langle \text{min} \rangle$, the shift has no effect on uniformity. The non-uniformity is bounded by $R/2^{56} < 2^{-24}$. It may be possible to speed up the code by dropping tiny terms such as $R_0 X_0$, but the analysis of non-uniformity proves too difficult.

To avoid the overflow when the computation yields $\langle \text{max} \rangle + 1$ with $\langle \text{max} \rangle = 2^{31} - 1$ (note that $R$ is then arbitrary), we compute the result in two pieces. Compute
(first) = ⟨min⟩ + R₂X₁^{2^{14}} if R₂ < 8 or (⟨min⟩ + 8X₁^{2^{14}} + (R₂ - 8)X₁^{2^{14}} if R₂ ≥ 8, the expressions being chosen to avoid overflow. Compute (second) = R₀Y₁ + R₁X₁ + ediv(…), at most R₂^{2^{14}} + R₀^{2^{14}} + R₀ ≤ 2^{28} + 15 × 2^{14} − 1, not at risk of overflowing. We have (first) + (second) = (max) + 1 = (min) + R if and only if (second) = R₂^{2^{14}} + R₀ + R₂^{2^{14}} and 2^{14}R₂X₁ = 2^{28}R₂ − 2^{14}R₂ (namely R₂ = 0 or X₁ = 2^{14} − 1). In that case, return (min), otherwise return (first) + (second), which is safe because it is at most (max). Note that the decision of what to return does not need (first) explicitly so we don’t actually compute it, just put it in an integer expression in which (second) is eventually added (or not).

- To get a floating point number in [0, 1) just call the R = 10000 ≤ 2^{17} − 1 procedure above to produce four blocks of four digits.
- To get an integer floating point number in a range (whose size can be up to 2 × 10^{16} − 1), work with fixed-point numbers: get six times four digits to build a fixed point number, multiply by R and add ⟨min⟩. This requires some care because l3fp-extended only supports non-negative numbers.

\c_kernel_randint_max_int

Constant equal to 2^{17} − 1, the maximal size of a range that \int_range:nn can do with its “simple” algorithm.

\cs_new:Npn \c__kernel_randint_max_int { 131071 }

(End definition for \c__kernel_randint_max_int.)

\_kernel_randint:n

Used in an integer expression, \_kernel_randint:n \{R\} gives a random number 1 + [(R \text{random}(2^{14}) + \text{random}(R))/2^{14}] that is in [1, R]. Previous code was computing \(p/2^{14}\) as ediv(p − 2^{13}, 2^{14}) but that wrongly gives −1 for p = 0.

\cs_new:Npn \_kernel_randint:n #1
\{ #1 \* \tex_uniformdeviate:D 16384
+ \tex_uniformdeviate:D #1 + 8192 ) / 16384
\}

(End definition for \_kernel_randint:n.)

\_fp_rand_myriads:n
\_fp_rand_myriads_loop:w
\_fp_rand_myriads_get:w

Used as \_fp_rand_myriads:n \{XXX\} with one letter X (specifically) per block of four digit we want; it expands to ; followed by the requested number of brace groups, each containing four (pseudo-random) digits. Digits are produced as a random number in [10000, 19999] for the usual reason of preserving leading zeros.
77.2 Random floating point

First we check that \texttt{random} was called without argument. Then get four blocks of four digits and convert that fixed point number to a floating point number (this correctly sets the exponent). This has a minor bug: if all of the random numbers are zero then the result is correctly 0 but it raises the \texttt{underflow} flag; it should not do that.

```latex
\texttt{\cs_new:Npn \__fp_rand_o:Nw ? #1 \@}
\{\tl_if_empty:nTF {#1}{\exp_after:wN \__fp_rand_o:w \exp:w \exp_end_continue_f:w}
\__fp_rand_myriads:n \{ XXXX \} \{ 0000 \} \{ 0000 \}; 0}
\}
\msg_expandable_error:nnnnn
\{ \texttt{fp} \}
\{ \texttt{num-args} \}
\{ \texttt{rand()} \}
\{ 0 \}
\{ 0 \}
\exp_after:wN \c_nan_fp
\}
```

(End definition for \__fp_rand_o:Nw and \__fp_rand_o:w.)

77.3 Random integer

Enforce that there is one argument (then add first argument 1) or two arguments. Call \__fp_randint_badarg:w on each; this function inserts 1 \exp_stop_f: to end the \exp_after:wN statement if either the argument is not an integer or if its absolute value is $\geq 10^{16}$. Also bail out if \__fp_compare_back:ww yields 1, meaning that the bounds are not in the right order. Otherwise an auxiliary converts each argument times $10^{-16}$ (hence the shift in exponent) to a 24-digit fixed point number (see \texttt{l3fp-extended}). Then compute the number of choices, $\langle \text{max} \rangle + 1 - \langle \text{min} \rangle$. Create a random 24-digit fixed-point number with \__fp_randint_myriads:n, then use a fused multiply-add instruction to multiply the number of choices to that random number and add it to $\langle \text{min} \rangle$. Then truncate to 16 digits (namely select the integer part of $10^{16}$ times the result) before converting back to a floating point number (\__fp_sanitize:Nw takes care of zero). To avoid issues with negative numbers, add 1 to all fixed point numbers (namely $10^{16}$ to the integers they represent), except of course when it is time to convert back to a float.

```latex
\texttt{\cs_new:Npn \__fp_randint_o:Nw ? \@}
\{\__fp_parse_function_one_two:nnw
\{ \texttt{randint} \}
\}
\exp_after:wN \__fp_randint_default:w \__fp_randint_o:w \}
\cs_new:Npn \__fp_randint_default:w #1 { \exp_after:wN #1 \c_one_fp }
```
\begin{verbatim}
\end{verbatim}

\texttt{\textbackslash int\_rand:nn} Evaluate the argument and filter out the case where the lower bound \texttt{#1} is more than the upper bound \texttt{#2}. Then determine whether the range is narrower than \texttt{\c__kernel\_randint\_max\_int}; \texttt{#2-#1} may overflow for very large positive \texttt{#2} and negative \texttt{#1}. If the
range is narrow, call \( \texttt{\_kernel_randint:n} \langle \texttt{choices} \rangle \) where \( \langle \texttt{choices} \rangle \) is the number of possible outcomes. If the range is wide, use somewhat slower code.

\begin{verbatim}
\cs_new:Npn \int_rand:nn #1#2
{\int_eval:n
 {\exp_after:wN \__fp_randint:ww
 \int_value:w \int_eval:n {#1} \exp_after:wN ;
 \int_value:w \int_eval:n {#2} ;}
}
\cs_new:Npn \__fp_randint:ww #1; #2;
{\if_int_compare:w #1 > #2 \exp_stop_f:
 \msg_expandable_error:nnnn
 { kernel } { randint-backward-range } {#1} {#2}
 \__fp_randint:ww #2; #1;
 \else:
 \if_int_compare:w \__fp_int_eval:w #2
 \if_int_compare:w #1 > \c_zero_int
 - #1 < \__fp_int_eval:w #1 +
 \fi:
 \c__kernel_randint_max_int
 \__fp_int_eval_end:
 \__kernel_randint:n
 {\__fp_int_eval:w #2 - #1 + 1 \__fp_int_eval_end: }
 - 1 + #1
 \else:
 \__kernel_randint:nn {#1} {#2}
 \fi:
 \fi:
}
\end{verbatim}

(End definition for \texttt{\int_rand:nn} and \texttt{\_fp_randint:ww}. This function is documented on page 165.)

Any \( n \in [-2^{31} + 1, 2^{31} - 1] \) is uniquely written as \( 2^{14} n_1 + n_2 \) with \( n_1 \in [-2^{17}, 2^{17} - 1] \) and \( n_2 \in [0, 2^{14} - 1] \). Calling \( \__fp_randint_split_o:Nw \) \( n \); gives \( n_1; n_2 \); and expands the next token once. We do this for two random numbers and apply \( \__fp_randint_split_o:Nw \) twice to fully decompose the range \( R \). One subtlety is that we compute \( R - 2^{31} = \langle \max \rangle - \langle \min \rangle - (2^{31} - 1) \in [-2^{31} + 1, 2^{31} - 1] \) rather than \( R \) to avoid overflow.

Then we have \( \__fp_randint_wide_aux:w \langle X_1 \rangle; \langle X_0 \rangle; \langle Y_1 \rangle; \langle Y_0 \rangle; \langle R_2 \rangle; \langle R_1 \rangle; \langle R_0 \rangle ; \) and we apply the algorithm described earlier.

\begin{verbatim}
\cs_new:Npn \__kernel_randint:nn #1#2
{\__fp_randint_split_o:Nw \__fp_randint_split_aux:w \__fp_randint_wide_aux:w}
\end{verbatim}

1177
\text{uniformdeviate:D} 268435456 ;
\exp_after:wN \_fp_randint_split_o:Nw
\int_value:w \_fp_int_eval:w 131072 +
\exp_after:wN \_fp_randint_split_o:Nw
\int_value:w
\_kernel_int_add:nnn \{#2\} \{-#1\} \{-\c_max_int\};
.
\cs_new:Npn \_fp_randint_split_o:Nw #1#2 ;
\if_meaning:w 0 #1
0 \exp_after:wN ; \int_value:w 0
\else:
\exp_after:wN \_fp_randint_split_aux:w
\int_value:w \_fp_int_eval:w (#1#2 - 8192) / 16384 ;
+ #1#2
\fi:
\exp_after:wN ;
\cs_new:Npn \_fp_randint_split_aux:w #1 ;
#1 \exp_after:wN ;
\int_value:w \_fp_int_eval:w - #1 * 16384
\cs_new:Npn \_fp_randint_wide_aux:w #1;#2; #3;#4; #5;#6;#7; .
\exp_after:wN \_fp_randint_wide_auxii:w
\int_value:w \_fp_int_eval:w #5 * #3 + #6 * #1 +
(#5 * #4 + #6 * #3 + #7 * #1 +
(#5 * #2 + #7 * #3 +
(16384 * #6 + #7) * (16384 * #4 + #2) / 268435456) / 16384
) / 16384 \exp_after:wN ;
\int_value:w \_fp_int_eval:w (#5 + #6) * 16384 + #7 ;
#1 ; #5 ;
\cs_new:Npn \_fp_randint_wide_auxii:w #1; #2; #3; #4;
\if_int_odd:w 0
\if_int_compare:w #1 = #2 \else: \exp_stop_f: \fi:
\if_int_compare:w #4 = \c_zero_int 1 \fi:
\if_int_compare:w #3 = 16383 - 1 \fi:
\exp_stop_f:
\exp_after:wN \prg_break:
\fi:
\if_int_compare:w #4 < 8 \exp_stop_f:
+ #4 * #3 * 16384
\else:
+ 8 * #3 * 16384 + (#4 - 8) * #3 * 16384
\fi:
+ #1
\prg_break_point:

(End definition for \_kernel_randint:nn and others.)

1178
\int_rand:n
\__fp_randint:n Similar to \int_rand:nn, but needs fewer checks.
\cs_new:Npn \int_rand:n #1
\{ \int_eval:n
\{ \exp_args:Nf \__fp_randint:n \{ \int_eval:n \{#1\} \} \}
\}
\cs_new:Npn \__fp_randint:n #1
\{ \if_int_compare:w #1 < \c_one_int
\msg_expandable_error:nnnn
\{ kernel \} \{ randint-backward-range \} \{ 1 \} \{#1\}
\__fp_randint:ww #1; 1;
\else:\if_int_compare:w #1 > \c__kernel_randint_max_int
\__kernel_randint:nn \{ 1 \} \{#1\}
\else:
\__kernel_randint:n \{#1\}
\fi:
\fi:
\}
\}
\)
\(\text{End definition for \int_rand:n and \__fp_randint:n. This function is documented on page 165.}\)
\End the initial conditional that ensures these commands are only defined in engines that support random numbers.
\}
\)
\)
\)

1179
Chapter 78

\textbf{l3fparray implementation}

In analogy to \texttt{l3intarray} it would make sense to have \texttt{<@=fparray>}, but we need direct access to \texttt{\_\_fp\_parse:n} from \texttt{l3fp\_parse}, and a few other (less crucial) internals of the \texttt{l3fp} family.

78.1 Allocating arrays

There are somewhat more than \((2^{31} - 1)^2\) floating point numbers so we store each floating point number as three entries in integer arrays. To avoid having to multiply indices by three or to add 1 etc, a floating point array is just a token list consisting of three tokens: integer arrays of the same size.

\texttt{\_\_fp\_array\_int} Used to generate unique names for the three integer arrays.

\texttt{\_\_fp\_array\_loop\_int} Used to loop in \texttt{\_\_fp\_array\_gzero:N}.

\texttt{\_fp\_array\_new:Nn} \texttt{\_fp\_array\_new:cn} \texttt{\_fp\_array\_new:nNNN} Build a three-token token list, then define all three tokens to be integer arrays of the same size. No need to initialize the data: the integer arrays start with zeros, and three zeros denote precisely \texttt{\_\_c\_zero\_fp}, as we want.
\cs_new_protected:Npn \__fp_array_new:nnNNN #1#2#3#4#5
{
\int_compare:nNnTF {#1} < 0
{
\msg_error:nnn { kernel } { negative-array-size } {#1}
\cs_undefine:N #1
\int_gsub:Nn \g__fp_array_int { 3 }
}
{
\intarray_new:Nn #2 {#1}
\intarray_new:Nn #3 {#1}
\intarray_new:Nn #4 {#1}
}
}
\__fp_array_bounds:NNnTF\__fp_array_bounds_error:NNn
\fparray_count:N\fparray_count:c
\fparray_gset:Nnn\fparray_item:Nn
\cs_new:Npn \__fp_array_bounds:NNnTF #1#2#3#4#5
{
\if_int_compare:w 1 > #3 \exp_stop_f:
\__fp_array_bounds_error:NNn #1 {#2} {#3}
\else:
\if_int_compare:w #3 > \fparray_count:N #2 \exp_stop_f:
\__fp_array_bounds_error:NNn #1 {#2} {#3}
\else:
#4
\fi:
\fi:
\fi
\cs_new:Npn \__fp_array_bounds_error:NNn #1#2#3
{
#1 { kernel } { out-of-bounds }
{ \token_to_str:N #2 } {#3} { \fparray_count:N #2 }
}

78.2 Array items

See the \l3intarray analogue: only names change. The functions \fparray_gset:Nnn and \fparray_item:Nn share bounds checking. The T branch is used if #3 is within bounds of the array #2.

\cs_new:Npn \__fp_array_bounds:NNnTF \__fp_array_bounds_error:NNn
\cs_generate_variant:Nn \fparray_new:Nn { c }
\cs_new_protected:Npn \__fp_array_new:nNNNN #1#2#3#4#5#6
{
\int_compare:nNnTF {#1} < 0
{
\msg_error:nnn { kernel } { negative-array-size } {#1}
\cs_undefine:N #1
\int_gsub:Nn \g__fp_array_int { 3 }
}
{
\intarray_new:Nn #2 {#1}
\intarray_new:Nn #3 {#1}
\intarray_new:Nn #4 {#1}
\intarray_new:Nn #6 {#1}
}

\fparray_count:N\fparray_count:c
Size of any of the intarrays, here we pick the third.

(End definition for \fparray_new:Nn and \__fp_array_new:nNNNN. This function is documented on page 260.)

(End definition for \fparray_count:N. This function is documented on page 260.)
Evaluate, then store exponent in one intarray, sign and 8 digits of mantissa in the next, and 8 trailing digits in the last.

\(\text{\texttt{\_fp_array_gset:Nnn}}\)

\(\text{\texttt{\_fp_array_gset:ccn}}\)

\(\text{\texttt{\_fp_array_gset:NNNNw}}\)

\(\text{\texttt{\_fp_array_gset_recover:Nw}}\)

\(\text{\texttt{\_fp_array_gset_special:nnNN}}\)

\(\text{\texttt{\_fp_array_gset_normal:w}}\)
End definition for \texttt{\textbackslash fparray_gzero:N} and others. This function is documented on page 260.

\begin{verbatim}
\cs_new_protected:Npn \fparray_gzero:N #1
    { \int_zero:N \l__fp_array_loop_int
        \prg_replicate:nn { \fparray_count:N #1 } 
        { \int_incr:N \l__fp_array_loop_int
            \exp_after:wN \__fp_array_gset_special:nnNNN
            \exp_after:wN 0
            \exp_after:wN \l__fp_array_loop_int
            #1 } }
\cs_generate_variant:Nn \fparray_gzero:N { c }
\end{verbatim}

(End definition for \texttt{\textbackslash fparray_gzero:N}. This function is documented on page 260.)

\begin{verbatim}
\cs_new:Npn \fparray_item:Nn #1#2
    { \exp_after:wN \__fp_array_item:NwN
        \exp_after:wN #1
        \int_value:w \int_eval:n {#2} ;
        \__fp_to_decimal:w }
\cs_generate_variant:Nn \fparray_item:Nn { c }
\cs_new:Npn \fparray_item_to_tl:Nn #1#2
    { \exp_after:wN \__fp_array_item:NwN
        \exp_after:wN #1
        \int_value:w \int_eval:n {#2} ;
        \__fp_to_tl:w }
\cs_generate_variant:Nn \fparray_item_to_tl:Nn { c }
\cs_new:Npn \__fp_array_item:NwN #1#2 ; #3
    { \__fp_array_bounds:NNnTF #1 {#2} { \exp_after:wN \__fp_array_item:NNNnN #1 {#2} #3 } { \exp_after:wN #3 \c_nan_fp } }
\cs_new:Npn \__fp_array_item:NNNnN #1#2#3#4
    { \exp_after:wN \__fp_array_item:N
        \int_value:w \__kernel_intarray_item:Nn #2 {#4} \exp_after:wN ;
        \int_value:w \__kernel_intarray_item:Nn #3 {#4} \exp_after:wN ;
        \int_value:w \__kernel_intarray_item:Nn #1 {#4} ;
    }
\cs_new:Npn \__fp_array_item:N #1
    { \if_meaning:w 0 #1 \exp_after:wN \__fp_array_item_special:w \fi:
        \__fp_array_item:w #1 }
\end{verbatim}

1183
\cs_new:Npn \_fp_array_item:w #1 #2#3#4#5 #6 ; 1 #7 ;
{
\exp_after:wN \_fp_array_item_normal:w
\int_value:w \if_meaning:w #1 1 0 \else: 2 \fi: \exp_stop_f:
#7 ; {#2#3#4#5} {#6} ;
}
\cs_new:Npn \_fp_array_item_special:w #1 ; #2 ; #3 ; #4
{
\exp_after:wN #4
\exp:w \exp_end_continue_f:w
\if_case:w #3 \exp_stop_f:
\exp_after:wN \c_zero_fp
\or: \exp_after:wN \c_nan_fp
\or: \exp_after:wN \c_minus_zero_fp
\or: \exp_after:wN \c_inf_fp
\else: \exp_after:wN \c_minus_inf_fp
\fi:
}
\cs_new:Npn \_fp_array_item_normal:w #1 #2#3#4#5 #6 ; #7 ; #8 ; #9
{ #9 \_fp \_fp_chk:w 1 #1 {#8} {#7} {#2#3#4#5} {#6} ; }

(End definition for \fparray_item:Nn and others. These functions are documented on page 260.)
Chapter 79

\13cctab implementation

As \LaTeX{} offers engine support for category code tables, and this is entirely lacking from the other engines, we need two complementary approaches. (Some future Xe\LaTeX{} may add support, at which point the conditionals below would be different.)

79.1 Variables

\begin{itemize}
\item \texttt{\g__cctab_stack_seq}
  List of catcode tables saved by nested \cctabbegin:N, to restore catcodes at the matching \cctabend:. When popped from the \texttt{\g__cctab_stack_seq} the table numbers are stored in \texttt{\g__cctab_unused_seq} for later reuse.
\item \texttt{\g__cctab_group_seq}
  A stack to store the group level when a catcode table started.
\item \texttt{\g__cctab_allocate_int}
  Integer to keep track of what category code table to allocate. In \LaTeX{} it is only used in format mode to implement \texttt{\cctabnew:N}. In other engines it is used to make csnames for dynamic tables.
\item \texttt{\l__cctab_internal_a_tl} \texttt{\l__cctab_internal_b_tl}
  Scratch space. For instance, when popping \texttt{\g__cctab_stack_seq/\g__cctab_unused_seq}, consists of the catcodetable number (integer denotation) in \LaTeX{}, or of an intarray variable (as a single token) in other engines.
\end{itemize}

\begin{verbatim}
\seq_new:N \g__cctab_stack_seq
\seq_new:N \g__cctab_unused_seq
\seq_new:N \g__cctab_group_seq
\int_new:N \g__cctab_allocate_int
\tl_new:N \l__cctab_internal_a_tl
\tl_new:N \l__cctab_internal_b_tl
\end{verbatim}
In LuaTeX we store the \endlinechar associated to each \catcodetable in a property list, unless it is the default value 13.

\prop_new:N \g__cctab_endlinechar_prop

(End definition for \g__cctab_endlinechar_prop.)

79.2 Allocating category code tables

The \cctab_new:N auxiliary allocates a new catcode table but does not attempt to set its value consistently across engines. It is used both in \cctab_new:N, which sets catcodes to \iniTeX values, and in \cctab_begin:N/\cctab_end: for dynamically allocated tables.

First, the LuaTeX case. Creating a new category code table is done like other registers. In ConTeXt, \newcatcodetable does not include the initialisation, so that is added explicitly.

\sys_if_engine_luatex:TF
\cs_new_protected:Npn \cctab_new:N #1
\__kernel_chk_if_free_cs:N #1
\__cctab_new:N #1
\cs_new_protected:Npn \__cctab_new:N #1
\newcatcodetable #1
\tex_initcatcodetable:D #1
\sys_if_engine_luatex:FN

Now the case for other engines. Here, each table is an integer array. Following the LuaTeX pattern, a new table starts with \iniTeX codes. The index base is out-by-one, so we have an internal function to handle that. The \iniTeX \endlinechar is 13.

\sys_if_engine_luatex:TF
\cs_new_protected:Npn \cctab_new:N #1
\__kernel_chk_if_free_cs:N #1
\__cctab_new:N #1
\int_step_inline:nn { 256 }
\{ \__cctab_gstore:Nnn #1 {##1} { 12 } \}
\int_step_inline:nnn { 65 } { 90 }
\{ \__cctab_gstore:Nnn #1 {##1} { 11 } \}
\__cctab_gstore:Nnn #1 { 92 } { 0 }
\int_step_inline:nnn { 97 } { 122 }
\{ \__cctab_gstore:Nnn #1 {##1} { 11 } \}
\__cctab_gstore:Nnn #1 { 127 } { 15 }
\sys_if_engine_luatex:FN
79.3 Saving category code tables

In various functions we need to save the current catcodes (globally) in a table. In LuaTeX, saving the catcodes is a primitives, but the \endlinechar needs more work: to avoid filling \g__cctab_endlinechar_prop with many entries we special-case the default value 13. In other engines we store 256 current catcodes and the \endlinechar in an intarray variable.

\cs_new_protected:Npn \__cctab_gset:n #1
\{ \exp_args:Nf \__cctab_gset_aux:n \{ \int_eval:n \{#1\} \} \}

\cs_new_protected:Npn \__cctab_gset_aux:n #1
\{ \tex_savecatcodetable:D #1 \scan_stop:
\int_compare:nNnTF { \tex_endlinechar:D } = { 13 }
\{ \prop_gremove:Nn \g__cctab_endlinechar_prop \{#1\} \}
\{ \prop_gput:NnV \g__cctab_endlinechar_prop \{#1\} \tex_endlinechar:D \}
\}

\cs_new_protected:Npn \cctab_gset:Nn \cctab_gset:cn
\{ \__cctab_chk_if_valid:NT #1
\group_begin:
\cctab_select:N \c_initex_cctab
\}

\cs_new_protected:Npn \__cctab_gset:n #1 #2
\{ \__cctab_chk_if_valid:NT #1
\group_begin:
\cctab_gset:cn \cctab_gset:Nn \{\#1\} \{\#2\}
\}

\cs_new_protected:Npn \__cctab_gset_aux:n #1
\{ \int_step_inline:nn { 256 }
\__kernel_intarray_gset:Nnn #1 {##1} \{ \char_value_catcode:n {##1 - 1}\}
\}

\cs_new_protected:Npn \_\_cctab_gset:n \_\_cctab_gset_aux:n
\{ \cs_generate_variant:Nn \cctab_new:N { c } \}

(End definition for \cctab_new:N, \_\_cctab_new:N, and \_\_cctab_gstore:Nn. This function is documented on page 261.)

Category code tables are always global, so only one version of assignments is needed. Simply run the setup in a group and save the result in a category code table \#1, provided it is valid. The internal function is defined above depending on the engine.

\cs_new_protected:Npn \cctab_gset:Nn \cctab_gset:cn
\{ \__cctab_chk_if_valid:NT \#1
\group_begin:
\cctab_gset:cn \cctab_gset:Nn \c_initex_cctab
\}

1187
79.4 Using category code tables

In \L{}\TeX{}, we must ensure that the saved tables are read-only. This is done by applying the saved table, then switching immediately to a scratch table. Any later catcode assignment will affect that scratch table rather than the saved one. If we simply switched to the saved tables, then \texttt{\stringchar_set_catcode_other:N} in the example below would change \texttt{\stringc_document_cctab} and a later use of that table would give the wrong category code to \_,.

\begin{verbatim}
\use:n {
\cctab_begin:N \c_document_cctab
  \char_set_catcode_other:N \_ \
\cctab_end:
\cctab_begin:N \c_document_cctab
\int_compare:nTF { \char_value_catcode:n { \_ } = 8 } {
  \TRUE }
\ERROR
\cctab_end:
}
\end{verbatim}

We must also make sure that a scratch table is never reused in a nested group: in the following example, the scratch table used by the first \texttt{\cctab_begin:N} would be changed globally by the second one issuing \texttt{\savecatcodetable}, and after \texttt{\group_end:} the wrong category codes (those of \texttt{\c_str_cctab}) would be imposed. Note that the inner \texttt{\cctab_end:} restores the correct catcodes only locally, so the problem really comes up because of the different grouping level. The simplest is to use a scratch table labeled by the \texttt{\currentgrouplevel}. We initialize one of them as an example.

\begin{verbatim}
\use:n {
\cctab_begin:N \c_document_cctab
  \group_begin:
  \cctab_begin:N \c_str_cctab
  \cctab_end:
  \group_end:
  \cctab_end:
}
\sys_if_engine_luatex:T
{
  \__cctab_new:N \g__cctab_internal_cctab
  \cs_new:Npn \__cctab_internal_cctab_name:
{ }
}
\end{verbatim}

(End definition for \texttt{\cctab_gset:Nn}. This function is documented on page 261.)
The public function simply checks the \texttt{cctab var} exists before using the engine-dependent \texttt{\_cctab_select:N}. Skipping these checks would result in low-level engine-dependent errors. First, the \LaTeX{} case. In other engines, selecting a catcode table is a matter of doing 256 catcode assignments and setting the \texttt{endlinechar}.

\begin{verbatim}
g__cctab_internal \tex_romannumeral:D \tex_currentgrouplevel:D _cctab \end{verbatim}

\texttt{\_cctab select:N}

\texttt{\_cctab select:c}

\texttt{\_cctab select:N}

(End definition for \texttt{\_cctab internal \_cctab} and \texttt{\_cctab internal \_cctab name}.)

\texttt{\_cctab select:N}

\texttt{\_cctab select:c}

\texttt{\_cctab select:N}

The public function simply checks the \texttt{cctab var} exists before using the engine-dependent \texttt{\_cctab select:N}. Skipping these checks would result in low-level engine-dependent errors. First, the \LaTeX{} case. In other engines, selecting a catcode table is a matter of doing 256 catcode assignments and setting the \texttt{endlinechar}.

\begin{verbatim}
cctab select:N

cctab select:c
\__cctab select:N
\end{verbatim}

The public function simply checks the \texttt{cctab var} exists before using the engine-dependent \texttt{\_cctab select:N}. Skipping these checks would result in low-level engine-dependent errors. First, the \LaTeX{} case. In other engines, selecting a catcode table is a matter of doing 256 catcode assignments and setting the \texttt{endlinechar}.

\begin{verbatim}
cctab select:N

cctab select:c
\__cctab select:N
\end{verbatim}

(End definition for \texttt{\_cctab internal \_cctab} and \texttt{\_cctab internal \_cctab name}. This function is documented on page \texttt{262}.)
\cs_new_protected:Npn \__cctab_begin_aux:
\{
  \__cctab_new:N \g__cctab_next_cctab
  \tl_set:NV \l__cctab_internal_a_tl \g__cctab_next_cctab
  \cs_undefine:N \g__cctab_next_cctab
\}

\cs_new_protected:Npn \__cctab_begin_aux:
\{
  \int_gincr:N \g__cctab_allocate_int
  \exp_args:Nc \__cctab_new:N { \g__cctab_ \int_use:N \g__cctab_allocate_int _cctab }
  \exp_args:NNc \tl_set:Nn \l__cctab_internal_a_tl { \g__cctab_ \int_use:N \g__cctab_allocate_int _cctab }
\}

(End definition for \g__cctab_next_cctab and \__cctab_begin_aux.)

\cctab_begin:N \cctab_begin:c
Check the \textit{cctab var} exists, to avoid low-level errors. Get in \l__cctab_internal_a_tl the number/name of a dynamic table, either from \g__cctab_unused_seq where we save tables that are not currently in use, or from \__cctab_begin_aux: if none are available. Then save the current catcodes into the table (pointed to by) \l__cctab_internal_a_tl and save that table number in a stack before selecting the desired catcodes.

\cs_new_protected:Npn \cctab_begin:N #1
\{
  \__cctab_chk_if_valid:NT #1
  { \seq_gpop:NNF \g__cctab_unused_seq \l__cctab_internal_a_tl
    { \__cctab_begin_aux: }
    \exp_args:Nx \__cctab_chk_group_begin:n { \__cctab_nesting_number:N \l__cctab_internal_a_tl }
    \seq_gpush:NV \g__cctab_stack_seq \l__cctab_internal_a_tl
    \exp_args:NV \__cctab_gset:n \l__cctab_internal_a_tl
    \__cctab_select:N #1
  }
\}
\cs_generate_variant:Nn \cctab_begin:N { c }

(End definition for \cctab_begin:N. This function is documented on page 261.)

\cctab_end: \cctab_end:
Make sure a \cctab_begin:N was used some time earlier, get in \l__cctab_internal_a_tl the code table number/name in which the prevailing catcodes were stored, then restore these catcodes. The dynamic table is now unused hence stored in \g__cctab_unused_seq for recycling by later \cctab_begin:N.

\cs_new_protected:Npn \cctab_end:
\{
  \seq_gpop:NTF \g__cctab_unused_seq \l__cctab_internal_a_tl
  { \__cctab_stack_seq \l__cctab_internal_a_tl
    \exp_args:Nx \__cctab_chk_group_end:n { \__cctab_nesting_number:N \l__cctab_internal_a_tl }
  }
\}

(End definition for \cctab_end:)
Catcode tables are not allowed to be intermixed with groups, so here we check that they are properly nested regarding \TeX groups. \_\_\_cctab chk group begin:n stores the current group level in a stack, and locally defines a dummy control sequence \_\_\_cctab - group{\langle cctab-level \rangle }_chk:
\_\_\_cctab chk group end:n pops the stack, and compares the returned value with \text currentgrouplevel:D. If they differ, \cctab end: is in a different grouping level than the matching \cctab begin:N. If they are the same, both happened at the same level, however a group might have ended and another started between \cctab begin:N and \cctab end::

\group_begin:
\cctab begin:N \c_document_cctab
\group_end:
\cctab_end:
\group_end:

In this case checking \text currentgrouplevel:D is not enough, so we locally define \_\_\_cctab group{\langle cctab-level \rangle }_chk:, and then check if it exist in \cctab end:. If it doesn’t, we know there was a group end where it shouldn’t.

The \langle cctab-level \rangle in the sentinel macro above cannot be replaced by the more convenient \text currentgrouplevel:D because with the latter we might be tricked. Suppose:

\group_begin:
\cctab begin:N \c_code_cctab % A
\group_end:
\group_begin:
\cctab begin:N \c_code_cctab % B
\cctab end: % C
\cctab end: % D
\group_end:

The line marked with A would start a cctab with a sentinel token named \_\_\_cctab - group_1 chk:, which would disappear at the \group_end: that follows. But B would create the same sentinel token, since both are at the same group level. Line C would end the cctab from line B correctly, but so would line D because line B created the same sentinel token. Using \langle ctab-level \rangle works correctly because it signals that certain ctab level was activated somewhere, but if it doesn’t exist when the \cctab_end: is reached, we had a problem.

Unfortunately these tests only flag the wrong usage at the \cctab_end:, which might be far from the \cctab begin:N. However it isn’t possible to signal the wrong usage at the \group_end: without using \text aftergroup:D, which is unsafe in certain types of groups.

The three cases checked here just raise an error, and no recovery is attempted: usually interleaving groups and catcode tables will work predictably.
\cs_new_protected:Npn \_cctab_chk_group_begin:n #1
\seq_gpush:Nx \g__cctab_group_seq \int_use:N \tex_currentgrouplevel:D \cs_set_eq:cN { __cctab_group_ #1 _chk: } \prg_do_nothing:
\cs_new_protected:Npn \_cctab_chk_group_end:n #1
\seq_gpop:NN \g__cctab_group_seq \l__cctab_internal_b_tl \bool_lazy_and:nnF
\int_compare_p:nNn { \tex_currentgrouplevel:D } = { \l__cctab_internal_b_tl }
\cs_if_exist_p:c { __cctab_group_ #1 _chk: } \msg_error:nnx { cctab } { group-mismatch }
\int_sign:n { \tex_currentgrouplevel:D - \l__cctab_internal_b_tl }
\cs_undefine:c { __cctab_group_ #1 _chk: }

(End definition for \_cctab_chk_group_begin:n and \_cctab_chk_group_end:n.)

\_cctab_nesting_number:N \_cctab_nesting_number:w

This macro returns the numeric index of the current catcode table. In Lua\TeX\ this is just the argument, which is a count reference to a \catcodetable register. In other engines, the number is extracted from the cctab variable.

\sys_if_engine_luatex:TF
\cs_new:Npn \_cctab_nesting_number:N #1 {#1}
\cs_new:Npn \_cctab_nesting_number:w
\exp_after:wN \exp_after:wN \exp_after:wN \_cctab_nesting_number:w
\token_to_str:N #1

(End definition for \_cctab_nesting_number:N and \_cctab_nesting_number:w.)

Finally, install some code at the end of the \TeX{} run to check that all \cctab_begin:N were ended by some \cctab_end:.

\cs_if_exist:NT \hook_gput_code:nnn
\hook_gput_code:nnn { enddocument/end } { cctab }
\seq_if_empty:NF \g__cctab_stack_seq
\msg_error:nn { cctab } { missing-end }

1192
\cctab_item:Nn
\cctab_item:cn

Evaluate the integer argument only once. In most engines the \cctab variable only has 256 entries so we only look up the catcode for these entries, otherwise we use the current catcode. In particular, for out-of-range values we use whatever fall-back \char_value_catcode:n. In \luatex, we use the \tex.getcatcode function.

\cs_new:Npn \cctab_item:Nn #1#2
{ \exp_args:Nf \__cctab_item:nN { \int_eval:n {#2} } #1 }
\sys_if_engine_luatex:TF
{ \cs_new:Npn \__cctab_item:nN #1#2
{ \lua_now:e { tex.print(-2, tex.getcatcode(\int_use:N #2, #1)) } } }
{ \cs_new:Npn \__cctab_item:nN #1#2
{ \int_compare:nNnTF {#1} < { 256 }
{ \intarray_item:Nn #2 { #1 + 1 } }
{ \char_value_catcode:n {#1} } } }
\cs_generate_variant:Nn \cctab_item:Nn { c }

79.5 Category code table conditionals

\cctab_if_exist_p:N\cctab_if_exist_p:c
\cctab_if_exist:NTF\cctab_if_exist:c\cctab_if_exist:c

Checks whether a \langle cctab var \rangle is defined.

\prg_new_eq_conditional:NNn \cctab_if_exist:N \cs_if_exist:N
{ TF , T , F , p }
\prg_new_eq_conditional:NNn \cctab_if_exist:c \cs_if_exist:c
{ TF , T , F , p }

(End definition for \cctab_if_exist:NTF. This function is documented on page 262.)

\__cctab_chk_if_valid:N\__cctab_chk_if_valid_aux:NTF

Checks whether the argument is defined and whether it is a valid \langle cctab var \rangle. In \luatex the validity of the \langle cctab var \rangle is checked by the engine, which complains if the argument is not a \chardef ed constant. In other engines, check if the given command is an intarray variable (the underlying definition is a copy of the cmr10 font).

\prg_new_protected_conditional:Npn \__cctab_chk_if_valid:N #1
{ TF , T , F }
\cctab_if_exist:NTF #1
{ \__cctab_chk_if_valid_aux:NTF #1
{ \prg_return_true: }
{ \msg_error:nnx { cctab } { invalid-cctab }
{ \token_to_str:N #1 }
\prg_return_false: }
}

1193
\sys_if_engine_luatex:TF
\cs_new_protected:Npn \__cctab_chk_if_valid_aux:NTF #1
{ \int_compare:nNnTF {#1-1} < { \e@alloc@ccodetable@count } }
\cs_if_exist:NT \c_syst_catcodes_n
{ \cs_gset_protected:Npn \__cctab_chk_if_valid_aux:NTF #1
{ \int_compare:nTF { #1 <= \c_syst_catcodes_n } }
}
{ \cs_new_protected:Npn \__cctab_chk_if_valid_aux:NTF #1
{ \exp_args:Nf \str_if_in:nnTF { \cs_meaning:N #1 } { select-font-cmr10-at- } }
}

(End definition for \__cctab_chk_if_valid:NTF and \__cctab_chk_if_valid_aux:NTF.)

\cctab_const:Nn
\cctab_const:cn
Creates a new ⟨cctab var⟩ then sets it with the current and user-supplied codes.
\cs_new_protected:Npn \cctab_const:Nn \cctab_const:Nn #1#2
{ \cctab_new:N #1 \cctab_gset:Nn #1 {#2} }
\cs_generate_variant:Nn \cctab_const:Nn { c }

(End definition for \cctab_const:Nn. This function is documented on page 261.)

\c_initex_cctab
\c_other_cctab
\c_str_cctab
Creating category code tables means thinking starting from iniTEX. For all-other and the standard “string” tables that’s easy.
\cctab_new:N \c_initex_cctab
\cctab_const:Nn \c_other_cctab
{ \cctab_select:N \c_initex_cctab
\int_set:Nn \tex_endlinechar:D { -1 }
\int_step_inline:nnn { 0 } { 127 }
{ \char_set_catcode_other:n {#1} }
}

79.6 Constant category code tables
To pick up document-level category codes, we need to delay setup to the end of the format, where that’s possible. Also, as there are a lot of category codes to set, we avoid using the official interface and store the document codes using internal code. Depending on whether we are in the hook or not, the catcodes may be code or document, so we explicitly set up both correctly.

```latex
\cs_if_exist:NTF \@expl@finalise@setup@@
{ \tl_gput_right:Nn \@expl@finalise@setup@@ }
{ \use:n }
{ \__cctab_new:N \c_code_cctab
  \group_begin:
  \int_set:Nn \tex_endlinechar:D { 32 }
  \bool_lazy_or:nnTF
  { \sys_if_engine_xetex_p: } { \sys_if_engine_luatex_p: }
  { \int_step_function:nN { 31 } \char_set_catcode_invalid:n }
  { \int_step_function:nN { 31 } \char_set_catcode_active:n }
  { \int_step_function:nnN { 33 } { 64 } \char_set_catcode_letter:n }
  { \int_step_function:nnN { 97 } { 122 } \char_set_catcode_letter:n }
  \char_set_catcode_ignore:n { 9 } % tab
  \char_set_catcode_other:n { 10 } % lf
  \char_set_catcode_active:n { 12 } % ff
  \char_set_catcode_end_line:n { 13 } % cr
  \char_set_catcode_ignore:n { 32 } % space
  \char_set_catcode_parameter:n { 35 } % hash
  \char_set_catcode_active:n { 36 } % dollar
  \char_set_catcode_comment:n { 37 } % percent
  \char_set_catcode_alignment:n { 38 } % ampersand
  \char_set_catcode_letter:n { 58 } % colon
  \char_set_catcode_escape:n { 92 } % backslash
  \char_set_catcode_math_superscript:n { 94 } % circumflex
  \char_set_catcode_letter:n { 95 } % underscore
  \char_set_catcode_group_begin:n { 123 } % left brace
  \char_set_catcode_other:n { 124 } % pipe
  \char_set_catcode_group_end:n { 125 } % right brace
  \char_set_catcode_space:n { 126 } % tilda
  \char_set_catcode_invalid:n { 127 } % ^^?
  \bool_lazy_or:nnF
  { \sys_if_engine_xetex_p: } { \sys_if_engine_luatex_p: }
  { \int_step_function:nnN { 128 } { 255 } \char_set_catcode_active:n }
  \__cctab_gset:n { \c_code_cctab }
  \group_end:
\__cctab_gset:n { \c_code_cctab }
\__cctab_gset:n { \c_document_cctab }
```

(End definition for \c_initex_cctab, \c_other_cctab, and \c_str_cctab. These variables are documented on page 262.)
79.7 Messages

\msg_new:nnnn { cctab } { stack-full }
{ The-category-code-table-stack-is-exhausted. }
\msg_new:nnnn { cctab } { extra-end }
{ Extra-\iow_char:N\cctab_end: ignored. }
\msg_new:nnnn { cctab } { missing-end }
{ Missing-\iow_char:N\cctab_end: before end of TeX run. }
\msg_new:nnnn { cctab } { invalid-cctab }
{ Invalid-\iow_char:N\catcode-table. }
\msg_new:nnnn { cctab } { group-mismatch }
{ Catcode-tables-and-groups-must-be-properly-nested, but-
you-tried-to-interleave-them.-LaTeX-will-try-to-proceed,-
but-results-may-be-unexpected.

\prop_gput:Nnn \g_msg_module_name_prop { cctab } { LaTeX3 }
\prop_gput:Nnn \g_msg_module_type_prop { cctab } { }

(/package)
Chapter 80

l3unicode implementation

Case changing both for strings and “text” requires data from the Unicode Consortium. Some of this is build in to the format (as \lc and \ucc values) but this covers only the simple one-to-one situations and does not fully handle for example case folding.

As only the data needs to remain at the end of this process, everything is set up inside a group. The only thing that is outside is creating a stream: they are global anyway and it is best to force a stream for all engines. For performance reasons, some of the code here is very low-level: the material is read during loading expl3 in package mode.

{\ior_new:N \g__char_data_ior}
{\bool_lazy_or:nnTF { \sys_if_engine_luatex_p: } { \sys_if_engine_xetex_p: } }
{\group_begin:}
Access the primitive but suppress further expansion: active chars are otherwise an issue.
{\cs_set:Npn \__char_generate_char:n #1}
{\exp_not:N \tex_unexpanded:D \exp_not:N \exp_after:wN
{\exp_not:N \tex_uchar:D ~ #1 ~}
\tex_catcode:D #1 ~}
A fast local implementation for generating characters; the chars may be active, so we prevent further expansion.
{\cs_set:Npx \__char_generate:n #1}
{\exp_not:N \tex_unexpanded:D \exp_not:N \exp_after:wN
{\exp_not:N \tex_uchar:D #1 ~}
\tex_catcode:D #1 ~}

Parse the main Unicode data file for two things. First, we want the titlecase exceptions: the one-to-one lower- and uppercase mappings it contains are all be covered by the \TeX data. Second, we need normalization data: at present, just the canonical NFD mappings. Those all yield either one or two codepoints, so the split is relatively easy.
{\ior_open:Nn \g__char_data_ior \UnicodeData.txt }
{\cs_set_protected:Npn \__char_data_auxi:w}
The other data files all use C-style comments so we have to worry about # tokens (and reading as strings). The set up for case folding is in two parts. For the basic (core) mappings, folding is the same as lower casing in most positions so only store the differences. For the more complex foldings, always store the result, splitting up the two or three code points in the input as required.
For upper- and lowercasing special situations, there is a bit more to do as we also have
title casing to consider, plus we need to stop part-way through the file.
\ior_open:Nn \g__char_data_ior { SpecialCasing.txt }
\cs_set_protected:Npn \__char_data_auxi:w #1 - #2 - #3 - #4 \q_stop
\tl_const:cx { c__char_foldcase_ \__char_generate_char:n {#1} _tl }
\__char_generate:n { "#2 } \__char_generate:n { "#3 } \tl_if_blank:nF {#4}
\__char_generate:n { \int_value:w "#4 } }
\ior_str_map_inline:Nn \g__char_data_ior
\reverse_if:N \if:w \c_hash_str \tl_head:w #1 \c_hash_str \q_stop
\__char_data_auxii:w #1 \c_hash_str \q_stop
\tl_if_empty:nF {#4}
\tl_const:cx { c__char_ #2 case_ \__char_generate_char:n {#1} _tl }
\__char_generate:n { "#3 }
\__char_generate:n { "#4 }
\tl_if_blank:nF {#5}
\__char_generate:n { "#5 } }
\ior_str_map_inline:Nn \g__char_data_ior
For the 8-bit engines, open the stream and close again: this keeps file records the same.

\{ 
\ior_open:Nn \g__char_data_ior { UnicodeData.txt }
\ior_close:N \g__char_data_ior
\}
Chapter 81

\textbf{l3text implementation}

81.1 Internal auxiliaries

\texttt{\_\_text\_stop} Internal scan marks.
\begin{verbatim}
\texttt{scan\_new:N \_\_text\_stop}
\end{verbatim}
(End definition for \texttt{\_\_text\_stop}).

\texttt{\_\_text\_nil} Internal quarks.
\begin{verbatim}
\texttt{quark\_new:N \_\_text\_nil}
\end{verbatim}
(End definition for \texttt{\_\_text\_nil}).

\texttt{\_\_text\_quark\_if\_nil\_p:n} Branching quark conditional.
\begin{verbatim}
\texttt{\_\_kernel\_quark\_new\_conditional:Nn \_\_text\_quark\_if\_nil\_p:n { TF }}
\end{verbatim}
(End definition for \texttt{\_\_text\_quark\_if\_nil\_p:nTF}).

\texttt{\_\_text\_recursion\_tail} \texttt{\_\_text\_recursion\_stop} Internal recursion quarks.
\begin{verbatim}
\texttt{quark\_new:N \_\_text\_recursion\_tail}
\texttt{quark\_new:N \_\_text\_recursion\_stop}
\end{verbatim}
(End definition for \texttt{\_\_text\_recursion\_tail} and \texttt{\_\_text\_recursion\_stop}).

\texttt{\_\_text\_use\_i\_delimit\_by\_q\_recursion\_stop:nw} Functions to gobble up to a quark.
\begin{verbatim}
\texttt{cs\_new:Npm \_\_text\_use\_i\_delimit\_by\_q\_recursion\_stop:nw #1 #2 \_\_text\_recursion\_stop \{#1\}}
\end{verbatim}
(End definition for \texttt{\_\_text\_use\_i\_delimit\_by\_q\_recursion\_stop:nw}).

\texttt{\_\_text\_if\_recursion\_tail\_stop\_do:Nn} Functions to query recursion quarks.
\begin{verbatim}
\texttt{\_\_kernel\_quark\_new\_test:N \_\_text\_if\_recursion\_tail\_stop\_do:Nn}
\end{verbatim}
(End definition for \texttt{\_\_text\_if\_recursion\_tail\_stop\_do:Nn}).
81.2 Utilities

The idea here is to take a token and ensure that if it’s an implicit char, we output the explicit version. Otherwise, the token needs to be unchanged. First, we have to split between control sequences and everything else.

\begin{verbatim}
\group_begin:
\char_set_catcode_active:n { 0 }
\cs_new:Npn \__text_token_to_explicit:N #1
{ \if_catcode:w \exp_not:N #1
  \if_catcode:w \scan_stop: \exp_not:N #1
  \scan_stop:
  \else:
    \exp_not:N ^^@
  \fi:
  \exp_after:wN \__text_token_to_explicit_cs:N
  \else:
  \exp_after:wN \__text_token_to_explicit_char:N
  \fi:
  \__text_token_to_explicit:n #1
}\group_end:
\end{verbatim}

For control sequences, we can check for macros versus other cases using \if_meaning:w, then explicitly check for \chardef and \mathchardef.

\begin{verbatim}
\cs_new:Npn \__text_token_to_explicit_cs:N #1
{ \exp_after:wN \if_meaning:w \exp_not:N #1 #1
  \exp_after:wN \use:nn \exp_after:wN
  \__text_token_to_explicit_cs_aux:N
  \else:
  \exp_after:wN \__text_token_to_explicit_char:N
  \fi:
  \__text_token_to_explicit:n #1
]\cs_new:Npn \__text_token_to_explicit_cs_aux:N #1
{ \bool_lazy_or:nnTF
  \token_if_chardef_p:N #1
  \token_if_mathchardef_p:N #1
  { \char_generate:nn {#1}
    { \char_value_catcode:n {#1} }
  }
  \__text_token_to_explicit:n #1
}\end{verbatim}

For character tokens, we need to filter out the implicit characters from those that are explicit. That’s done here, then if necessary we work out the category code and generate the char. To avoid issues with alignment tabs, that one is done by elimination rather than looking up the code explicitly. The trick with finding the charcode is that the \TeX messages are either the ⟨something⟩ character ⟨char⟩ or the ⟨type⟩ ⟨char⟩.

\begin{verbatim}
\cs_new:Npn \__text_token_to_explicit_char:N #1
{ \__text_token_to_explicit:n #1
}\end{verbatim}

1203
\__text_char_catcode:N

An idea from l3char: we need to get the category code of a specific token, not the general case.

\cs_new:Npn \__text_char_catcode:N #1

{ \if_catcode:w #1 \c_math_toggle_token 3 \else: \if_catcode:w #1 12 \else: 4 \fi: \fi: \fi: \fi: \fi: \fi: \fi: \fi: \fi: \exp_after:wN ; \token_to_meaning:N #1 \s__text_stop

(End definition for \__text_token_to_explicit:N and others.)
\textbf{Test for tokens that make sense to expand here: that is more restrictive than the engine view.}

\begin{verbatim}
\prg_new_conditional:Npnn \__text_if_expandable:N #1 { T , F , TF } {
  \token_if_protected_macro_p:N #1
  \token_if_protected_long_macro_p:N #1
  \token_if_eq_meaning_p:NN \q__text_recursion_tail #1
  \prg_return_false: }
\end{verbatim}
81.3 Configuration variables

\texttt{l\_text\_accents\_tl} Special cases for accents and letter-like symbols, which in some cases will need to be
converted further.

\begin{verbatim}
\tl_new:N \l_text_accents_tl
\tl_set:Nn \l_text_accents_tl { \^ \~ \= \u \. " \l \v \d \c \k \b \t }
\tl_new:N \l_text_letterlike_tl
\tl_set:Nn \l_text_letterlike_tl { \AA \aa \AE \ae \DH \dh \DJ \dj \IJ \ij \L \l \NG \ng \O \o \OE \oe \SS \ss \TH \th }
\end{verbatim}

(End definition for \texttt{l\_text\_accents\_tl} and \texttt{l\_text\_letterlike\_tl}. These variables are documented
on page 268.)

\texttt{l\_text\_case\_exclude\_arg\_tl} Non-text arguments.

\begin{verbatim}
\tl_new:N \l_text_case_exclude_arg_tl
\tl_set:Nn \l_text_case_exclude_arg_tl { \begin \cite \end \label \ref }
\end{verbatim}

(End definition for \texttt{l\_text\_case\_exclude\_arg\_tl}. This variable is documented on page 268.)

\texttt{l\_text\_math\_arg\_tl} Math mode as arguments.

\begin{verbatim}
\tl_new:N \l_text_math_arg_tl
\tl_set:Nn \l_text_math_arg_tl { \ensuremath }
\end{verbatim}

(End definition for \texttt{l\_text\_math\_arg\_tl}. This variable is documented on page 268.)

\texttt{l\_text\_math\_delims\_tl} Paired math mode delimiters.

\begin{verbatim}
\tl_new:N \l_text_math_delims_tl
\tl_set:Nn \l_text_math_delims_tl { $ $ \( \) }
\end{verbatim}

(End definition for \texttt{l\_text\_math\_delims\_tl}. This variable is documented on page 268.)

\texttt{l\_text\_expand\_exclude\_tl} Commands which need not to expand. We start with a somewhat historical list, and tidy
up if possible.

\begin{verbatim}
\tl_new:N \l_text_expand_exclude_tl
\tl_set:Nn \l_text_expand_exclude_tl { \begin \cite \end \label \ref }
\bool_lazy_and:nnT
{ \str_if_eq_p:Vn \fmtname { LaTeX2e } }
{ \tl_if_exist_p:N \@expl@finalise@setup@@ }
{ \tl_gput_right:Nn \@expl@finalise@setup@@ }
\tl_new:N \l_text_math_delims_tl
\tl_set:Nn \l_text_math_delims_tl { $ $ \( \) }
\end{verbatim}

1206
\tl_gput_right:Nn \@kernel@after@begindocument
{
  \group_begin:
  \cs_set_protected:Npn \__text_tmp:w #1
  {
    \tl_clear:N \l_text_expand_exclude_tl
    \tl_map_inline:nn {#1}
    {
      \bool_lazy_any:nF
      {
        \token_if_protected_macro_p:N \##1
      } \token_if_protected_long_macro_p:N \##1
      {
        \str_if_eq_p:ee
        { \cs_replacement_spec:N \##1 }
        { \exp_not:n {\protect \##1 } \c_space_tl }
      }
    }
    \tl_put_right:Nn \l_text_expand_exclude_tl {\##1}
  }
  \exp_args:NV \__text_tmp:w \l_text_expand_exclude_tl
  \exp_args:NNNV \group_end:
  \tl_set:Nn \l_text_expand_exclude_tl \l_text_expand_exclude_tl
}\group_end:

(End definition for \l_text_expand_exclude_tl. This variable is documented on page 268.)

\l__text_math_mode_tl

Used to control math mode output: internal as there is a dedicated setter.

\tl_new:N \l__text_math_mode_tl

(End definition for \l__text_math_mode_tl.)

81.4 Expansion to formatted text

Markers for implicit char handling.

\tex_chardef:D \c__text_chardef_space_token = \textbackslash %
\tex_chardef:D \c__text_mathchardef_space_token = \textbackslash %
\tex_chardef:D \c__text_chardef_group_begin_token = \{ \%\}
\tex_chardef:D \c__text_mathchardef_group_begin_token = \textbackslash \{ \% \} \textbackslash \{ \\
\tex_chardef:D \c__text_chardef_group_end_token = \textbackslash \}
\tex_chardef:D \c__text_mathchardef_group_end_token = \textbackslash \}

(End definition for \c__text_chardef_space_token and others.)

After precautions against \& tokens, start a simple loop: that of course means that “text” cannot contain the two recursion quarks. The loop here must be f-type expandable; we have arbitrary user commands which might be protected and take arguments, and if the expansion code is used in a typesetting context, that will otherwise explode. (The same issue applies more clearly to case changing: see the example there.)

\cs_new:Npn \text_expand:n \#1
The approach to making the code f-type expandable is to use a marker result token and to shuffle the collected tokens.

The main loop is a standard “tl action”; groups are handled recursively, while spaces are just passed through. Thus all of the action is in handling N-type tokens.
Before we get into the real work, we have to watch out for problematic implicit characters: spaces and grouping tokens. Converting these to explicit characters later would lead to real issues as they are not N-type. A space is the easy case, so it’s dealt with first: just insert the explicit token and continue the loop.

Implicit {/} offer two issues. First, the token could be an implicit brace character: we need to avoid turning that into a brace group, so filter out the cases manually. Then we handle the case where an implicit group is present. That is done in an “open-ended” way: there’s the possibility the closing token is hidden somewhere.
The first step in dealing with N-type tokens is to look for math mode material: that needs to be left alone. The starting function has to be split into two as we need \quark-if_recursion_tail_stop:N first before we can trigger the search. We then look for matching pairs of delimiters, allowing for the case where math mode starts but does not end. Within math mode, we simply pass all the tokens through unchanged, just checking the N-type ones against the end marker.

\cs_new:Npn \__text_expand_N_type_auxiii:N #1
\exp_after:wN \__text_expand_math_search:NNN
\exp_after:wN #1 \l_text_math_delims_tl
\q__text_recursion_tail \q__text_recursion_tail
\q__text_recursion_stop
\cs_new:Npn \__text_expand_math_search:NNN #1#2#3
\__text_if_recursion_tail_stop_do:Nn #2
\__text_expand_explicit:N #1
\token_if_eq_meaning:NNTF #1 #2
\__text_use_i_delimit_by_q_recursion_stop:nw
\__text_expand_store:n {#1}
\__text_expand_math_loop:Nw #3
\__text_expand_math_search:NNN #1
\cs_new:Npn \__text_expand_math_loop:Nw #1#2 \q__text_recursion_stop
\__text_if_head_is_N_type:NTF {#2}
\__text_expand_math_N_type:NN }
\__text_if_head_is_group:nTF {#2}
\__text_expand_math_group:Nn }
\__text_expand_math_space:Nw }
\text_recursion_stop
\text_recursion_stop
\text_recursion_stop
At this stage, either we have a control sequence or a simple character: split and handle.

Next we exclude math commands: this is mainly as there might be an \ensuremath.
Accents.

Another list of exceptions: these ones take no arguments so are easier to handle.
\LaTeX's \protect makes life interesting. Where possible, we simply remove it and replace with the “parent” command; of course, the \protect might be explicit, in which case we need to leave it alone. That includes the case where it’s not even followed by an N-type token. There is also the case of a straight \@protected@testopt to cover.

\begin{verbatim}
\cs_new:Npx \__text_expand_cs:N #1
\begin{Verbatim}
\exp_not:N \str_if_eq:nnTF {#1} \protect \__text_expand_protect:w
\begin{Verbatim}
\bool_lazy_and:nnTF \cs_if_exist_p:N \fmtname \str_if_eq_p:Vn \fmtname { LaTeX2e }
\__text_expand_testopt:N #1 \__text_expand_replace:N #1
\end{Verbatim}
\end{Verbatim}
\__text_expand_protect:w #1 \q__text_recursion_stop
\begin{Verbatim}
\tl_if_head_is_N_type:nTF {#1}
\__text_expand_protect:N \begin{Verbatim}
\__text_if_recursion_tail_stop_do:Nn #1 \protect \__text-expand_end:w
\begin{Verbatim}
\exp_args:Ne \__text_expand_protect:nN \cs_to_str:N #1 #1
\end{Verbatim}
\end{Verbatim}
\__text-expand_loop:w
\end{Verbatim}
\__text-expand_loop:w
\cs_new:Npx \__text-expand_testopt:N #1
\begin{Verbatim}
\token_if_eq_meaning:NNTF #1 \@protected@testopt \__text-expand_testopt:NNn
\begin{Verbatim}
\__text-expand_encoding:N #1
\end{Verbatim}
\end{Verbatim}
\end{Verbatim}
\end{verbatim}
Deal with encoding-specific commands

See if there is a dedicated replacement, and if there is, insert it.

Finally, expand any macros which can be: this then loops back around to deal with what they produce. The only issue is if the token is \exp_not:n, as that must apply to the following balanced text.

Since \exp_not:n is actually a primitive, it allows a strange syntax and it particular the primitive expands what follows and discards spaces and \scan_stop: until finding a braced argument (the opening brace can be implicit but we will not support this
Here, we repeatedly f-expand after such an \exp_not:n, and test what follows. If it is a brace group, then we found the intended argument of \exp_not:n. If it is a space, then the next f-expansion will eliminate it. If it is an N-type token then \_text:_expand_unexpanded:N leaves the token to be expanded if it is expandable, and otherwise removes it, assuming that it is \scan_stop:. This silently hides errors when \exp_not:n is incorrectly followed by some non-expandable token other than \scan_stop:, but this should be pretty rare, and there is no good error recovery anyways.

\cs_new:Npn \_text_expand_unexpanded:w
\begin{Verbatim}
\exp_after:wN \_text_expand_unexpanded_test:w
\exp:w \exp_end_continue_f:w
\end{Verbatim}
\cs_new:Npn \_text_expand_unexpanded_test:w #1 \q__text_recursion_stop
\begin{Verbatim}
\tl_if_head_is_group:nTF {#1}
{ \_text_expand_unexpanded:n }
{
\_text_expand_unexpanded:w
\tl_if_head_is_N_type:nT {#1} { \_text_expand_unexpanded:N }
}
#1 \q__text_recursion_stop
\end{Verbatim}
\cs_new:Npn \_text_expand_unexpanded:N #1
\begin{Verbatim}
\exp_after:wN \if_meaning:w \exp_not:N #1 #1
\else:
\exp_after:wN #1
\fi:
\end{Verbatim}
\cs_new:Npn \_text_expand_unexpanded:N #1
\begin{Verbatim}
\_text_expand_store:n {#1}
\_text_expand_loop:w
\end{Verbatim}

(End definition for \text_expand:n and others. This function is documented on page 265.)

Create equivalents to allow replacement.

\text_declare_expand_equivalent:Nn
\text_declare_expand_equivalent:cn
\begin{Verbatim}
\cs_new_protected:Npn \text_declare_expand_equivalent:Nn #1#2
\begin{Verbatim}
\tl_clear_new:c { l__text_expand_ \token_to_str:N #1 _tl }
\tl_set:cn { l__text_expand_ \token_to_str:N #1 _tl } {#2}
\end{Verbatim}
\cs_generate_variant:Nn \text_declare_expand_equivalent:Nn { c }
\end{Verbatim}

(End definition for \text_declare_expand_equivalent:Nn. This function is documented on page 265.)
Chapter 82

\text-case implementation

\text-case_check_letter_bool

Needed to determine the route used in titlecasing.

\bool_new:N \l_text_case_check_letter_bool
\bool_set_true:N \l_text_case_check_letter_bool

(End definition for \l_text_case_check_letter_bool. This variable is documented on page 268.)

The user level functions here are all wrappers around the internal functions for case changing.

\cs_new:Npn \text_lowercase:n #1
\cs_new:Npn \text_uppercase:n #1
\cs_new:Npn \text_titlecase:n #1
\cs_new:Npn \text_titlecase_first:n #1

\cs_new:Npn \text_lowercase:nn #1#2
\cs_new:Npn \text_uppercase:nn #1#2
\cs_new:Npn \text_titlecase:nn #1#2
\cs_new:Npn \text_titlecase_first:nn #1#2

(End definition for \text_lowercase:n and others. These functions are documented on page 266.)

As for the expansion code, the business end of case changing is the handling of \N-type tokens. First, we expand the input fully (so the loops here don’t need to worry about awkward look-aheads and the like). Then we split into the different paths.

The code here needs to be \F-type expandable to deal with the situation where case changing is applied in running text. There, we might have case changing as a document command and the text containing other non-expandable document commands.
If we use an e-type expansion and wrap each token in \exp_not:n, that would explode: the document command grabs \exp_not:n as an argument, and things go badly wrong. So we have to wrap the entire result in exactly one \exp_not:n, or rather in the kernel version.

As for expansion, collect up the tokens for future use.

The main loop is the standard tl action type.
For a group, we could worry about whether this contains a character or not. However, that would make life very complex for little gain: exactly what a first character is is rather weakly-defined anyway. So if there is a group, we simply assume that a character has been seen, and for title case we switch to the “rest of the tokens” situation. To avoid having too much testing, we use a two-step process here to allow the titlecase functions to be separate.

\cs_new:Npn \__text_change_case_group_lower:nnn #1#2#3
{\__text_change_case_store:o
{\exp_after:wN
{\exp:w
\__text_change_case_aux:nnn {#3} {#1} {#2}
}
\__text_change_case_loop:nnw {lower} {#2}
\__text_change_case_break:w
}
\use:x
{\cs_new:Npn \exp_not:N \__text_change_case_space:nnw ##1##2 \c_space_tl
}
}
The first step of handling N-type tokens is to filter out the end-of-loop. That has to be done separately from the first real step as otherwise we pick up the wrong delimiter. The loop here is the same as the expand one, just forcing the additional data long. If no close-math token is found then the final clean-up is forced (i.e. there is no assumption of well-behaved input in terms of math mode).
Once potential math-mode cases are filtered out the next stage is to test if the token
grabbed is a control sequence: the two routes the code may take are then very different.

To deal with a control sequence there is first a need to test if it is on the list which
indicate that case changing should be skipped. That’s done using a loop as for the other
special cases. If a hit is found then the argument is grabbed and passed through as-is.
Letter-like commands may still be present: they are set up using a simple lookup approach, so can easily be handled with no loop. If there is no hit, we are at the end of the process: we loop around. Letter-like chars are all available only in upper- and lowercase, so titlecasing maps to the uppercase version.

For upper- and lowercase changes, once we get to this stage there are only a couple of questions remaining: is there a language-specific mapping and is there the special case of a terminal sigma. If not, then we pass to a simple character mapping.
If the current character is an uppercase sigma, a check is made on the next item in the input. If it is N-type and not a control sequence then there is a look-ahead phase: the logic here is simply based on letters. The one exception is Dutch: see below.

In the 8-bit engines, we have to look ahead once we find the first byte of the possible hit.
For titlecasing, we need to fully expand the new character to see if it is a letter (or active) but that means looking ahead in the 8-bit case, so we have to grab the required tokens up-front. Life is a lot easier for Unicode \TeX's, where we just have one token to worry about. The one wrinkle here is that for look-ahead we'd get into trouble: luckily, only Dutch has that issue.
For Unicode engines we can handle all characters directly. However, for the 8-bit engines the aim is to deal with (a subset of) Unicode (UTF-8) input. They deal with that by making the upper half of the range active, so we look for that and if found work out how many UTF-8 octets there are to deal with. Those can then be grabbed to reconstruct the full Unicode character, which is then used in a lookup. (As will become obvious below, there is no intention here of covering all of Unicode.) For (u)p\-TEX there are a limited number of tokens we can touch.

\begin{verbatim}
\cs_new:Npn \_text_change_case_char:nnnN #1#2#3#4
{ \token_if_active:NTF #4
  { \_text_change_case_store:n {#4} }
  { \_text_change_case_store:e \use:c { char_#1 case:N } #4 }
} \use:c { __text_change_case_char_next_#2:nn } {#2} {#3} #4
\bool_lazy_or:nnF \sys_if_engine_luatex_p: \sys_if_engine_xetex_p:
{ \cs_new_eq:NN \_text_change_case_char_auxi:nnnN \_text_change_case_char:nnnN
  \cs_gset:Npn \_text_change_case_char:nnnN #1#2#3#4
  { \int_compare:nNnTF { '#4 } > { "80 }
    { \int_compare:nNnTF { '#4 } < { "E0 }
      { \_text_change_case_char_UTFviii:nnnNN }
      { \_text_change_case_char_UTFviii:nnnNNN }
    }
    { \_text_change_case_char_utfauxi:nnnN }
    {#1} {#2} {#3} #4 }
} \sys_if_engine_pdftex:TF
{ \cs_new:Npn \_text_change_case_char_utfauxi:nnnN #1#2#3#4
  { \int_compare:nNnTF { '#4 } < { "F0 }
    { \_text_change_case_char_UTFviii:nnnNN }
    {#1} {#2} {#3} #4 }
} \OutputLine
\end{verbatim}
A simple alternative version for German.

\bool_lazy_or:nnTF
  { \sys_if_engine_luatex_p: }
  { \sys_if_engine_xetex_p: }
  { \cs_new:cpn { __text_change_case_upper_de-xeszett:nnnN } #1#2#3#4
    \int_compare:nNnTF { '#4 } = { "00DF }
    { \__text_change_case_store:e
      { \char_generate:nn { "1E9E } { \__text_char_catcode:N #4 } }
      \use:c { __text_change_case_char_next_ #2 :nn } {#2} {#3} }
For Greek uppercasing, we need to know if characters in the Greek range have accents. That means doing a nfd conversion first, then starting a search. As described by the Unicode CLDR, Greek accents need to be found after any U+0308 (diaeresis) and are done in two groups to allow for the canonical ordering. The implementation here follows the data and examples from ICU (https://icu.unicode.org/design/case/greek-upper), although necessarily the implementation is somewhat different. The ypogegrammeni is filtered out here as it is not actually in the Greek range, so gets lost if we leave until later.
At this stage we have the first NFD codepoint as #3. What we need to know is whether after that we have another character token, either from the NFD or directly in the input. If not, we store the changed character at this stage.

Now, we check the detail of the next codepoint: again we filter out the not-a-char cases, before checking if it’s an dialytika, accent or diacritic. (The latter do not have the same hiatus behavior as accents.) There is additional work if the codepoint can take a ypogegrammeni: there, we need to move any ypogegrammeni to after accents (in case the input is not normalised). The ypogegrammeni itself is handled separately.
We handle dialytika in parts as it’s also needed for the hiatus. We know only two letters take it, so we can shortcut here on the second part of the tests.

Adding a hiatus needs some of the same ideas, but if there is not one we skip this code point, hence needing a separate function.

For clearing out trailing combining marks after we have dealt with the first one.
Luckily the Greek range is limited and clear.

We follow ICU in adding a few extras to the accent list here.
\prg_return_true:
\else:
  \if_int_compare:w #1 = "0303 \exp_stop_f:
    \prg_return_true:
  \else:
    \if_int_compare:w #1 = "0311 \exp_stop_f:
      \prg_return_true:
    \else:
      \prg_return_false:
    \fi:
  \fi:
  \fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
\fi:
There is one special case in Greek that needs to be picked up based on being an isolated letter. We do that using a test similar to final sigma, but it has to fire off from the space grabber.
Titlecasing retains accents, but to prevent the uppercasing code from kicking in, there has to be an explicit function here. For Lithuanian, the issue to be dealt with is dots over lower case letters: these should be present if there is another accent. The first step is a simple match attempt: look for the three uppercase accented letters which should gain a dot- above char in their lowercase form.
If there was a hit, output the result with the dot-above and move on. Otherwise, look for one of the three letters that can take a combining accent: I, J and I-ogonek.

Again, branch depending on a hit. If there is one, we output the character then need to look for a combining accent: as usual, we need to be aware of the loop situation.
\bool_lazy_and:nnT
{ ! \token_if_cs_p:N #3 }
{ \bool_lazy_any_p:n
{ \int_compare_p:nNn { '#3 } = { "0300 } }
{ \int_compare_p:nNn { '#3 } = { "0301 } }
{ \int_compare_p:nNn { '#3 } = { "0303 } }
}
{ \__text_change_case_store:e
{ \char_generate:nn { "0307 } { \__text_char_catcode:N #3 } }
}
\__text_change_case_loop:nnw {#1} {#2} #3

(End definition for \text_change_cases_lower_lt:nnnN and others.)

The uppercasing version: first find i/j/i-o-gonek, then look for the combining char: drop it if present.

\bool_lazy_or:nnT
{ \sys_if_engine_luatex_p: }
{ \sys_if_engine_xetex_p: }
{ \cs_new:Npn \__text_change_case_upper_lt:nnnN #1#2#3#4
\exp_args:Ne \__text_change_case_upper_lt_aux:nnnN
{ \int_case:nn { '#4 }
{ "0069 } { "0049 }
{ "006A } { "004A }
{ "012F } { "012E }
}
{#2} {#3} #4
}
\cs_new:Npn \__text_change_case_upper_lt_aux:nnnN #1#2#3#4
{ \tl_if_blank:nTF {#1}
{ \__text_change_case_char:nnnN { upper } {#2} {#3} #4 }
{ \__text_change_case_store:e
{ \char_generate:n {#1} { \__text_char_catcode:N #4 } }
{ \__text_change_case_upper_lt:nnnN #2#3 }
}
\tl_if_head_is_N_type:nTF {#3}
{ \__text_change_case_upper_lt:nnnN #1#2#3 }
{ \use:c { __text_change_case_char_next_ #1 :nn } }
}
For Dutch, there is a single look-ahead test for ij when title casing. If the appropriate letters are found, produce IJ and gobble the j/J.

\cs_new:Npn \__text_change_case_title_nl:nnnN \#1#2#3#4
\begin{Verbatim}
\texttt{\textbf{\bool_lazy_or:nnTF \int_compare_p:nNn { \#4 } = { "0049 } \int_compare_p:nNn { \#4 } = { "0069 }}
\__text_change_case_store:e
\__text_char_catcode:N \#3
\use:c { \text_change_case_char_next_ \#1 } {#1} {#2} \\use:c { \text_change_case_char_next_ \#1 } {#1} {#2} \#3
\end{Verbatim}

\cs_new:Npn \__text_change_case_title_nl:nnw #1#2#3
\begin{Verbatim}
\tl_if_head_is_N_type:nTF {#3}
\__text_change_case_title_nl:nnN
\end{Verbatim}

\cs_new:Npn \__text_change_case_title_nl:nnN #1#2#3
\begin{Verbatim}
\textbf{\bool_lazy_and:nnTF \int_compare_p:nNn { \#3 } \int_compare_p:nNn { \#3 }
\__text_change_case_store:e
\__text_char_catcode:N \#3
\use:c { \text_change_case_char_next_ \#1 } {#1} {#2} \\use:c { \text_change_case_char_next_ \#1 } {#1} {#2} \#3
\end{Verbatim}

(End definition for \textchange, \textchange, and others.)
The Turkic languages need special treatment for dotted-i and dotless-i. The lower casing rule can be expressed in terms of searching first for either a dotless-I or a dotted-I. In the latter case the mapping is easy, but in the former there is a second stage search.

After a dotless-I there may be a dot-above character. If there is then a dotted-i should be produced, otherwise output a dotless-i. When the combination is found both the dotless-I and the dot-above char have to be removed from the input.
For 8-bit engines, dot-above is not available so there is a simple test for an upper-case I. Then we can look for the UTF-8 representation of an upper case dotted-I without the combining char. If it’s not there, preserve the UTF-8 sequence as-is. With 8bit engines, we cannot completely preserve category codes, so we have to make some assumptions: output a “normal” i for the dotted case. As the original character here is catcode-13, we have to make a choice about handling of i: generate a “normal” one.

\cs_new:Npn \__text_change_case_lower_tr:nnnN #1#2#3#4
{\int_compare:nNnTF { '#4 } = { "0049 }
{\__text_change_case_store:V \c__text_dotless_i_tl \__text_change_case_loop:nnw {#1} {#3} }
{ {\int_compare:nNnTF { '#4 } = { "00C4 }
{ \__text_change_case_lower_tr:nnnNN {#1} {#2} {#3} #4 }
{ \__text_change_case_char:nnnN {#1} {#2} {#3} #4 }
}
}
\cs_new:Npn \__text_change_case_lower_tr:nnnNN #1#2#3#4#5
{\int_compare:nNnTF { '#5 } = { "00B0 }
{ \__text_change_case_store:e
{ \char_generate:nn { "0069 }
{ \char_value_catcode:n { "0069 } }
}
\__text_change_case_loop:nnw {#1} {#3} }
{ \__text_change_case_char:nnnN {#1} {#2} {#3} #4#5 }
}

(End definition for \__text_change_case_lower_tr:nnnN and others.)

\__text_change_case_upper_tr:nnnN

Uppercasing is easier: just one exception with no context.
\cs_new:Npx \__text_change_case_upper_tr:nnnN #1#2#3#4
{\exp_not:N \int_compare:nNnTF { '#4 } = { "0069 }
{ \bool_lazy_or:nnTF
{ \sys_if_engine_luatex_p: }
{ \sys_if_engine_xetex_p: }
{ \exp_not:N \__text_change_case_store:e
{ \exp_not:N \char_generate:nn { "0130 }
{ \exp_not:N \__text_char_catcode:N #4 }
}
}
{ \exp_not:N \__text_change_case_char:nnnN {#1} {#2} {#3} #4#5 }
}
\exp_not:N \__text_change_case_upper_tr:nnnN V

1238
Straight copies.

82.2 Case changing data for 8-bit engines

For cases where there is an 8-bit option in the T1 font set up, a variant is provided in both cases. There are also a few extras for LGR.
For 8-bit engines we now need to define the case-change data for the multi-octet mappings. This data is here not in the char module as the multi-byte nature means they are never N-type. As this data is needed both for text and string work, it is stored as kernel: it remains internal as it is not really suitable for wider exposure.

The first set of codepoints are those in T1: all of these are two bytes.

```latex
\begin{verbatim}
\_text_tmp:w \c_text_sigma_tl \{ 03C3 \}
\_text_tmp:w \c_text_grosses_Eszett_tl \{ 1E9E \}
\group_end:

(End definition for \c_text_dotless_i_tl and others.)
```

```latex
\begin{verbatim}
\begin{group}
\bool_lazy_or:nnF { \sys_if_engine_luatex_p: } { \sys_if_engine_xetex_p: }
{ \cs_set_protected:Npn \__text_loop:nn #1#2
{ \quark_if_recursion_tail_stop:n {#1}
 \use:x 
 { \__text_tmp:w \char_to_utfviii_bytes:n { "#1 } 
 \char_to_utfviii_bytes:n { "#2 } }
 \__text_loop:nn }
\cs_set_protected:Npn \__text_tmp:nnnn #1#2#3#4#5
{ \tl_const:cx { \c_kernel_ #1 case_ \char_generate:nn {#2} { 12 } \char_generate:nn {#3} { 12 } _tl }
 \exp_after:wN \exp_after:wN \exp_after:wN \exp_not:N \char_generate:nn {#4} { 13 }
 \exp_after:wN \exp_after:wN \exp_after:wN \exp_not:N \char_generate:nn {#5} { 13 }
 }
\cs_set_protected:Npn \__text_tmp:w #1#2#3#4#5#6#7#8
{ \tl_const:cx 
 { \c_kernel_lowercase_ \char_generate:nn {#1} { 12 } \char_generate:nn {#2} { 12 } _tl }
 \exp_after:wN \exp_after:wN \exp_after:wN \exp_not:N \char_generate:nn {#3} { 13 }
 \exp_after:wN \exp_after:wN \exp_after:wN \exp_not:N \char_generate:nn {#4} { 13 }
 \exp_after:wN \exp_after:wN \exp_after:wN \exp_not:N \char_generate:nn {#5} { 13 }
 \exp_after:wN \exp_after:wN \exp_after:wN \exp_not:N \char_generate:nn {#6} { 13 }
 \exp_after:wN \exp_after:wN \exp_after:wN \exp_not:N \char_generate:nn {#7} { 13 }
 \exp_after:wN \exp_after:wN \exp_after:wN \exp_not:N \char_generate:nn {#8} { 13 }
 }
\cs_set_protected:Npn \__text_loop:nnn #1#2#3#4#5
{ \tl_const:cx 
 { \c_kernel_ #1 case_ \char_generate:nn {#2} { 12 } \char_generate:nn {#3} { 12 } _tl }
 \exp_after:wN \exp_after:wN \exp_after:wN \exp_not:N \char_generate:nn {#4} { 13 }
 \exp_after:wN \exp_after:wN \exp_after:wN \exp_not:N \char_generate:nn {#5} { 13 }
 }
\cs_set_protected:Npn \__text_loop:nnn #1#2#3#4#5#6#7#8
{ \tl_const:cx 
 { \c_kernel_lowercase_ \char_generate:nn {#1} { 12 } \char_generate:nn {#2} { 12 } _tl }
 \exp_after:wN \exp_after:wN \exp_after:wN \exp_not:N \char_generate:nn {#3} { 13 }
 \exp_after:wN \exp_after:wN \exp_after:wN \exp_not:N \char_generate:nn {#4} { 13 }
 \exp_after:wN \exp_after:wN \exp_after:wN \exp_not:N \char_generate:nn {#5} { 13 }
 \exp_after:wN \exp_after:wN \exp_after:wN \exp_not:N \char_generate:nn {#6} { 13 }
 \exp_after:wN \exp_after:wN \exp_after:wN \exp_not:N \char_generate:nn {#7} { 13 }
 \exp_after:wN \exp_after:wN \exp_after:wN \exp_not:N \char_generate:nn {#8} { 13 }
 }
\end{group}
\end{verbatim}
```

1240
{ 0120 } { 0121 }
{ 0122 } { 0123 }
{ 0124 } { 0125 }
{ 0128 } { 0129 }
{ 012A } { 012B }
{ 012C } { 012D }
{ 012E } { 012F }
{ 0132 } { 0133 }
{ 0134 } { 0135 }
{ 0136 } { 0137 }
{ 0139 } { 013A }
{ 013B } { 013C }
{ 013E } { 013F }
{ 0141 } { 0142 }
{ 0143 } { 0144 }
{ 0145 } { 0146 }
{ 0147 } { 0148 }
{ 014A } { 014B }
{ 014C } { 014D }
{ 014E } { 014F }
{ 0150 } { 0151 }
{ 0152 } { 0153 }
{ 0154 } { 0155 }
{ 0156 } { 0157 }
{ 0158 } { 0159 }
{ 015A } { 015B }
{ 015C } { 015D }
{ 015E } { 015F }
{ 0160 } { 0161 }
{ 0162 } { 0163 }
{ 0164 } { 0165 }
{ 0168 } { 0169 }
{ 016A } { 016B }
{ 016C } { 016D }
{ 016E } { 016F }
{ 0170 } { 0171 }
{ 0172 } { 0173 }
{ 0174 } { 0175 }
{ 0176 } { 0177 }
{ 0178 } { 00FF }
{ 0179 } { 017A }
{ 017B } { 017C }
{ 017D } { 017E }
{ 01CD } { 01CE }
{ 01CF } { 01D0 }
{ 01D1 } { 01D2 }
{ 01D3 } { 01D4 }
{ 01E2 } { 01E3 }
{ 01E6 } { 01E7 }
{ 01EA } { 01EB }
{ 01F4 } { 01F5 }
{ 0218 } { 0219 }
{ 021A } { 021B }
Add T2 (Cyrillic): again two bytes.

Greek support: everything in the two-octet range.
\texttt{\textbackslash q\_recursion\_tail} ?
Odds and ends for Greek; mainly symbols that are for compatibility, but also things like the terminal sigma. Almost all are uppercase mappings, but there is one that is not!

```
\cs_set_protected:Npn \_text_tmp:w #1#2#3
{\group_begin:
  \cs_set_protected:Npn \_text_tmp:w ##1##2##3##4##5##6##7##8
  {\tl_const:cx
    \c__kernel_ #3 case_
    \char_generate:nn {##1} { 12 }
    \char_generate:nn {##2} { 12 }
    _tl
  }
  \use:x
  \__text_tmp:w
  \char_to_utfviii_bytes:n { "#1 }
  \char_to_utfviii_bytes:n { "#2 }
\group_end:
\__text_tmp:w { 0345 } { 0399 } { upper }
\__text_tmp:w { 03C2 } { 03A3 } { upper }
\__text_tmp:w { 03D0 } { 0392 } { upper }
\__text_tmp:w { 03D1 } { 0398 } { upper }
\__text_tmp:w { 03D5 } { 03A6 } { upper }
\__text_tmp:w { 03D6 } { 03A0 } { upper }
\__text_tmp:w { 03F0 } { 039A } { upper }
\__text_tmp:w { 03F1 } { 03A1 } { upper }
\__text_tmp:w { 03F4 } { 03B5 } { lower }
\__text_tmp:w { 03FS } { 0395 } { upper }
```

Odds and ends that are not simple one-to-one mappings. These are still two-octet code points.
Greek support: the three-octet code points.

\cs_set_protected:Npn \__text_tmp:nnnnnn #1#2#3#4#5#6#7
{\tl_const:cx
  \group_end:
  \__text_tmp:w \char_to_utfviii_bytes:n { "#1 } }

\__text_tmp:w { 00DF } { SS } { upper }
\__text_tmp:w { 00DF } { Ss } { title }
\__text_tmp:w { 0131 } { I } { upper }

\cs_set_protected:Npn \__text_tmp:nnnnnn #1#2#3#4#5#6#7#8
{\tl_const:cx
  \group_end:
  \__text_tmp:w #1#2#3#4#5#6#7#8
}

\cs_set_protected:Npn \__text_loop:nn { 1F08 } { 1F00 }
{\text_loop:nn { 1F08 } { 1F00 }
<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1F0A</td>
<td>1F02</td>
</tr>
<tr>
<td>1F0B</td>
<td>1F03</td>
</tr>
<tr>
<td>1F0C</td>
<td>1F04</td>
</tr>
<tr>
<td>1F0D</td>
<td>1F05</td>
</tr>
<tr>
<td>1F0E</td>
<td>1F06</td>
</tr>
<tr>
<td>1F0F</td>
<td>1F07</td>
</tr>
<tr>
<td>1F10</td>
<td>1F11</td>
</tr>
<tr>
<td>1F12</td>
<td>1F13</td>
</tr>
<tr>
<td>1F14</td>
<td></td>
</tr>
<tr>
<td>1F15</td>
<td></td>
</tr>
<tr>
<td>1F20</td>
<td></td>
</tr>
<tr>
<td>1F21</td>
<td></td>
</tr>
<tr>
<td>1F22</td>
<td></td>
</tr>
<tr>
<td>1F23</td>
<td></td>
</tr>
<tr>
<td>1F24</td>
<td></td>
</tr>
<tr>
<td>1F25</td>
<td></td>
</tr>
<tr>
<td>1F31</td>
<td></td>
</tr>
<tr>
<td>1F32</td>
<td></td>
</tr>
<tr>
<td>1F33</td>
<td></td>
</tr>
<tr>
<td>1F34</td>
<td></td>
</tr>
<tr>
<td>1F35</td>
<td></td>
</tr>
<tr>
<td>1F36</td>
<td></td>
</tr>
<tr>
<td>1F37</td>
<td></td>
</tr>
<tr>
<td>1F40</td>
<td></td>
</tr>
<tr>
<td>1F41</td>
<td></td>
</tr>
<tr>
<td>1F42</td>
<td></td>
</tr>
<tr>
<td>1F43</td>
<td></td>
</tr>
<tr>
<td>1F44</td>
<td></td>
</tr>
<tr>
<td>1F45</td>
<td></td>
</tr>
<tr>
<td>1F51</td>
<td></td>
</tr>
<tr>
<td>1F53</td>
<td></td>
</tr>
<tr>
<td>1F55</td>
<td></td>
</tr>
<tr>
<td>1F57</td>
<td></td>
</tr>
<tr>
<td>1F60</td>
<td></td>
</tr>
<tr>
<td>1F61</td>
<td></td>
</tr>
<tr>
<td>1F62</td>
<td></td>
</tr>
<tr>
<td>1F63</td>
<td></td>
</tr>
<tr>
<td>1F64</td>
<td></td>
</tr>
<tr>
<td>1F65</td>
<td></td>
</tr>
<tr>
<td>1F66</td>
<td></td>
</tr>
<tr>
<td>1F67</td>
<td></td>
</tr>
<tr>
<td>1F71</td>
<td></td>
</tr>
<tr>
<td>1F72</td>
<td></td>
</tr>
<tr>
<td>1F73</td>
<td></td>
</tr>
<tr>
<td>1F74</td>
<td></td>
</tr>
<tr>
<td>1F75</td>
<td></td>
</tr>
<tr>
<td>1F76</td>
<td></td>
</tr>
<tr>
<td>1F77</td>
<td></td>
</tr>
</tbody>
</table>
One three-octet special case for Greek: it also moves to two-octets!

\cs_set_protected:Npn \__text_tmp:w #1#2#3
\group_begin:
\cs_set_protected:Npn \__text_tmp:w ##1##2##3##4##5##6##7##8
{\tl_const:cx
  \__kernel_ #3 case_
\char_generate:nn {##1} { 12 }
\char_generate:nn {##2} { 12 }
\char_generate:nn {##3} { 12 }\group_stop
The (fixed) look-up mappings for letter-like control sequences.

To deal with possible encoding-specific extensions to \@ucilclist, we check at the end of the preamble. This will therefore only apply to \LaTeX2ε package mode.
{ \tl_gput_right:Nn \@kernel@after@begindocument
    \group_begin:
    \cs_set_protected:Npn \_text_change_case_setup:Nn #1#2
    { \quark_if_recursion_tail_stop:N #1
        \tl_if_single_token:nT {#2}
        { \cs_if_exist:cF
            \tl_const:cn { \_text_uppercase_ \token_to_str:N #1 _tl }
            \cs_if_exist:cF
            \tl_const:cn { \_text_lower_case_ \token_to_str:N #2 _tl }
            \_text_change_case_setup:Nn
            \exp_after:wN \_text_change_case_setup:Nn \@ucclist
            \q_recursion_stop
            \group_end:
        }
        \tl_gput_right:Nn
        \_text_change_case_setup:Nn
        \exp_after:wN \_text_change_case_setup:Nn \@ucclist
        \q_recursion_stop
        \group_end:
    }
\} ⟨/package⟩
Chapter 83

\text-purify implementation

83.1 Purifying text

Functions to query recursion quarks.

As in the other parts of the module, we start off with a standard “action” loop, with expansion applied up-front.

As for expansion, collect up the tokens for future use.
The main loop is a standard “tl action”. Unlike the expansion or case changing, here any groups have to be run inline. Most of the business end is as before in the \textit{N}-type token processing.

```latex
\cs_new:Npn \__text_purify_loop:w #1 \q__text_recursion_stop
\{\tl_if_head_is_N_type:nTF {#1}
\{ \__text_purify_N_type:N \}
\{ \tl_if_head_is_group:nTF {#1}
\{ \__text_purify_group:n \}
\{ \__text_purify_space:w \}
\}
\#1 \q__text_recursion_stop\}
\cs_new:Npn \__text_purify_group:n #1 { \__text_purify_loop:w #1 }
\exp_last_unbraced:NNo \cs_new:Npn \__text_purify_space:w \c_space_tl\q__text_recursion_tail \q__text_recursion_stop
\cs_new:Npn \__text_purify_math_search:NNN #1#2#3
\{ \__text_if_recursion_tail_stop_do:Nn #2
\{ \__text_purify_math_cmd:N #1 \}
\token_if_eq_meaning:NNTF #1 #2
\{ \__text_use_i_delimit_by_q_recursion_stop:nw \__text_purify_math_start:NNw #2 #3 \}
\}
\cs_new:Npn \__text_purify_math_start:NNw #1#2#3 \q__text_recursion_stop
\{ \__text_purify_math_loop:NNw #1#2#3 \q__text_recursion_stop
\\__text_purify_math_result:n \}
\cs_new:Npn \__text_purify_math_end:NN \q__text_recursion_stop
```

The first part of handling math mode is exactly the same as in the other functions: look for a start-of-math mode token and if found start a new loop tracking the closing token.

```latex
\cs_new:Npn \__text_purify_N_type:N #1
\{ \__text_if_recursion_tail_stop_do:Nn #1 { \__text_purify_end:w }
\__text_purify_N_type_aux:N #1\}
\cs_new:Npn \__text_purify_N_type_aux:N #1
\{ \exp_after:wN \__text_purify_math_search:NNN \exp_after:wN #1 \l_text_math_delims_tl \q__text_recursion_tail \q__text_recursion_stop \}
```

```latex
\cs_new:Npn \__text_purify_math_search:NNN #1#2#3
\{ \__text_if_recursion_tail_stop_do:Nn #2
\{ \__text_purify_math_cmd:N #1 \}
\token_if_eq_meaning:NNTF #1 #2
\{ \__text_use_i_delimit_by_q_recursion_stop:nw \__text_purify_math_start:NNw #2 #3 \}
\}
\cs_new:Npn \__text_purify_math_start:NNw #1#2#3 \q__text_recursion_stop
\{ \__text_purify_math_loop:NNw #1#2#3 \q__text_recursion_stop
```

```latex
\cs_new:Npn \__text_purify_math_end:NN \q__text_recursion_stop
```
Then handle math mode as an argument: same outcomes, different input syntax.
For \(N\)-type tokens, we first look for a string-context replacement before anything else: this can therefore cover anything. Assuming we don’t find one, check to see if we can expand control sequences: if not, they have to be dropped. We also allow for \LaTeX{}_{2e} \protect: there’s an assumption that we don’t have \protect \{ \oops \} or similar, but that’s also in the expansion code and seems like a reasonable balance.

\begin{verbatim}
\cs_new:Npn \__text_purify_replace:N #1 {
\__text_purify_loop:w #1 }
\cs_new:Npn \__text_purify_expand:N #1 {
\str_if_eq:NnTF {#1} { \protect }
\__text_purify_protect:N }
\cs_new:Npn \__text_if_recursion_tail_stop_do:Nn #2 #1 }
\cs_if_eq:NNTF #2 #1 {
\__text_use_i_delimit_by_q_recursion_stop:nw
\__text_purify_math_cmd:n }
\__text_purify_math_cmd:NN #1 }
\cs_new:Npn \__text_purify_math_cmd:n #1 {
\__text_purify_end:w \__text_purify_math_result:n {#1} }
\cs_new:Npn \__text_purify_math_end:w \__text_purify_math_result:n {#1} }
\cs_new:Npn \__text_purify_protect:N #1 {
\__text_if_recursion_tail_stop_do:Nn #1 { 
\__text_purify_end:w }
\__text_purify_lo
Now pre-define a range of standard commands that need dedicated definitions in purified text. First handle font-related stuff: all of this needs to be disabled.

\map\inline\{ \fontencoding \fontfamily \fontseries \fontshape \}

(End definition for \text\purify equivalent. This function is documented on page 267.)

\text\purify equivalent:
\text\purify equivalent:

Now pre-define a range of standard commands that need dedicated definitions in purified text. First handle font-related stuff: all of this needs to be disabled.

\map\inline\{ \char\fontencoding \char\fontfamily \char\fontseries \char\fontshape \}

(End definition for \text\purify equivalent. This function is documented on page 267.)
Environments have to be handled by pure expansion.

\begin{verbatim}
\_text_end_env:n \text_declare_purify_equivalent:Nn \begin { \use:c }
\text_declare_purify_equivalent:Nn \end { \_text_end_env:n }
\cs_new:Nn \_text_end_env:n #1 { \cs:w end #1 \cs_end: }
\end{verbatim}

(End definition for \_text_end_env:n.)

Some common symbols and similar ideas.

\begin{verbatim}
\_text_declare_purify_equivalent:Nn \_ \{ \}
\_tl_map_inline:nn
\_text_declare_purify_equivalent:Nx #1 { \_cs_to_str:N #1 }
\end{verbatim}

Cross-referencing.

\begin{verbatim}
\_text_declare_purify_equivalent:Nn \label { \use_none:n }
\end{verbatim}

Spaces.

\begin{verbatim}
\group_begin:
\char_set_catcode_active:N \-
\use:n
\end{verbatim}
83.2 Accent and letter-like data for purifying text

In contrast to case changing, both 8-bit and Unicode engines need information for text purification to handle accents and letter-like functions: these all need to be removed. However, the results are of course engine-dependent.

For the letter-like commands, life is relatively easy: they all simply need to be added as standard exceptions. The only oddity is \texttt{SS}, which gets converted to two letters. (At some stage an alternative version can presumably be added to babel or similar.)
Accent handling is a little more complex. Accents may exist as pre-composed codepoints or as independent glyphs. The former are all saved as single token lists, whilst for the latter the combining accent needs to be re-ordered compared to the character it applies to.

First set up the combining accents.
Now we handle the pre-composed accents: the list here is taken from `puenc.def`. All of the precomposed cases take a single letter as their second argument. We do not try to cover the case where an accent is added to a “real” dotless-i or -j, or a æ/Æ. Rather, we assume that if the UTF-8 character is used, it will have the real accent character too.
\texttt{\_	extunderscore\_text\_purify\_} \texttt{\token\_to\_str:N \#1 \token\_to\_str:N \#2 _tl}
\texttt{\_text\_tmp:n \#3}
\texttt{\_text\_loop:NNn}
\texttt{\textbackslash \_text\_loop:NNn}
\texttt{\'} A \{ 00c0 \}
\texttt{\'} A \{ 00c1 \}
\texttt{\'} A \{ 00c2 \}
\texttt{\'} A \{ 00c3 \}
\texttt{\'} A \{ 00c4 \}
\texttt{\'} A \{ 00c5 \}
\texttt{\'} A \{ 00c7 \}
\texttt{\'} A \{ 00c8 \}
\texttt{\'} A \{ 00c9 \}
\texttt{\'} A \{ 00ca \}
\texttt{\'} A \{ 00cb \}
\texttt{\'} A \{ 00cc \}
\texttt{\'} A \{ 00cd \}
\texttt{\'} A \{ 00ce \}
\texttt{\'} A \{ 00cf \}
\texttt{\'} A \{ 00d1 \}
\texttt{\'} A \{ 00d2 \}
\texttt{\'} A \{ 00d3 \}
\texttt{\'} A \{ 00d4 \}
\texttt{\'} A \{ 00d5 \}
\texttt{\'} A \{ 00d6 \}
\texttt{\'} A \{ 00d9 \}
\texttt{\'} A \{ 00da \}
\texttt{\'} A \{ 00db \}
\texttt{\'} A \{ 00dc \}
\texttt{\'} A \{ 00dd \}
\texttt{\'} A \{ 00e0 \}
\texttt{\'} A \{ 00e1 \}
\texttt{\'} A \{ 00e2 \}
\texttt{\'} A \{ 00e3 \}
\texttt{\'} A \{ 00e4 \}
\texttt{\'} A \{ 00e5 \}
\texttt{\'} A \{ 00e7 \}
\texttt{\'} A \{ 00e8 \}
\texttt{\'} A \{ 00e9 \}
\texttt{\'} A \{ 00ea \}
\texttt{\'} A \{ 00eb \}
\texttt{\'} A \{ 00ec \}
\texttt{\'} A \{ 00ed \}
\texttt{\'} A \{ 00ee \}
\texttt{\'} A \{ 00ef \}
\texttt{\'} A \{ 00f0 \}
\texttt{\'} A \{ 00f1 \}
\texttt{\'} A \{ 00f2 \}
\texttt{\'} A \{ 00f3 \}
\texttt{\'} A \{ 00f4 \}
\H U { 0170 }
\H u { 0171 }
\k U { 0172 }
\k u { 0173 }
\^ W { 0174 }
\^ w { 0175 }
\^ Y { 0176 }
\^ y { 0177 }
\^ U { 0178 }
\^ u { 0179 }
\^ Z { 017A }
\^ z { 017B }
\. Z { 017B }
\. z { 017C }
\v Z { 017D }
\v z { 017E }
\v A { 01CD }
\v a { 01CE }
\v I { 01CF }
\v \i { 01D0 }
\v i { 01D0 }
\v O { 01D1 }
\v o { 01D2 }
\v U { 01D3 }
\v u { 01D4 }
\v G { 01E6 }
\v g { 01E7 }
\v K { 01E8 }
\v k { 01E9 }
\k O { 01EA }
\k o { 01EB }
\v \j { 01F0 }
\v j { 01F0 }
\v \G { 01F4 }
\v g { 01F5 }
\v \N { 01F8 }
\v n { 01F8 }
\v \AE { 01FC }
\v \ae { 01FD }
\v \O { 01FE }
\v \o { 01FF }
\v H { 021E }
\v h { 021F }
\. A { 0226 }
\. a { 0227 }
\c E { 0228 }
\c e { 0229 }
\. O { 022E }
\. o { 022F }
\= Y { 0232 }
\= y { 0233 }
\q_recursion_tail ? { }
\q_recursion_stop
\group_end:
Chapter 84

l3box implementation

84.1 Support code

\__box_dim_eval:w Evaluating a dimension expression expandably. The only difference with \dim_eval:n is the lack of \dim_use:N, to produce an internal dimension rather than expand it into characters.

\cs_new_eq:NN \__box_dim_eval:w \tex_dimexpr:D
\cs_new:Npn \__box_dim_eval:n #1 { \__box_dim_eval:w #1 \scan_stop: }

(End definition for \__box_dim_eval:w and \__box_dim_eval:n.)

\__kernel_kern:n We need kerns in a few places. At present, we don’t have a module for this concept, so it goes in at first use: here. The idea is to avoid repeated use of the bare primitive.

\cs_new_protected:Npn \__kernel_kern:n #1 { \tex_kern:D \__box_dim_eval:n {#1} }

(End definition for \__kernel_kern:n.)

84.2 Creating and initialising boxes

The following test files are used for this code: m3box001.lvt.

\box_new:N Defining a new \langle box \rangle register: remember that box 255 is not generally available.
\box_new:c

Defining a new \langle box \rangle register: remember that box 255 is not generally available.

\cs_new_protected:Npn \box_new:N #1
\{ \__kernel_chk_if_free_cs:N #1
\cs:w newbox \cs_end: #1
\}
\cs_generate_variant:Nn \box_new:N { c }

1265
Clear a ⟨box⟩ register.
\begin{verbatim}
\cs_new_protected:N \box_clear:N #1
{ \box_set_eq:NN \empty_box \c_empty_box }
\cs_new_protected:N \box_clear:N #1
{ \box_gset_eq:NN \empty_box \c_empty_box }
\cs_generate_variant:Nn \box_clear:N { c }
\cs_generate_variant:Nn \box_gclear:N { c }
\end{verbatim}

Clear or new.
\begin{verbatim}
\cs_new_protected:N \box_clear_new:N #1
{ \box_if_exist:NTF \box_clear:N \box_new:N #1 }
\cs_new_protected:N \box_gclear_new:N #1
{ \box_if_exist:NTF \box_gclear:N \box_new:N #1 }
\cs_generate_variant:Nn \box_clear_new:N { c }
\cs_generate_variant:Nn \box_gclear_new:N { c }
\end{verbatim}

Assigning the contents of a box to be another box.
\begin{verbatim}
\cs_new_protected:N \box_set_eq:NN #1 \tex_ht:D
\cs_new_protected:N \box_set_eq:NN #1 \tex_dp:D
\cs_generate_variant:Nn \box_set_eq:NN { c }
\cs_generate_variant:Nn \box_set_eq:NN { c , c , c }
\end{verbatim}

Assigning the contents of a box to be another box, then drops the original box.
\begin{verbatim}
\cs_new_protected:N \box_set_eq_drop:NN #1 \tex_ht:D
\cs_new_protected:N \box_set_eq_drop:NN #1 \tex_dp:D
\cs_generate_variant:Nn \box_set_eq_drop:NN { c , c , c }
\end{verbatim}

Copies of the \cs functions defined in \texttt{l3basics}.
\begin{verbatim}
\prg_new_eq_conditional:Nnn \box_if_exist:N \cs_if_exist:N { T , F , p }
\end{verbatim}

84.3 Measuring and setting box dimensions

Accessing the height, depth, and width of a ⟨box⟩ register.
\begin{verbatim}
\cs_new_eq:NN \box_ht:N \tex_ht:D
\cs_new_eq:NN \box_dp:N \tex_dp:D
\cs_generate_variant:Nn \box_ht:N { c }
\cs_generate_variant:Nn \box_dp:N { c }
\end{verbatim}

The \texttt{\box_ht:N} and \texttt{\box_dp:N} primitives do not expand but rather are suitable for use after \texttt{\the} or inside dimension expressions. Here we obtain the same behaviour by using \texttt{\_\_\_\_\_\_\_dim_eval:n} (basically \texttt{\dimexpr}) rather than \texttt{\dim_eval:n} (basically \texttt{\the\dimexpr}).
Setting the size whilst respecting local scope requires copying; the same issue does not come up when working globally. When debugging, the dimension expression #2 is surrounded by parentheses to catch early termination.

Using boxes

Move box material in different directions. When debugging, the dimension expression #1 is surrounded by parentheses to catch early termination.
84.5 Box conditionals

The primitives for testing if a \langle box \rangle is empty/void or which type of box it is.

\if_hbox:N \if_vbox:N \if_box_empty:N

\box_if_horizontal_p:N \box_if_vertical_p:N \box_if_empty_p:N

\prg_new_conditional:Npnn \box_if_horizontal:N #1 { p , T , F , TF }
\prg_new_conditional:Npnn \box_if_vertical:N #1 { p , T , F , TF }
\prg_new_conditional:Npnn \box_if_empty:N #1 { p , T , F , TF }

Testing if a \langle box \rangle is empty/void.

\box_set_to_last:N \box_gset_to_last:N

\cs_new_protected:Npn \box_set_to_last:N #1
\cs_new_protected:Npn \box_gset_to_last:N #1

\cs_generate_variant:Nn \box_set_to_last:N { c }
\cs_generate_variant:Nn \box_gset_to_last:N { c }

(End definition for \box_set_to_last:N and \box_gset_to_last:N. These functions are documented on page 273.)

84.6 The last box inserted

\c_empty_box

A box we never use.

\box_new:N \c_empty_box

(End definition for \c_empty_box. This variable is documented on page 273.)
84.8 Scratch boxes

Scratch boxes.

(End definition for \_tmpa_box and others. These variables are documented on page 273.)

84.9 Viewing box contents

\TeX’s \showbox is not really that helpful in many cases, and it is also inconsistent with other \TeX3 show functions as it does not actually shows material in the terminal. So we provide a richer set of functionality.

Essentially a wrapper around the internal function, but evaluating the breadth and depth arguments now outside the group.

Getting \TeX to write to the log without interruption the run is done by altering the interaction mode. For that, the \TeX3 extensions are needed.

The internal auxiliary to actually do the output uses a group to deal with breadth and depth values. The \use:n here gives better output appearance. Setting \tracingonline and \errorcontextlines is used to control what appears in the terminal.
\int_set:Nn \tex_showboxbreadth:D {#3}
\int_set:Nn \tex_showboxdepth:D {#4}
\int_set:Nn \tex_tracingonline:D {#1}
\int_set:Nn \tex_errorcontextlines:D { -1 }
\tex_showbox:D \use:n {#2}
\group_end:
{
    \msg_error:nnx { kernel } { variable-not-defined }
    { \token_to_str:N #2 }
}
\cs_generate_variant:Nn \__box_show:NNnn { NNff }
(End definition for \__box_show:NNnn.)

\hbox:n
(The test suite for this command, and others in this file, is m3box002.ltx.)
Put a horizontal box directly into the input stream.
\cs_new_protected:Npn \hbox:n #1
{ \tex_hbox:D \scan_stop: { \color_group_begin: #1 \color_group_end: } }
(End definition for \hbox:n. This function is documented on page 274.)
\hbox_set:Nn
\hbox_set:cn
\hbox_gset:Nn
\hbox_gset:cn
\cs_new_protected:Npn \hbox_set:Nn #1#2
{ \tex_setbox:D #1 \tex_hbox:D to \__box_dim_eval:n {#2}
\color_group_begin: #2 \color_group_end: }
\cs_new_protected:Npn \hbox_gset:Nn #1#2
{ \tex_global:D \tex_setbox:D #1 \tex_hbox:D to \__box_dim_eval:n {#2}
\color_group_begin: #2 \color_group_end: }
\cs_generate_variant:Nn \hbox_set:Nn { c }
\cs_generate_variant:Nn \hbox_gset:Nn { c }
(End definition for \hbox_set:Nn and \hbox_gset:Nn. These functions are documented on page 274.)
\hbox_set_to_wd:Nnn
\hbox_set_to_wd:cn
\hbox_gset_to_wd:Nnn
\hbox_gset_to_wd:cn
Storing material in a horizontal box with a specified width. Again, put the dimension expression in parentheses when debugging.
\cs_new_protected:Npn \hbox_set_to_wd:Nnn #1#2#3
{ \tex_setbox:D #1 \tex_hbox:D to \__box_dim_eval:n {#2}
\color_group_begin: #3 \color_group_end: }
\cs_new_protected:Npn \hbox_gset_to_wd:Nnn #1#2#3
{ \tex_global:D \tex_setbox:D #1 \tex_hbox:D to \__box_dim_eval:n {#2}
\color_group_begin: #3 \color_group_end: }
\cs_generate_variant:Nn \hbox_set_to_wd:Nnn { c }
\cs_generate_variant:Nn \hbox_gset_to_wd:Nnn { c }

84.10 Horizontal mode boxes

(End definition for \__box_show:NNnn.)
Storing material in a horizontal box. This type is useful in environment definitions.

\hbox_set:Nw \hbox_set:cnw \hbox_gset:Nw \hbox_gset:cnw Combining the above ideas.

\hbox_to_wd:nn \hbox_to_zero:n Put a horizontal box directly into the input stream.

(End definition for \hbox_set_to_wd:Nnw and \hbox_gset_to_wd:Nnw. These functions are documented on page 275.)

(End definition for \hbox_set:Nw and others. These functions are documented on page 275.)

(End definition for \hbox_set_to_wd:NNw and \hbox_gset_to_wd:NNw. These functions are documented on page 275.)
```
\end{definition}
\begin{flushleft}
\textbf{Put a zero-sized box with the contents pushed against one side (which makes it stick out on the other) directly into the input stream.}
\end{flushleft}

\begin{verbatim}
\cs_new_protected:Npn \hbox_overlap_center:n #1
  \{ \hbox_to_zero:n { \tex_hss:D #1 \tex_hss:D } \}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \hbox_overlap_left:n #1
  \{ \hbox_to_zero:n { \tex_hss:D #1 } \}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \hbox_overlap_right:n #1
  \{ \hbox_to_zero:n { #1 \tex_hss:D } \}
\end{verbatim}

\begin{flushleft}
\textbf{Put a zero-sized box with the contents pushed against one side (which makes it stick out on the other) directly into the input stream.}
\end{flushleft}

\begin{verbatim}
\cs_new_protected:Npn \vbox:n #1
  \{ \tex_vbox:D { \color_group_begin: #1 \par \color_group_end: } \}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \vbox_to_ht:nn #1#2
  \{ \tex_vbox:D to \__box_dim_eval:n \{ #1 \}
  \{ \color_group_begin: #2 \par \color_group_end: \}
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \vbox_to_zero:n #1
  \{ \tex_vbox:D to \c_zero_dim
  \{ \color_group_begin: #1 \par \color_group_end: \}
\end{verbatim}

\subsection{Vertical mode boxes}

\TeX{} ends these boxes directly with the internal \texttt{end_graf} routine. This means that there is no \texttt{\par} at the end of vertical boxes unless we insert one. Thus all vertical boxes include a \texttt{\par} just before closing the color group.

\begin{verbatim}
\cs_new_protected:Npn \vbox:n \vbox_top:n
\begin{verbatim}
\cs_new_protected:Npn \vbox_to_ht:nn \vbox_to_zero:n
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \vbox_to_ht:nn \vbox_to_zero:n
\end{verbatim}
```

84.11 Vertical mode boxes

\TeX{} ends these boxes directly with the internal \texttt{end_graf} routine. This means that there is no \texttt{\par} at the end of vertical boxes unless we insert one. Thus all vertical boxes include a \texttt{\par} just before closing the color group.

\begin{verbatim}
\end{verbatim}

\begin{verbatim}
\end{verbatim}

\begin{verbatim}
\end{verbatim}

\begin{verbatim}
\end{verbatim}

\subsection{Vertical mode boxes}

\TeX{} ends these boxes directly with the internal \texttt{end_graf} routine. This means that there is no \texttt{\par} at the end of vertical boxes unless we insert one. Thus all vertical boxes include a \texttt{\par} just before closing the color group.

\begin{verbatim}
\end{verbatim}

\begin{verbatim}
\end{verbatim}

\begin{verbatim}
\end{verbatim}

\begin{verbatim}
\end{verbatim}

\subsection{Vertical mode boxes}

\TeX{} ends these boxes directly with the internal \texttt{end_graf} routine. This means that there is no \texttt{\par} at the end of vertical boxes unless we insert one. Thus all vertical boxes include a \texttt{\par} just before closing the color group.

\begin{verbatim}
\end{verbatim}

\begin{verbatim}
\end{verbatim}

\begin{verbatim}
\end{verbatim}

\begin{verbatim}
\end{verbatim}

\subsection{Vertical mode boxes}

\TeX{} ends these boxes directly with the internal \texttt{end_graf} routine. This means that there is no \texttt{\par} at the end of vertical boxes unless we insert one. Thus all vertical boxes include a \texttt{\par} just before closing the color group.

\begin{verbatim}
\end{verbatim}

\begin{verbatim}
\end{verbatim}

\begin{verbatim}
\end{verbatim}

\subsection{Vertical mode boxes}

\TeX{} ends these boxes directly with the internal \texttt{end_graf} routine. This means that there is no \texttt{\par} at the end of vertical boxes unless we insert one. Thus all vertical boxes include a \texttt{\par} just before closing the color group.

\begin{verbatim}
\end{verbatim}

\begin{verbatim}
\end{verbatim}

\begin{verbatim}
\end{verbatim}

\subsection{Vertical mode boxes}

\TeX{} ends these boxes directly with the internal \texttt{end_graf} routine. This means that there is no \texttt{\par} at the end of vertical boxes unless we insert one. Thus all vertical boxes include a \texttt{\par} just before closing the color group.
Storing material in a vertical box with a natural height.

\begin{verbatim}
\cs_new_protected:Npn \vbox_set:Nn #1#2
\{
\tex_setbox:D #1 \tex_vbox:D
\{ \color_group_begin: #2 \par \color_group_end: \}
\}
\cs_new_protected:Npn \vbox_gset:Nn #1#2
\{
\tex_global:D \tex_setbox:D #1 \tex_vbox:D
\{ \color_group_begin: #2 \par \color_group_end: \}
\}
\cs_generate_variant:Nn \vbox_set:Nn { c }
\cs_generate_variant:Nn \vbox_gset:Nn { c }
\end{verbatim}

(End definition for \vbox_set:Nn and \vbox_gset:Nn. These functions are documented on page 276.)

Storing material in a vertical box with a natural height and reference point at the baseline of the first object in the box.

\begin{verbatim}
\cs_new_protected:Npn \vbox_set_top:Nn #1#2
\{
\tex_setbox:D #1 \tex_vtop:D
\{ \color_group_begin: #2 \par \color_group_end: \}
\}
\cs_new_protected:Npn \vbox_gset_top:Nn #1#2
\{
\tex_global:D \tex_setbox:D #1 \tex_vtop:D
\{ \color_group_begin: #2 \par \color_group_end: \}
\}
\cs_generate_variant:Nn \vbox_set_top:Nn { c }
\cs_generate_variant:Nn \vbox_gset_top:Nn { c }
\end{verbatim}

(End definition for \vbox_set_top:Nn and \vbox_gset_top:Nn. These functions are documented on page 276.)

Storing material in a vertical box with a specified height.

\begin{verbatim}
\cs_new_protected:Npn \vbox_set_to_ht:Nnn #1#2#3
\{
\tex_setbox:D #1 \tex_vbox:D to \__box_dim_eval:n {#2}
\{ \color_group_begin: #3 \par \color_group_end: \}
\}
\cs_new_protected:Npn \vbox_gset_to_ht:Nnn #1#2#3
\{
\tex_global:D \tex_setbox:D #1 \tex_vbox:D to \__box_dim_eval:n {#2}
\{ \color_group_begin: #3 \par \color_group_end: \}
\}
\cs_generate_variant:Nn \vbox_set_to_ht:Nnn { c }
\cs_generate_variant:Nn \vbox_gset_to_ht:Nnn { c }
\end{verbatim}

(End definition for \vbox_set_to_ht:Nnn and \vbox_gset_to_ht:Nnn. These functions are documented on page 276.)
Storing material in a vertical box. This type is useful in environment definitions.

\begin{verbatim}
\vbox_set:Nw \vbox_set:cw \vbox_gset:Nw \vbox_gset:cw \vbox_set_end: \vbox_gset_end:
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \vbox_set:Nw \vbox_set:cw
\{ \tex_setbox:D \tex_vbox:D \c_group_begin_token
\color_group_begin:
\}
\cs_new_protected:Npm \vbox_gset:Nw \vbox_gset:cw
\{ \tex_global:D \tex_setbox:D \tex_vbox:D \c_group_begin_token
\color_group_begin:
\}
\cs_generate_variant:Nn \vbox_set:Nw { c }
\cs_generate_variant:Nn \vbox_gset:Nw { c }
\cs_new_protected:Npm \vbox_set_end:
\par \color_group_end:
\c_group_end_token
\cs_new_eq:NN \vbox_gset_end: \vbox_set_end:
\end{verbatim}

(End definition for \vbox_set:Nw and others. These functions are documented on page 276.)

\begin{verbatim}
\vbox_to_ht:Nnw \vbox_to_ht:cnw \vbox_gset_to_ht:Nnw \vbox_gset_to_ht:cnw
\end{verbatim}

A combination of the above ideas.

\begin{verbatim}
\vbox_set_to_ht:Nnw \vbox_set_to_ht:cnw \vbox_gset_to_ht:Nnw \vbox_gset_to_ht:cnw
\{ \tex_setbox:D \tex_vbox:D to \_box_dim_eval:n \{\#2\}
\c_group_begin_token
\color_group_begin:
\}
\cs_new_protected:Npm \vbox_gset_to_ht:Nnw \vbox_gset_to_ht:cnw
\{ \tex_global:D \tex_setbox:D \tex_vbox:D to \_box_dim_eval:n \{\#2\}
\c_group_begin_token
\color_group_begin:
\}
\cs_generate_variant:Nn \vbox_set_to_ht:Nnw { c }
\cs_generate_variant:Nn \vbox_gset_to_ht:Nnw { c }
\end{verbatim}

(End definition for \vbox_set_to_ht:Nnw and \vbox_gset_to_ht:Nnw. These functions are documented on page 277.)

\begin{verbatim}
\vbox_unpack:N \vbox_unpack:c \vbox_unpack_drop:N \vbox_unpack_drop:c
\end{verbatim}

Unpacking a box and if requested also clear it.

\begin{verbatim}
\cs_new_eq:NN \vbox_unpack:N \tex_unvcopy:D
\cs_new_eq:NN \vbox_unpack_drop:N \tex_unvbox:D
\{ \tex_unvcopy:D \vbox_unpack:N \vbox_unpack_drop:N \vbox_unpack_drop:c
\}{ \vbox_unpack:N \vbox_unpack_drop:N \vbox_unpack_drop:c
\end{verbatim}

(End definition for \vbox_unpack:N and \vbox_unpack_drop:N. These functions are documented on page 277.)
\vbox_set_split_to_ht:NNn\n\vbox_set_split_to_ht:cNn\n\vbox_set_split_to_ht:Ncn\n\vbox_set_split_to_ht:ccn\n\vbox_gset_split_to_ht:NNn\n\vbox_gset_split_to_ht:cNn\n\vbox_gset_split_to_ht:Ncn\n\vbox_gset_split_to_ht:ccn

Splitting a vertical box in two.

\cs_new_protected:Npn \vbox_set_split_to_ht:NNn #1#2#3
\{ \tex_setbox:D #1 \tex_vsplit:D #2 to \_\_box_dim_eval:n \{#3\} \}
\cs_generate_variant:Nn \vbox_set_split_to_ht:NNn { c , Nc , cc }
\cs_new_protected:Npn \vbox_gset_split_to_ht:NNn #1#2#3
\{ \tex_global:D \tex_setbox:D #1 \tex_vsplit:D #2 to \_\_box_dim_eval:n \{#3\} \}
\cs_generate_variant:Nn \vbox_gset_split_to_ht:NNn { c , Nc , cc }

(End definition for \vbox_set_split_to_ht:NNn and \vbox_gset_split_to_ht:NNn. These functions are documented on page 277.)

84.12 Affine transformations

\l__box_angle_fp When rotating boxes, the angle itself may be needed by the engine-dependent code. This is done using the \texttt{fp} module so that the value is tidied up properly.
\fp_new:N \l__box_angle_fp
(End definition for \l__box_angle_fp.)

\l__box_cos_fp \l__box_sin_fp These are used to hold the calculated sine and cosine values while carrying out a rotation.
\fp_new:N \l__box_sin_fp
\fp_new:N \l__box_cos_fp
(End definition for \l__box_sin_fp and \l__box_sin_fp.)

\l__box_top_dim These are the positions of the four edges of a box before manipulation.
\l__box_bottom_dim\l__box_left_dim\l__box_right_dim
\dim_new:N \l__box_top_dim
\dim_new:N \l__box_bottom_dim
\dim_new:N \l__box_left_dim
\dim_new:N \l__box_right_dim
(End definition for \l__box_top_dim and others.)

\l__box_top_new_dim These are the positions of the four edges of a box after manipulation.
\l__box_bottom_new_dim\l__box_left_new_dim\l__box_right_new_dim
\dim_new:N \l__box_top_new_dim
\dim_new:N \l__box_bottom_new_dim
\dim_new:N \l__box_left_new_dim
\dim_new:N \l__box_right_new_dim
(End definition for \l__box_top_new_dim and others.)

\l__box_internal_box Scratch space, but also needed by some parts of the driver.
\box_new:N \l__box_internal_box
(End definition for \l__box_internal_box.)
Rotation of a box starts with working out the relevant sine and cosine. The actual rotation is in an auxiliary to keep the flow slightly clearer.

Rotation of a box starts with working out the relevant sine and cosine. The actual rotation is in an auxiliary to keep the flow slightly clearer.

\def\box_rotate#1#2{
  \__box_rotate#1{#2}{\hbox_set}
}\cs_generate_variant{\box_rotate}{c}
\def\box_grotate#1#2{
  \__box_rotate#1{#2}{\hbox_gset}
}\cs_generate_variant{\box_grotate}{c}
\def\__box_rotate#1#2#3{
  \fp_set{\l__box_angle_fp}{#2}
  \fp_set{\l__box_sin_fp}{\sind{\l__box_angle_fp}}
  \fp_set{\l__box_cos_fp}{\cosd{\l__box_angle_fp}}
  \__box_rotate#1
}\cs_new_protected:Npn \__box_rotate:N #1\#2\#3

The edges of the box are then recorded: the left edge is always at zero. Rotation of the four edges then takes place: this is most efficiently done on a quadrant by quadrant basis.

\def\__box_rotate:N #1
  \dim_set{\l__box_top_dim}{\box_ht{#1}}
  \dim_set{\l__box_bottom_dim}{-\box_dp{#1}}
  \dim_set{\l__box_right_dim}{\box_wd{#1}}
  \dim_zero{\l__box_left_dim}
\cs_new_protected:Npn \__box_rotate:N #1

\def\__box_rotate:N #1\#2\#3

The next step is to work out the x and y coordinates of vertices of the rotated box in relation to its original coordinates. The box can be visualized with vertices B, C, D and E is illustrated (Figure 1). The vertex O is the reference point on the baseline, and in this implementation is also the centre of rotation. The formulae are, for a point \( P \) and angle \( \alpha \):

\[
\begin{align*}
P_x'' &= P_x - O_x \\
P_y'' &= P_y - O_y \\
P_x''' &= (P_x' \cos(\alpha)) - (P_y' \sin(\alpha)) \\
P_y''' &= (P_x' \sin(\alpha)) + (P_y' \cos(\alpha)) \\
P_x''' &= P_x'' + O_x + L_x \\
P_y''' &= P_y'' + O_y
\end{align*}
\]

\begin{figure}
\centering
\begin{tikzpicture}
  \draw (0,0) -- (2,0) -- (2,2) -- (0,2) -- cycle;
  \draw[ dashed] (0,0) -- (0,2);
  \draw[ dashed] (2,0) -- (2,2);
  \node at (0,0) {O}; \node at (2,2) {A}; \node at (0,2) {B}; \node at (2,0) {C}; \node at (1,1) {D}; \node at (1,1.5) {E}; \\
\end{tikzpicture}
\caption{Co-ordinates of a box prior to rotation.}
\end{figure}
The “extra” horizontal translation $L_x$ at the end is calculated so that the leftmost point of the resulting box has $x$-coordinate 0. This is desirable as \TeX boxes must have the reference point at the left edge of the box. (As $O$ is always $(0,0)$, this part of the calculation is omitted here.)

\begin{verbatim}
\fp_compare:nNnTF \l__box_sin_fp > \c_zero_fp
{ \fp_compare:nNnTF \l__box_cos_fp > \c_zero_fp
  { \__box_rotate_quadrant_one: }
  { \__box_rotate_quadrant_two: }
}
{ 
  \fp_compare:nNnTF \l__box_cos_fp < \c_zero_fp
  { \__box_rotate_quadrant_three: }
  { \__box_rotate_quadrant_four: }
}
\end{verbatim}

The position of the box edges are now known, but the box at this stage be misplaced relative to the current \TeX reference point. So the content of the box is moved such that the reference point of the rotated box is in the same place as the original.

\begin{verbatim}
\hbox_set:Nn \l__box_internal_box { \box_use:N #1 }
\hbox_set:Nn \l__box_internal_box
{ \__kernel_kern:n { -\l__box_left_new_dim } \hbox:n
  { \__box_backend_rotate:Nn \l__box_internal_box \l__box_angle_fp }
}
\end{verbatim}

Tidy up the size of the box so that the material is actually inside the bounding box. The result can then be used to reset the original box.

\begin{verbatim}
\box_set_ht:Nn \l__box_internal_box { \l__box_top_new_dim }
\box_set_dp:Nn \l__box_internal_box { -\l__box_bottom_new_dim }
\box_set_wd:Nn \l__box_internal_box { \l__box_right_new_dim - \l__box_left_new_dim }
\box_use_drop:N \l__box_internal_box
\end{verbatim}

These functions take a general point $(#1,#2)$ and rotate its location about the origin, using the previously-set sine and cosine values. Each function gives only one component of the location of the updated point. This is because for rotation of a box each step needs only one value, and so performance is gained by avoiding working out both $x'$ and $y'$ at the same time. Contrast this with the equivalent function in the \texttt{l3coffins} module, where both parts are needed.

\begin{verbatim}
\cs_new_protected:Npn \__box_rotate_xdir:n { \fp_to_dim:n { \l__box_cos_fp * \dim_to_fp:n {#1} - \l__box_sin_fp * \dim_to_fp:n {#2} }
\end{verbatim}
Rotation of the edges is done using a different formula for each quadrant. In every case, the top and bottom edges only need the resulting $y$-values, whereas the left and right edges need the $x$-values. Each case is a question of picking out which corner ends up at with the maximum top, bottom, left and right value. Doing this by hand means a lot less calculating and avoids lots of comparisons.

\begin{verbatim}
\cs_new_protected:Npn \__box_rotate_quadrant_one: 
  { 
    \__box_rotate_ydir:nnN \l__box_right_dim \l__box_top_dim 
    \l__box_right_new_dim 
    \__box_rotate_ydir:nnN \l__box_left_dim \l__box_bottom_dim 
    \l__box_left_new_dim 
    \__box_rotate_xdir:nnN \l__box_right_dim \l__box_top_dim 
    \l__box_left_new_dim 
    \__box_rotate_xdir:nnN \l__box_left_dim \l__box_bottom_dim 
    \l__box_right_new_dim 
  }
\cs_new_protected:Npn \__box_rotate_quadrant_two: 
  { 
    \__box_rotate_ydir:nnN \l__box_right_dim \l__box_bottom_dim 
    \l__box_top_new_dim 
    \__box_rotate_ydir:nnN \l__box_left_dim \l__box_top_dim 
    \l__box_bottom_new_dim 
    \__box_rotate_xdir:nnN \l__box_right_dim \l__box_bottom_dim 
    \l__box_left_new_dim 
    \__box_rotate_xdir:nnN \l__box_left_dim \l__box_top_dim 
    \l__box_right_new_dim 
  }
\cs_new_protected:Npn \__box_rotate_quadrant_three: 
  { 
    \__box_rotate_ydir:nnN \l__box_left_dim \l__box_bottom_dim 
    \l__box_top_new_dim 
    \__box_rotate_ydir:nnN \l__box_right_dim \l__box_top_dim 
    \l__box_bottom_new_dim 
    \__box_rotate_xdir:nnN \l__box_right_dim \l__box_bottom_dim 
    \l__box_left_new_dim 
    \__box_rotate_xdir:nnN \l__box_left_dim \l__box_top_dim 
    \l__box_right_new_dim 
  }
\cs_new_protected:Npn \__box_rotate_quadrant_four: 
  { 
    \__box_rotate_ydir:nnN \l__box_left_dim \l__box_bottom_dim 
    \l__box_top_new_dim 
    \__box_rotate_ydir:nnN \l__box_right_dim \l__box_top_dim 
    \l__box_bottom_new_dim 
    \__box_rotate_xdir:nnN \l__box_right_dim \l__box_bottom_dim 
    \l__box_left_new_dim 
    \__box_rotate_xdir:nnN \l__box_left_dim \l__box_top_dim 
    \l__box_right_new_dim 
  }
\end{verbatim}
Definitions for \_\_box_rotate, \_\_box_scale, and \_\_box_resize:

\_\_box_rotate_ydir:nnN \l__box_left_dim \l__box_top_dim \l__box_top_new_dim \l__box_right_dim \l__box_bottom_dim \l__box_bottom_new_dim
\_\_box_rotate_xdir:nnN \l__box_left_dim \l__box_bottom_dim \l__box_left_new_dim \l__box_right_dim \l__box_top_dim \l__box_top_new_dim

\_\_box_scale_x_fp \l__box_scale_y_fp

Scaling is potentially-different in the two axes.

\fp_new:N \l__box_scale_x_fp
\fp_new:N \l__box_scale_y_fp

Resizing a box starts by working out the various dimensions of the existing box.

\cs_new_protected:Npn \box_resize_to_wd_and_ht_plus_dp:Nnn #1#2#3
\{ \__box_resize_to_wd_and_ht_plus_dp:NnnN #1 {#2} {#3} \hbox_set:Nn \}
\cs_generate_variant:Nn \box_resize_to_wd_and_ht_plus_dp:Nnn { c }
\cs_new_protected:Npn \box_gresize_to_wd_and_ht_plus_dp:Nnn #1#2#3
\{ \__box_resize_to_wd_and_ht_plus_dp:NnnN #1 {#2} {#3} \hbox_gset:Nn \}
\cs_generate_variant:Nn \box_gresize_to_wd_and_ht_plus_dp:Nnn { c }
\cs_new_protected:Npn \__box_resize_to_wd_and_ht_plus_dp:NnnN #1#2#3#4
\{ #4 #1 \{ \__box_resize_set_corners:N #1 \}
\cs_new_protected:Npm \__box_resize:N \__box_resize:NNN
\cs_new_protected:Npn \__box_resize_set_corners:N #1
\{ \}
\cs_generate_variant:Nn \__box_resize_set_corners:N { c }
\cs_new_protected:Npm \__box_resize_to_wd_and_ht_plus_dp:NnnN \__box_resize_set_corners:N \__box_resize:N
\__box_resize:N

The \textit{x}-scaling and resulting box size is easy enough to work out: the dimension is that given as \#2, and the scale is simply the new width divided by the old one.

\fp_set:Nn \l__box_scale_x_fp
{ \dim_to_fp:n {#2} / \dim_to_fp:n { \l__box_right_dim } }

The \textit{y}-scaling needs both the height and the depth of the current box.

\fp_set:Nn \l__box_scale_y_fp
{ \dim_to_fp:n {\l__box_right_dim} / \dim_to_fp:n { \l__box_top_dim - \l__box_bottom_dim } }

Hand off to the auxiliary which does the rest of the work.

\cs_new_protected:Npm \__box_resize:N \__box_resize_set_corners:N \__box_resize:N

With at least one real scaling to do, the next phase is to find the new edge co-ordinates. In the x direction this is relatively easy: just scale the right edge. In the y direction, both dimensions have to be scaled, and this again needs the absolute scale value. Once that is all done, the common resize/rescale code can be employed.

Scaling to a (total) height or to a width is a simplified version of the main resizing operation, with the scale simply copied between the two parts. The internal auxiliary is called using the scaling value twice, as the sign for both parts is needed (as this allows the same internal code to be used as for the general case).
When scaling a box, setting the scaling itself is easy enough. The new dimensions are also relatively easy to find, allowing only for the need to keep them positive in all cases. Once that is done then after a check for the trivial scaling a hand-off can be made to the common code. The code here is split into two as this allows sharing with the auto-resizing functions.

\begin{verbatim}
\cs_new_protected:Npn \box_scale:Nnn #1#2#3
{ \__box_scale:NnnN #1 {#2} {#3} \hbox_set:Nn }
\cs_generate_variant:Nn \box_scale:Nnn { c }
\cs_new_protected:Npn \box_gscale:Nnn #1#2#3
{ \__box_scale:NnnN #1 {#2} {#3} \hbox_gset:Nn }
\cs_generate_variant:Nn \box_gscale:Nnn { c }
\cs_new_protected:Npn \__box_scale:NnnN #1#2#3#4
{ #4 #1
  { \fp_set:Nn \l__box_scale_x_fp {#2} }
  { \fp_set:Nn \l__box_scale_y_fp {#3} }
  \__box_scale:N #1 }
\cs_new_protected:Npn \__box_scale:N #1
{ \dim_set:Nn \l__box_top_dim { \box_ht:N #1 }
  \dim_set:Nn \l__box_bottom_dim { -\box_dp:N #1 }
  \dim_set:Nn \l__box_right_dim { \box_wd:N #1 }
  \dim_zero:N \l__box_left_dim
  \dim_set:Nn \l__box_top_new_dim { \fp_abs:n { \l__box_scale_y_fp } \l__box_top_dim }
  \dim_set:Nn \l__box_bottom_new_dim { \fp_abs:n { \l__box_scale_y_fp } \l__box_bottom_dim }
  \dim_set:Nn \l__box_right_new_dim { \fp_abs:n { \l__box_scale_x_fp } \l__box_right_dim }
  \__box_resize_common:N #1 }
\end{verbatim}

Although autosizing a box uses dimensions, it has more in common in implementation with scaling. As such, most of the real work here is done elsewhere.
\hbox_set:Nn
\cs_generate_variant:Nn \box_autosize_to_wd_and_ht_plus_dp:Nnn { c }
\cs_new_protected:Np \box_gautosize_to_wd_and_ht_plus_dp:Nnn { \hbox_gset:Nn }
\cs_generate_variant:Nn \box_gautosize_to_wd_and_ht_plus_dp:Nnn { c }
\cs_new_protected:Npn \__box_autosize:NnnnN #1 #2 #3 #4 #5
\__box_resize_common:N
The main resize function places its input into a box which start off with zero width, and includes the handles for engine rescaling.
\cs_new_protected:Np \__box_resize_common:N #1
\hbox_set:Nn \l__box_internal_box
\__box_backend_scale:Nnn \l__box_scale_x_fp \l__box_scale_y_fp
\fp_compare:nNnTF \l__box_scale_y_fp > \c_zero_fp
\box_set_ht:Nn \l__box_internal_box \l__box_top_new_dim
\box_set_dp:Nn \l__box_internal_box \l__box_bottom_new_dim
\fp_compare:nNnTF \l__box_scale_x_fp < \c_zero_fp
\hbox_to_wd:nn \l__box_right_new_dim
The new height and depth can be applied directly.
\fp_compare:nNnTF \l__box_scale_y_fp > \c_zero_fp
\box_set_ht:Nn \l__box_internal_box \l__box_top_new_dim
\box_set_dp:Nn \l__box_internal_box \l__box_bottom_new_dim
\fp_compare:nNnTF \l__box_scale_x_fp < \c_zero_fp
\hbox_to_wd:nn \l__box_right_new_dim

(End definition for \box_autosize_to_wd_and_ht:#nn and others. These functions are documented on page 270.)
\{ \__kernel_kern:n \{ \l__box_right_new_dim \} \box_use_drop:N \l__box_internal_box \tex_hss:D \}

\}

\}

\box_set_wd:Nn \l__box_internal_box \{ \l__box_right_new_dim \} \hbox:n

\}

\}

\}

\}

\}

\}

\}

\}

\}

\}

\}

\}

\}

\}

\}

\}

\}

\}

\}

\}

(End definition for \__box_resize_common:N.)

(//package)
Chapter 85

\l3coffins Implementation

85.1 Coffins: data structures and general variables

\l__coffin_internal_box  Scratch variables.
\l__coffin_internal_dim
\l__coffin_internal_tl

(End definition for \l__coffin_internal_box, \l__coffin_internal_dim, and \l__coffin_internal_tl.)

\c__coffin_corners_prop  The “corners”; of a coffin define the real content, as opposed to the \TeX{ bounding box. They all start off in the same place, of course.

(End definition for \c__coffin_corners_prop.)

\c__coffin_poles_prop  Pole positions are given for horizontal, vertical and reference-point based values.

(End definition for \c__coffin_poles_prop.)
\l__coffin_slope_A_fp
\l__coffin_slope_B_fp

Used for calculations of intersections.

\fp_new:N \l__coffin_slope_A_fp
\fp_new:N \l__coffin_slope_B_fp

For propagating errors so that parts of the code can work around them.

\bool_new:N \l__coffin_error_bool

The offset between two sets of coffin handles when typesetting. These values are corrected from those requested in an alignment for the positions of the handles.

\dim_new:N \l__coffin_offset_x_dim
\dim_new:N \l__coffin_offset_y_dim

Needed for finding the intersection of two poles.

\tl_new:N \l__coffin_pole_a_tl
\tl_new:N \l__coffin_pole_b_tl

For calculating intersections and so forth.

\dim_new:N \l__coffin_x_dim
\dim_new:N \l__coffin_y_dim
\dim_new:N \l__coffin_x_prime_dim
\dim_new:N \l__coffin_y_prime_dim

\l__coffin_poles_prop

85.2 Basic coffin functions

There are a number of basic functions needed for creating coffins and placing material in them. This all relies on the following data structures.

\__coffin_to_value:N

Coffins are a two-part structure and we rely on the internal nature of box allocation to make everything work. As such, we need an interface to turn coffin identifiers into numbers. For the purposes here, the signature allowed is \text{N} despite the nature of the underlying primitive.

\cs_new_eq:NN \__coffin_to_value:N \tex_number:D
Several of the higher-level coffin functions would give multiple errors if the coffin does not exist. A cleaner way to handle this is provided here: both the box and the coffin structure are checked.

\begin{verbatim}
\texttt{\prg_new_conditional:Nnnn \coffin_if_exist:N #1 { p , T , F , TF }}
{ \cs_if_exist:NTF #1 }
{ \cs_if_exist:cTF \{ coffin \ - \_coffin_to_value:N #1 - poles \}
{ \prg_return_true: }
{ \prg_return_false: }
}
{ \prg_return_false: }
\prg_generate_conditional_variant:Nnn \coffin_if_exist:N
\{ c \} \{ p , T , F , TF \}
\end{verbatim}

(End definition for \texttt{\coffin_if_exist:NTF}. This function is documented on page 282.)

\_coffin_if_exist:NT Several of the higher-level coffin functions would give multiple errors if the coffin does not exist. So a wrapper is provided to deal with this correctly, issuing an error on erroneous use.

\begin{verbatim}
\texttt{\cs_new_protected:Npn \_coffin_if_exist:NT #1#2}
{ \coffin_if_exist:NTF #1 }
{ #2 }
{ \msg_error:nx \{ coffin \} \{ unknown \}
{ \token_to_str:N #1 }
}
\end{verbatim}

(End definition for \texttt{\_coffin_if_exist:NT}.)

\coffin_clear:N Clearing coffins means emptying the box and resetting all of the structures.

\begin{verbatim}
\texttt{\cs_new_protected:Npn \coffin_clear:N #1}
\{ \_coffin_if_exist:NT #1 }
\{ box_clear:N #1 \_coffin_reset_structure:N #1 \}
\end{verbatim}

(End definition for \texttt{\coffin_clear:N} and \texttt{\coffin_gclear:N}. These functions are documented on page 282.)
Creating a new coffin means making the underlying box and adding the data structures. The \debug_suspend: and \debug_resume: functions prevent \prop_gclear_new: from writing useless information to the log file.

```
\cs_new_protected:Npn \coffin_new:N #1
{ \box_new:N #1 \debug_suspend: \prop_gclear_new:c { coffin ~ \__coffin_to_value:N #1 ~ corners } \prop_gclear_new:c { coffin ~ \__coffin_to_value:N #1 ~ poles } \prop_gset_eq:cN { coffin ~ \__coffin_to_value:N #1 ~ corners } \c__coffin_corners_prop \prop_gset_eq:cN { coffin ~ \__coffin_to_value:N #1 ~ poles } \c__coffin_poles_prop \debug_resume: }
\cs_generate_variant:Nn \coffin_new:N { c }
```

(End definition for \coffin_new:N. This function is documented on page 282.)

Horizontal coffins are relatively easy: set the appropriate box, reset the structures then update the handle positions.

```
\hcoffin_set:Nn \hcoffin_set:cn \hcoffin_gset:Nn \hcoffin_gset:cn
\cs_new_protected:Npm \hcoffin_set:Nn #1#2
{ \__coffin_if_exist:NT #1
  { \hbox_set:Nn #1
    { \color_ensure_current: #2
      \__coffin_update:N #1
    }
    \__coffin_update:N #1
  }
}\cs_generate_variant:Nn \hcoffin_set:Nn { c }
\cs_new_protected:Npm \hcoffin_gset:Nn #1#2
{ \__coffin_if_exist:NT #1
  { \hbox_gset:Nn #1
    { \color_ensure_current: #2
      \__coffin_gupdate:N #1
    }
    \__coffin_gupdate:N #1
  }
}\cs_generate_variant:Nn \hcoffin_gset:Nn { c }
```

(End definition for \hcoffin_set:Nn and \hcoffin_gset:Nn. These functions are documented on page 283.)

Setting vertical coffins is more complex. First, the material is typeset with a given width. The default handles and poles are set as for a horizontal coffin, before finding the top baseline using a temporary box. No \color_ensure_current: here as that would add a
whatsit to the start of the vertical box and mess up the location of the T pole (see \TeX by Topic for discussion of the \vtop primitive, used to do the measuring).

\begin{verbatim}
\cs_new_protected:Npn \vcoffin_set:Nnn #1#2#3 {
  \__coffin_set_vertical:NnnNN #1 {#2} {#3}
  \vbox_set:Nn \__coffin_update:N
}
\cs_generate_variant:Nn \vcoffin_set:Nnn { c }
\cs_new_protected:Npn \vcoffin_gset:Nnn #1#2#3 {
  \__coffin_set_vertical:NnnNN #1 {#2} {#3}
  \vbox_gset:Nn \__coffin_gupdate:N
}
\cs_generate_variant:Nn \vcoffin_gset:Nnn { c }
\cs_new_protected:Npx \__coffin_set_vertical:NnnNN #1#2#3#4#5 {
  \__coffin_if_exist:NT #1 {
    #4 #1
    {
      \dim_set:Nn \tex_hsize:D {#2}
      \__coffin_set_vertical_aux: #3
    }
  }
  #5 #1
  \vbox_set_top:Nn \l__coffin_internal_box { \vbox_unpack:N #1 }
  \__coffin_set_pole:Nnx #1 { T }
  {
    \dim_eval:n { \box_ht:N #1 - \box_ht:N \l__coffin_internal_box }
    { 1000pt }
    { Opt }
  }
  \box_clear:N \l__coffin_internal_box
}
\cs_new_protected:Npx \__coffin_set_vertical_aux:
  \bool_lazy_and:nnT { \cs_if_exist_p:N \fmtname }
  { \str_if_eq_p:Vn \fmtname { LaTeX2e } }
  {
    \dim_set_eq:NN \exp_not:N \linewidth \tex_hsize:D
    \dim_set_eq:NN \exp_not:N \columnwidth \tex_hsize:D
  }
\end{verbatim}
(End definition for \vcoffin_set:Nnn and others. These functions are documented on page 283.)

These are the “begin”/“end” versions of the above: watch the grouping!

\begin{verbatim}
\hcoffin_set:Nw \hcoffin_set:cw
\hcoffin_gset:Nw \hcoffin_gset:cw
\hcoffin_set_end:
\hcoffin_gset_end:
\end{verbatim}
The same for vertical coffins.

(End definition for \hcoffin_set:Nw and others. These functions are documented on page 283.)
33144  #5
33145  #6 #1
33146  \vbox_set_top:Nn \l__coffin_internal_box { \vbox_unpack:N #1 }
33147  \__coffin_set_pole:Nnx #1 { T }
33148  {  
33149  {  
33150  \dim_eval:n  
33151  {  \box_ht:N #1 - \box_ht:N \l__coffin_internal_box  }
33152  }  
33153  {  1000pt  }
33154  {  Opt  }
33155  }  
33156  \box_clear:N \l__coffin_internal_box  
33157  }  
33158  \cs_new_protected:Npn \vcoffin_set_end: { } 
33159  \cs_new_protected:Npn \vcoffin_gset_end: { } 
33160  (End definition for \vcoffin_set:Nnw and others. These functions are documented on page 283.)
33161  \coffin_set_eq:NN \coffin_set_eq:Nc \coffin_set_eq:cN \coffin_set_eq:cc
33162  \coffin_gset_eq:NN \coffin_gset_eq:Nc \coffin_gset_eq:cN \coffin_gset_eq:cc
33163  Setting two coffins equal is just a wrapper around other functions.
33164  \cs_new_protected:Npm \coffin_set_eq:NN #1#2
33165  {  
33166  \__coffin_if_exist:NT #1  
33167  {  
33168  \box_set_eq:NN #1 #2  
33169  \prop_set_eq:cc { coffin - \__coffin_to_value:N #1 - corners }  
33170  { coffin - \__coffin_to_value:N #2 - corners }  
33171  \prop_set_eq:cc { coffin - \__coffin_to_value:N #1 - poles }  
33172  { coffin - \__coffin_to_value:N #2 - poles }  
33173  }  
33174  }  
33175  \cs_generate_variant:Nn \coffin_set_eq:NN { c , Nc , cc }
33176  \cs_new_protected:Npm \coffin_gset_eq:NN #1#2
33177  {  
33178  \__coffin_if_exist:NT #1  
33179  {  
33180  \box_gset_eq:NN #1 #2  
33181  \prop_gset_eq:cc { coffin - \__coffin_to_value:N #1 - corners }  
33182  { coffin - \__coffin_to_value:N #2 - corners }  
33183  \prop_gset_eq:cc { coffin - \__coffin_to_value:N #1 - poles }  
33184  { coffin - \__coffin_to_value:N #2 - poles }  
33185  }  
33186  }  
33187  \cs_generate_variant:Nn \coffin_gset_eq:NN { c , Nc , cc }
33188  (End definition for \coffin_set_eq:NN and \coffin_gset_eq:NN. These functions are documented on page 282.)
33189  \c_empty_coffin \l__coffin_aligned_coffin \l__coffin_aligned_internal_coffin
33190  Special coffins: these cannot be set up earlier as they need \coffin_new:N. The empty coffin is set as a box as the full coffin-setting system needs some material which is not yet available. The empty coffin is created entirely by hand: not everything is in place yet.
85.3 Measuring coffins

Coffins are just boxes when it comes to measurement. However, semantically a separate set of functions are required.

85.4 Coffins: handle and pole management

A simple wrapper around the recovery of a coffin pole, with some error checking and recovery built-in.
Setting the pole of a coffin at the user/designer level requires a bit more care. The idea here is to provide a reasonable interface to the system, then to do the setting with full expansion. The three-argument version is used internally to do a direct setting.

```latex
\cs_new_protected:Npn \coffin_set_horizontal_pole:Nnn #1#2#3
  { \__coffin_set_horizontal_pole:NnnN #1 {#2} {#3} \prop_put:cnx }
\cs_generate_variant:Nn \coffin_set_horizontal_pole:Nnn { c }
\cs_new_protected:Npn \coffin_gset_horizontal_pole:Nnn #1#2#3
  { \__coffin_set_horizontal_pole:NnnN #1 {#2} {#3} \prop_gput:cnx }
\cs_generate_variant:Nn \coffin_gset_horizontal_pole:Nnn { c }
\cs_new_protected:Npn \__coffin_set_horizontal_pole:NnnN #1#2#3#4
  { \__coffin_if_exist:NT #1
  { \__coffin_if_exist:NT #2
    { \__coffin_if_exist:NT #3
      { \__coffin_if_exist:NT #4
        \__coffin_set_pole:Nnn #1#2#3
        \__coffin_set_pole:Nnx }
      { 0pt } { \dim_eval:n {#3} }
      { 1000pt } { 0pt }
    }
  }
  {\__coffin_if_exist:NT #4
    \__coffin_set_pole:Nnn #1#2#3
    \__coffin_set_pole:Nnx }
}
```

```latex
\cs_new_protected:Npm \__coffin_greset_structure:N #1
  { \prop_gset_eq:cN { coffin \__coffin_to_value:N #1 \corners } \c__coffin_corners_prop }
\cs_new_protected:Npn \__coffin_greset_structure:N #1
  { \prop_gset_eq:cN { coffin \__coffin_to_value:N #1 \poles } \c__coffin_poles_prop }
```

(End definition for \__coffin_reset_structure:N and \__coffin_greset_structure:N.)
\__coffin_update:N\cs_new_protected:Npm \__coffin_update:N #1
\__coffin_reset_structure:N #1 \__coffin_update_corners:N #1 \__coffin_update_poles:N #1

\__coffin_gupdate:N #1 \__coffin_greset_structure:N #1 \__coffin_gupdate_corners:N #1 \__coffin_gupdate_poles:N #1

\__coffin_update_corners:N\cs_new_protected:Npm \__coffin_update_corners:N #1 #2
\exp_args:Nc \__coffin_update_corners:NNN { coffin \__coffin_to_value:N #1 corners #1 #2 { tl } { 0pt \dim_eval:n { \box_ht:N #2 } } #1 #2 { tr } { \dim_eval:n { \box_wd:N #2 } \dim_eval:n { \box_ht:N #2 } } #1 #2 { bl } { 0pt \dim_eval:n { -\box_dp:N #2 } } #1 #2 { br } { \dim_eval:n { \box_wd:N #2 } \dim_eval:n { -\box_dp:N #2 } } }\__coffin_gupdate_corners:NNN #1#2#3

Simple shortcuts.

Updating the corners of a coffin is straightforward as at this stage there can be no rotation. So the corners of the content are just those of the underlying \TeX box.
This function is called when a coffin is set, and updates the poles to reflect the nature of size of the box. Thus this function only alters poles where the default position is dependent on the size of the box. It also does not set poles which are relevant only to vertical coffins.

\cs_new_protected:Npn \__coffin_update_poles:N \#1
\cs_new_protected:Npn \__coffin_gupdate_poles:N \#1
\cs_new_protected:Npn \__coffin_update_poles:NN \#1\#2
\cs_new_protected:Npn \__coffin_update_poles:NNN \#1\#2\#3

(End definition for \__coffin_update_corners:N and others.)
85.5 Coffins: calculation of pole intersections

The lead off in finding intersections is to recover the two poles and then hand off to the auxiliary for the actual calculation. There may of course not be an intersection, for which an error trap is needed.

\cs_new_protected:Npn \__coffin_calculate_intersection:Nnn #1#2#3
\__coffin_get_pole:NnN #1 {#2} \l__coffin_pole_a_tl
\__coffin_get_pole:NnN #1 {#3} \l__coffin_pole_b_tl
\bool_set_false:N \l__coffin_error_bool
\exp_last_two_unbraced:Noo
\__coffin_calculate_intersection:nnnnnnnn
\l__coffin_pole_a_tl \l__coffin_pole_b_tl
\bool_if:NT \l__coffin_error_bool
\msg_error:nn { coffin } { no-pole-intersection }
\dim_zero:N \l__coffin_x_dim
\dim_zero:N \l__coffin_y_dim
}

The two poles passed here each have four values (as dimensions), \((a, b, c, d)\) and \((a', b', c', d')\). These are arguments 1–4 and 5–8, respectively. In both cases \(a\) and \(b\) are the co-ordinates of a point on the pole and \(c\) and \(d\) define the direction of the pole. Finding the intersection depends on the directions of the poles, which are given by \(d/c\) and \(d'/c'\). However, if one of the poles is either horizontal or vertical then one or more of \(c, d, c'\) and \(d'\) are zero and a special case is needed.

\cs_new_protected:Npn \__coffin_calculate_intersection:nnnnnnnn
\#1#2#3#4#5#6#7#8
\dim_compare:nNnTF {#3} = \c_zero_dim

The case where the first pole is vertical. So the \(x\)-component of the interaction is at \(a\). There is then a test on the second pole: if it is also vertical then there is an error.

\dim_set:Nn \l__coffin_x_dim {#1}
\dim_compare:nNnTF {#7} = \c_zero_dim
\bool_set_true:N \l__coffin_error_bool

The second pole may still be horizontal, in which case the \(y\)-component of the intersection is \(b'\). If not,
\[
y = \frac{d'}{c'} (a - a') + b'
\]
with the \(x\)-component already known to be \#1.

\dim_compare:nNnTF {#8} = \c_zero_dim
\#6
\dim_compare:nNnTF {#8} = \c_zero_dim
\#6

If the first pole is not vertical then it may be horizontal. If so, then the procedure is essentially the same as that already done but with the $x$- and $y$-components interchanged.

Now we deal with the case where the second pole may be vertical, or if not we have

$$x = \frac{c'}{d'} (b - b') + a'$$

which is again handled by the same auxiliary.

The first pole is neither horizontal nor vertical. To avoid even more complexity, we now work out both slopes and pass to an auxiliary.

Assuming the two poles are not parallel, then the intersection point is found in two steps. First we find the $x$-value with

$$x = \frac{sa - s'a' - b + b'}{s - s'}$$

1297
and then finding the $y$-value with

$$y = s(x - a) + b$$

85.6 Affine transformations

\l__coffin_sin_fp
\l__coffin_cos_fp

Used for rotations to get the sine and cosine values.

(End definition for \l__coffin_sin_fp and \l__coffin_cos_fp.)

\l__coffin_bounding_prop

A property list for the bounding box of a coffin. This is only needed during the rotation, so there is just the one.

(End definition for \l__coffin_bounding_prop.)

\l__coffin_corners_prop \l__coffin_poles_prop

Used to avoid needing to track scope for intermediate steps.

(End definition for \l__coffin_corners_prop and \l__coffin_poles_prop.)
The shift of the bounding box of a coffin from the real content.
\dim_new:N \l__coffin_bounding_shift_dim

These are used to hold maxima for the various corner values: these thus define the minimum size of the bounding box after rotation.
\dim_new:N \l__coffin_left_corner_dim
\dim_new:N \l__coffin_right_corner_dim
\dim_new:N \l__coffin_bottom_corner_dim
\dim_new:N \l__coffin_top_corner_dim

Rotating a coffin requires several steps which can be conveniently run together. The sine and cosine of the angle in degrees are computed. This is then used to set \l__coffin_sin_fp and \l__coffin_cos_fp, which are carried through unchanged for the rest of the procedure.
\cs_new_protected:Npn \coffin_rotate:Nn #1#2
{ \__coffin_rotate:NnNNN #1 {#2} \box_rotate:Nn \prop_set_eq:cN \hbox_set:Nn }
\cs_generate_variant:Nn \coffin_rotate:Nn { c }
\cs_new_protected:Npn \coffin_grotate:Nn #1#2
{ \__coffin_rotate:NnNNN #1 {#2} \box_grotate:Nn \prop_gset_eq:cN \hbox_gset:Nn }
\cs_generate_variant:Nn \coffin_grotate:Nn { c }
\cs_new_protected:Npn \__coffin_rotate:NnNNN #1#2#3#4#5
{\fp_set:Nn \l__coffin_sin_fp { sind ( #2 ) }
 \fp_set:Nn \l__coffin_cos_fp { cosd ( #2 ) }
 \prop_map_inline:Nn \l__coffin_corners_prop
 { \__coffin_rotate_corner:Nnnn #1 {##1} ##2 }
 \prop_map_inline:Nn \l__coffin_poles_prop
 { \__coffin_rotate_pole:Nnnnnn #1 {##1} ##2 }
 \__coffin_set_bounding:N #1
 \prop_map-inline:Nn \l__coffin_bounding_prop
 { \__coffin_rotate_bounding:nnn {##1} ##2 }
 \__coffin_find_corner_maxima:N #1
 \__coffin_find_bounding_shift:
        #3 #1 {#2}

The corners and poles of the coffin can now be rotated around the origin. This is best achieved using mapping functions.
\prop_map_inline:Nn \l__coffin_corners_prop
 { \coffin - \__coffin_to_value:N #1 - corners }
\prop_map_inline:Nn \l__coffin_poles_prop
 { \coffin - \__coffin_to_value:N #1 - poles }

The bounding box of the coffin needs to be rotated, and to do this the corners have to be found first. They are then rotated in the same way as the corners of the coffin material itself.
\prop_map-inline:Nn \l__coffin_bounding_prop
 { \__coffin_rotate_bounding:nnn {##1} ##2 }

At this stage, there needs to be a calculation to find where the corners of the content and the box itself will end up.
\prop_map-inline:Nn \l__coffin_find_corner_maxima:N #1
\prop_map-inline:Nn \l__coffin_find_bounding_shift:
        #3 #1 {#2}
The correction of the box position itself takes place here. The idea is that the bounding box for a coffin is tight up to the content, and has the reference point at the bottom-left. The $x$-direction is handled by moving the content by the difference in the positions of the bounding box and the content left edge. The $y$-direction is dealt with by moving the box down by any depth it has acquired. The internal box is used here to allow for the next step.

\begin{verbatim}
\hbox_set:Nn \l__coffin_internal_box
{ \_kernel_kern:n
  { \_coffin_bounding_shift_dim - \_coffin_left_corner_dim }
\box_move_down:nn { \_coffin_bottom_corner_dim }
  { \box_use:N #1 }
}
\end{verbatim}

If there have been any previous rotations then the size of the bounding box will be bigger than the contents. This can be corrected easily by setting the size of the box to the height and width of the content. As this operation requires setting box dimensions and these transcend grouping, the safe way to do this is to use the internal box and to reset the result into the target box.

\begin{verbatim}
\box_set_ht:Nn \l__coffin_internal_box
{ \_coffin_top_corner_dim - \_coffin_bottom_corner_dim }
\box_set_dp:Nn \l__coffin_internal_box { 0pt }
\box_set_wd:Nn \l__coffin_internal_box
{ \_coffin_right_corner_dim - \_coffin_left_corner_dim }
5 #1 { \box_use_drop:N \l__coffin_internal_box }
\end{verbatim}

The final task is to move the poles and corners such that they are back in alignment with the box reference point.

\begin{verbatim}
\prop_map_inline:Nn \l__coffin_corners_prop
{ \_coffin_shift_corner:Nnnn #1 {##1} ##2 }
\prop_map_inline:Nn \l__coffin_poles_prop
{ \_coffin_shift_pole:Nnnnnn #1 {##1} ##2 }
\end{verbatim}

Update the coffin data.

\begin{verbatim}
#4 { coffin - \_coffin_to_value:N #1 - corners }
\_coffin_corners_prop
#4 { coffin - \_coffin_to_value:N #1 - poles }
\_coffin_poles_prop
\end{verbatim}

(End definition for \coffin_rotate:Nn, \coffin_grotate:Nn, and \__coffin_rotate:NNNN. These functions are documented on page 284.)

\_coffin_set_bounding:N

The bounding box corners for a coffin are easy enough to find: this is the same code as for the corners of the material itself, but using a dedicated property list.

\begin{verbatim}
\cs_new_protected:Nmp \_coffin_set_bounding:N #1
{ \prop_put:Nnx \_coffin_bounding_prop { tl }
  { 0pt } { \dim_eval:n { \box_ht:N #1 } } }
\prop_put:Nnx \_coffin_bounding_prop { tr }
{ \dim_eval:n { \box_wd:N #1 } }
\prop_put:Nnx \_coffin_bounding_prop { bl }
{ \dim_eval:n { \box_ht:N #1 } }
\prop_put:Nnx \_coffin_bounding_prop { br }
{ \dim_eval:n { \box_wd:N #1 } }
\end{verbatim}
Rotating the position of the corner of the coffin is just a case of treating this as a vector from the reference point. The same treatment is used for the corners of the material itself and the bounding box.

Rotating a single pole simply means shifting the co-ordinate of the pole and its direction. The rotation here is about the bottom-left corner of the coffin.

A rotation function, which needs only an input vector (as dimensions) and an output space. The values \texttt{\l__coffin_cos_fp} and \texttt{\l__coffin_sin_fp} should previously have been set up correctly. Working this way means that the floating point work is kept to a minimum: for any given rotation the sin and cosine values do no change, after all.
\__coffin_find_corner_maxima:N
\__coffin_find_corner_maxima_aux:nn

The idea here is to find the extremities of the content of the coffin. This is done by looking for the smallest values for the bottom and left corners, and the largest values for the top and right corners. The values start at the maximum dimensions so that the case where all are positive or all are negative works out correctly.

\__coffin_find_bounding_shift:
\__coffin_find_bounding_shift_aux:nn

The approach to finding the shift for the bounding box is similar to that for the corners. However, there is only one value needed here and a fixed input property list, so things are a bit clearer.
Shifting the corners and poles of a coffin means subtracting the appropriate values from the \(x\) and \(y\)-components. For the poles, this means that the direction vector is unchanged.

\[
\begin{align*}
&\text{\_\text{coffin\_shift\_corner:Nnnn}} \\
&\text{\_\text{coffin\_shift\_pole:Nnnnnn}}
\end{align*}
\]

When scaling, the values given have to be turned into absolute values.

\[
\begin{align*}
&\text{\_\text{coffin\_scaled\_total\_height\_dim}} \\
&\text{\_\text{coffin\_scaled\_width\_dim}}
\end{align*}
\]

Resizing a coffin begins by setting up the user-friendly names for the dimensions of the coffin box. The new sizes are then turned into scale factor. This is the same operation as takes place for the underlying box, but that operation is grouped and so the same calculation is done here.

\[
\begin{align*}
&\text{\_\text{coffin\_resize:Nnn}} \\
&\text{\_\text{coffin\_gresize:Nnn}} \\
&\text{\_\text{coffin\_resize:NNN}}
\end{align*}
\]
The poles and corners of the coffin are scaled to the appropriate places before actually resizing the underlying box.

Negative $x$-scaling values place the poles in the wrong location: this is corrected here.

For scaling, the opposite calculation is done to find the new dimensions for the coffin. Only the total height is needed, as this is the shift required for corners and poles. The
scaling is done the \TeX{} way as this works properly with floating point values without needing to use the \texttt{fp} module.

\begin{verbatim}
\cs_new_protected:Npn \coffin_scale:Nnn #1#2#3
\{ \__coffin_scale:NnnNN #1 {#2} {#3} \box_scale:Nnn \prop_set_eq:cN \}
\cs_generate_variant:Nn \coffin_scale:Nnn { c }
\cs_new_protected:Npn \coffin_gscale:Nnn #1#2#3
\{ \__coffin_scale:NnnNN #1 {#2} {#3} \box_gscale:Nnn \prop_gset_eq:cN \}
\cs_generate_variant:Nn \coffin_gscale:Nnn { c }
\cs_new_protected:Npn \__coffin_scale:NnnNN #1#2#3#4#5
\{
\fp_set:Nn \l__coffin_scale_x_fp {#2}
\fp_set:Nn \l__coffin_scale_y_fp {#3}
\#4 #1 { \l__coffin_scale_x_fp } { \l__coffin_scale_y_fp }
\dim_set:Nn \l__coffin_internal_dim
{ \coffin_ht:N #1 + \coffin_dp:N #1 }
\dim_set:Nn \l__coffin_scaled_total_height_dim
{ \fp_abs:n { \l__coffin_scale_y_fp } \l__coffin_internal_dim }
\dim_set:Nn \l__coffin_scaled_width_dim
{ \fp_abs:n { \l__coffin_scale_x_fp } \coffin_wd:N #1 }
\__coffin_resize_common:NnnN #1
{ \l__coffin_scaled_width_dim } { \l__coffin_scaled_total_height_dim }
#5 \}
\end{verbatim}

(End definition for \texttt{coffin_scale:Nnn}, \texttt{coffin_gscale:Nnn}, and \texttt{coffin_scale:NnnNN}. These functions are documented on page 284.)

\texttt{__coffin_scale_vector:nnNN}

This functions scales a vector from the origin using the pre-set scale factors in \( x \) and \( y \). This is a much less complex operation than rotation, and as a result the code is a lot clearer.

\begin{verbatim}
\cs_new_protected:Npn \__coffin_scale_vector:nnNN #1#2#3#4
\{
\dim_set:Nn #3 \{ \fp_to_dim:n { \dim_to_fp:n {#1} * \l__coffin_scale_x_fp } \}
\dim_set:Nn #4 \{ \fp_to_dim:n { \dim_to_fp:n {#2} * \l__coffin_scale_y_fp } \}
\end{verbatim}

(End definition for \texttt{__coffin_scale_vector:nnNN}.)

\texttt{__coffin_scale_corner:Nnnn}
\texttt{__coffin_scale_pole:Nnnnnn}

Scaling both corners and poles is a simple calculation using the preceding vector scaling.

\begin{verbatim}
\cs_new_protected:Npn \__coffin_scale_corner:Nnnn #1#2#3#4
\{ \__coffin_scale_vector:nnNN \#3 \#4 \l__coffin_x_dim \l__coffin_y_dim
\prop_put:Nnx \l__coffin_corners_prop {#2}
{ \dim_use:N \l__coffin_x_dim } { \dim_use:N \l__coffin_y_dim } \}
\cs_new_protected:Npn \__coffin_scale_pole:Nnnnnn #1#2#3#4#5#6
\{ \__coffin_scale_vector:nnNN \#3 \#4 \l__coffin_x_dim \l__coffin_y_dim
\prop_put:Nnx \l__coffin_poles_prop {#2}
{ \dim_use:N \l__coffin_x_dim } { \dim_use:N \l__coffin_y_dim }
{#5} {#6}
\end{verbatim}

1305
These functions correct for the $x$ displacement that takes place with a negative horizontal scaling.

The command joins two coffins, using a horizontal and vertical pole from each coffin and making an offset between the two. The result is stored as the as a third coffin, which has all of its handles reset to standard values. First, the more basic alignment function is used to get things started.

Correct the placement of the reference point. If the $x$-offset is negative then the reference point of the second box is to the left of that of the first, which is corrected using a kern. On the right side the first box might stick out, which would show up if it is wider than the sum of the $x$-offset and the width of the second box. So a second kern may be needed.
The coffin structure is reset, and the corners are cleared: only those from the two parent coffins are needed.

\__coffin_reset_structure:N \l__coffin_aligned_coffin
\prop_clear:c
{
  coffin - \__coffin_to_value:N \l__coffin_aligned_coffin
  c_space_tl corners
}
\__coffin_update_poles:N \l__coffin_aligned_coffin

The structures of the parent coffins are now transferred to the new coffin, which requires that the appropriate offsets are applied. That then depends on whether any shift was needed.

\dim_compare:nNnTF \l__coffin_offset_x_dim < \c_zero_dim
\__coffin_offset_poles:Nnn #1 { -\l__coffin_offset_x_dim } { 0pt }
\__coffin_offset_poles:Nnn #4 { 0pt } { \l__coffin_offset_y_dim }
\__coffin_offset_corners:Nnn #1 { -\l__coffin_offset_x_dim } { 0pt }
\__coffin_offset_corners:Nnn #4 { 0pt } { \l__coffin_offset_y_dim }
\__coffin_update_vertical_poles:NNN #1 #4 \l__coffin_aligned_coffin
#9 #1 \l__coffin_aligned_coffin
\__coffin_update_poles:NNN #1 #4 \l__coffin_aligned_coffin
\__coffin_offset_poles:NNN #1 { 0pt } { 0pt }
\__coffin_offset_poles:NNN #4 \l__coffin_offset_x_dim { \l__coffin_offset_y_dim }
\__coffin_offset_corners:NNN #1 { 0pt } { \l__coffin_offset_y_dim }
\__coffin_offset_corners:NNN #4 \l__coffin_offset_x_dim { \l__coffin_offset_y_dim }
\__coffin_update_vertical_poles:NNN #1 #4 \l__coffin_aligned_coffin
#9 #1 \l__coffin_aligned_coffin

(End definition for \coffin_join:NnnNnnnn, \coffin_gjoin:NnnNnnnn, and \__coffin_join:NnnNnnnnN. These functions are documented on page 285.)

\coffin_attach:NnnNnnnn
\coffin_attach:cnnNnnnn
\coffin_attach:Nnncnnnn
\coffin_attach:cnncnnnn
\coffin_gattach:NnnNnnnn
\coffin_gattach:cnnNnnnn
\coffin_gattach:Nnncnnnn
\coffin_gattach:cnncnnnn
\__coffin_attach:NnnNnnnnN
\__coffin_attach_mark:NnnNnnnn

A more simple version of the above, as it simply uses the size of the first coffin for the new one. This means that the work here is rather simplified compared to the above code. The function used when marking a position is hear also as it is similar but without the structure updates.

\cs_new_protected:Npn \coffin_attach:NnnNnnnn #1#2#3#4#5#6#7#8
\{ 
  \__coffin_attach:NnnNnnnnN #1 {#2} {#3} {#4} {#5} {#6} {#7} {#8}
  \coffin_set_eq:NN 
\}
\cs_generate_variant:Nn \coffin_attach:NnnNnnnn { c , Nnnc , cnnc }
\cs_new_protected:Npn \coffin_gattach:Nnnnnnnn #1#2#3#4#5#6#7#8
{ \__coffin_attach:NnnnNnnnnN #1 {#2} {#3} #4 {#5} {#6} {#7} {#8} \coffin_gset_eq:NN }
\cs_generate_variant:Nn \coffin_gattach:Nnnnnnnn { c , Nnnc , cnnc }
\cs_new_protected:Npn \__coffin_attach:NnnnNnnnnN #1#2#3#4#5#6#7#8#9
{ \__coffin_align:NnnnNnnnnN #1 {#2} {#3} #4 {#5} {#6} {#7} {#8} \l__coffin_aligned_coffin
\box_set_ht:Nn \l__coffin_aligned_coffin { \box_ht:N #1 }
\box_set_dp:Nn \l__coffin_aligned_coffin { \box_dp:N #1 }
\box_set_wd:Nn \l__coffin_aligned_coffin { \box_wd:N #1 }
\__coffin_reset_structure:N \l__coffin_aligned_coffin
\prop_set_eq:cc { coffin ~ \__coffin_to_value:N \l__coffin_aligned_coffin~ corners }
{ coffin ~ \__coffin_to_value:N #1 ~ corners }
\__coffin_update_poles:N \l__coffin_aligned_coffin
\__coffin_offset_poles:Nnn #1 { 0pt } { 0pt }
\__coffin_offset_poles:Nnn #4
{ \l__coffin_offset_x_dim } { \l__coffin_offset_y_dim }
\__coffin_update_vertical_poles:NNN #1 #4 \l__coffin_aligned_coffin #9 #1 \l__coffin_aligned_coffin
\cs_new_protected:Npn \__coffin_attach_mark:NnnnNnnnnN #1#2#3#4#5#6#7#8#9
{ \__coffin_align:NnnnNnnnnN #1 {#2} {#3} #4 {#5} {#6} {#7} {#8} \l__coffin_aligned_coffin
\box_set_ht:Nn \l__coffin_aligned_coffin { \box_ht:N #1 }
\box_set_dp:Nn \l__coffin_aligned_coffin { \box_dp:N #1 }
\box_set_wd:Nn \l__coffin_aligned_coffin { \box_wd:N #1 }
\__coffin_offset_poles:Nnn #1 { 0pt } { 0pt }
\__coffin_offset_poles:Nnn #4
{ \l__coffin_offset_x_dim } { \l__coffin_offset_y_dim }
\__coffin_update_vertical_poles:NNN #1 #4 \l__coffin_aligned_coffin #9 #1 \l__coffin_aligned_coffin
\cs_new_protected:Npm \__coffin_attach_mark:NNnnnnn #1#2#3#4#5#6#7#8#9
{ \__coffin_align:NNnnnnnn #1{#2}{#3}{#4}{#5}{#6}{#7}{#8} \l__coffin_aligned_coffin
\box_set_ht:Nn \l__coffin_aligned_coffin { \box_ht:N #1 }
\box_set_dp:Nn \l__coffin_aligned_coffin { \box_dp:N #1 }
\box_set_wd:Nn \l__coffin_aligned_coffin { \box_wd:N #1 }
\__coffin_offset_poles:Nnn #1 { 0pt } { 0pt }
\__coffin_offset_poles:Nnn #4
{ \l__coffin_offset_x_dim } { \l__coffin_offset_y_dim }
\__coffin_update_vertical_poles:NNN #1 #4 \l__coffin_aligned_coffin #9 #1 \l__coffin_aligned_coffin
\cs_new_protected:Npm \__coffin_attach:NNnnnnnnn #1#2#3#4#5#6#7#8#9
{ \__coffin_attach:NNnnnnnn #1{#2}{#3}{#4}{#5}{#6}{#7}{#8} \l__coffin_aligned_coffin
\box_set_ht:Nn \l__coffin_aligned_coffin { \box_ht:N #1 }
\box_set_dp:Nn \l__coffin_aligned_coffin { \box_dp:N #1 }
\box_set_wd:Nn \l__coffin_aligned_coffin { \box_wd:N #1 }
\__coffin_offset_poles:Nnn #1 { 0pt } { 0pt }
\__coffin_offset_poles:Nnn #4
{ \l__coffin_offset_x_dim } { \l__coffin_offset_y_dim }
\__coffin_update_vertical_poles:NNN #1 #4 \l__coffin_aligned_coffin #9 #1 \l__coffin_aligned_coffin
\cs_new_protected:Npm \__coffin_align:NNnnnnnnn #1#2#3#4#5#6#7#8#9
{ \__coffin_calculate_intersection:NNn #1{#2}{#3} \l__coffin_aligned_coffin
\dim_set:Nn \l__coffin_x_prime_dim { \l__coffin_x_dim }
\dim_set:Nn \l__coffin_y_prime_dim { \l__coffin_y_dim }
\dim_set:Nn \l__coffin_calculate_intersection:NNn #1{#2}{#3} \l__coffin_aligned_coffin
\dim_set:Nn \l__coffin_offset_x_dim
The internal function aligns the two coffins into a third one, but performs no corrections on the resulting coffin poles. The process begins by finding the points of intersection for the poles for each of the input coffins. Those for the first coffin are worked out after those for the second coffin, as this allows the ‘primed’ storage area to be used for the second coffin. The ‘real’ box offsets are then calculated, before using these to re-box the input coffins. The default poles are then set up, but the final result depends on how the bounding box is being handled.
Transferring structures from one coffin to another requires that the positions are updated by the offset between the two coffins. This is done by mapping over the property list of the source coffins, moving as appropriate and saving to the new coffin data structures. The test for \texttt{\textendash} means that the structures from the parent coffins are uniquely labelled and do not depend on the order of alignment. The pay off for this is that \texttt{\textendash} should not be used in coffin pole or handle names, and that multiple alignments do not result in a whole set of values.

Saving the offset corners of a coffin is very similar, except that there is no need to worry about naming: every corner can be saved here as order is unimportant.
The T and B poles need to be recalculated after alignment. These functions find the larger absolute value for the poles, but this is of course only logical when the poles are horizontal.

```latex
\cs_new_protected:Npn \__coffin_update_vertical_poles:NNN #1#2#3
\{ \__coffin_get_pole:NNN #3 { #1 -T } \l__coffin_pole_a_tl
\__coffin_get_pole:NNN #3 { #2 -T } \l__coffin_pole_b_tl
\exp_last_two_unbraced:Noo \__coffin_update_T:nnnnnnnnN \l__coffin_pole_a_tl \l__coffin_pole_b_tl #3
\}
\cs_new_protected:Npn \__coffin_update_T:nnnnnnnnN #1#2#3#4#5#6#7#8#9
\{ \dim_compare:nNnTF {#2} < {#6}
\{ \__coffin_set_pole:Nnx #9 { T } { { 0pt } {#6} { 1000pt } { 0pt } } \}
\}
\cs_new_protected:Npn \__coffin_update_B:nnnnnnnnN #1#2#3#4#5#6#7#8#9
\{ \dim_compare:nNnTF {#2} < {#6}
\{ \__coffin_set_pole:Nnx #9 { B } { { 0pt } {#2} { 1000pt } { 0pt } } \}
\}
```

(End definition for \__coffin_offset_corners:NN and \__coffin_offset_corner:NNnn.)

\begin{itemize}
\item \texttt{\_coffin_update_vertical_poles:NNN}
\item \texttt{\_coffin_update_T:nnnnnnnnN}
\item \texttt{\_coffin_update_B:nnnnnnnnN}
\end{itemize}
An empty-but-horizontal coffin.

Typesetting a coffin means aligning it with the current position, which is done using a coffin with no content at all. As well as aligning to the empty coffin, there is also a need to leave vertical mode, if necessary.

85.8 Coffin diagnostics

This property list is used to print coffin handles at suitable positions. The offsets are expressed as multiples of the basic offset value, which therefore acts as a scale-factor.
\prop_put:Nnn \l__coffin_display_handles_prop { Thc } { \{ \{ t \} \{ hc \} \{ 0 \} \{ -1 \} \} }
\prop_put:Nnn \l__coffin_display_handles_prop { Tr } { \{ \{ t \} \{ 1 \} \{ 1 \} \{ -1 \} \} }
\prop_put:Nnn \l__coffin_display_handles_prop { Hl } { \{ \{ vc \} \{ r \} \{ -1 \} \{ 1 \} \} }
\prop_put:Nnn \l__coffin_display_handles_prop { Hhc } { \{ \{ vc \} \{ hc \} \{ 0 \} \{ 1 \} \} }
\prop_put:Nnn \l__coffin_display_handles_prop { Hr } { \{ \{ vc \} \{ 1 \} \{ 1 \} \{ 1 \} \} }
\prop_put:Nnn \l__coffin_display_handles_prop { Bhc } { \{ \{ b \} \{ r \} \{ -1 \} \{ -1 \} \} }
\prop_put:Nnn \l__coffin_display_handles_prop { Hr } { \{ \{ b \} \{ hc \} \{ 0 \} \{ -1 \} \} }
\prop_put:Nnn \l__coffin_display_handles_prop { Br } { \{ \{ b \} \{ 1 \} \{ 1 \} \{ -1 \} \} }

(End definition for \l__coffin_display_handles_prop.)

\l__coffin_display_offset_dim The standard offset for the label from the handle position when displaying handles.
\dim_new:N \l__coffin_display_offset_dim
\dim_set:Nn \l__coffin_display_offset_dim { 2pt }

(End definition for \l__coffin_display_offset_dim.)

\l__coffin_display_x_dim \l__coffin_display_y_dim As the intersections of poles have to be calculated to find which ones to print, there is a need to avoid repetition. This is done by saving the intersection into two dedicated values.
\dim_new:N \l__coffin_display_x_dim
\dim_new:N \l__coffin_display_y_dim

(End definition for \l__coffin_display_x_dim and \l__coffin_display_y_dim.)

\l__coffin_display_poles_prop A property list for printing poles: various things need to be deleted from this to get a "nice" output.
\prop_new:N \l__coffin_display_poles_prop

(End definition for \l__coffin_display_poles_prop.)

\l__coffin_display_font_tl Stores the settings used to print coffin data: this keeps things flexible.
\tl_new:N \l__coffin_display_font_tl
\bool_lazy_and:nnT { \cs_if_exist_p:N \fmtname } { \str_if_eq_p:Vn \fmtname { LaTeX2e } } { \tl_set:Nn \l__coffin_display_font_tl { \sffamily \tiny } }

(End definition for \l__coffin_display_font_tl.)

\_coffin_rule:nn Abstract out creation of rules here until there is a higher-level interface.
\cs_new_protected:Npn \_coffin_rule:nn #1#2
{ \mode_leave_vertical:
\hbox:n { \tex_vrule:D width #1 height #2 \scan_stop: } }

1312
Marking a single handle is relatively easy. The standard attachment function is used, meaning that there are two calculations for the location. However, this is likely to be okay given the load expected. Contrast with the more optimised version for showing all handles which comes next.

1313 \cs_new_protected:Npn \coffin_mark_handle:Nnnn #1#2#3#4
  \hcoffin_set:Nn \l__coffin_display_pole_coffin
  \color_select:n {#4}
  \__coffin_rule:nn { 1pt } { 1pt }
  \__coffin_attach_mark:NnnNnnn #1 {#2} {#3}
  \l__coffin_display_pole_coffin { hc } { vc } { Opt } { Opt }
  \hcoffin_set:Nn \l__coffin_display_coords_coffin
  \color_select:n {#4}
  \l__coffin_display_font_tl
  \tl_to_str:n { #2 , #3 } ( \tl_to_str:n { #2 , #3 } )
  \prop_get:NnN \l__coffin_display_handles_prop
  \l__coffin_internal_tl
  \quark_if_no_value:NTF \l__coffin_internal_tl
  \prop_get:NnN \l__coffin_display_handles_prop
  \l__coffin_internal_tl
  \quark_if_no_value:NTF \l__coffin_internal_tl
  \__coffin_attach_mark:NnnnNnnnNnnn #1 {#2} {#3}
  \l__coffin_display_coords_coffin { l } { vc }
  \ipt { Opt }
  \l__coffin_display_offset_dim
  \exp_last_unbraced:No \__coffin_mark_handle_aux:nnnnnnn
  \l__coffin_internal_tl #1 {#2} {#3}
  \exp_last_unbraced:No \__coffin_mark_handle_aux:nnnnnnn
  \l__coffin_internal_tl #1 {#2} {#3}

1313 \cs_new_protected:Npn \__coffin_mark_handle_aux:nnnnnnn #1#2#3#4#5#6#7
  \__coffin_attach_mark:NnnnNnnnNnnn #5 {#6} {#7}
  \l__coffin_display_coords_coffin {#1} {#2}
  \l__coffin_display_offset_dim
  \exp_last_unbraced:No \__coffin_mark_handle_aux:nnnnn
  \l__coffin_internal_tl #1 {#2} {#3}

1313 \cs_generate_variant:Nn \coffin_mark_handle:Nnnn \{ c \}

(End definition for \coffin_mark_handle:Nnnn and \__coffin_mark_handle_aux:nnnnnnn. This function is documented on page 286.)
Printing the poles starts by removing any duplicates, for which the $H$ poles is used as the definitive version for the baseline and bottom. Two loops are then used to find the combinations of handles for all of these poles. This is done such that poles are removed during the loops to avoid duplication.

For each pole there is a check for an intersection, which here does not give an error if none is found. The successful values are stored and used to align the pole coffin with the main coffin for output. The positions are recovered from the preset list if available.
This is a dedicated version of \coffin_attach:NnNnnn with a hard-wired first coffin. As the intersection is already known and stored for the display coffin the code simply uses it directly, with no calculation.
\coffin_show_structure:N
\coffin_show_structure:c
\coffin_log_structure:N
\coffin_log_structure:c
\_.coffin_show_structure:N

For showing the various internal structures attached to a coffin in a way that keeps things relatively readable. If there is no apparent structure then the code complains.

\cs_new_protected:Npn \coffin_show_structure:N { \__coffin_show_structure:NN \msg_show:nnxxxx }
\cs_generate_variant:Nn \coffin_show_structure:N { c }
\cs_new_protected:Npn \coffin_log_structure:N { \__coffin_show_structure:NN \msg_log:nnxxxx }
\cs_generate_variant:Nn \coffin_log_structure:N { c }
\cs_new_protected:Npn \__coffin_show_structure:NN #1#2
{
  \__coffin_if_exist:NT #2 {

    \coffin:nn { coffin } { show } { \token_to_str:N #2 }

  }

}
\prop_map_function:cN { \_coffin_to_value:N #2 - poles }
\msg_show_item_unbraced:nn
\}
\}
\)

\coffin_show:N
\coffin_show:c
\coffin_log:N
\coffin_log:c
\coffin_show:Nnn
\coffin_show:cnn
\coffin_log:Nnn
\coffin_log:cnn
\_.coffin_show:NNnn

Essentially a combination of \coffin_show_structure:N and \box_show:N. We need to avoid having two prompts, so we use \msg_term:nxxxxx in the show case.

\cs_new_protected:Npn \coffin_show:N #1
{
  \_coffin_show:NNnn \msg_term:nxxxxx \box_show:Nnn
}\}
\}
\)

(End definition for \coffin_display_handles:Nn and others. This function is documented on page 285.)

(End definition for \coffin_show_structure:Nn, \coffin_log_structure:Nn, and \_coffin_show_structure:Nn. These functions are documented on page 286.)
85.9 Messages

\msg_new:nnnn { coffin } { no-pole-intersection } { No-intersection-between-coffin-poles. }
\msg_new:nnn { coffin } { unknown } { Unknown-coffin-’#1’. }
\msg_new:nnnn { coffin } { unknown-pole } { Pole-’#1'-unknown-for-coffin-’#2’. }
\msg_new:nnn { coffin } { show } { Size-of-coffin-#1 : #2
Poles-of-coffin-#1 : #3 . }
Chapter 86

I3color Implementation

86.1 Basics

\l__color_current_tl

The color currently active for foreground (text, etc.) material. This is stored in the form of a color model followed by one or more values. There are four pre-defined models, three of which take numerical values in the range [0, 1]:

- gray ⟨gray⟩ Grayscale color with the ⟨gray⟩ value running from 0 (fully black) to 1 (fully white)
- cmyk ⟨cyan⟩ ⟨magenta⟩ ⟨yellow⟩ ⟨black⟩
- rgb ⟨red⟩ ⟨green⟩ ⟨blue⟩

Notice that the value are separated by spaces. There is a fourth pre-defined model using a string value and a numerical one:

- spot ⟨name⟩ ⟨tint⟩ A pre-defined spot color, where the ⟨name⟩ should be a pre-defined string color name and the ⟨tint⟩ should be in the range [0, 1].

Additional models may be created to allow mixing of spot colors. The number of data entries these require will depend on the number of colors to be mixed.

\TeXhackers note: The content of \l__color_current_tl comprises two brace groups, the first containing the color model and the second containing the value(s) applicable in that model.

(End definition for \l__color_current_tl.)

\color_group_begin: \color_group_end:

Grouping for color is the same as using the basic \group_begin: and \group_end: functions. However, for semantic reasons, they are renamed here.

(End definition for \color_group_begin: and \color_group_end: These functions are documented on page 288.)
A driver-independent wrapper for setting the foreground color to the current color “now”.

\color\_ensure\_current:
\begin{verbatim}
\cs_new_protected:Npn \color\_ensure\_current:
\{ \__color\_backend\_pickup:N \l__color\_current\_tl
\__color\_select:N \l__color\_current\_tl
\}
\end{verbatim}

(End definition for \color\_ensure\_current:. This function is documented on page 288.)

\s\_color\_stop

Internal scan marks.

(End definition for \s\_color\_stop.)

\__color\_select:N
\__color\_select\_math:N
\__color\_select:nn

Take an internal color specification and pass it to the driver. This code is needed to ensure the current color but will also be used by the higher-level material.

\begin{verbatim}
\cs_new_protected:Npn \__color\_select:N \#1
{ \exp\_after:wN \__color\_select:nn \#1
\group\_insert\_after:N \__color\_backend\_reset:
\}
\cs_new_protected:Npn \__color\_select\_math:N \#1
{ \exp\_after:wN \__color\_select:nn \#1 }
\cs_new_protected:Npn \__color\_select:nn \#1\#2
{ \use:c { __color\_backend\_select\_:n } \{ \#2 \} }
\end{verbatim}

(End definition for \__color\_select:N, \__color\_select\_math:N, and \__color\_select:nn.)

\l\_color\_current\_tl

The current color, with the model and

\begin{verbatim}
\tl\_new:N \l\_color\_current\_tl
\tl\_set:Nn \l\_color\_current\_tl \{ \{ \text{gray} \} \{0\} \}
\end{verbatim}

(End definition for \l\_color\_current\_tl.)

86.2 Predefined color names

The ability to predefine colors with a name is a key part of this module and means there has to be a method for storing the results. At first sight, it seems natural to follow the usual expl3 model and create a \texttt{color} variable type for the process. That would then allow both local and global colors, constant colors and the like. However, these names need to be accessible in some form at the user level, for selection of colors either simply by name or as part of a more complex expression. This does not require that the full name is exposed but does require that they can be looked up in a predictable way. As such, it is more useful to expose just the color names as part of the interface, with the result that only local color names can be created. (This is also seen for example in key creation in \texttt{l3keys}.) As a result, color names are declarative (no \texttt{new} functions).

Since there is no need to manipulate colors \textit{en masse}, each is stored in a two-part structure: a \texttt{prop} for the colors themselves, and a \texttt{tl} for the default model for each color.
86.3 Setup

\l__color_internal_int
\l__color_internal_tl

\int_new:N \l__color_internal_int
\tl_new:N \l__color_internal_tl

(End definition for \l__color_internal_int and \l__color_internal_tl.)

\s__color_mark
Internal scan marks. \s__color_stop is already defined in l3color-base.
\scan_new:N \s__color_mark

(End definition for \s__color_mark.)

86.4 Utility functions

\__color_if_defined:nTF
A simple wrapper to avoid needing to have the lookup repeated in too many places. To
guard against a color created in a group, we need to test for entries in the prop.
\prg_new_conditional:Npnn \__color_if_defined:n #1 { T, F, TF } { \prop_if_exist:cTF { l__color_named_ #1 _prop } { \prop_if_empty:cTF { l__color_named_ #1 _prop } { \prg_return_false: \prg_return_true: } \prg_return_false: } }

(End definition for \__color_if_defined:nTF.)

\__color_model:N
\__color_values:N
Simple abstractions.
\cs_new:Npn \__color_model:N #1 { \exp_after:wN \use_i:nn #1 }
\cs_new:Npn \__color_values:N #1 { \exp_after:wN \use_ii:nn #1 }

(End definition for \__color_model:N and \__color_values:N.)

\__color_extract:nNN \__color_extract:VNN
Recover the values for the standard model for a color.
\cs_new_protected:Npn \__color_extract:nNN #1#2#3 { \tl_set_eq:Nc #2 { l__color_named_ #1 _tl } \prop_get:cVN { l__color_named_ #1 _prop } #2 #3 \cs_generate_variant:Nn \__color_extract:nNN { V } }

(End definition for \__color_extract:nNN.)
86.5 Model conversion

Model conversion is carried out using standard formulae for base models, as described in the manual for \texttt{xcolor} (see also the PostScript Language Reference Manual). For other models direct conversion might not be defined, so we go through the fallback models if necessary.

```latex
\cs_new_protected:Npn \__color_convert:nnN #1#2#3
\cs_generate_variant:Nn \__color_convert:nnN { VV }
\cs_generate_variant:Nn \exp_last_unbraced:Nf { c }
\cs_new_protected:Npn \__color_convert:nnnN #1#2#3#4
\tl_set:Nx #4 { \cs_if_exist_use:cTF { __color_convert_ #1 _ #2 :w } { #3 \s__color_stop } { \cs_if_exist:cTF { __color_convert_ \use:c { c__color_fallback_ #1 _tl } _ #2 :w } { \exp_last_unbraced:cf { __color_convert_ \use:c { c__color_fallback_ #1 _tl } _ #2 :w } \s__color_stop } { \exp_last_unbraced:cf { __color_convert_ \use:c { c__color_fallback_ #2 _tl } _ #2 :w } { \cs_if_exist_use:cTF { __color_convert_ #1 _ \use:c { c__color_fallback_ #2 _tl } :w } #3 \s__color_stop } \s__color_stop } }
\s__color_stop }
\cs_generate_variant:Nn \__color_convert:nnnN { nV , nnV }
\cs_new:Npn \__color_convert_gray_gray:w #1 \s__color_stop { #1 }
\cs_new:Npn \__color_convert_gray_rgb:w #1 \s__color_stop { #1 ~ #1 ~ #1 }
\cs_new:Npn \__color_convert_gray_cmyk:w #1 \s__color_stop { 0 ~ 0 ~ 0 ~ \fp_eval:n { 1 - #1 } }
\cs_new:Npn \__color_convert_rgb_gray:w #1 ~ #2 ~ #3 \s__color_stop { \fp_eval:n { 0.3 * #1 + 0.59 * #2 + 0.11 * #3 } }
```

These rather odd values are based on NTSC television: the set are used for the \texttt{cmyk} conversion.

```latex
\cs_new:Npn \__color_convert_rgb_cmyk:w #1 - #2 - #3 \s__color_stop { \fp_eval:n { 0.3 * #1 + 0.59 * #2 + 0.11 * #3 } }
```
The conversion from rgb to cmyk is the most complex: a two-step procedure which requires black generation and undercolor removal functions. The PostScript reference describes them as device-dependent, but following xcolor we assume they are linear. Moreover, as the likelihood of anyone using a non-unitary matrix here is tiny, we simplify and treat those two concepts as no-ops.

\documentclass{article}
\usepackage{xcolor}
\begin{document}
\newcommand{\__color_convert_rgb_cmyk}[4]{\fp_eval:n { min ( 1 - #1 , 1 - #2 , 1 - #3 ) } \c_space_tl \fp_eval:n { min ( 1 , max ( 0 , 1 - #2 - #1 ) ) } \c_space_tl \fp_eval:n { min ( 1 , max ( 0 , 1 - #3 - #1 ) ) } \c_space_tl #1} 
\newcommand{\__color_convert_cmyk_rgb}[4]{\fp_eval:n { 1 - min ( 1 , #1 + #4 ) } \c_space_tl \fp_eval:n { 1 - min ( 1 , #2 + #4 ) } \c_space_tl \fp_eval:n { 1 - min ( 1 , #3 + #4 )}} 
\newcommand{\__color_convert_cmyk_gray}[4]{\fp_eval:n { 1 - min ( 1 , 0.3 * #1 + 0.59 * #2 + 0.11 * #3 + #4 ) }} 
\end{document}

86.6 Color expressions

Working space to store the color data whilst doing calculations: keeping it on the stack is attractive but gets tricky (return is non-trivial).

\texttt{\_\_color_model_tl} \texttt{\_\_color_value_tl} \texttt{\_\_color_next_model_tl} \texttt{\_\_color_next_value_tl}

The main function for parsing color expressions removes actives but otherwise expands, then starts working through the expression itself. At the end, we apply the payload.

\texttt{\_\_color_parse:nN} \texttt{\_\_color_parse_aux:nN} \texttt{\_\_color_parse_eq:nN} \texttt{\_\_color_parse_eq:NNn} \texttt{\_\_color_parse_mix:NNW} \texttt{\_\_color_parse_mix:NNWV} \texttt{\_\_color_parse_mix:NNWn} \texttt{\_\_color_parse_mix:NNnN} \texttt{\_\_color_parse_mix:NNnNn} \texttt{\_\_color_parse_mix:NNnNnN} \texttt{\_\_color_parse_mix:NNnNnNnN} \texttt{\_\_color_parse_mix:NNnNnNnNnN} 

\texttt{\_\_color_parse_mix:NNnNnNnNnN}
Before going to all of the effort of parsing an expression, these two precursor functions look for a pre-defined name, either on its own or with a trailing ! (which is the same thing).

Here, we have to allow for the case where there is a fixed model: that can’t be swept up by generic conversion as we are dealing with a named color.
Once we establish that a full parse is needed, the next job is to get the detail of the first color. That will determine the model we use for the calculation: splitting here makes checking that a bit easier.

\cs_new_protected:Npn \__color_parse_loop_init:Nnn #1#2#3
\group_begin:
\__color_extract:nNN {#2} \l__color_model_tl \l__color_value_tl
\__color_parse_loop:w #3 ! ! ! ! \s__color_stop
\tl_set:Nx \l__color_internal_tl { { \l__color_model_tl } { \l__color_value_tl } }
\exp_args:NNNV \group_end:
\tl_set:Nn #1 \l__color_internal_tl
\__color_parse_end:

This is the loop proper: there can be an open-ended set of colors to parse, separated by ! tokens. There are a few cases to look out for. At the end of the expression and with we find a mix of 100 then we simply skip the next color entirely (we can’t stop the loop as there might be a further valid color to mix in). On the other hand, if we get a mix of 0 then drop everything so far and start again. There is also a trailing white to “read in” if the final explicit data is a mix. Those conditions are separate from actually looping, which is therefore sorted out by checking if we have further data to process: in contrast to xcolor, we don’t allow !! so the test can be simplified.

\cs_new_protected:Npn \__color_parse_loop:w #1 ! #2 ! #3 ! #4 ! #5 \s__color_stop
\bool_lazy_or:nnF
\tl_if_blank_p:n {#1}
\int_compare_p:nNn {#1} = { 100 }
\tl_if_blank:nTF {#2}
\int_compare:nNNTF {#1} = { 0 }
{ \tl_if_blank:nTF {#2}
{ \tl_if_blank:nTF { white } }
{ \tl_if_blank:nTF {#2} }
\l__color_model_tl \l__color_value_tl
}
{ \use:x
{ \__color_parse_loop:nn {#1}
{ \tl_if_blank:nTF {#2} { white } {#2} }
}
}
\tl_if_blank:nF {#3}
{ \__color_parse_loop:ww #3 ! #4 ! #5 \s__color_stop }
\__color_parse_end:

The “payload” of calculation in the loop first. If the model for the upcoming color is different from that of the existing (partial) color, convert the model. For gray the two are flipped round so that the outcome is something with “real” color. We are then in
a position to do the actual calculation itself. The two auxiliaries here give us a way to
break the loop should an invalid name be found.

\cs_new_protected:Npn \_color_parse_loop:nn #1#2
\__color_if_defined:nTF {#2}
\__color_extract:nNN {#2} \l__color_next_model_tl \l__color_next_value_tl
\tl_if_eq:NNF \l__color_model_tl \l__color_next_model_tl
{ \str_if_eq:VnTF \l__color_model_tl { gray }
  \__color_parse_gray:n {#2} }
  \__color_parse_std:n {#2} }
\tl_set:Nx \l__color_value_tl
\__color_parse_mix:NVVn\l__color_model_tl \l__color_value_tl \l__color_next_value_tl {#1}
\__color_parse_break:w

The \texttt{gray} model needs special handling: the models need to be swapped: we do that
using a dedicated function.
\cs_new_protected:Npn \_color_parse_gray:n #1
\tl_set_eq:NN \l__color_model_tl \l__color_next_model_tl
\tl_set:Nn \l__color_next_model_tl { gray }
\exp_args:NnV \__color_convert:nnN { gray } \l__color_model_tl \l__color_value_tl
\prop_get:cVN { l__color_named_ #1 _prop } \l__color_model_tl \l__color_next_value_tl
\__color_parse_break:w

Do the vector arithmetic: mainly a question of shuffling input, along with one pre-
calculation to keep down the use of division.
\cs_new:Npn \__color_parse_mix:Nnnn #1#2#3#4
\exp_args:Nf \__color_parse_mix:nNnn
\fp_eval:n { #4 / 100 } \#1 \{ #2 \} \{ #3 \}
\cs_generate_variant:Nn \__color_parse_mix:Nnnn { NVV }
\cs_new:Npn \__color_parse_mix:nNnn #1#2#3#4
\use:c { \__color_parse_mix_ #2 :nw } { #1 } \#3 \s__color_mark #4 \s__color_stop
\cs_new:Npn \__color_parse_mix_gray:nw #1#2 \s__color_mark #3 \s__color_stop
\fp_eval:n { #2 * #1 + #3 * ( 1 - #1 )}
\cs_new:Npn \__color_parse_mix_rgb:nw #1#2 ~ #3 ~ #4 \s__color_mark #5 ~ #6 ~ #7 \s__color_stop
\fp_eval:n { #2 * #1 + #5 * ( 1 - #1 )} \c_space_tl
\fp_eval:n { #3 * #1 + #6 * ( 1 - #1 )} \c_space_tl
\fp_eval:n { #4 * #1 + #7 * ( 1 - #1 )}
\cs_new:Npn \__color_parse_mix_cmyk:nw #1#2 ~ #3 ~ #4 ~ #5 \s__color_mark #6 ~ #7 ~ #8 ~ #9 \s__color_stop
\fp_eval:n { #2 * #1 + #6 * ( 1 - #1 )} \c_space_tl
\fp_eval:n { #3 * #1 + #7 * ( 1 - #1 )} \c_space_tl
\fp_eval:n { #4 * #1 + #8 * ( 1 - #1 )} \c_space_tl
\fp_eval:n { #5 * #1 + #9 * ( 1 - #1 )}

(End definition for \__color_parse:nN and others.)
\__color_parse_model_gray:w
\__color_parse_model_rgb:w
\__color_parse_model_cmyk:w
\__color_parse_number:n
\__color_parse_number:w
\cs_new:Npn \__color_parse_model_gray:w #1 , #2 \s__color_stop
\__color_parse_number:n \{ \#1 \}
\cs_new:Npn \__color_parse_model_rgb:w #1 , #2 , #3 , #4 \s__color_stop
\__color_parse_number:n \{ \#1 \} ~ \__color_parse_number:n \{ \#2 \} ~ \__color_parse_number:n \{ \#3 \}
\cs_new:Npn \__color_parse_model_cmyk:w #1 , #2 , #3 , #4 , #5 \s__color_stop
\__color_parse_number:n \{ \#1 \} ~ \__color_parse_number:n \{ \#2 \} ~ \__color_parse_number:n \{ \#3 \} ~ \__color_parse_number:n \{ \#4 \} ~ \__color_parse_number:n \{ \#5 \}

Turn the input into internal form, also tidying up the number quickly.
\cs_new:Npn \__color_parse_number:n #1 \\
{ \__color_parse_number:w #1 . 0 . \s__color_stop }
\cs_new:Npn \__color_parse_number:w #1 . #2 . #3 \s__color_stop \\
{ \tl_if_blank:nTF {#1} { 0 } {#1} . #2 }

(End definition for \__color_parse_model_gray:w and others.)

\__color_parse_model_Gray:w \__color_parse_model_HSB:w \\
\__color_parse_model_RGB:w
\__color_parse_model_hsb:w \__color_parse_model_hsb:nnn \\
\__color_parse_model_model_HSB:w \__color_parse_model_model_RGB:w
\__color_parse_model_wave:w \__color_parse_model_wave:nnn \\
\__color_parse_model_wave_auxi:nn \__color_parse_model_wave_auxii:nn
\__color_parse_model_wave_rho:n

The conversion here is non-trivial but is described at length in the xcolor manual. For ease, we calculate the integer and fractional parts of the hue first, then use them to work out the possible values for r, g and b before putting them in the correct places.
Following the description in the xcolor manual. As we always use rgb, there is no need to find the sixth, we just pass the information straight to the hsb auxiliary defined earlier.
\fp_compare:nNnTF {#1} < \{ 490 \}
{ \__color_parse_model_wave_auxii:nn
  \{ 4 - \__color_parse_model_wave_rho:n \{ (#1 - 440) / 50 \} \}
  \{#2\}
}
{ \fp_compare:nNnTF {#1} < \{ 510 \}
  \{ \__color_parse_model_wave_auxii:nn
    \{ 2 + \__color_parse_model_wave_rho:n \{ (#1 - 510) / -20 \} \}
    \{#2\}
  }
  \{ \fp_compare:nNnTF {#1} < \{ 580 \}
    \{ \__color_parse_model_wave_auxii:nn
      \{ 2 - \__color_parse_model_wave_rho:n \{ (#1 - 510) / 70 \} \}
      \{#2\}
    }
    \{ \fp_compare:nNnTF {#1} < \{ 645 \}
      \{ \__color_parse_model_wave_auxii:nn
        \{ \__color_parse_model_wave_rho:n \{ (#1 - 645) / -65 \} \}
        \{#2\}
      }
      \{ \__color_parse_model_wave_auxii:nn \{ 0 \} \{#2\} \}
    }
    \}\}
  \}
{ \__color_check_model:N \__color_check_model:nn
  \exp_args:Neee \__color_parse_model_hsb_aux:nnn
  \{ \fp_eval:n \{#1\} \}
  \{ 1 \}
  \{ \__color_parse_model_wave_rho:n \{#2\} \}
}
\cs_new:Npn \__color_parse_model_wave_auxii:nn #1#2
{\exp_args:Neee \__color_parse_model_hsb_aux:nnn
{ \fp_eval:n \{#1\} }
{ 1 }
{ \__color_parse_model_wave_rho:n \{#2\} }
}
\cs_new:Npn \__color_check_model:nn
{ \fp_eval:n \{ min(1, max(0,#1)) \} }
\end{definition}

(End definition for \__color_parse_model_Gray:w and others.)

86.7 Selecting colors (and color models)

\l_color_fixed_model_tl For selecting a single fixed model.
\tl_new:N \l_color_fixed_model_tl
(End definition for \l_color_fixed_model_tl. This variable is documented on page 291.)

\__color_check_model:N \__color_check_model:nn
Check that the model in use is the one required.
\_\_color_check_model:N #1
\tl_if_empty:NF \l__color_fixed_model_tl
  \exp_after:wN \__color_check_model:nn #1
  \tl_if_eq:NNF \l__color_model_tl \l__color_fixed_model_tl
    \__color_convert:VVN \l__color_model_tl \l__color_fixed_model_tl
    \l__color_value_tl
  \tl_set:Nx #1 { { \l__color_fixed_model_tl } { \l__color_value_tl } }
\}
\cs_new_protected:Npn \__color_check_model:nn #1#2
  \tl_set:Nn \l__color_model_tl {#1}
  \tl_set:Nn \l__color_value_tl {#2}
\)
(End definition for \_\_color_check_model:N and \_\_color_check_model:nn.)
\_\_color_finalise_current:
A backend-neutral location for “last minute” manipulations before handing off to the backend code. We set the special \_\_color_syntax here: this will therefore always be available. The finalisation is separate from the main function so it can also be applied to \textit{e.g.} page color.
\cs_new_protected:Npx \_\_color_finalise_current:
  \tl_set:Nx \exp_not:c { l__color_named_ . _tl }
  \prop_clear:N \exp_not:c { l__color_named_ . _prop }
  \prop_put:NVx \exp_not:c { l__color_named_ . _prop }
  \exp_not:c { l__color_named_ . _tl }
  \exp_not:N \__color_values:N \exp_not:N \l__color_current_tl
\)
(End definition for \_\_color_finalise_current.:)
\color_select:n
\color_select:mm
\color_select_main:Nw
\color_select_loop:Nw
\color_select:nnN
\color_select_swap:Nnn
Parse the input expressions then get the backend to actually activate them. The main complexity here is the need to check through multiple models. That is done “locally” here as the approach is subtly different to when different models are being stored.
\cs_new_protected:Npx \color_select:n #1
  \__color_parse:nN {#1} \l__color_current_tl
  \__color_finalise_current:
  \__color_select:N \l__color_current_tl
\)
\cs_new_protected:Npx \color_select:mm #1#2
  \__color_select_main:Nw \l__color_current_tl
  \_1 / / \_s__color_mark #2 / / \_s__color_stop
  \__color_finalise_current:
  \__color_select:N \l__color_current_tl
\)
1330
If the first color model is the fixed one, or if there is no fixed model, we don’t need most of the data: just set up and apply the backend function.

If a fixed model applies, we need to check each possible value in order. If there is no hit at all, fall back on the generic formula-based interchange.

(End definition for \color_select:n and others. These functions are documented on page 291.)

86.8 Math color

The approach here is the same as for the \LaTeX \texttt{mathcolor} command, but as we are working at the expl3 level we can make some minor changes.

\l_color_math_active_tl Tokens representing active sub/superscripts.

(End definition for \l_color_math_active_tl. This function is documented on page 292.)
Not all engines have multiple color stacks, and at the same time we are not expecting breaking within a colored math fragment. So we track the color stack ourselves.

\seq_new:N \__color_math_seq

(End definition for \__color_math_seq.)

The basic set up here is relatively simple: store the current color, parse the new color as-normal, then switch color before inserting the tokens we are asked to change. The tricky part is right at the end, handling the reset.

\cs_new_protected:Npn \color_math:nn #1#2
{\__color_math:nn {#2}{ \__color_parse:nN {#1} \l__color_current_tl }}
\cs_new_protected:Npn \color_math:nnn #1#2#3
{\__color_math:nn {#3}{ \__color_select_main:Nw \l__color_current_tl #1 / / \s__color_mark #2 / / \s__color_stop}}
\cs_new_protected:Npn \__color_math:nn #1#2
{\seq_gpush:NV \g__color_math_seq \l__color_current_tl #2 \__color_select_math:N \l__color_current_tl #1 \__color_math_scan:w}

(End definition for \color_math:nn, \color_math:nnn, and \__color_math:nn. These functions are documented on page 292.)

The complication when changing the color back is due to the fact that the \color_math:nn(n) may be followed by \^ or \_ or the hidden superscript (for example ') and its argument may end in a \mathop in which case the sub- and superscripts may be attached as \limits instead of after the material. All cases need separate treatment. To avoid repeatedly collecting the same token, we first check for an alignment tab: assuming we don’t have one of those, we can “recycle” \l_peek_token safely.

\cs_new_protected:Npn \__color_math_scan:w
  \peek_remove_filler:n
  \peek_catcode:NTF \c_alignment_token
  { \__color_math_scan_end: }
  { \__color_math_scan_auxi: }
\cs_new_protected:Npn \__color_math_scan_auxi:
\cs_new_protected:Npn \__color_math_scan_end:

Dealing with literal _ and ^ is easy, and as we have exactly two cases, we can hard-code this. We use a hard-coded list for limits: these are all primitives. The \use_none:n here also removes the test token so it is left just in the right place.
The one final case to handle is math-active tokens, most obviously \', as these won't be covered earlier.

\cs_new_protected:Npn \__color_math_scan_auxii:
\tl_map_inline:Nn \l_color_math_active_tl
\token_if_eq_meaning:NNT \l_peek_token \##1
\tl_map_break:n
\use_i:nn
\__color_math_scan_auxiii:N \##1
\__color_math_scan_end:
\cs_new_protected:Npn \__color_math_scan_auxiii:N #1
\exp_after:wN \exp_after:wN \exp_after:wN \__color_math_scan:w
\char_generate:nn { '#1 } { 13 }
\cs_new_protected:Npn \__color_backend_reset:
\seq_gpop:NN \g__color_math_seq \l__color_current_tl
\__color_math_scripts:Nw
\__color_math_script_aux:N

The tricky part of handling sub and superscripts is that we have to reset color to the one that is on the stack but reset it back to what it was before to allow for cases like

\[ \color_math:n \{ red \} \{ a + \sum \}_-{ \ i = 1 } ^{ n } \]
Here, \TeX{} constructs a \texttt{\vbox} stacking subscript, summation sign, and superscript. So technically the superscript comes first and the \texttt{\sum} that should get colored red is the middle.

The approach here is to set up a brace group immediately after the script token, then to set the color appropriately in that argument. We need an extra group to keep the color contained, and as we need to allow for an explicit closing brace in the source, the inner group also is a brace one rather than \texttt{\group_begin:-} based. At the end of the outer group we need to insert \texttt{\_\_\_color_math_scan:w} to continue the search for a second script token.

Notice that here we \textit{don't} need to use the math-specific color selector as we can allow the \texttt{\group_insert_after:N \@@\_backend\_reset:} to operate normally.

\begin{verbatim}
\cs_new_protected:Npn \__color_math_scripts:Nw #1
\end{verbatim}

Deal with the case where we do not have an explicit brace pair in the source.

\begin{verbatim}
\cs_new_protected:Npn \__color_math_script_aux:N #1 { #1 \c_group_end_token }
\end{verbatim}

\textit{(End definition for \_\_\_color_math_scripts:Nw and \_\_\_color_math_script_aux:N.)}

\section{Fill and stroke color}

\begin{verbatim}
\cs_new_protected:Npm \color_fill:n #1
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \color_stroke:n #1
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \color_fill:nn #1#2
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npm \color_stroke:nn #1#2
\end{verbatim}
86.10 Defining named colors

Defining named colors means working through the model list and saving both the “main” color and any equivalents in other models. Even if there is only one model, we store a \prop{} as well as a \tl, as there could be grouping weirdness, etc. When setting using an expression, we need to avoid any fixed model issues, which is done without a group as in \l3keys{}.

When setting an expression-based color, there could be multiple model data available for one or more of the input colors. Where that is true for the first named color in an expression, we re-parse the expression when they are also parameter-based: only \cmyk,
gray and rgb make any sense here. There is a bit of a performance hit but this should be rare and taking place during set-up.

\begin{verbatim}
\cs_new_protected:Npn \__color_set:nnw #1#2#3 ! #4 \s__color_stop
  \clist_map_inline:nn { cmyk , gray , rgb }
  \prop_get:cnNT { l__color_named_ #3 _prop } {##1} \l__color_internal_tl
  \prop_if_in:cnF { l__color_named_ #1 _prop } {##1}
  \group_begin:
  \tl_set:cn { l__color_named_ #3 _tl } {##1}
  \__color_parse:nN {#2} \l__color_internal_tl
  \exp_args:NNNV \group_end:
  \tl_set:Nn \l__color_internal_tl \l__color_internal_tl
  \prop_put:cxx { l__color_named_ #1 _prop }
  { \__color_model:N \l__color_internal_tl }
  { \__color_values:N \l__color_internal_tl }
\end{verbatim}
\__color_select:nnN \l__color_named_tl
\tl_set:Nx \l__color_internal_tl \{ \__color_model:N \l__color_named_tl \}
\tl_if_empty:cT \{ \l__color_named_#1_tl \}
\prop_put:cVx \{ \l__color_named_#1_prop \} \l__color_internal_tl
\{ \__color_values:N \l__color_named_tl \}
\__color_set_loop:nw \l__color_sub #1 \s__color_mark \s__color_stop
}
}

\cs_new_protected:Npn \color_set_eq:nn #1#2
{\__color_if_defined:nTF \l__color_named_tl \{
\__color_if_defined:nTF \l__color_named_#1_tl { \l__color_named_#1_tl }
\prop_clear_new:c \l__color_named_#1_prop
\str_if_eq:nnTF {#2} { . } { \l__color_named_#2_tl }
\prop_set_eq:cc \l__color_named_#1_prop \l__color_named_#2_prop
} \msg_error:nnn \{ color \} \{ unknown-color \} {#2} }

(End definition for \color_set:nn and others. These functions are documented on page 290.)

A small set of colors are always defined.
\color_set:nnn \{ black \} \{ gray \} \{ 0 \}
\color_set:nnn \{ white \} \{ gray \} \{ 1 \}
\color_set:nnn \{ cyan \} \{ cmyk \} \{ 1 , 0 , 0 , 0 \}
\color_set:nnn \{ magenta \} \{ cmyk \} \{ 0 , 1 , 0 , 0 \}
\color_set:nnn \{ yellow \} \{ cmyk \} \{ 0 , 0 , 1 , 0 \}
\color_set:nnn \{ red \} \{ rgb \} \{ 1 , 0 , 0 \}
\color_set:nnn \{ green \} \{ rgb \} \{ 0 , 1 , 0 \}
\color_set:nnn \{ blue \} \{ rgb \} \{ 0 , 0 , 1 \}

\l__color_named_.prop
\l__color_named_.tl

A special named color: this is always defined though not fixed in definition.
\prop_new:c \l__color_named_.prop
\tl_new:c \l__color_named_.tl
\tl_set:cx \l__color_named_.tl \{ \__color_model:N \l__color_current_tl \}

(End definition for \l__color_named_.prop and \l__color_named_.tl.)

86.11 Exporting colors
\color_export:nnN
\color_export:nnnN
\__color_export:nN
\__color_export:nnnN

1337
\cs_new_protected:Npn \color_export:nnN #1#2#3
\group_begin:
\tl_if_exist:cT { c__color_export_ #2 _tl }
\tl_set_eq:Nc \l_color_fixed_model_tl { c__color_export_ #2 _tl }
\__color_parse:nN {#1} #3
\__color_export:nN {#2} #3
\exp_args:NNNV \group_end:
\tl_set:Nn #3 #3
\end
\cs_new_protected:Npn \color_export:nnnN #1#2#3#4
\__color_select_main:Nw #4
#1 / / \s__color_mark #2 / / \s__color_stop
\__color_export:nN {#3} #4
\end
\cs_new_protected:Npn \__color_export:nN #1#2
\exp_after:wN \__color_export:nnnN #2 {#1} #2
\exp_after:wN #4
\end
\cs_if_exist_use:cF { __color_export_format_ #3 :nnN }
{ \msg_error:nnn { color } { unknown-export-format } {#3}
\use_none:nnn
}
{#1} {#2} #4
\end
(End definition for \color_export:nnN and others. These functions are documented on page 293.)

\__color_export_format_backend:nnN Simple.
\cs_new_protected:Npn \__color_export_format_backend:nnN #1#2#3
\cs_if_eq:nnTF {#2} {#1}
{ #5 #4 #3 \s__color_stop }
{ \__color_convert:nnnN {#2} {#1} {#3} #4
\exp_after:wN \s__color_stop
\exp_after:wN #4
\end
(End definition for \__color_export_format_backend:nnN.)

\__color_export:nnnN A generic auxiliary for cases where only one model is appropriate.
\cs_new_protected:Npn \__color_export:nnnN #1#2#3#4#5
\str_if_eq:nnTF {#2} {#1}
{ #5 #4 #3 \s__color_stop }
{ \__color_convert:nnnN {#2} {#1} {#3} #4
\exp_after:wN \s__color_stop
\exp_after:wN #4
\end
(End definition for \__color_export:nnnN.)
\c__color_export_comma-sep-cmyk_tl
\c__color_export_comma-sep-rgb_tl
\c__color_export_HTML_tl
\c__color_export_space-sep-cmyk_tl
\c__color_export_space-sep-rgb_tl
\tl_const:cn { c__color_export_comma-sep-cmyk_tl } { cmyk }
\tl_const:cn { c__color_export_comma-sep-rgb_tl } { rgb }
\tl_const:Nn \c__color_export_HTML_tl { rgb }
\tl_const:cn { c__color_export_space-sep-cmyk_tl } { cmyk }
\tl_const:cn { c__color_export_space-sep-rgb_tl } { rgb }
1338
\group_begin:
\cs_set_protected:Npn \__color_tmp:w #1#2
\cs_new_protected:cpx { \color_export_format_ #1 :nnN } ##1##2##3
\exp_not:N \color_export:nnnNN {#2} {##1} {##2} ##3
\exp_not:c { \color_export_ #1 :Nw }
\__color_tmp:w { comma-sep-cmyk } { cmyk }
\__color_tmp:w { comma-sep-rgb } { rgb }
\__color_tmp:w { space-sep-cmyk } { cmyk }
\__color_tmp:w { space-sep-rgb } { rgb }
\group_end:
\__color_export_space-sep-cmyk:Nw \__color_export_comma-sep-cmyk:Nw
\cs_new_protected:cpn { \color_export_comma-sep-cmyk:Nw }
\#1\#2 \#3 \#4 \#5 \s__color_stop
\tl_set:Nn \#1 { \#2 , \#3 , \#4 , \#5 }
\cs_new_protected:cpn { \color_export_space-sep-cmyk:Nw } \#1\#2 \s__color_stop
\tl_set:Nn \#1 { \#2 }
\endinput
86.12 Additional color models

For every colorspace, we define one of the base colorspaces as a fallback. The base colorspaces themselves are their own fallback.

Mapping from names to colorants.

Whitepoint data for the CIELAB profiles.

The range for CIELAB color spaces.
For tracking the alternative model set up for separations, etc.

\prop_new:N \g__color_alternative_model_prop
\clist_map_inline:nn { cyan , magenta , yellow , black }
\prop_gput:Nnn \g__color_alternative_model_prop { #1 } { cmyk }
\clist_map_inline:nn { red , green , blue }
\prop_gput:Nnn \g__color_alternative_model_prop { #1 } { rgb }

(End definition for \g__color_alternative_model_prop.)

Same for the values: a bit more involved.

\prop_new:N \g__color_alternative_values_prop
\prop_gput:Nnn \g__color_alternative_values_prop { cyan } { 1 , 0 , 0 , 0 }
\prop_gput:Nnn \g__color_alternative_values_prop { magenta } { 0 , 1 , 0 , 0 }
\prop_gput:Nnn \g__color_alternative_values_prop { yellow } { 0 , 0 , 1 , 0 }
\prop_gput:Nnn \g__color_alternative_values_prop { black } { 0 , 0 , 0 , 1 }
\prop_gput:Nnn \g__color_alternative_values_prop { red } { 1 , 0 , 0 , 0 }
\prop_gput:Nnn \g__color_alternative_values_prop { green } { 0 , 1 , 0 , 0 }
\prop_gput:Nnn \g__color_alternative_values_prop { blue } { 0 , 0 , 1 , 0 }

(End definition for \g__color_alternative_values_prop.)

Set up a new model: in general this has to be handled by a family-dependent function.
To avoid some “interesting” questions with casing, we fold the case of the family name.
The key–value list should always be present, so we convert it up-front to a \prop, then
deal with the detail on a per-family basis.

\cs_new_protected:Npn \color_model_new:nnn #1#2#3
\exp_args:Nee \__color_model_new:nnn
\tl_to_str:n {#1}
\str_foldcase:n {#2} {#3}

\cs_new_protected:Npn \__color_model_new:nnn #1#2#3
\cs_if_exist:cTF { __color_parse_model_ #1 :w }
\msg_error:n.nn { color } { model-already-defined } {#1}
\{ \}
\cs_if_exist:cTF { __color_model_ #2 :n }
\prop_set_from_keyval:Nn \l__color_internal_prop {#3}
\use:c { __color_model_ #2 :n } {#1}
\{ \}
\msg_error:n.nn { color } { unknown-model-type } {#2}
\}
\}

(End definition for \color_model_new:nnn and \__color_model_new:nnn. This function is documented
on page 294.)

A shared auxiliary to do the basics of setting up a new model: reserve a number, create
a white-equivalent, set up links to the backend.
Separations must have a “real” name, which is pretty easy to find.

We have two keys to find at this stage: the alternative space model and linked values.
As each alternative space leads to a different requirement for conversion, and as there are only a small number of choices, we manually split the data and then set up. Notice that mixing tints is really just the same as mixing gray. The white color is special, as it allows tints to be adjusted without an additional color space. To make sure the data is set for that at all group levels, we need to work on a per-level basis. Within the output, only the setup needs the “real” name of the colorspace: we use a simple tracking number for general usage as this is a clear namespace without issues of escaping chars.
Generic model conversion \textit{via} an alternative intermediate.

Setting up for CIELAB needs a bit more work: there is the illuminant and the need for an appropriate object.

If a CIELAB space is being set up, we need the illuminant, then create the appropriate set up. At present, this doesn’t include BlackPoint or Range data, but that may be added later. As CIELAB colors cannot be converted to anything else, we fallback to
producing black in the gray colorspace: the user should set up a second model for colors
set up this way.

\cs_new_protected:Npn \__color_model_separation_CIELAB:nnnnnnn #1#2#3#4#5#6#7
\{ 
  \tl_if_exist:cTF { c__color_model_whitepoint_CIELAB_ #1 _tl } 
  \{ 
    \__color_backend_separation_init_CIELAB:nnn {#1} {#3} { #4 ~ #5 ~ #6 } 
    \tl_const:cn { c__color_fallback_ #2 _tl } { gray } 
    \cs_new:cpn { __color_convert_ #2 _gray:w } ##1 \s__color_stop 
    { 0 } 
    \cs_new:cpn { __color_convert_gray_ #2 :w } ##1 \s__color_stop 
    { 1 } 
  \} 
  \{ 
    \msg_error:nnn { color } 
    { unknown-CIELAB-illuminant } {#1} 
  \} 
\}

(End definition for \_color_model_separation:n and others.)

\__color_model_devicen:n
\__color_model_devicen:nn
\__color_model_devicen:nnn
\__color_model_devicen:nnnn
\__color_model_devicen_parse_1:nn
\__color_model_devicen_parse_2:nn
\__color_model_devicen_parse_3:nn
\__color_model_devicen_parse_4:nn
\__color_model_devicen_parse_generic:nn
\__color_model_devicen_init:nnn
\__color_model_devicen_init:nnnn
\__color_model_devicen_trasform:w
\__color_model_devicen_trasform_1:nnnnn
\__color_model_devicen_trasform_3:nnnnn
\__color_model_devicen_trasform_4:nnnnn
\__color_model_devicen_colorant:n
\__color_model_devicen_convert:nnn
\__color_model_devicen_convert_cmyk:n
\__color_model_devicen_convert_rgb:n
\__color_model_devicen_convert:nnnn
\__color_model_devicen_convert:n
\__color_model_devicen_convert_aux:n
\__color_convert_devicen_cmyk:nnnnnn
\__color_convert_devicen_cmyk_aux:nnnnnn
\__color_convert_devicen_gray:nnnnnn
\__color_convert_devicen_gray_aux:nnnnnn
\__color_convert_devicen_rgb:nnnnnn
\__color_convert_devicen_rgb_aux:nnnnnn

We require a list of component names here: one might call them colorants, but it’s
convenient to use \TeX names instead so we slightly adjust the terminology.

\cs_new_protected:Npn \__color_model_devicen:n #1
\{ 
  \prop_get:NnNTF \l__color_internal_prop { names } \l__color_internal_tl 
  \exp_args:NV \__color_model_devicen:nn \l__color_internal_tl {#1} 
\}
\msg_error:nnn { color } 
{ DeviceN-requires-names } {#1} 

All valid models will have an alternative listed, either hard-coded for the core device
ones, or dynamically added for Separations, etc.

\cs_new_protected:Npn \__color_model_devicen:nn #1#2
\{ 
  \tl_clear:N \l__color_model_tl 
  \clist_map_inline:nn {#1} 
  \prop_get:NnNTF \g__color_alternative_model_prop {##1} \l__color_internal_tl 
  \exp_args:NV \__color_model_devicen:nn \l__color_internal_tl {#1} 
  \msg_error:nnn { color } 
  { DeviceN-inconsistent-alternative } {#1} 
\}

\tl_if_empty:NTF \l__color_model_tl 
\{ \tl_set_eq:NN \l__color_model_tl \l__color_internal_tl 
\} 
\str_if_eq:VVF \l__color_model_tl \l__color_internal_tl 
\{ 
  \msg_error:nnn { color } 
  { DeviceN-inconsistent-alternative } 
\}
We now complete the data we require by first finding out how many colorants there are, then moving on to begin constructing the function required to map to the alternative color space.

\cs_new_protected:Npn \__color_model_devicen:nnn #1#2#3
\exp_args:NV \__color_model_devicen:nnnn { \clist_count:n {#2} } {#1} {#2} {#3}

At this stage, we have checked everything is in place, so we can set up the \TeX{} and backend data structures. As for separations, it’s not really possible in general to have a fallback, so we simply provide “black” for each element.

\cs_new_protected:Npn \__color_model_devicen:nnnn #1#2#3#4
\__color_model_init:nnnx {#4} {#1} { devicen } \prg_replicate:nn { #1 - 1 } { ~ 0 }
\cs_if_exist_use:cF { __color_model_devicen_parse_ #1 :nn } \{ \__color_model_devicen_parse_generic:nn \{#4\} {#1} \}
\__color_model_devicen_init:nnn {#1} {#2} {#3} \prg_replicate:nn { #1 - 1 } { ~ 1 }

For short lists of DeviceN colors, we can use hand-tuned parsing. This lines up with other models, where we allow for up to four components. For larger spaces, rather than limit artificially, we use a somewhat slow approach based on open-ended commas-lists.

\cs_new_protected:cpn { __color_model_devicen_parse_1:nn } #1#2
To construct the tint transformation, we have to use PostScript. The aim is to have the 
final tint for each device colorant as

\[ 1 - \prod_{n} (1 - X_nD_n) \]

where \( X \) is a DeviceN colorant and \( D \) is the amount of device colorant that the DeviceN 
colorant maps to. At the start of the process, the PostScript stack will contain the 
\( X_n \) values, whilst we have the \( D \) values on a per-DeviceN colorant basis. The more 
convenient approach for us is therefore to take each DeviceN colorant in turn and find 
the value \( 1 - X_nD_n \), multiplying as we go, and finalise with the subtraction. That 
contrasts to colorspace: it splits the process up by process color, which works better 
when you have a fixed list of colorants. (colorspace only supports up to 4 DeviceN colors, 
and only cmyk as the alternative space.) To set this up, we first need to know the number 
of values in the target color space: this is easily handled as there are a very small range 
of possibles. Once we have that information, it’s relatively easy to build the required 
PostScript using some generic code.

\[
\begin{align*}
\text{\texttt{cs_new:pn \_\_color_model_devicen_init:nnn \#1\#2\#3}} \\
\text{\texttt{\str_case:nn \#2}} \\
\text{\texttt{\{ cmyk \} \{ 4 \}}} \\
\text{\texttt{\{ gray \} \{ 1 \}}} \\
\text{\texttt{\{ rgb \} \{ 3 \}}}
\end{align*}
\]
As we always need to split the alternative values into parts, we use a shared auxiliary and only use a minimal difference between code paths. Construction of the tint transformation is as far as possible done using loops, which means there are some inefficiencies for device colors in the DeviceN space: we roll the stack one-at-a-time even if there is a potential shortcut. However, that way there is nothing to special-case. Once this is sorted, we can write the tint transform object, which will remain as the last object until we sort out the final step: the colorant list.

\cs_new_protected:Npn \_color_model_devicen_init:nnnn \#1 \#2 \#3 \#4
\{\tl_set:Nx \l__color_internal_tl \prg_replicate:nn {\#1} { 1.0 ~ } \tl_zero:N \l__color_internal_int \clist_map_inline:nn {\#4} \{ \tl_intcr:N \l__color_internal_int \prop_get:NnN \g__color_alternative_values_prop {##1} \l__color_value_tl \exp_after:wN \_color_model_devicen_transform:w \l__color_value_tl , 0 , 0 , 0 , \s__color_stop \#1 \#2 \} \tl_put_right:Nx \l__color_internal_tl \prg_replicate:nn {\#1} { neg ~ 1.0 ~ add ~ #1 ~ -1 ~ roll ~ } \int_eval:n { #2 + #1 } ~ #1 ~ roll \prg_replicate:nn {\#2} { ~ pop } ~ \#1 ~ 1 ~ roll \} \use:x \{ \_color_backend_devicen_init:nnn \{ \clist_map_function:nN {\#4} \_color_model_devicen_colorant:n \} \str_case:nn {\#3} \{ \prg_replicate:nn {\#1} { /DeviceCMYK } \prg_replicate:nn {\#2} { /DeviceGray } \prg_replicate:nn {\#3} { /DeviceRGB } \} \exp_not:V \l__color_internal_tl \} \cs_new_protected:Npn \_color_model_devicen_transform:w \#1 , \#2 , \#3 , \#4 , \#5 \s__color_stop \#6 \#7 \{
Here we need to set up conversion from the DeviceN space to the alternative at the \TeX\ level. This also means supplying methods for inter-converting to other parameter-based spaces. Essentially the approach is exactly the same as the PostScript, just expressed in \TeX\ terms.
\__color_model_devicen_convert:nnnnn \{#1\} \{ rgb \} \{ 3 \} \{#2\}
\cs_new_protected:Npn \__color_model_devicen_convert:nnnn #1#2#3#4#5
\cs_new:cpx \{ \_color_convert_ \_ #2 : \#1 \} \s__color_stop \{#5\}
\cs_new:cpx \{ \_color_convert_ \#1 : \_ #2 \} \s__color_stop
\exp_not:c \{ \_color_convert_devicen_ \#2 : \prg_replicate:nn \{#3\} \{ n \} \w \}
\exp_not:N \s__color_mark
\clist_map_function:nN \{#4\} \__color_model_devicen_convert:n
\exp_not:N \s__color_stop
\cs_new:Npn \__color_model_devicen_convert:n \#1
\exp_args:Ne \__color_model_devicen_convert_aux:n
\prop_item:Nn \g__color_alternative_values_prop \{#1\}
\cs_new:Npn \__color_model_devicen_convert_aux:n \#1
\tl_if_blank:nF \{#2\}
{ \#2 }
\tl_if_blank:nF \{#3\}
{ \#3 }
{ \#4 }
\tl_if_blank:nF \{#4\}
{ \#4 }
\use:e
\__color_convert_devicen_cmyk:nnnnn
\fp_eval:n \{ #2 * (1 - (#1 * #6)) \}
\fp_eval:n \{ #3 * (1 - (#1 * #7)) \}
\fp_eval:n \{ #4 * (1 - (#1 * #8)) \}
\fp_eval:n \{ #5 * (1 - (#1 * #9)) \}
\exp_not:N \s__color_mark
\use:e
\__color_convert_devicen_cmyk_cmyk:nnnnw
#1#2#3#4#5#6#7#8#9
\use:e
\exp_not:N \__color_convert_devicen_cmyk_aux:nnnnw
\fp_eval:n \{ #2 * (1 - (#1 * #6)) \}
\fp_eval:n \{ #3 * (1 - (#1 * #7)) \}
\fp_eval:n \{ #4 * (1 - (#1 * #8)) \}
\fp_eval:n \{ #5 * (1 - (#1 * #9)) \}
\exp_not:N \s__color_mark
\cs_new:Npn \__color_convert_devicen_cmyk_aux:nnnnw #1#2#3#4 #5 \s__color_mark #6 \s__color_stop
\tl_if_blank:nTF {#5}
\{
  \fp_eval:n { 1 - #1 } -
  \fp_eval:n { 1 - #2 } -
  \fp_eval:n { 1 - #3 } -
  \fp_eval:n { 1 - #4 }
\}
\}
\__color_convert_devicen_cmyk:nnnnw {#1} {#2} {#3} {#4}
#5 \s__color_mark #6 \s__color_stop
\}
\cs_new:Npn \__color_convert_devicen_gray:nw #1#2 \s__color_mark #3#4 \s__color_stop
\__color_convert_devicen_gray:nnn {#2} {#1} #4
#3 \s__color_mark #5 \s__color_stop
\}
\cs_new:Npn \__color_convert_devicen_gray:nnn #1#2#3
\exp_arsgs:Ne \__color_convert_devicen_gray_aux:nw
{ \fp_eval:n { #2 * (1 - (#1 * #3)) } }
\}
\cs_new:Npn \__color_convert_devicen_gray_aux:nw #1 #2 \s__color_mark #3 \s__color_stop
\tl_if_blank:nTF {#2}
\{
  \fp_eval:n { 1 - #1 }
\}
\__color_convert_devicen_gray:nw {#1}
#2 \s__color_mark #3 \s__color_stop
\}
\cs_new:Npn \__color_convert_devicen_rgb:nnnw #1#2#3#4 #5 \s__color_mark #6#7 \s__color_stop
\use:e
\exp_not:N \__color_convert_devicen_rgb_aux:nnnnw
{ \fp_eval:n { #2 * (1 - (#1 * #5)) } }
{ \fp_eval:n { #3 * (1 - (#1 * #6)) } }
{ \fp_eval:n { #4 * (1 - (#1 * #7)) } }
\}
\cs_new:Npn \__color_convert_devicen_rgb:nnnnw #1#2#3#4 #5 \s__color_mark #6 \s__color_stop
The signatures in the ICC file header indicating the underlying colorspace. We map it to three values: The number of components, the values corresponding to white, and the range.

\begin{verbatim}
\prop_const_from_keyval:Nn \c__color_icc_colorspace_signatures_prop
\end{verbatim}

For an ICC profile, we need a file name and a number of components. The file name is processed here so the backend can treat it as a string.
Here, we can use the same internals as for DeviceN approach as we know the number of components. No conversion is possible, so there is no need to worry about that at all.

\cs_new_protected:Npn \__color_model_iccbased_aux:nnnnnn #1#2#3#4#5#6
\begin{verbatim}
\__color_model_init:nnnn {#5} {#1} { iccbased } {#3}
\tl_const:cn { c__color_fallback_ #5 _tl } { gray }
\cs_new:cpn { __color_convert_ #5 _gray:w } ##1 \s__color_stop { 0 }
\cs_new:cpn { __color_convert_gray_ #5 :w } ##1 \s__color_stop { #2 }
\use:c { __color_model_devicen_parse_ #1 :nn } {#5} {#1}
\exp_args:Nx \__color_backend_iccbased_init:nnn
{ \file_full_name:n {#6} } {#1} {#4}
\end{verbatim}

(End definition for \__color_model_iccbased:n and others.)

\subsection{Applying profiles}

With a limited range of outcomes, this is largely about getting data to the backend.
86.14 Diagnostics

Extract the information about a color and format for the user: the approach is similar to the keys module here.

\begin{verbatim}
\cs_new_protected:Npn \color_show:n { \__color_show:Nn \msg_show:nnxxxx }
\cs_new_protected:Npn \color_log:n { \__color_show:Nn \msg_log:nnxxxx }
\cs_new_protected:Npn \__color_show:Nn #1#2
{ #1 { color } { show }
{#2}
{ \__color_if_defined:nT {#2}
{ \exp_args:Nv \__color_show:n { l__color_named_ #2 _tl }
\prop_map_function:cN
{ l__color_named_ #2 _prop }
\msg_show_item_unbraced:nn
}
{ }
{ }
}
\cs_new:Npn \__color_show:n #1
{ \msg_show_item_unbraced:nn { model } {#1}
{ }
{ }
}
\end{verbatim}

(End definition for \color_show:n and others. These functions are documented on page 291.)

86.15 Messages

\begin{verbatim}
\msg_new:nnnn { color } { CIELAB-requires-illuminant }
\msg_new:nnnn { CIELAB-color-space-’#1’-require-an-illuminant. }
{ \LaTeX\ has\ been\ asked\ to\ create\ a\ separation\ color\ space\ using\ CIELAB-specifications,\ but\ no-\\
\iow_indent:n { illuminant=\-<\basis> }
\key\ was\ given\ with\ the\ correct\ information. \LaTeX\ will\ use\ illuminant-
’d50’\-for-recovery.
\}
\msg_new:nnnn { color } { conversion-not-available }
\msg_new:nnnn { color } { DeviceN-inconsistent-alternative }
\end{verbatim}

1355
{ DeviceN-color-spaces-require-a-single-alternative-space. }

{ LaTeX-has-been-asked-to-create-a-DeviceN-color-space-'#1',-
  but-the-constituent-colors-do-not-have-a-common-alternative-
  color. }

\msg_new:nnnn { color } { DeviceN-no-alternative }

{ DeviceN-color-spaces-require-an-alternative-space. }

{ LaTeX-has-been-asked-to-create-a-DeviceN-color-space-'#1',-
  but-the-constituent-colors-do-not-all-have-a-device-based-alternative. }

\msg_new:nnnn { color } { DeviceN-requires-names }

{ DeviceN-color-space-#'1'-require-a-list-of-names. }

{ LaTeX-has-been-asked-to-create-a-DeviceN-color-space,-
  but-no-
  \iow_indent:n { names=--<names> } }

key-was-given-with-the-correct-information.

\msg_new:nnnn { color } { ICC-Device-unknown }

{ Unknown-device-color-space-'#1'. }

{ LaTeX-has-been-asked-to-apply-an-ICC-profile-but-the-device-color-space-
  '##1'-is-unknown. }

\msg_new:nnnn { color } { ICCBased-unsupported-colorspace }

{ ICCBased-color-space-#'1'-uses-an-unsupported-data-color-space. }

{ LaTeX-has-been-asked-to-create-a-ICCBased-colorspace,-but-the-
  used-data-colorspace-is-not-supported.-ICC-profiles-used-for-
  defining-a-ICCBased-colorspace-should-use-a-Lab,-RGB,-or-
  CMYK-data-colorspace.-LaTeX-will-ignore-this-request. }

\msg_new:nnnn { color } { ICCBased-requires-file }

{ ICCBased-color-space-#'1'-require-an-file. }

{ LaTeX-has-been-asked-to-create-an-ICCBased-colorspace,-but-no-
  \iow_indent:n { file=--<name> } }

key-was-given-with-the-correct-information.-LaTeX-will-ignore-this-
request.

\msg_new:nnnn { color } { model-already-defined }

{ Color-model-#'1'-already-defined. }

{ LaTeX-was-asked-to-define-a-new-color-model-called-#'1',-but-
  this-color-model-already-exists. }

\msg_new:nnnn { color } { separation-alternative-model }

{ Separation-color-space-#'1'-require-an-alternative-model. }

{ LaTeX-has-been-asked-to-create-a-separation-color-space,-
  \iow_indent:n { alternative-color-space=--<names> } }

key-was-given-with-the-correct-information.-LaTeX-will-ignore-this-
request.
\texttt{\msg_new:nnnn \{ color \} \{ separation-alternative-values \}}
\texttt{\{ \texttt{Separation-color-space-}'#1'\texttt{-require-values-for-the-alternative-space. \}}}
\texttt{\{ \LaTeX\texttt{-has-been-asked-to-create-a-separation-color-space,-}}
\texttt{\texttt{but-no-\}}\texttt{\{ alternative-values=\texttt{-<model> \}}}
\texttt{\texttt{\texttt{\}}\texttt{\texttt{\}}}
\texttt{\texttt{key-was-given-with-the-correct-information. \}}

\texttt{\msg_new:nnnn \{ color \} \{ separation-requires-name \}}
\texttt{\{ \texttt{Separation-color-space-}'#1'\texttt{-require-a-formal-name. \}}}
\texttt{\{ \LaTeX\texttt{-has-been-asked-to-create-a-separation-color-space,-}}
\texttt{\texttt{but-no-\}}\texttt{\{ name=\texttt{-<formal-name> \}}}
\texttt{\texttt{\}}\texttt{\texttt{\}}
\texttt{\texttt{key-was-given-with-the-correct-information. \}}

\texttt{\msg_new:nnnn \{ color \} \{ unknown-color \}}
\texttt{\{ Unknown-color-}'#1'\texttt{. \}}
\texttt{\{ \LaTeX\texttt{-has-been-asked-to-use-a-color-named-}'#1',-}}
\texttt{\texttt{but-this-has-never-been-defined. \}}

\texttt{\msg_new:nnnn \{ color \} \{ unknown-alternative-model \}}
\texttt{\{ \texttt{Separation-color-space-}'#1'\texttt{-require-an-valid-alternative-space. \}}}
\texttt{\{ \LaTeX\texttt{-has-been-asked-to-create-a-separation-color-space,-}}
\texttt{\texttt{but-the-model-given-as\}}\texttt{\{ alternative-model=\texttt{-<model> \}}}
\texttt{\texttt{\}}\texttt{\texttt{\}}
\texttt{\texttt{is-unknown. \}}

\texttt{\msg_new:nnnn \{ color \} \{ unknown-export-format \}}
\texttt{\{ Unknown-export-format-}'#1'\texttt{. \}}
\texttt{\{ \LaTeX\texttt{-has-been-asked-to-export-a-color-in-format-}'#1',-}}
\texttt{\texttt{but-this-has-never-been-defined. \}}

\texttt{\msg_new:nnnn \{ color \} \{ unknown-CIELAB-illuminant \}}
\texttt{\{ Unknown-illuminant-model-}'#1'\texttt{. \}}
\texttt{\{ \LaTeX\texttt{-has-been-asked-to-use-a-color-space-using-CIELAB-}}
\texttt{\texttt{illuminant-}'#1',-but-this-does-not-exist. \}}

\texttt{\msg_new:nnnn \{ color \} \{ unknown-model \}}
\texttt{\{ Unknown-color-model-}'#1'\texttt{. \}}
\texttt{\{ \LaTeX\texttt{-has-been-asked-to-use-a-color-model-called-}'#1',-\}}
but-this-model-is-not-set-up.
}
\msg_new:nnn \{ color \} \{ unknown-model-type \}
\{ Unknown-color-model-type-\'#1'. \}
\{
LaTeX-has-been-asked-to-create-a-new-color-model-called-\'#1',-
but-this-type-of-model-was-never-set-up.
\}
\prop_gput:Nnn \g_msg_module_name_prop \{ color \} \{ LaTeX3 \}
\prop_gput:Nnn \g_msg_module_type_prop \{ color \} \{ \}
\msg_new:nnn \{ color \} \{ show \}
\{
  The-color-\'#1-
  \tl_if_empty:nTF \{#2\}
  \{ is-undefined. \}
  \{ has-the-properties: \#2 \}
\}
(/package)
Chapter 87

l3pdf implementation

\l3pdf_stop Internal scan marks.
\scan_new:N \l3pdf_stop
(End definition for \l3pdf_stop.)
\g__pdf_init_bool A flag so we have some chance of avoiding setting things we are not allowed to. As we are potentially early in the format, we have to work a bit harder than ideal.
\bool_new:N \g__pdf_init_bool
\bool_lazy_and:nnT
{ \str_if_eq_p:Vn \fmtname { LaTeX2e } }
{ \tl_if_exist_p:N \@expl@finalise@setup@@ }
{ \tl_gput_right:Nn \@expl@finalise@setup@@
  { \tl_gput_right:Nn \@kernel@after@begindocument
    { \bool_gset_true:N \g__pdf_init_bool } } }
(End definition for \g__pdf_init_bool.)

87.1 Compression
\pdf_uncompress: Simple to do.
\cs_new_protected:Npn \pdf_uncompress:
{ \bool_if:NF \g__pdf_init_bool
  { \__pdf_backend_compresslevel:n { 0 }
    \__pdf_backend_compress_objects:n { \c_false_bool } } }
(End definition for \pdf_uncompress:. This function is documented on page 297.)
87.2 Objects

Simple to do.

```latex
\cs_new_protected:Npn \pdf_object_new:nn #1 #2
\{ \__pdf_backend_object_new:nn {#1} {#2} \}
\prg_new_conditional:Npnn \pdf_object_if_exist:n #1 { p , T , F , TF }
\{ \int_if_exist:cTF { c__pdf_backend_object_ \tl_to_str:n {#1} _int }
\{ \prg_return_true: \}
\{ \prg_return_false: \}
\}
\cs_new_protected:Npn \pdf_object_write:nn #1 #2
\{ \__pdf_backend_object_write:nn {#1} {#2}
\bool_gset_true:N \g__pdf_init_bool
\}
\cs_generate_variant:Nn \pdf_object_write:nn { nx }
\cs_new:Npn \pdf_object_ref:n #1 { \__pdf_backend_object_ref:n {#1} }
\cs_new_protected:Npn \pdf_object_unnamed_write:nn #1 #2
\{ \__pdf_backend_object_now:nn {#1} {#2}
\bool_gset_true:N \g__pdf_init_bool
\}
\cs_generate_variant:Nn \pdf_object_unnamed_write:nn { nx }
\cs_new:Npn \pdf_object_ref_last: { \__pdf_backend_object_last: }
```

(End definition for \pdf_object_new:nn and others. These functions are documented on page 295.)

```latex
\cs_new:Npn \pdf_pageobject_ref:n #1
\{ \__pdf_backend_pageobject_ref:n {#1} \}
```

(End definition for \pdf_pageobject_ref:n. This function is documented on page 296.)

87.3 Version

To compare version, we need to split the given value then deal with both major and minor version

```latex
\prg_new_conditional:Npnn \pdf_version_compare:Nn #1 #2 \{ p , T , F , TF 
\{ \use:c { __pdf_version_compare_ #1 :w } #2 \dots \s__pdf_stop \}
\cs_new:cpn { __pdf_version_compare_=:w } #1 . #2 . #3 \s__pdf_stop
\{ \bool_lazy_and:nnTF
\{ \int_compare_p:nNn \__pdf_backend_version_major: = {#1} \}
\{ \int_compare_p:nNn \__pdf_backend_version_minor: = {#2} \}
\{ \prg_return_true: \}
\{ \prg_return_false: \}
\}
\cs_new:cpn { __pdf_version_compare_<:w } #1 . #2 . #3 \s__pdf_stop
\{ \bool_lazy_or:nnTF
\{ \int_compare_p:nNn \__pdf_backend_version_major: < {#1} \}
\{ \int_compare_p:nNn \__pdf_backend_version_minor: < {#2} \}
\{ \prg_return_true: \}
```

1360
\bool_lazy_and_p:nn\{
  \int_compare_p:nNn \__pdf_backend_version_major: = {#1}
  \int_compare_p:nNn \__pdf_backend_version_minor: < {#2}
}\
  \prg_return_true: \}
\bool_lazy_or:nnTF\{
  \int_compare_p:nNn \__pdf_backend_version_major: > {#1}
  \bool_lazy_and_p:nn\{
  \int_compare_p:nNn \__pdf_backend_version_major: = {#1}
  \int_compare_p:nNn \__pdf_backend_version_minor: > {#2}
}\
  \prg_return_true: \}
\prg_return_false: \}
\cs_new:cpn\{ \__pdf_version_compare_>:w \} #1 . #2 . #3 \s__pdf_stop
\bool_lazy_or:nnTF\{
  \int_compare_p:nNn \__pdf_backend_version_major: > {#1}
  \bool_lazy_and_p:nn\{
  \int_compare_p:nNn \__pdf_backend_version_major: = {#1}
  \int_compare_p:nNn \__pdf_backend_version_minor: > {#2}
}
  \prg_return_true: \}
\prg_return_false: \}

{End definition for \pdf_version_compare:Nn and others. This function is documented on page ??.)

Split the version and set.

\cs_new:cpn\{ \__pdf_version_gset:w \} #1 . #2 . #3 \s__pdf_stop
\bool_lazy_or:nnTF\{
  \int_compare_p:nNn \__pdf_backend_version_major: > {#1}
  \bool_lazy_and_p:nn\{
  \int_compare_p:nNn \__pdf_backend_version_major: = {#1}
  \int_compare_p:nNn \__pdf_backend_version_minor: > {#2}
}
  \prg_return_true: \}
\prg_return_false: \}

{End definition for \pdf_version_gset:n, \pdf_version_min_gset:n, and \__pdf_version_gset:w.
These functions are documented on page 296.)

Wrappers.

\cs_new:Npn \pdf_version:\{
  \__pdf_backend_version_major_gset:n \{#1\}
  \__pdf_backend_version_minor_gset:n \{#2\}
}\}

{End definition for \pdf_version:, \pdf_version_major:, and \pdf_version_minor:.
These functions are documented on page 296.)
87.4 Destinations

\pdf_destination:nn
\cs_new_protected:Npn \pdf_destination:nn #1#2
{ \__pdf_backend_destination:nn (#1) (#2) }
(End definition for \pdf_destination:nn. This function is documented on page 297.)

\pdf_destination:nnnn
\cs_new_protected:Npn \pdf_destination:nnnn #1#2#3#4
{ \hbox_to_zero:n { \__pdf_backend_destination:nnnn (#1) (#2) (#3) (#4) } }
(End definition for \pdf_destination:nnnn. This function is documented on page 298.)

{/package}
Chapter 88

\l3candidates Implementation

\section{Additions to l3box}

\subsection{Viewing part of a box}

\box_clip:N
\box_clip:c
\box_gclip:N
\box_gclip:c

A wrapper around the driver-dependent code.

\begin{verbatim}
\cs_new_protected:Npn \box_clip:N #1
  { \hbox_set:Nn #1 { \__box_backend_clip:N #1 } }
\cs_generate_variant:Nn \box_clip:N { c }
\cs_new_protected:Npn \box_gclip:N #1
  { \hbox_gset:Nn #1 { \__box_backend_clip:N #1 } }
\cs_generate_variant:Nn \box_gclip:N { c }
\end{verbatim}

Trimming from the left- and right-hand edges of the box is easy: kern the appropriate parts off each side.

\begin{verbatim}
\box_set_trim:Nnnnn
\box_set_trim:cnnnn
\box_gset_trim:Nnnnn
\box_gset_trim:cnnnn
\__box_set_trim:Nnnnn
\hbox_set:Nn \l__box_internal_box
\__kernel_kern:n { -#2 }
\box_use:N #1
\__kernel_kern:n { -#4 }
\end{verbatim}

For the height and depth, there is a need to watch the baseline is respected. Material always has to stay on the correct side, so trimming has to check that there is enough material to trim. First, the bottom edge. If there is enough depth, simply set the depth, or if not move down so the result is zero depth. \texttt{\box_move_down:nn} is used in both
cases so the resulting box always contains a \texttt{lower} primitive. The internal box is used here as it allows safe use of \texttt{\textbackslash box\_set\_dp:Nn}.

\begin{verbatim}
\dim_compare:nNnTF { \box_dp:N #1 } > {#3}
  \hbox_set:Nn \l__box_internal_box
  \box_move_down:nn \c_zero_dim
  \box_use_drop:N \l__box_internal_box
\end{verbatim}

\begin{verbatim}
\box_set_dp:Nn \l__box_internal_box { \box_dp:N #1 - (#3) }
\end{verbatim}

Same thing, this time from the top of the box.

\begin{verbatim}
\dim_compare:nNnTF { \box_ht:N \l__box_internal_box } > {#5}
  \hbox_set:Nn \l__box_internal_box
  \box_move_up:nn \c_zero_dim
  \box_use_drop:N \l__box_internal_box
\end{verbatim}

\begin{verbatim}
\box_set_ht:Nn \l__box_internal_box { \box_ht:N \l__box_internal_box - (#5) }
\end{verbatim}

(End definition for \texttt{\box\_set\_trim:Nnnn}, \texttt{\box\_gset\_trim:Nnnn}, and \texttt{\_\_box\_set\_trim:Nnnnn}. These functions are documented on page 301.)

\texttt{\box\_set\_viewport:Nnnn} \texttt{\box\_gset\_viewport:Nnnn}

The same general logic as for the trim operation, but with absolute dimensions. As a result, there are some things to watch out for in the vertical direction.

\begin{verbatim}
\cs_new_protected:Npn \box_set_viewport:Nnnn \l__box_viewport:Nnnnn
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \box_gset_viewport:Nnnn \l__box_viewport:Nnnnn
\end{verbatim}

\begin{verbatim}
\cs_new_protected:Npn \__box_set_viewport:NnnnnN
\end{verbatim}

\begin{verbatim}
\cs_generate_variant:Nn \box_set_viewport:Nnnn { c }
\cs_generate_variant:Nn \box_gset_viewport:Nnnn { c }
\cs_generate_variant:Nn \__box_set_viewport:Nnnnn { c }
\end{verbatim}

1364
88.2 Additions to \texttt{l3flag}

(End definition for \texttt{\box_set_viewport:Nnnn}, \texttt{\box_gset_viewport:Nnnn}, and \texttt{\box_viewport:NnnnN}. These functions are documented on page 301.)
It might be faster to just call the “trap” function in all cases but conceptually the function name suggests we should only run it if the flag is zero in case the “trap” made customizable in the future.

```
\cs_new:Npn \flag_raise_if_clear:n #1
{ \if\cs:exist:w flag-#1-0 \cs:ende: \else: \cs:w flag-#1 \cs:ende: 0 ; \fi: }
```

(End definition for \flag_raise_if_clear:n. This function is documented on page 302.)

### 88.3 Additions to l3msg

A short-hand used for \int_show:n and similar functions that passes to \tl_show:n the result of applying #1 (a function such as \int_eval:n) to the expression #2. The use of f-expansion ensures that #1 is expanded in the scope in which the show command is called, rather than in the group created by \iow_wrap:nnN. This is only important for expressions involving the \currentgrouplevel or \currentgrouptype. On the other hand we want the expression to be converted to a string with the usual escape character, hence within the wrapping code.

```
\cs_new_protected:Npn \msg_show_eval:Nn #1#2
{ \exp_args:Nf \__msg_show_eval:nnN {#1} {#2} \tl_show:n }
\cs_new_protected:Npn \msg_log_eval:Nn #1#2
{ \exp_args:Nf \__msg_show_eval:nnN {#1} {#2} \tl_log:n }
\cs_new_protected:Npn \__msg_show_eval:nnN #1#2#3 {#3}
```

(End definition for \msg_show_eval:Nn, \msg_log_eval:Nn, and \__msg_show_eval:nnN. These functions are documented on page 303.)

Each item in the variable is formatted using one of the following functions. We cannot use \ and so on because these short-hands cannot be used inside the arguments of messages, only when defining the messages.

```
\cs_new:Npx \msg_show_item:n #1
{ \exp_not:N \tl_to_str:n { {#1} } }
\cs_new:Npx \msg_show_item_unbraced:n #1
{ \exp_not:N \tl_to_str:n {#1} }
```

```
\cs_new:Npx \msg_show_item:nn #1#2
{ \use:nn { ~ } { ~ } \exp_not:N \tl_to_str:n { {#1} } \use:nn { ~ } { ~ } \exp_not:N \tl_to_str:n { {#2} } }
\cs_new:Npx \msg_show_item_unbraced:nn #1#2
{ \use:nn { ~ } { ~ } \exp_not:N \tl_to_str:n {#1} \use:nn { ~ } { ~ } \exp_not:N \tl_to_str:n {#2} }
```

1366
88.4 Additions to \texttt{l3prg}

\begin{verbatim}
\bool_set_inverse:N \bool_set_inverse:c
\bool_gset_inverse:N \bool_gset_inverse:c
\end{verbatim}

Set to false or true locally or globally.

\begin{verbatim}
cs_new_protected:Npn \bool_set_inverse:N \bool_set_inverse:c
| \cs_generate_variant:Nn \bool_set_inverse:N { c }
| { \bool_if:NTF #1 { \bool_set_false:N } { \bool_set_true:N } #1 }
| \cs_generate_variant:Nn \bool_gset_inverse:N { c }
| { \bool_if:NTF #1 { \bool_gset_false:N } { \bool_gset_true:N } #1 }
\end{verbatim}

(End definition for \texttt{\bool_set_inverse:N} and \texttt{\bool_gset_inverse:N}. These functions are documented on page \texttt{303}.)

\begin{verbatim}
\s__bool_mark \s__bool_stop
\end{verbatim}

Internal scan marks.

\begin{verbatim}
cs_new:N \s__bool_mark \s__bool_stop
\end{verbatim}

(End definition for \texttt{\s__bool_mark} and \texttt{\s__bool_stop}.)

\begin{verbatim}
\bool_case_true:n \bool_case_true:nT \bool_case_true:nF \bool_case_true:n
\bool_case_false:nTF \bool_case_false:nT \bool_case_false:nF \bool_case_false:n
\__bool_case:NnTF \__bool_case_true:w \__bool_case_false:w \__bool_case_end:nw
\end{verbatim}

For boolean cases the overall idea is the same as for \texttt{\tl_case:nnTF} as described in \texttt{l3tl}.

\begin{verbatim}
cs_new:Npn \bool_case_true:nTF \bool_case_true:nT \bool_case_true:nF \bool_case_true:n
| \bool_case_false:nTF \bool_case_false:nT \bool_case_false:nF \bool_case_false:n
| \__bool_case:NnTF \__bool_case_true:w \__bool_case_false:w \__bool_case_end:nw
\end{verbatim}

(End definition for \texttt{\bool_case_true:n} and \texttt{\bool_case_false:n}.)

1367
88.5 Additions to \texttt{l3prop}

\texttt{\_prop_use_i_delimit_by_s_stop:nw}

Functions to gobble up to a scan mark.

\texttt{\_prop_rand_key_value:N}

Contrarily to \texttt{clist}, \texttt{seq} and \texttt{tl}, there is no function to get an item of a \texttt{prop} given an integer between 1 and the number of items, so we write the appropriate code. There is no bounds checking because \texttt{\int_rand:nn} is always within bounds. The initial \texttt{\int_-value:w} is stopped by the first \texttt{\s__prop} in \texttt{#1}.

\texttt{\_prop_rand_item:w}

The idea is to first expand both sequences, adding the usual \{ ? \texttt{\prg_break:} \} \}

to the end of each one. This is most conveniently done in two steps using an auxiliary function. The mapping then throws away the first tokens of \texttt{#2} and \texttt{#5}, which for items in both sequences are \texttt{\s__seq \_seq_item:n}. The function to be mapped are then be applied to the two entries. When the code hits the end of one of the sequences, the break material stops the entire loop and tidy up. This avoids needing to find the count of the two sequences, or worrying about which is longer.
\cs_new:Npn \seq_mapthread_function:NNN #1#2#3
  \exp_after:wN \__seq_mapthread_function:wNN #2 \s__seq_stop #1 #3
\cs_new:Npn \__seq_mapthread_function:wNN \s__seq #1 \s__seq_stop #2#3
  \exp_after:wN \__seq_mapthread_function:wNw #2 \s__seq_stop #3
\cs_new:Npn \__seq_mapthread_function:wNw \s__seq #1 \s__seq_stop #2
  \__seq_mapthread_function:Nnnwnn #2 #1 { ? \prg_break: } { }
\prg_break_point:
\cs_new:Npn \__seq_mapthread_function:Nnnwnn #1#2#3#4 \s__seq_stop #5#6
  \use_none:n #2 \use_none:n #5
  #1 {#3} {#6}
\__seq_mapthread_function:Nnnwnn \s__seq #3 \__seq_item:n
\__seq_pop_item_def:
\cs_generate_variant:Nn \seq_mapthread_function:NNN { Nc , c , cc }

(End definition for \seq_mapthread_function:NNN and others. This function is documented on page 305.)

\seq_set_filter:NNn \seq_gset_filter:NNn \__seq_set_filter:NNNn

Similar to \seq_map_inline:Nn, without a \prg_break_point: because the user’s code is performed within the evaluation of a boolean expression, and skipping out of that would break horribly. The \__seq_wrap_item:n function inserts the relevant \__seq_item:n without expansion in the input stream, hence in the x-expanding assignment.

\cs_new_protected:Npn \seq_set_filter:NNn \seq_gset_filter:NNn \__seq_set_filter:NNNn
  \__seq_set_filter:NNNn \__kernel_tl_set:Nx
\cs_new_protected:Npn \seq_gset_filter:NNn \__seq_set_filter:NNNn \__kernel_tl_gset:Nx
\cs_new_protected:Npn \__seq_set_filter:NNNn #1#2#3#4
  \__seq_push_item_def:n \__seq_pop_item_def:

(End definition for \seq_set_filter:NNn, \seq_gset_filter:NNn, and \__seq_set_filter:NNNn. These functions are documented on page 305.)

\seq_set_from_inline_x:Nnn \seq_gset_from_inline_x:Nnn \__seq_set_from_inline_x:NNnn

Set \__seq_item:n then map it using the loop code.

\cs_new_protected:Npn \seq_set_from_inline_x:Nnn \seq_gset_from_inline_x:Nnn \__seq_set_from_inline_x:NNnn
  \__seq_set_from_inline_x:NNnn \__kernel_tl_set:Nx
\cs_new_protected:Npn \seq_gset_from_inline_x:Nnn \__seq_set_from_inline_x:NNnn \__kernel_tl_gset:Nx
\cs_new_protected:Npn \__seq_set_from_inline_x:NNnn #1#2#3#4
  \__seq_push_item_def:n \__seq_pop_item_def:

Set \__seq_item:n then map it using the loop code.
Reuse \seq_set_from_inline_x:Nnn.

\seq_set_from_function:NnN
\seq_gset_from_function:NnN

End definition for \seq_set_from_inline_x:Nnn, \seq_gset_from_inline_x:Nnn, and \__seq_set_from_inline_x:NNnn. These functions are documented on page 305.

\__seq_set_item:Nnn
\__seq_set_item:nnnn
\__seq_set_item:wwnn
\__seq_set_item_end:w

Useful to more quickly go through items.

The conditionals are distinguished from the Nnn versions by the last argument \use_ii:nn vs \use_i:nn.

Save the item to be stored and evaluate the position and the sequence length only once. Then depending on the sign of the position, check that it is not bigger than the length (in absolute value) nor zero.

(End definition for \__seq_set_item:NN.)
If the position is not ok, \_\_seq_set_item\_false:nnNNNN calls an error or returns false (depending on the \use_i:nn vs \use_ii:nn argument mentioned above).

\cs_new_protected:Npn \_\_seq_set_item\_false:nnNNNN #1#2#3#4#5#6
\begin{verbatim}
{ #6
 \msg_error:nxxx { \_seq } { item-too-large }
 \token_to_str:N #3 \{#2\} \{#1\}
 \prg_return_false: }
\end{verbatim}
\msg_new:nnnn { \_seq } { item-too-large }
\{ Sequence-'#1'-does-not-have-an-item-#3 \}
\begin{verbatim}
{ An-attempt-was-made-to-push-or-pop-the-item-at-position-#3-
of-'#1','-but-this-
 \int_compare:nTF { #3 = 0 } \{ position-does-not-exist. \} \int_compare:nF { #2 = 1 } \{#2\}. }
\end{verbatim}

If the position is ok, \_\_seq_set_item:nNnnNNNN makes the assignment and returns true (in the case of conditionnals). Here \#1 is an integer expression (position minus one), it needs to be evaluated. The sequence \#5 starts with \_s\_seq (even if empty), which stops the integer expression and is absorbed by it. The \if_meaning:w test is slightly faster than an integer test (but only works when testing against zero, hence the offset we chose in the position). When we are done skipping items, insert the saved item \l__-_seq\_internal\_a_tl. For put functions the last argument of \_\_seq_set_item\_end:w is \use_none:nn and it absorbs the item \#2 that we are removing: this is only useful for the pop functions.

\cs_new_protected:Npn \_\_seq_set_item:nNnnNNNN #1#2#3#4#5#6#7#8
\begin{verbatim}
{ #7 #5
 \_s\_seq
 \exp_after:wN \_\_seq_set_item:wn
 \int_value:w \_\_seq_int_eval:w #1 #5 \_s\_seq_stop #6
}
\end{verbatim}
\cs_new:Npn \_\_seq_set_item:wn #1 \_seq_item:n #2
\begin{verbatim}
{ \if_meaning:w 0 #1 \_\_seq_set_item\_end:w \fi:
 \exp_not:n \{ \_\_seq_item:n \{#2\} \}
 \exp_after:wN \_\_seq_set_item:wn
 \int_value:w \_\_seq_int_eval:w #1 - 1 \_s\_seq
}
\end{verbatim}
\cs_new:Npn \_\_seq_set_item\_end:w #1 \_\_seq_set_item\_end:wn #1 #2 \_\_seq_item:n #2
\begin{verbatim}
{ \if_meaning:w 0 #1 \_\_seq_set_item\_end:w \fi:
 \exp_not:n \{ \_\_seq_item:n \{#2\} \}
 \exp_after:wN \_\_seq_set_item:wn
 \\int_value:w \_\_seq_int_eval:w #1 - 1 \_s\_seq
}
\end{verbatim}
\cs_new:Npn \_\_seq_set_item\_end:wn #1 \_\_seq_set_item\_end:end:wn #1 \_\_seq_item:n #2 \_\_seq_stop #5
\begin{verbatim}
{ \_s\_seq
 \exp_not:o \l__-_seq\_internal\_a_tl
 \exp_not:n \{#4\}
\end{verbatim}

1371
The \texttt{NnN} versions simply call the conditionals, for which we will rely on the internals of \texttt{\seq_set_item:Nnn}. The \texttt{NnN} auxiliary eventually inserts \texttt{\l__seq_internal_a_tl} in place of the item found in the sequence, so we empty that. Instead of the last argument \texttt{\use_i:n} or \texttt{\use_ii:n} used for put functions, we introduce \texttt{\__seq_pop_item:nn}, which stores \texttt{\q_no_value} before calling its second argument (\texttt{\prg_return_true:/false:}) to end the conditional. The item found is passed to \texttt{\__seq_pop_item_aux:w}, which interrupts the \texttt{x}-expanding sequence assignment and stores the item using the assignment function in \texttt{\l__seq_internal_b_tl}.

\begin{verbatim}
\cs_new_protected:Npn \__seq_pop_item:NnNN #1#2#3#4
\tl_clear:N \l__seq_internal_a_tl
\tl_set:Nn \l__seq_internal_b_tl { \__kernel_tl_set:Nx #3 }
\exp_args:Nff \__seq_set_item:nnNNNN
{ \int_eval:n {#2} } { \seq_count:N #1 }
#1 \__seq_pop_item_aux:w #4 \__seq_pop_item:nn
\end{verbatim}

(End definition for \texttt{\seq_set_item:Nnn} and others. These functions are documented on page 306.)
Additions to i3sys

Various different engines, various different ways to extract the data!

```
\c_sys_engine_version_str
\str_const:Nx \c_sys_engine_version_str
{ \str_case:on \c_sys_engine_str
    { pdftex }
        { \fp_eval:n { round(\int_use:N \tex_pdftexversion:D / 100 , 2) }
            . \tex_pdftexrevision:D
        }
    { ptxe }
        { \cs_if_exist:NT \tex_ptexversion:D
            { \int_use:N \tex_ptexversion:D
                . \int_use:N \tex_ptexminorversion:D
                \tex_ptexrevision:D
                - \int_use:N \tex_epTeXversion:D
            }
        }
    { lualatex }
        { \fp_eval:n { round(\int_use:N \tex_luatexversion:D / 100 , 2) }
            . \tex_luatexrevision:D
        }
    { uptex }
        { \cs_if_exist:NT \tex_ptexversion:D
            { \int_use:N \tex_ptexversion:D
                . \int_use:N \tex_ptexminorversion:D
                \tex_ptexrevision:D
                - u \int_use:N \tex_uptexversion:D
                \tex_uptexrevision:D
                - \int_use:N \tex_epTeXversion:D
            }
        }
    { xetex }
        { \int_use:N \tex_XeTeXversion:D
```
Actually much easier than either the standard open or input versions! When calling \_\_kernel_ior_open:Nn the file the pipe is added to signal a shell command, but the quotes are not added yet—they are added later by \_\_kernel_file_name_quote:n.

\cs_new_protected:Npn \ior_shell_open:Nn #1#2
\{\sys_if_shell:TF
\{ \exp_args:No \_\_ior_shell_open:nN \{ tl_to_str:n \{#2\} \} #1 \}
\{ \msg_error:nn \{ ior \} \{ pipe-failed \} \}
\}
\cs_new_protected:Npn \_\_ior_shell_open:nN #1#2
\{\tl_if_in:nnTF \{#1\} \{ " \}
\{ \msg_error:nmx
\{ \kernel \} \{ quote-in-shell \} \{#1\}
\}
\{ \_\_kernel_ior_open:Nn \{ |#1\} \}
\msg_new:nnnn \{ ior \} \{ pipe-failed \}
\{ Cannot-run-piped-system-commands. \}
\{ LaTeX-tried-to-call-a-system-process-but-this-was-not-possible.\}
\{ Try-the-"--shell-escape"-(or-"--enable-pipes")-option. \}
\}

88.9 Additions to \l3tl

88.9.1 Building a token list

Between \tl_build_begin:N\tl_var and \tl_build_end:N\tl_var, the \tl_var has the structure
\exp_end: ... \exp_end: \_\_tl_build_last:NNn \{ assignment\} \{ next tl\}
\{(left)\} \{right\}

where \{right\} is not braced. The “data” it represents is \{left\} followed by the “data” of \{next tl\} followed by \{right\}. The \{next tl\} is a token list variable whose name is that of \tl_var followed by \'. There are between 0 and 4 \exp_end: to keep track of when \{left\}
and \(\text{right}\) should be put into the \(\langle\text{next tl}\rangle\). The \(\langle\text{assignment}\rangle\) is \cs_set_nopar:Npx if the variable is local, and \cs_gset_nopar:Npx if it is global.

First construct the \(\langle\text{next tl}\rangle\): using a prime here conflicts with the usual expl3 convention but we need a name that can be derived from \#1 without any external data such as a counter. Empty that \(\langle\text{next tl}\rangle\) and setup the structure. The local and global versions only differ by a single function \cs_(g)set_nopar:Npx used for all assignments: this is important because only that function is stored in the \(\langle\text{tl var}\rangle\) and \(\langle\text{next tl}\rangle\) for subsequent assignments. In principle \__tl_build_begin:NNN could use \tl_(g)clear_new:N to empty \#1 and make sure it is defined, but logging the definition does not seem useful so we just do \#3 \#1 \{} to clear it locally or globally as appropriate.

\begin{verbatim}
\cs_new_protected:Npn \tl_build_begin:N #1
  \__tl_build_begin:NN \cs_set_nopar:Npx #1
\cs_new_protected:Npn \tl_build_gbegin:N #1
  \__tl_build_begin:NN \cs_gset_nopar:Npx #1
\cs_new_protected:Npn \__tl_build_begin:NN #1#2
  \exp_args:Nc \__tl_build_begin:NNN { \cs_to_str:N #2 ' } #2 #1
\cs_new_protected:Npn \__tl_build_begin:NNN #1#2#3
  #3 #1 \{} \#3 #2
  \exp_not:n { \exp_end: \exp_end: \exp_end: \exp_end: }
  \exp_not:n { \__tl_build_last:NNn #3 \#1 \{} }
\__tl_build_last:NNn
\__tl_build_put:nn
\__tl_build_put:nw
\end{verbatim}

(End definition for \tl_build_begin:N and others. These functions are documented on page 307.)

The \begin and \gbegin functions already clear enough to make the token list variable effectively empty. Eventually the \begin and \gbegin functions should check that \#1' is empty or undefined, while the \clear and \gclear functions ought to empty \#1', \#1'', and so on, similar to \tl_build_end:N. This only affects memory usage.

\begin{verbatim}
\cs_new_eq:NN \tl_build_clear:N \tl_build_begin:N
\cs_new_eq:NN \tl_build_gclear:N \tl_build_gbegin:N
\end{verbatim}

(End definition for \tl_build_clear:N and \tl_build_gclear:N. These functions are documented on page 307.)

Similar to \tl_put_right:Nn, but apply \exp:w to \#1. Most of the time this just removes one \exp_end:. When there are none left, \__tl_build_last:NNn is expanded instead. It resets the definition of the \langle\text{tl var}\rangle by ending the \exp_not:n and the definition early. Then it makes sure the \langle\text{next tl}\rangle (its argument \#1) is set-up and starts a new definition. Then \__tl_build_put:nn and \__tl_build_put:nw place the \langle\text{left}\rangle part of the original \langle\text{tl var}\rangle as appropriate for the definition of the \langle\text{next tl}\rangle (the \langle\text{right}\rangle part is left in the right place without ever becoming a macro argument). We use \exp_after:wN rather than some \exp:args:No to avoid reading arguments that are likely very long token lists. We use \cs_(g)set_nopar:Npx rather than \tl_(g)set:Nx partly for the same reason and partly because the assignments are interrupted by brace tricks, which implies that the assignment does not simply set the token list to an x-expansion of the second argument.

\begin{verbatim}
\cs_new_protected:Npn \tl_build_put_right:Nn \tl_build_put_right:Nx
\tl_build_gput_right:Nx \tl_build_gput_right:Nx
\__tl_build_put:nn \__tl_build_put:nw
\__tl_build_last:NNn \__tl_build_put:nn \__tl_build_put:nw
\end{verbatim}
See \texttt{tl\_build\_put\_right:Nn} for all the machinery. We could easily provide \texttt{tl\_build\_put\_left\_right:Nnn}, by just add the \langle right\rangle material after the \langle left\rangle in the \texttt{x}\texttt{-expanding assignment}.

(End definition for \texttt{tl\_build\_put\_right:Nn} and others. These functions are documented on page 308.)
The idea is to expand the \langle tl \rangle then the \langle next tl \rangle and so on, all within an x-expanding assignment, and wrap as appropriate in \exp_not:n. The various \langle left \rangle parts are left in the assignment as we go, which enables us to expand the \langle next tl \rangle at the right place. The various \langle right \rangle parts are eventually picked up in one last \exp_not:n, with a brace trick to wrap all the \langle right \rangle parts together.

\cs_new_protected:Npn \tl_build_get:NN { \__tl_build_get:NNN \__kernel_tl_set:Nx }
\cs_new_protected:Npn \__tl_build_get:NNN #1 #2 #3 { #1 #3 { \if_false: { \fi: \exp_after:wN \__tl_build_get:w #2 } } }
\cs_new:Npn \__tl_build_get:w #1 \__tl_build_last:NNn #2 #3 #4 { \exp_not:n {#4} \if_meaning:w \c_empty_tl #3 \exp_after:wN \__tl_build_get_end:w \fi: \exp_after:wN \__tl_build_get:w #3 }
\cs_new:Npn \__tl_build_get_end:w #1 #2 #3 { \exp_after:wN \exp_not:n \exp_after:wN { \if_false: } \fi: }

Get the data then clear the \langle next tl \rangle recursively until finding an empty one. It is perhaps wasteful to repeatedly use \cs_to_sr:N. The local/global scope is checked by \tl_set:Nx or \tl_gset:Nx.

\cs_new_protected:Npn \tl_build_end:N #1 { \__tl_build_get:NNN \__kernel_tl_set:Nx #1 #1 \exp_args:Nc \__tl_build_end_loop:NN { \cs_to_str:N #1 ' } \tl_clear:N }
\cs_new_protected:Npn \tl_build_gend:N #1 { \__tl_build_get:NNN \__kernel_tl_gset:Nx #1 #1 \exp_args:Nc \__tl_build_end_loop:NN { \cs_to_str:N #1 ' } \tl_gclear:N }
\cs_new_protected:Npn \__tl_build_end_loop:NN #1 #2 { \if_meaning:w \c_empty_tl #1 \exp_after:wN \use_none:nnnnnn \fi: \exp_after:wN \__tl_build_end_loop:NN { \cs_to_str:N #1 ' } #2 }

(End definition for \tl_build_end:N, \tl_build_gend:N, and \__tl_build_end_loop:NN. These functions are documented on page 308.)
88.9.2 Other additions to l3tl

For the braced version \__tl_range_braced:w sets up \__tl_range_collect_braced:w which stores items one by one in an argument after the semicolon. The unbraced version is almost identical. The version preserving braces and spaces starts by deleting spaces before the argument to avoid collecting them, and sets up \__tl_range_collect:nn with a first argument of the form \{ \langle collected \rangle \langle tokens \rangle \}, whose head is the collected tokens and whose tail is what remains of the original token list. This form makes it easier to move tokens to the \langle collected \rangle tokens.

88.10 Additions to l3token

While \char_generate:nn can produce active characters in some engines it cannot in general. It would be possible to simply change the catcode of space but then the code would need to avoid all spaces, making it quite unreadable. Instead we use the primitive \tex_lowercase:D trick.
Most of the work is done by \_\_peek_execute_branches\ldots:, which calls either \_\_peek_true:w or \_\_peek_false:w according to whether the next token \_\_peek_token matches the search token (stored in \_\_peek_search_token and \_\_peek_search_tl). Here, in the true case we run \_\_peek_collect_true:w, which generally calls \_\_peek_collect:N to store the peeked token into \_\_peek_collect_tl, except in special non-N-type cases (begin-group, end-group, or space), where a frozen token is stored. The true branch calls \_\_peek_execute_branches\ldots: to fetch more matching tokens. Once there are no more, \_\_peek_false_aux:n closes the safe-align group and runs the user's inline code.
\cs_new_protected:Npn \__peek_collect:N #1
\tl_put_right:Nn \l__peek_collect_tl {#1} \
\__peek_true_aux:w
\}
\cs_new_protected:Npn \__peek_collect_remove:nw #1
\tl_put_right:Nn \l__peek_collect_tl {#1}
\exp_after:wN \__peek_true_remove:w
\}

(End definition for \peek_catcode_collect_inline:Nn and others. These functions are documented on page 309.)

{/package}
Chapter 89

\textbf{l3deprecation implementation}

89.1 Patching definitions to deprecate

\texttt{\_\_\_kernel\_patch\_deprecation:nnNpn \{\langle date\rangle\} \{\langle replacement\rangle\} \{\langle definition\rangle\} \{\langle function\rangle\} \{\langle parameters\rangle\} \{\langle code\rangle\}}

defines the \texttt{\langle function\rangle} to produce a warning and run its \texttt{\langle code\rangle}, or to produce an error and not run any \texttt{\langle code\rangle}, depending on the expl3 date.

- If the expl3 date is less than the \texttt{\langle date\rangle} (plus 6 months in case \texttt{undo-recent-deprecations} is used) then we define the \texttt{\langle function\rangle} to produce a warning and run its code. The warning is actually suppressed in two cases:
  - if neither \texttt{undo-recent-deprecations} nor \texttt{enable-debug} are in effect we may be in an end-user’s document so it is suppressed;
  - if the command is expandable then we cannot produce a warning.

- Otherwise, we define the \texttt{\langle function\rangle} to produce an error.

In both cases we additionally make \texttt{\debug\_on:n \{deprecation\}} turn the \texttt{\langle function\rangle} into an \texttt{\outer} error, and \texttt{\debug\_off:n \{deprecation\}} restore whatever the behaviour was without \texttt{\debug\_on:n \{deprecation\}}.

In later sections we use the \texttt{l3doc} key \texttt{deprecated} with a date equal to that \texttt{\langle date\rangle} plus 6 months, so that \texttt{l3doc} will complain if we forget to remove the stale \texttt{\langle parameters\rangle} and \texttt{\langle code\rangle}}.

In the explanations below, \texttt{\langle definition\rangle} \texttt{\langle function\rangle} \texttt{\langle parameters\rangle} \texttt{\langle code\rangle}} or assignments that only differ in the scope of the \texttt{\langle definition\rangle} will be called “the standard definition”.

(The parameter text is grabbed using \texttt{\#5\#}.) The arguments of \texttt{\_\_\_kernel\_deprecation\_code:nn} are run upon \texttt{\debug\_on:n \{deprecation\}} and \texttt{\debug\_off:n \{deprecation\}}, respectively. In both scenarios we the \texttt{\langle function\rangle} may be \texttt{\outer} so we redefine it with \texttt{\tex\_let:D} before redefining it, with \texttt{\_\_\_kernel\_deprecation\_error:NNn} or with some code added shortly.
In case we want a warning, the ⟨function⟩ is defined to produce such a warning without grabbing any argument, then redefine itself to the standard definition that the ⟨function⟩ should have, with arguments, and call that definition. The \exp_not:n avoid needing to double the #, which we could not do anyways. We then deal with the code for \debug_off:n {deprecation}: presumably someone doing that does not need the warning so we simply do the standard definition.

In case we want neither warning nor error, the ⟨function⟩ is given its standard definition. Here #1 is \cs_new:Npn or \cs_new_protected:Npn and #2 is ⟨function⟩ ⟨parameters⟩ ⟨{code}⟩, so #1#2 performs the assignment. For \debug_off:n {deprecation} we want to use the same assignment but with a different scope, hence the \cs_if_eq:NNTF test.

(End definition for \__kernel_patch_deprecation:nnNNpn and others.)
The \texttt{outer} definition here ensures the command cannot appear in an argument. Use this auxiliary on all commands that have been removed since 2015.

```
\cs_new_protected:Npn \__kernel_deprecation_error:Nnn #1#2#3
\begin{verbatim}
  \tex_protected:D \tex_outer:D \tex_edef:D #1
  \exp_not:N \msg_expandable_error:nmmm
  { depression } { deprecated-command }
  { \tl_to_str:n {#3} } { \token_to_str:N #1 } { \tl_to_str:n {#2} }
  \exp_not:N \msg_error:nnxxx
  { depression } { deprecated-command }
  { \tl_to_str:n {#3} } { \token_to_str:N #1 } { \tl_to_str:n {#2} }
\end{verbatim}
```

(End definition for \texttt{\__kernel_deprecation_error:Nnn}.)

```
\msg_new:nnn { deprecation } { deprecated-command }
\begin{verbatim}
  \tl_if_blank:nF {#3} { Use \tl_trim_spaces:n {#3} -not- }
  #2-deprecated-on-#1.
\end{verbatim}
```

### 89.2 Removed functions

Short-hands for old commands whose definition does not matter any more as they were removed.

```
\cs_new_protected:Npn \__deprecation_old_protected:Nnn #1#2#3
\begin{verbatim}
  \__kernel_patch_deprecation:nnNNpn {#3} {#2}
  \cs_gset_protected:Npn #1 { }
\end{verbatim}
```

```
\cs_new_protected:Npn \__deprecation_old:Nnn #1#2#3
\begin{verbatim}
  \__kernel_patch_deprecation:nnNNpn {#3} {#2}
  \cs_gset:Npn #1 { }
\end{verbatim}
```

\__deprecation_old_protected:Nnn \box_gset_eq_clear:NN 2021-07-01
\__deprecation_old_protected:Nnn \box_set_eq_clear:NN 2021-07-01
\__deprecation_old_protected:Nnn \box_resize:Nnn 2019-01-01
\__deprecation_old_protected:Nnn \box_use_clear:N 2019-01-01
\__deprecation_old_protected:Nnn \c_job_name_tl \c_sys_jobname_str 2017-01-01
\__deprecation_old_protected:Nnn \c_minus_one { -1 } 2019-01-01
\__deprecation_old_protected:Nnn \c_zero { 0 } 2020-01-01
\__deprecation_old_protected:Nnn \c_one { 1 } 2020-01-01
```
\_\_deprecation\_old:Nnn \c_two
{ 2 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_three
{ 3 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_four
{ 4 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_five
{ 5 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_six
{ 6 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_seven
{ 7 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_eight
{ 8 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_nine
{ 9 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_ten
{ 10 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_eleven
{ 11 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_twelve
{ 12 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_thirteen
{ 13 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_fourteen
{ 14 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_fifteen
{ 15 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_sixteen
{ 16 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_thirty_two
{ 32 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_one_hundred
{ 100 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_two_hundred_fifty_five
{ 255 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_two_hundred_fifty_six
{ 256 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_one_thousand
{ 1000 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_ten_thousand
{ 10000 } { 2020-01-01 }
\_\_deprecation\_old:Nnn \c_term_ior
{ -1 } { 2021-07-01 }
\_\_deprecation\_old:Nnn \dim_case:nnn
{ \dim_case:nF } { 2015-07-14 }
\_\_deprecation\_old_protected:Nnn \file_add_path:nN
{ \file_get_full_name:nN } { 2019-01-01 }
\_\_deprecation\_old_protected:Nnn \file_if_exist_input:nT
{ \file_if_exist:nT and \file_input:n } { 2018-03-05 }
\_\_deprecation\_old_protected:Nnn \file_if_exist_input:nTF
{ \file_if_exist:nT and \file_input:n } { 2018-03-05 }
\_\_deprecation\_old_protected:Nnn \file_list:
{ \file_log_list: } { 2019-01-01 }
89.3 Deprecated l3str functions

\str_lower_case:n
\str_lower_case:f
\str_upper_case:n
\str_upper_case:f
\str_fold_case:n
\str_fold_case:V

\strDeclareEightBitEncoding:nnn

This command was made internal, with one more argument. There is no easy way to compute a reasonable value for that extra argument so we take a value that is big enough to accommodate all of Unicode.
89.4 Deprecated \texttt{l3seq} functions

\seq_indexed_map_inline:Nn
\seq_indexed_map_function:NN

(End definition for \texttt{\_\_str\_declare\_eight\_bit\_encoding:nn}. This function is documented on page ??.)

89.5 Deprecated \texttt{l3sys} functions

\sys_load_deprecation:

(End definition for \texttt{\_\_kernel\_patch\_deprecation:nnNNpn}. This function is documented on page ??.)

89.6 Deprecated \texttt{l3tl} functions

\tl_lower_case:n
\tl_upper_case:n
\tl_mixed_case:n

(End definition for \texttt{\_\_kernel\_patch\_deprecation:nnNNpn}. This function is documented on page ??.)
89.7 Deprecated \l3token functions

\char_lower_case:N
\char_upper_case:N
\char_fold_case:N
\char_str_lower_case:N
\char_str_upper_case:N
\char_str_fold_case:N

__(End definition for \char_lower_case:N and others. These functions are documented on page ??.)

A little extra fun here to deal with the expansion.

}``

1389
(End definition for \texttt{\texttt{peek.catcode.ignore.spaces}:NTF and others. These functions are documented on page ??.)

```
</package>
```
Index

The italic numbers denote the pages where the corresponding entry is described, numbers underlined point to the definition, all others indicate the places where it is used.

Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>!</td>
<td>252</td>
</tr>
<tr>
<td>&quot;</td>
<td>18690, 18693, 29540, 32018, 32042, 32048, 32052, 32058, 32062, 32068, 32074, 32081, 32082, 32088, 32092, 32094, 32203</td>
</tr>
<tr>
<td>#</td>
<td>10509, 13811, 18690, 31887</td>
</tr>
<tr>
<td>$</td>
<td>5394, 13810, 18690, 31887</td>
</tr>
<tr>
<td>%</td>
<td>10511, 13812, 18690, 31887</td>
</tr>
<tr>
<td>&amp;</td>
<td>9318, 13803, 18690, 18693, 20692, 31899</td>
</tr>
<tr>
<td>+</td>
<td>13410, 13433, 19071, 19073, 19077, 19085</td>
</tr>
<tr>
<td>*</td>
<td>252</td>
</tr>
<tr>
<td>**</td>
<td>252</td>
</tr>
<tr>
<td>-</td>
<td>20692, 31899</td>
</tr>
<tr>
<td>\</td>
<td>251, 252</td>
</tr>
<tr>
<td>*</td>
<td>20692, 31899</td>
</tr>
<tr>
<td>-</td>
<td>251, 252</td>
</tr>
<tr>
<td>\</td>
<td>29540, 32017, 32105, 32106, 32115, 32116, 32125, 32126, 32143, 32155, 32156, 32206, 32207, 32237, 32238, 32241, 32242</td>
</tr>
<tr>
<td>/</td>
<td>252</td>
</tr>
<tr>
<td>/</td>
<td>188, 4273</td>
</tr>
<tr>
<td>:</td>
<td>13809</td>
</tr>
<tr>
<td>::</td>
<td>42, 381, 408, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2288, 2292, 2293, 2296, 2299, 2306, 2309, 2312, 2318, 2405, 2481, 2483, 2488, 2495, 2499, 2502, 2507, 2522, 2563, 2564, 2565, 2566, 2577</td>
</tr>
<tr>
<td>:=</td>
<td>42, 2284, 2405, 2566</td>
</tr>
<tr>
<td>:V</td>
<td>42, 2312</td>
</tr>
<tr>
<td>:V_unbraced</td>
<td>42, 2480</td>
</tr>
<tr>
<td>:c</td>
<td>42, 2286</td>
</tr>
<tr>
<td>:e</td>
<td>42, 2290, 2405</td>
</tr>
<tr>
<td>:e_unbraced</td>
<td>42, 2480, 2522</td>
</tr>
</tbody>
</table>
Index

\belowdisplayshortskip .......... 205
\belowdisplayskip .......... 206
\bseries ................. 31861
\binoppenalty .......... 207
\bodydir ................. 793
\bodydirection ............. 794
bool commands:
  \bool_case_false:n ........ 304, 36132
  \bool_case_false:nTF ........ 304, 36132, 36142, 36144
  \bool_case_true:n ........ 304, 36132
  \bool_case_true:nTF ........ 304, 36132, 36134, 36136
  \bool_const:Nn ............. 65, 8391
  \bool_do_until:Nn ........... 68, 8599
  \bool_do_until:n ............ 69, 8605
  \bool_do_while:Nn ........... 69, 8599
  \bool_do_while:n ............ 69, 8605
  \bool_get:N ............... 223, 21258
  \bool_get:Nn ............... 65, 8413
  \bool_get_eq:NN ............ 65, 6417, 6768, 8400
  \bool_get_false:N ........... 65, 6715, 8397, 14018, 14027, 36128
  \bool_get_inverse:N .......... 223, 21226
  \bool_get_inverse:N ........ 303, 36124
  \bool_get_true:N ............ 65, 6781, 8397, 8862, 14008, 35884, 35906, 35913, 36128
  \bool_if:N ............... 65, 145, 2966,
  5509, 5518, 5599, 6114, 6200, 6218,
  6236, 6390, 6609, 6617, 6850, 7453,
  7476, 7549, 7771, 7941, 7947, 7988,
  8355, 8360, 8430, 8439, 8594, 8596,
  8600, 8602, 8660, 10729, 10736,
  14022, 14031, 19715, 19723, 20729,
  20822, 20901, 21139, 21148, 21194,
  21410, 21412, 21414, 21460, 21462,
  21464, 21502, 21504, 21506, 21522,
  21524, 21526, 21573, 21614, 21633,
  21635, 21640, 21647, 21711, 21740,
  21750, 21778, 30303, 33361, 34056,
  35889, 35962, 36125, 36128, 36150
  \bool_if:nTF ... 65, 67, 69, 882, 6203,
  8441, 8473, 8544, 8551, 8570, 8577,
  8586, 8607, 8616, 8620, 8629, 8702,
  11908, 11913, 36155, 36161, 36215
  \bool_if:nTF ........... 66, 8469, 20924
  \bool_if:nTF ........... 66, 8469
  \bool_if:n ............... 65, 8430
  \bool_if:n ............... 65, 8430
  \bool_if:n ............... 65, 8430
  \bool_if:n ............... 65, 571, 8394, 8416, 8421,
  8473, 8481, 8551, 8577, 8583, 8587
  \bool_lazy_all:nTF ........ 67, 68, 5792, 8531
  \bool_lazy_all:n ........ 68, 8531
  \bool_lazy_and:nnTF ........ 67, 68, 8548, 8774,
  8990, 10311, 11494, 29044, 29565,
  29582, 29914, 30849, 30902, 30930,
  31771, 33085, 33969, 35877, 35923
  \bool_lazy_and:n .......... 67, 68, 8548, 30292, 35934, 35946
  \bool_lazy_any:nTF ........ 67, 68,
  8557, 11109, 13860, 14132, 14156,
  14178, 14348, 29527, 29579, 29663
  \bool_lazy_any:n ........ 67, 68, 8557, 30852
  \bool_lazy_or:nTF ........ 67, 68, 3805,
  8574, 8757, 8830, 15132, 19047,
  29169, 29195, 29258, 29435, 29680,
  29688, 29905, 30231, 30289, 30305,
  30310, 30345, 30406, 30444, 30526,
  30570, 30609, 30733, 30747, 30781,
  30788, 30866, 30911, 30944, 30975,
  31022, 31045, 31081, 31806, 31900,
  31985, 34377, 34682, 35931, 35943
  \bool_lazy_or:n ........... 67, 68, 8574,
  29917, 30612, 30750, 30933, 31774
  \bool_log:N ................ 66, 8446
  \bool_log:n ................ 66, 8442
  \bool_new:N ............... 65, 4474, 4904, 4684, 4685, 6687,
  6688, 6689, 8339, 8465, 8466, 8467,
  8468, 8857, 10457, 13871, 20556,
  20795, 20796, 20803, 20804, 20808,
  20811, 20924, 29977, 32958, 35876
  \bool_not:p:n ............. 68, 8583
  \bool_set:N ................ 223, 21218
  \bool_set:N ................ 65, 565, 8413
  \bool_set_eq:NN ............ 65, 4611, 4929, 8409, 19733, 19735
  \bool_set_false:N .......... 65, 160, 5483, 5688,
  6659, 6730, 6744, 6807, 6849, 8397,
  10563, 10705, 10713, 10721, 10731,
  10738, 20843, 21404, 21405, 21406,
  21456, 21457, 21461, 21497, 21505,
  21507, 21516, 21517, 21527, 21546,
  21556, 2161, 33357, 34054, 36125
  \bool_set_inverse:n ........ 223, 21226
  \bool_set_inverse:n ........ 303, 36124
  \bool_set_true:N ........... 65, 174, 5488, 5692, 6653,
  6847, 6928, 8397, 10691, 20838,
  21411, 21413, 21415, 21455, 21463,
  21465, 21498, 21499, 21503, 21518,
  21523, 21525, 21543, 21551, 21628,
  29978, 33375, 33397, 33428, 36125
  \bool_shown ............... 66, 8446
  \bool_show:n .............. 66, 8442
Index

\__bool_case_true:w 36132
\__bool_case_end:nw 8514
\__bool_)_0: 8514
\__bool_&_1: 8514
\__bool_&_0: 8502
\__bool_!:Nw 8531
\__bool_lazy_all:n 8531
\__bool_lazy_any:n 8557
\__bool_p:Nw 8512
\__bool_show:NN 8446
\__bool_use_i_delimit_by_q:- 8446
\_recursion_stop:nw 8427, 8545, 8571
\_bool_|_0: 8514
\_bool_|_1: 8514
\_bool_|_2: 8514
\_botmark 208
\_botmarks 514
\_box 209
\_box autosize_to_wd_and_ht:Nnn 279, 32864
\_box autosize_to_wd_and_ht_plus_-dp:Nn 279, 32864
\_box_clear:N 279, 32864
\_box_clear_new:N 271, 32208
\_box_clip:N 301, 35982
\_box_dp:N 301, 35982
\_box_gset_viewport:Nnnnn 327, 32290, 32297, 32302, 32306, 32609, 3278, 32853, 32872, 32878, 33194, 33195, 33301, 33306, 33334, 33348, 33519, 33797, 33818, 34121, 36002, 36009, 36014
\_box_gset_to_last:N 329, 32864
\_box_gset_ht:Nn 329, 32864
\_box_gset_to_wd_and_ht:Nnn 329, 32864
\_box_gset_dp:N 301, 35982
\_box_gset_to_wd_and_ht_-plus_dp:Nnn 329, 32864
\_box_gclear:N 327, 32265, 32271, 34002
\_box_gclear_new:N 327, 32268
\_box_gclip:N 301, 35982
\_box_gsize_to_ht:Nn 329, 32757
\_box_gsize_to_ht_plus_dp:Nn 3280, 32757
\_box_gsize_to_wd:Nn 3280, 32757
\_box_gsize_to_wd_and_ht:Nnn 3280, 32757
\_box_gsize_to_wd_and_ht_plus_-dp:Nnn 3280, 32757
\_box_gsize_to_wd_and_ht:-plus_dp:N 3280, 32708, 33633
\_box_grotate:N 3281, 32590, 33468
\_box_gscale:Nn 3281, 32835, 33675
\_box_gset_dp:N 3272, 32299
\_box_gset_eq:NN 3271, 32265, 32274, 33419, 35992, 36043
\_box_gset_eq_clear:N 36674
\_box_gset_eq_drop:NN 278, 32280, 36675
\_box_gset_ht:Nn 3272, 32299
\_box_gset_last:N 3273, 32353
\_box_gset_trim:NNnn 301, 35988
\_box_gsetviewport:Nnnnn 301, 35939
\_box_gset_wd:Nn 3272, 32299

\_box_autostr:N .............. 65, 8439
\_box_unstr:N ................ 65, 8439, 8443, 8445
\_box_until_do:Nn ............. 69, 8593
\_box_until_do:nn .............. 69, 8605
\_box_while_do:Nn .............. 69, 8593
\_box_use_i_delimit_by_q:- 8446
\_recursion_stop:nw 8427, 8545, 8571
\_box_or:nTF .................. 68, 8584
\_box_xor:n .................... 68, 8584
\c_false_bool 26, 64, 366, 401, 550, 564, 567-569, 1640, 1692, 1693, 1724, 1746, 1751, 1783, 1802, 2003, 2010, 2940, 3214, 5169, 5187, 5381, 5428, 5727, 5929, 5946, 5959, 6137, 6273, 6778, 7876, 7894, 7904, 7966, 7974, 8389, 8400, 8404, 8457, 8522, 8545, 8551, 8569, 8713, 19717, 19729, 21162, 21164, 21171, 21176, 35892, 36141, 36143, 36145, 36147
\g_tmpb_bool .................. 66, 8465
\l_tmpb_bool ................... 66, 8465
\g_tmpb_bool ................... 66, 8465
\l_tmpb_bool ................... 66, 8465
\c_true_bool ................... 66, 8465
\c_false_bool 26, 64, 366, 401, 550, 564, 567-569, 1640, 1692, 1693, 1724, 1746, 1751, 1783, 1802, 2003, 2010, 2940, 3214, 5169, 5187, 5381, 5428, 5727, 5929, 5946, 5959, 6137, 6273, 6778, 7876, 7894, 7904, 7966, 7974, 8389, 8400, 8404, 8457, 8522, 8545, 8551, 8569, 8713, 19717, 19729, 21162, 21164, 21171, 21176, 35892, 36141, 36143, 36145, 36147
\l_tapa_bool ................... 66, 8465
\l_tampb_bool .................. 66, 8465
\g_tapa_bool ................... 66, 8465
\g_tampb_bool .................. 66, 8465
bool internal commands:
\_bool_!:Nw 8502
\_bool_&_0: 8514
\_bool_&_1: 8514
\_bool_&_2: 8514
\_bool_!:Nw 8507
\_bool_&_0: 8514
\_bool_&_1: 8514
\_bool_&_2: 8514
\_bool_case:NnTF 36132
\_bool_case_end:nw 36132
\_bool_case_false:w 36132
\_bool_case_true:w 36132
\_bool_choose:NNN .. 8509, 8513, 8514
\_bool_get_next:NNN .............. 567, 568, 8489, 8492, 8504, 8510, 8525, 8526, 8527, 8528, 8529, 8530
\_bool_if:p .................... 8481
\_bool_if:p_aux:w .............. 567, 8481
\_bool_if_recursion_tail_stop_-do:nn 8429, 8543, 8569
Index

\c_code_cctab 262, 14089, 29161
\c_document_cctab 262, 1188, 29161
\c_initex_cctab 262, 28945, 29148
\c_other_cctab 262, 29148

cttab internal commands:
\g__cctab_allocate_int 28871, 29003, 29005, 29007
\_cctab_begin_aux 1189, 1190, 28991, 29015
\_cctab_chk_group_begin:n 1191, 29016, 29035
\_cctab_chk_group_end:n 1191, 29029, 29035
\_cctab_chk_if_valid:NTF 28942, 28963, 29012, 29018
\_cctab_chk_if_valid_aux:NTF 29010
\g__cctab_endlinechar_prop 1187, 28874, 28921, 28923, 28970
\g__cctab_group_seq 28870, 29037, 29043
\_cctab_gset:n 28913, 28947, 29019, 29198
\_cctab_gset_aux:n 28913
\_cctab_gstore:Nn 28875
\l__cctab_internal_a_tl 1189, 1190, 28872, 28970, 28971, 28996, 29006, 29014, 29017, 29018, 29019, 29026, 29029, 29030, 29031
\l__cctab_internal_b_tl 28872, 29043, 29047, 29054
\g__cctab_internal_cctab 28952
\_cctab_internal_cctab_name 28952, 28973, 28974, 28975, 28976
\_cctab_item:NN 28902, 28958, 29089
\_cctab_nesting_number:N 29017, 29030, 29059
\_cctab_nesting_number:w 29059
\_cctab_new:Nw 1186, 1189, 28875, 28954, 28974, 28995, 29004, 29165
\g__cctab_next_cctab 28991
\_cctab_select:NN 1189, 28962, 29020, 29031
\g__cctab_stack_seq 1185, 28868, 29018, 29026, 29077
\g__cctab_unused_seq 1185, 1190, 28868, 29014, 29028

cell 253
\char 213, 19250

char commands:
\l_char_active_seq 85, 186, 18088
\char_fold_case:N 30920
\char_foldcase:N 183, 18933, 36926, 36927
\char_generate:nn 117, 182, 428, 446, 447, 533, 673, 733, 1378, 3751, 3752, 3753, 3754, 3755, 3757, 3758, 3759, 4319, 4335, 4347, 4765, 5802, 6158, 12150, 12166, 13932, 14187, 14203, 18715, 18926, 18930, 18961, 18990, 19045, 19057, 19062, 19439, 19477, 30247, 30257, 30258, 30415, 30471, 30575, 30574, 30765, 30776, 30823, 30824, 30825, 30836, 30861, 30889, 30916, 30939, 30956, 30968, 30980, 30985, 31010, 31028, 31057, 31059, 31063, 31101, 31102, 31107, 31108, 31117, 31118, 31123, 31125, 31345, 31346, 31350, 31353, 31358, 31383, 31400, 31401, 31402, 31407, 31409, 31411, 31419, 31420, 31421, 31426, 31428, 31430, 31542, 31543, 31545, 31549, 31551, 31909, 31930, 31932, 31991, 32005, 32007, 34783
\char_gset_active_eq:NN 182, 18994
\char_gset_active_eq:n 182, 18994
\char_lower_case:N 36920
\char_lowercase:N 183, 18993, 36920, 36921
\char_mixed_case:N 36925
\char_mixed_case:Nn 36920
\char_set_active_eq:n 182, 18247, 4275, 4276, 18694
\char_set_catcode:nn 184, 149, 150, 151, 152, 153, 154, 155, 156, 157, 18594, 18601, 18603, 18605, 18607, 18609, 18611, 18613, 18615, 18617, 18619, 18621, 18623, 18625, 18627, 18629, 18631, 18633, 18635, 18637, 18639, 18641, 18643, 18645, 18647, 18649, 18651, 18653, 18655, 18657, 18659, 18661, 18663, 28984
\char_set_catcode:active:N 183, 3799, 7285, 9318, 18600, 18695, 18748, 19085, 31891, 36552
\char_set_catcode:active:n 184, 18632, 18799, 20692, 20693, 29172, 29179, 29197, 29208, 29407
\char_set_catcode:alignment:N 183, 7337, 18600, 19073
\char_set_catcode:alignment:n 184, 167, 18632, 18783, 29185
\char_set_catcode:comment:N 183, 18600
\char_set_catcode:comment:n 184, 18632, 29184
\__coffin_y_dim ............ 32962, 33544, 33548, 33825, 33829, 34105, 34108
\__coffin_x_prime_dim ............ 32963, 33544
\__coffin_y_prime_dim ............ 32963, 33544, 33549, 33826, 33831, 34106, 34110
color commands:
color.sc .......................... 292
\color_ensure_current: .... 288, 289, 33026, 33039, 33099, 33110, 34197
\color_export:mmN ........... 293, 34953
\color_export:mmNNN .......... 293, 34953
\color_fill:n ............... 291, 34806
\color_fill:nn ............... 292, 34806
\__color_fixed_model_t1 291, 34328, 34330, 34369, 34399, 34424, 34464, 34648, 34683, 34684, 34690, 34709, 34711, 34837, 34841, 34843, 34957
\color_group_begin: .... 288, 32401, 32403, 32410, 32417, 32422, 32430, 32436, 32450, 32456, 32463, 32468, 32481, 32483, 32487, 32492, 32497, 32502, 32509, 32512, 32514, 32521, 32526, 32534, 32540, 32555, 32561, 34195
\color_group_end: ............ 288, 32401, 32405, 32410, 32417, 32422, 32442, 32463, 32468, 32481, 32483, 32487, 32492, 32497, 32502, 32509, 32514, 32521, 32526, 32547, 34195
\color_log:n ............... 291, 33706
\color_math:nn ............ 292, 34716
\color_math:nn(n) .......... 3392
\color_math:nn ............ 292, 34716
\__color_math_active_t1 ............ 292, 34716
\color_model_new:mmN ........... 293, 35075
\color_profile_apply:nn ... 294, 35677
\color_select:n ............ 291, 33985, 33992, 34030, 34065, 34665
\color_select:n(n) .......... 292
\color_set:mmN ............ 291, 34665
\color_set:mmN ............ 292
\color_set:mmNN ........... 290, 34834
\color_set:mmNNN .......... 290, 34834
\color_set:eq:mmN .......... 291, 35706
\color_set_eq:mmN .......... 291, 35706
\color_stroke:n .......... 292, 34806
\color_stroke:nn .......... 292, 34806
color internal commands:
\g__color_alternative_model_prop .................. 35062, 35174, 35272
\g__color_alternative_values_prop .................. 35067, 35189, 35203, 35213, 35427, 35529
\__color_backend_devicen-init:mmN ............ 35442
\__color_backend_iccbased_devicen-init:mmN .... 35694, 35699, 35704
\__color_backend_iccbased_devicen-init:mmNNN ... 35674
\__color_backend_pickup:mmN .................. 34199, 34310
\__color_backend_reset: .... 34206, 34787
\__color_backend_separation-init:mmNN .......... 35190, 35204, 35214
\__color_backend_separation-init_CIELAB:mmNN ... 35242
\__color_check_model:mmN ........ 34324, 34637
\__color_check_model:mmNN ........ 34637
\g__color_colorants_prop
\g__color_alternative_model_prop
\g__color_alternative_values_prop
\g__color_check_model:mmN
\__color_check_model:mmNN
\__color_backend_devicen-init:mmN
\__color_backend_iccbased_devicen-init:mmN
\__color_backend_iccbased_devicen-init:mmNNN
\__color_backend_pickup:mmN
\__color_backend_reset:
\__color_backend_separation-init:mmNN
\__color_backend_separation-init_CIELAB:mmNN
\__color_check_model:mmN
\__color_check_model:mmNN
\g__color_colorants_prop
\cs_gset_protected_nopar:Npx 

\cs_if_eq:NNTF \cs_if_exist:NTF \cs_if_exist_use:N \cs_if_exist_use:NTF \cs_if_free:NTF \cs_if_free_p:N \cs_log:N \cs_meaning:N \cs_new:Nn \cs_new:Npn 

\cs_if_eq:Npx  \cs_if_exist:N  \cs_if_exist:N  \cs_if_exist:NTF \cs_if_exist_p:N  \cs_if_exist_use:N \cs_if_exist_use:NTF \cs_if_free:NTF \cs_if_free_p:N 

\cs_log:N 

\cs_meaning:N 

\cs_new:Nn 

\cs_new:Npn 

\index
Index

1414

31884, 31962, 32252, 34227, 34228, 34272, 34274, 34276, 34278, 34280, 34282, 34287, 34294, 34296, 34302, 34446, 34453, 34458, 34460, 34467, 34475, 34477, 34486, 34496, 34498, 34500, 34502, 34504, 34509, 34517, 34523, 34535, 34537, 34538, 34539, 34540, 34541, 34542, 34543, 34544, 34551, 34553, 34562, 34571, 34585, 34672, 34634, 34971, 35029, 35170, 35181, 35188, 35196, 35202, 35210, 35212, 35244, 35246, 35254, 35311, 35330, 35342, 35355, 35370, 35383, 35397, 35488, 35514, 35525, 35532, 35534, 35547, 35553, 35564, 35579, 35585, 35590, 35609, 35606, 35616, 35671, 35672, 35726, 35909, 35916, 35917, 35921, 35929, 35941, 35968, 35970, 35971, 36092, 36132, 36134, 36136, 36138, 36140, 36142, 36144, 36146, 36148, 36153, 36159, 36165, 36168, 36169, 36179, 36187, 36189, 36195, 36201, 36204, 36301, 36335, 36495, 36503, 36523, 36525, 36526, 36529, 36531, 36533, 36535, 36543

\cs_new:Npx \cs_new_eq:NN \cs_new_nopar:Nn \cs_new_nopar:Npx \cs_new_eq:Nx \cs_new_eq:Nx \cs_new_eq:Nx

15, 18, 39, 399, 409, 1917, 1941, 1946, 2926, 3098, 4739, 4741, 4743, 4745, 4747, 4749, 10479, 10491, 10983, 15149, 21902, 22649, 23473, 24057, 27211, 27219, 27225, 27837, 29659, 29847, 3031, 30423, 30101, 30518, 30375, 35515, 36105, 36107, 36109, 36116

\cs_new_eq:Nx \cs_new_nopar:Nx \cs_new_nopar:Np \cs_new_nopar:Np \cs_new_nopar:Np \cs_new_eq:Nx

17, 2047, 2112

\cs_new_eq:Nx \cs_new_nopar:Np \cs_new_nopar:Np \cs_new_nopar:Np

15, 369, 370, 1917, 1933, 1939

\cs_new_nopar:NN \cs_new_nopar:NN \cs_new_nopar:NN

15, 1917, 1933, 1940

\cs_new_protected:Nn \cs_new_protected:Nx \cs_new_protected:Nx


\cs_new_protected:Nn \cs_new_protected:Nx

15, 1933, 1940
\texttt{\texttt{cs}} \texttt{internal \ commands:}

\texttt{\texttt{cs\_to\_str:N}} \quad \ldots \quad 1377
\texttt{\texttt{cs\_to\_str:N}} \quad 6, 21, 107, 120, 365, 366, 398, 720, 1786, 1801, 2704, 2870, 4440, 5941, 10435, 13803, 13804, 13805, 13806, 13807, 13808, 13809, 13810, 13811, 13812, 13813, 13814, 16258, 16280, 22864, 29877, 31888, 36422, 36508, 36513, 36521
\texttt{\texttt{cs\_undefine:N}} \quad \ldots \quad 20, 831, 932, 1975, 9455, 9456, 9457, 10127, 11611, 11612, 12836, 13090, 28713, 28997, 29057

\texttt{\texttt{cs\_internal \ commands:}}

\texttt{\texttt{\_cs\_generate\_variant:nnNN}} \quad \ldots \quad 3204
\texttt{\texttt{\_cs\_generate\_variant:nnNn}} \quad \ldots \quad 2905, 2938
\texttt{\texttt{\_cs\_generate\_variant:nnNn}} \quad \ldots \quad 3204
\texttt{\texttt{\_cs\_generate\_variant:NNn}} \quad \ldots \quad 2945, 2947
\texttt{\texttt{\_cs\_generate\_variant:w}} \quad \ldots \quad 3204
\texttt{\texttt{\_cs\_generate\_variant:ww}} \quad \ldots \quad 2915
\texttt{\texttt{\_cs\_generate\_variant:wwNn}} \quad \ldots \quad 401–403, 2954, 3071
\texttt{\texttt{\_cs\_generate\_variant:wwNn}} \quad \ldots \quad 2915
\texttt{\texttt{\_cs\_generate\_variant\_F\_form:nn}} \quad \ldots \quad 3204
\texttt{\texttt{\_cs\_generate\_variant\_loop:nnNn}} \quad \ldots \quad 401, 402, 2955, 2967
\texttt{\texttt{\_cs\_generate\_variant\_loop\_base:N}} \quad \ldots \quad 2967
\texttt{\texttt{\_cs\_generate\_variant\_loop\_end:nnnWn}} \quad \ldots \quad 401–403, 2957, 2967
\texttt{\texttt{\_cs\_generate\_variant\_loop\_invalid:nnnWn}} \quad \ldots \quad 402, 2967
\texttt{\texttt{\_cs\_generate\_variant\_loop\_long:nnnWn}} \quad \ldots \quad 402, 2960, 2967
\texttt{\texttt{\_cs\_generate\_variant\_loop\_same:w}} \quad \ldots \quad 402, 2967
\texttt{\texttt{\_cs\_generate\_variant\_loop\_special:nnnWnNn}} \quad \ldots \quad 2967, 3066
\texttt{\texttt{\_cs\_generate\_variant\_p\_form:nnn}} \quad \ldots \quad 3204
\texttt{\texttt{\_cs\_generate\_variant\_same:N}} \quad \ldots \quad 402, 3012, 3060
\texttt{\texttt{\_cs\_generate\_variant\_T\_form:nnn}} \quad \ldots \quad 3204
\texttt{\texttt{\_cs\_generate\_variant\_TF\_form:nnn}} \quad \ldots \quad 3204
\texttt{\texttt{\_cs\_get\_function\_name:N}} \quad \ldots \quad 364
\texttt{\texttt{\_cs\_get\_function\_signature:N}} \quad \ldots \quad 364
\texttt{\texttt{\_cs\_parm\_from\_arg\_count:\_test:nnN}} \quad \ldots \quad 1987
\texttt{\texttt{\_cs\_split\_function\_auxi:w}} \quad \ldots \quad 1795
\texttt{\texttt{\_cs\_split\_function\_auxi:w}} \quad \ldots \quad 1795
2141, 2142, 2143, 2919, 2937, 3079, 3087, 3105, 3150, 3266, 3273, 3274, 3275, 3276, 3277, 3278, 3279, 3280, 3281, 3282, 3283, 3284, 3285, 3286, 3287, 3288, 3289, 3290, 3291, 3292, 3293, 3294, 3295, 3296, 3297, 3298, 3299, 3300, 3301, 3302, 3303, 3304, 3305, 3306, 3307, 3308, 3309, 3310, 3311, 3312, 3313, 3314, 3315, 3316
\_cs_to_str:N ........... 365, 1786
\_cs_to_str:w .......... 365, 1786
\_cs_use_i_delimit_by_s_stop:kw ........... 2896, 3217
\_cs_use_none_delimit_by_q-_recursion_stop:w ........... 2896, 2943, 2950, 3230
\_cs_use_none_delimit_by_s_-stop:w .......... 2896, 3221
csc .......... 254
cscd .......... 254
casename .......... 636, 4, 21, 25, 30, 34, 47, 70, 71, 73, 82, 107, 111, 130, 224
\_debug_off: .......... 345\_debug_off:n ........... 29, 1381, 1382, 1566\_debug_on: .......... 345\_debug_on:n ........... 29, 1381, 1566\_debug_resume: .... 29, 1288, 1576, 33017\_debug_suspend: .... 29, 1288, 1576, 33010\_deprecation internal commands:

Index

D
\_d .......... 29540, 32022\_date .......... 345\_day .......... 225, 1322, 9007\_dd .......... 257\_deadcycles .......... 226\_debug commands:
\_debug\_off: .......... 345\_debug\_off:n ........... 29, 1381, 1382, 1566\_debug\_on: .......... 345\_debug\_on:n ........... 29, 1381, 1566\_debug\_resume: .... 29, 1288, 1576, 33017\_debug\_suspend: .... 29, 1288, 1576, 33010\_debug internal commands:
Index

25214, 28393, 29429, 29452, 29586, 29779, 29780, 29911, 29932, 30163, 30164, 30475, 31967, 31971, 31972, 33206, 35112, 35455, 36182, 36297, 36301, 36304, 36305, 36338, 36428, 36429, 36437, 36442, 36447, 36452, 36466, 36484, 36497, 36504, 36634

\exp_stop_f: ...............
...

exp internal commands:
\__exp_arg_last_unbraced:nnnn
\__exp_arg_next:Nnnn
\__exp_arg_next:nnnn
\__exp_e:N
\__exp_e:nnn
\__exp_e:end:nnn
\__exp_e:protected:nnn
\__exp_e:put:nnn

393, 395, 396, 2639, 2732, 2744, 2831
393, 395, 396, 2639, 2732, 2744, 2831
Index

26685, 26686, 26718, 26721, 26730, 26732, 26734, 26748, 26762, 26786, 26870, 26878, 26909, 26917, 26923, 26934, 26937, 26940, 26949, 26958, 26960, 26966, 26973, 26976, 26985, 26993, 27014, 27047, 27048, 27075, 27077, 27095, 27096, 27115, 27126, 27135, 27138, 27149, 27152, 27155, 27173, 27183, 27194, 27196, 27205, 27252, 27267, 27282, 27297, 27312, 27327, 27330, 27332, 27349, 27394, 27701, 27737, 27738, 27748, 27789, 27790, 27814, 27841, 27842, 27845, 27847, 27848, 27853, 27865, 27884, 27889, 27897, 27900, 27932, 27942, 27943, 27953, 27975, 27990, 28008, 28016, 28019, 28047, 28055, 28071, 28143, 28161, 28197, 28215, 28248, 28271, 28277, 28350, 28351, 28482, 28483, 28492, 28499, 28504, 28514, 28524, 28549, 28552, 28565, 28597, 28605, 28606, 28634, 28656, 28657, 28658, 28661, 28666, 28686, 28687, 28739, 28740, 28781, 28789, 28842, 28848, 28861, 29303, 29304, 29312, 29326, 29330, 29331, 29347, 29415, 29419, 29430, 29451, 29455, 29471, 29483, 29515, 29516, 29517, 29518, 29519, 29520, 29521, 29702, 29708, 29961, 30634, 30637, 30638, 30641, 30642, 30665, 30666, 30667, 30668, 30669, 30670, 30691, 30692, 30693, 30694, 30695, 30712, 30713, 30714, 30715, 30729, 30730, 30731, 32342, 32344, 32350, 36097, 36296, 36332, 36337, 36338, 36456, 36463, 36468, 36494, 36500, 36504, 36519, 36540, 36548, 36584, 36585, 36586, 36592
\c_fifteen .......................... 36716

file commands:
\file_add_path:nN .................. 36736
\file_compare_timestamp:NNn 97
\file_compare_timestamp:NNnTF .......... 97, 11224
\file_compare_timestamp_p:NNn .......... 97, 11224
\g_file_curr_dir_str ........... 97, 10763, 11322, 11328, 11341
\g_file_curr_ext_str ........... 97, 10763, 11324, 11330, 11343
\g_file_curr_name_str ........... 93, 9099, 9226, 10763, 11323, 11329, 11342, 36749
\g_file_current_name_tl ........ 36748
\file_full_name:n .............. 94, 10930, 11029, 11089, 11097, 11103, 11119, 11228, 11229, 35675, 35680
\file_get:nn ....................... 94, 10889, 36849, 36851, 36853, 36855
\file_get:nnTF ............ 94, 10889, 10891
\file_get:nnTF .................... 94, 346, 11020, 36737
\file_get_full_name:n ........... 10112, 10896, 11020, 11022, 11034, 11035, 11278, 11284, 11289, 11301
\file_get_hex_dump:n ........ 95, 11165
\file_get_hex_dump:nnn ........ 95, 11165
\file_get_hex_dump:nnnTF .......... 95, 11209, 11211
\file_get_hex_dump:nTF ........ 95, 11165, 11166
\file_get_mdfive_hash:nn ........ 96, 11167, 11175
\file_get_mdfive_hash:nn:file-get_size:n ........ 11165
\file_get_mdfive_hash:nn:file-get_size:nTF .......... 11165
\file_get_mdfive_hash:nn:nTF .......... 96, 11168
\file_get_size:n ............. 96, 11170
\file_get_size:nTF .......... 96, 11170
\file_get_timestamp:n ........ 96, 11165
\file_get_timestamp:nTF ........ 96, 11165, 11172
\file_hex_dump:n .............. 96, 11165, 11172
\file_hex_dump:nn ............ 95, 11100
\file_hex_dump:nnTF .......... 95, 11100, 11218, 35657
\file_if_exist:nTF ............ 94, 97, 11276, 11587, 11589, 11593, 14086, 36739, 36741
\file_if_exist:input:n .... 97, 11282
\file_if_exist:input:nTF ........ 97, 11282, 36738, 36740
\file_input:n ................... 97, 11290, 14090, 36739, 36741
\file_input_stop: ............... 97, 11293
\file_list: ................. 36742
\file_log_list: ............ 97, 11402, 36743
\file_mdfive_hash:n .......... 96, 11082
\file_parse_full_name:n 95, 648, 11345
\file_parse_full_name:nnn .......... 95, 11082
\file_parse_full_name:nnnTF .......... 95, 11082
\file_path_include:n ............. 95, 648, 11345, 11392
\file_path_remove:n ............. 36744
\l_file_search_path_seq ........ 94, 10797, 10944, 11044, 36745, 36747
\file_show_list: ............... 97, 11402
\_file\_parse\_full\_name\_auxi\_n \ . 11352, 11355
\_file\_parse\_full\_name\_base\_n \ . 649, 11363, 11366
\_file\_parse\_full\_name\_tidy\_n \ . 649, 11373, 11374, 11376, 11380
\_file\_parse\_version\_w \ . 11468
\_file\_quark\_if\_nil\_nTF \ . 10801, 10870, 10884, 10976, 10985
\_file\_quark\_if\_nil\_p\_n \ . 10801
\_g\_file\_record\_seq \ . 647, 650, 10791, 11309, 11412, 11426, 11427
\_file\_size\_n \ . 10929, 10938, 10958, 10993
\_g\_file\_stack\_seq \ . 647, 10766, 11320, 11335
\_file\_str\_cmp\_nn \ . 11223, 11253
\_file\_timestamp\_n \ . 11224
\_file\_tmp\_w \ . 10768, 10772, 10776, 10782, 10788
\_l\_file\_tmp\_seq \ . 10798, 11406, 11409, 11412, 11413, 11415, 11423, 11428, 11499, 11501, 11524, 11528
\_filedump \ . 775
\_filemoddate \ . 776
\_filesize \ . 777
\_finalhyphenemerges \ . 271
\_firstmark \ . 272
\_firstmarks \ . 529
\_firstvaliddiagram \ . 820
\_c\_five \ . 36096
flag commands:
\flag\_clear\_n \ 169, 170, 5759, 7626, 7627, 14104, 14192, 14286, 14315, 14384, 14432, 14433, 14485, 14486, 14722, 14723, 14724, 14725, 14726, 14827, 14922, 14923, 14924, 14925, 15080, 15081, 15082, 17849, 17862
\flag\_clear\_new\_n \ 170, 746, 14667, 14668, 14669, 14670, 14850, 14851, 14852, 15030, 15031, 17861
\flag\_height\_n \ 170, 7636, 7638, 14010, 17870, 17886, 17900
\flag\_if\_exist\_n \ . 170
\flag\_if\_exist\_nTF \ . 170, 17862, 17873
\flag\_if\_exist\_p\_n \ . 170, 17873
\flag\_if\_raised\_n \ . 170
\flag\_if\_raised\_nTF \ . 170, 5767, 14003, 14008, 14010, 14094, 14700, 14705, 14712, 14884, 14889, 14994, 15045, 15052, 17878
\flag\_if\_raised\_p\_n \ . 170, 17878
\flag\_log\_n \ . 170, 17863
\flag\_new\_n \ 169, 170, 746, 831, 5749, 7486, 7487, 13872, 13873, 17844, 17862, 22883, 22884, 22885, 22886
\flag\_raise\_n \ . 170, 7666, 7674, 14150, 14200, 14300, 14333, 14404, 14417, 14555, 14460, 14541, 14743, 14744, 14767, 14786, 14781, 14782, 14801, 14802, 14808, 14809, 14839, 14989, 14990, 15099, 15100, 15104, 15105, 15119, 15120, 17897
\flag\_raise\_if\_clear\_n \ . 302, 5786, 5808, 22917, 22926, 22934, 22951, 22960, 22991, 36002
\flag\_show\_n \ . 170, 17863
flag internal commands:
\_flag\_clear\_nTF \ . 17849
\_flag\_height\_end\_n \ . 17880
\_flag\_height\_loop\_n \ . 17863
\_flag\_show\_Nn \ . 17863
\float\_ing\_penalty \ . 273
\floor \ . 253
\_fname\_n \ . 106, 8755, 8759, 8765, 8771, 8781, 29566, 29583, 29854, 33086, 33907, 33970, 33971, 35878
\font \ . 274
\font\_char\_dp \ . 530
\font\_char\_ht \ . 531
\font\_char\_ic \ . 532
\font\_char\_wd \ . 533
\font\_dimen \ . 924, 275
\font\_encoding \ . 31829
\font\_family \ . 31830
\font\_id \ . 821
\font\_name \ . 276
\font\_series \ . 31831
\font\_shape \ . 31832
\font\_size \ . 31835
\footnote\_size \ . 31871
\force\_j\_tok\_\_n \ . 1177
\format\_name \ . 822
\_c\_four \ . 36694
\_c\_fourteen \ . 36714
fp commands:
\_c\_e\_fp \ . 247, 249, 24764
\flag\_fp\_division\_by\_zero \ . 248, 22883
\flag\_fp\_invalid\_operation \ . 248, 22883
\flag\_fp\_overflow \ . 248, 22883
\flag\_fp\_underflow \ . 248, 22883
\fp\_abs\_n \ . 252, 257, 1168, 28372, 32755, 32857, 32859, 32861, 33685, 33687
\fp\_add\_Nn \ . 240, 1168, 24727
\fp\_compare\_nTN \ . 247, 24764, 24805, 24946, 24952, 24967, 24965,
\_fp\_parse\_caseless\_infinity:N 24103
\_fp\_parse\_caseless\_nan:N 24103
\_fp\_parse\_compare:NNNNNNN 24470
\_fp\_parse\_compare\_auxi:NNNNNNN 24470
\_fp\_parse\_compare\_auxii:NNNNN 24470
\_fp\_parse\_compare\_end:NNNNW 24470
\_fp\_parse\_continue:NW 1006, 1007, 1034, 24180, 24193, 24369, 24568, 25176, 25186, 25194
\_fp\_parse\_compare\_compare:NNNNNNN 24561, 24576
\_fp\_parse\_digits\_i:N 23362
\_fp\_parse\_digits\_ii:N 23362
\_fp\_parse\_digits\_iii:N 23362
\_fp\_parse\_digits\_iv:N 23362
\_fp\_parse\_digits\_v:N 23362
\_fp\_parse\_digits\_vi:N 23362, 23697, 23745
\_fp\_parse\_digits\_vii:N 1019, 23623, 23684, 23734
\_fp\_parse\_excl\_error: 24470
\_fp\_parse\_expand:w 1010, 1011, 23359, 23361, 23371, 23411, 23471, 23515, 23524, 23527, 23531, 23568, 23602, 23640, 23642, 23661, 23663, 23685, 23702, 23715, 23735, 23765, 23793, 23809, 23820, 23843, 23872, 23882, 23889, 23903, 23919, 23939, 23950, 24032, 24055, 24067, 24142, 24151, 24159, 24172, 24290, 24336, 24360, 24386, 24434, 24454, 24523, 24536, 24572
\_fp\_parse\_exponent\_N: 1024
23470, 23676, 23825, 23892, 23894
\_fp\_parse\_exponent\_NW: 23700, 23713, 23762, 23790, 23841, 23870, 23889
\_fp\_parse\_exponent\_aux:NN 23894
\_fp\_parse\_exponent\_body:N 23921, 23925
\_fp\_parse\_exponent\_digits:N 23929, 23941
\_fp\_parse\_exponent\_keep:N 23952
\_fp\_parse\_exponent\_keep:NTF 23932, 23952
\_fp\_parse\_exponent\_sign:N 23911, 23915
\_fp\_parse\_function:NNN 23056, 23058, 23060, 23063, 24135, 24776, 24778, 27233, 27235, 27237, 27239, 28400, 28402
\_fp\_parse\_function\_all_fp_o:nnw 23190, 24578, 25074
\_fp\_parse\_function\_one_two:nnw . . . . . . 1158, 24590, 27827, 27833, 28469
\_fp\_parse\_function\_one_two\_aux:nnw 24590
\_fp\_parse\_function\_one_two\_auxi:nnw 24590
\_fp\_parse\_function\_one_two\_error_o:w 24590
\_fp\_parse\_infix:NN 1012, 1016, 1032, 1037, 23410, 23580, 23619, 24095, 24110, 24132, 24248, 24251, 24334
\_fp\_parse\_infix\_f:N 24470
\_fp\_parse\_infix\_f&:NW 24470
\_fp\_parse\_infix\_f:(_N 24470
\_fp\_parse\_infix\_f):N 24470
\_fp\_parse\_infix\_f*:N 24470
\_fp\_parse\_infix\_f+:N 24470
\_fp\_parse\_infix\_f:-N 24470
\_fp\_parse\_infix\_f:<N 24470
\_fp\_parse\_infix\_f:=N 24470
\_fp\_parse\_infix\_f>:N 24470
\_fp\_parse\_infix\_f?:N 24444
\_fp\_parse\_infix\_f:(operation):N 1010
\_fp\_parse\_infix\_f:-N 24476
\_fp\_parse\_infix\_after\_operand:NNW 1016, 23463, 23541, 24039, 24246
\_fp\_parse\_infix\_after\_paren:NN 24064, 24090, 24203
\_fp\_parse\_infix\_and:NN 24376, 24443
\_fp\_parse\_infix\_check:NNN 24260, 24279, 24311
\_fp\_parse\_infix\_comma:w 1034, 1038, 24341
\_fp\_parse\_infix\_end:NN 1034, 1038, 24160, 24165, 24173, 24322
\_fp\_parse\_infix\_juxt:N 1037, 24259, 24267, 24376
\_fp\_parse\_infix\_mark:NNN 24256, 24298, 24321
\_fp\_parse\_infix\_mul:N 1037, 1046, 24284, 24301, 24309, 24376, 24411, 24420
\_fp\_parse\_infix\_or:NN 24376, 24442
\_fp\_parse\_infix\_l:W 24427
\_fp\_parse\_large:N 1018, 23647, 23730
\int_{0}^{\infty} x^2 \, dx = 2\pi\int_{0}^{\infty} x^2 \, dx = 2\pi \left[ \frac{x^3}{3} \right]_{0}^{\infty} = 2\pi \frac{\infty^3}{3} = \infty

\int_{-\infty}^{\infty} e^{-x^2} \, dx = \sqrt{\pi}

\int_{0}^{1} e^x \, dx = e - 1

\int_{1}^{\infty} \frac{1}{x^2} \, dx = \lim_{b \to \infty} \left[ -\frac{1}{x} \right]_{1}^{b} = \lim_{b \to \infty} \left( -\frac{1}{b} + 1 \right) = 1
Index

22307, 22336, 22467, 28800, 35423
\_int_zero_new:N ............ 157, 17165
\c_max_char_int 166, 4758, 17825, 18733
\c_max_int ......................
  166, 236, 514, 966, 1172, 17824,
  28623, 32365, 32371, 34152, 34155
\c_max_register_int ............
  166, 413, 1482, 3339, 3364, 3401, 9916, 9918, 16544, 17020
\c_one_int .. 166, 3924, 4227, 6512,
  6523, 17192, 17194, 17196, 17198,
  17823, 19064, 22365, 22384, 25847,
  25851, 25855, 25909, 26500, 26643,
  26784, 27090, 27930, 28014, 28481,
  28485, 28492, 28677, 36537, 36545
\g_tapn_int ............. 166, 17835
\l_tapb_int .......... 4, 52, 166, 17835
\g_tapn_int ............. 166, 17835
\l_tapn_int ............. 4, 166, 17835
\c_zero_int ..................
  166, 354, 365, 680, 1481, 1792,
  1794, 3947, 3971, 4022, 4123, 4224,
  4225, 4226, 4447, 5043, 5543, 5939,
  6412, 6483, 6499, 6539, 6590, 7020,
  7159, 7230, 7239, 7291, 7308, 7311,
  7663, 7671, 8694, 8888, 9022, 9072,
  10909, 11265, 12117, 12960, 13009,
  13220, 13225, 13546, 13581, 14299,
  15098, 17116, 17117, 17340, 17316,
  17162, 17213, 17221, 17398, 17401,
  17823, 18732, 19061, 19062, 19063,
  20180, 20358, 22226, 22352, 22850,
  23076, 23080, 23082, 23086, 23090,
  23103, 23116, 23123, 23136, 23148,
  23160, 23169, 23172, 23183, 23289,
  23294, 24857, 24889, 25873, 25908,
  25910, 25911, 25912, 26677, 26744,
  26677, 26985, 27008, 27040, 27886,
  28246, 28464, 28493, 28593, 28657
int internal commands:
  \_int_abs:n ............ 17049
  \_int_case:nTF .......... 17287
  \_int_case:nw .......... 17287
  \_int_case_end:nw ...... 17287
  \_int_compare:nnN ...... 815, 17226
  \_int_compare:NNw ... 814, 815, 17226
  \_int_compare:Nw ...... 814, 815, 17226
  \_int_compare:w ....... 814, 17226
  \_int_compare!=:NNw .. 17226
  \_int_compare<:NNw .... 17226
  \_int_compare<=:NNw ... 17226
  \_int_compare=:NNw ... 17226
  \_int_compare>=:NNw .. 17226
  \_int_compare>:NNw ... 17226
  \_int_compare_end=:NNw .. 815, 17226
  \_int_compare_error:...... 813, 814, 17211, 17229, 17231
  \_int_compare_error:NNw ........ 813–815, 17211, 17251
  \_int_const:n .......... 17126
  \_int_constdef:N ....... 17126
  \_int_div_truncate:NNNw .... 17081
  \_int_eval:w ............. 344, 808, 809, 814, 17020, 17033,
  17034, 17038, 17052, 17060, 17061,
  17068, 17069, 17083, 17085, 17086,
  17103, 17106, 17107, 17108, 17115,
  17146, 17180, 17182, 17184, 17186,
  17204, 17206, 17229, 17263, 17281,
  17319, 17327, 17392, 17393, 17394,
  17420, 17603, 17630, 17636, 17663
  \_int_eval_end: .......... 17020, 17033,
  17038, 17052, 17087, 17103, 17109,
  17118, 17146, 17180, 17182, 17184,
  17186, 17204, 17206, 17281, 17319,
  17297, 17603, 17630, 17636, 17663
  \_int_from_alph:n .......... 826, 17720
  \_int_from_alph:nw ...... 826, 17720
  \_int_from_base:N ........ 826, 17720
  \_int_from_base:NNw ...... 826, 17720
  \_int_from_roman:NN ...... 17729
  \c__int_from_roman_C_int .... 17765
  \c__int_from_roman_c_int .... 17765
  \c__int_from_roman_D_int .... 17765
  \c__int_from_roman_d_int .... 17765
  \_int_from_roman_error:w .. 17779
  \c__int_from_roman_I_int .... 17765
  \c__int_from_roman_i_int .... 17765
  \_int_from_roman_M_int .... 17765
  \c__int_from_roman_m_int .... 17765
  \_int_from_roman_V_int .... 17765
  \c__int_from_roman_v_int .... 17765
  \c__int_from_roman_F_int .... 17765
  \c__int_from_roman_f_int .... 17765
  \c__int_from_roman_R_int .... 17765
  \c__int_from_roman_r_int .... 17765
  \_int_if_recursion_tail_stop:N
  ................................ 17030, 17792
  \_int_if_recursion_tail_stop-
  do:Nn ........ 17030, 17731, 17748, 17795
  \l_int_internal_a_int .......... 17839
  \l_int_internal_b_int .......... 17839
  \c__int_max_constdef_int .... 17126
  \_int_maxim:nw ........ 17049
  \_int_mod:w ........ 17081
  \_int_pass_signs:wn ........ 825, 17710, 17724, 17741
ior internal commands:
\ior_if_eof_p:N ......................... 87, 88, 302, 10222, 10291, 36767
\ior_if_eof:NTF ................. 88, 10222, 10223
\ior_if_eof:N ........... 302, 10243
\ior_if_eof_term:N ........... 88, 10262, 29343, 29373
\ior_if_eof:NN ........ 89, 10288
\ior_if_EOF .................. 93, 10104
\ior_if_EOF .................. 93, 10104
\ior_if_eof:NN .................. 10109, 10112, 10114
\ior_if_get:N .............. 10200, 10244, 10263
\ior_if_get:N .............. 10243
\ior_if_get:NN ............ 10180
\ior_if_map_inline:N ...... 10262
\ior_if_map_inline:N ...... 10262
\ior_if_map_inline_loop ...... 10262
\ior_if_map_variable:N ...... 10288
\ior_if_map_variable_loop:N ...... 10288
\ior_if_new:N .......... 614, 10120, 10137
\ior_if_new_aux:N .......... 10124, 10128
\ior_if_open_stream:N .......... 10131
\ior_if_shell_open:N .......... 36905
\ior_if_show:NN ................. 10164
\ior_if_str_get:NN .......... 10222, 10246, 10265
\ior_if_stream_tl ............ 10086, 10134, 10138, 10145
\ior_if_streams_prop .......... 615, 10088, 10146, 10157, 10172, 10187
\ior_if_streams_seq ............. 10086, 10134, 10158, 10159
\ior_if_term:iorn .......... 10085, 10102, 10154, 10160, 10196, 10253
\ior_if_term:noprompt:iorn .......... 10242, 10252

ior commands:
\ior_if_eof:N ................. 86, 302, 10133, 10152, 11055, 11068, 11502, 29316, 29349, 29386, 29392
\ior_get:NN .. 87, 88, 302, 10209, 10289
\ior_get:NNTF ........ 87, 10209, 10210
\ior_get_str:N ........ 36766
\ior_get_term:N ........ 302, 10243
\ior_if_eof:N .......... 618
\ior_if_eof:NTF ......... 90, 10193, 10215, 10235, 10275, 10294, 11065, 11079
\ior_if_eof_p:n ........ 90, 10193
\ior_list_streams: ..... 36768
\ior_log:N ........ 86, 10164
\ior_log_list: ....... 87, 10180, 36771
\ior_log_streams: .... 36770
\ior_map_break: .... 89, 10258, 10276, 10283, 10295, 10301, 29311, 29382
\ior_map_break:n .... 89, 10258
\ior_map-inline:N ...... 88, 10262, 11500
\ior_map_variable:Nn ...... 88, 10288, 29308
\ior_new:N ................... 86, 10102, 10104, 10105, 11081, 29257
\ior_open:NN ............ 86, 10106, 29272, 29317, 29350, 29391
\ior_open:NnF ........... 86, 10107, 10110
\ior_shell_open:NN .... 302, 11489, 36395
\ior_show:N .......... 86, 10164
\ior_show_list: ...... 87, 10180, 36769

ior external commands:
\l_ior_file_name_tl ................. 10086, 10109, 10112, 10114
\l_ior_get:NN .............. 10200, 10244, 10263
\l_ior_get:N .............. 10243
\l_ior_internal_tl ............. 10084, 10172, 10175, 10281, 10285
\l_ior_list:N ............ 10180
\l_ior_map_inline:N ...... 10262
\l_ior_map_inline:N ...... 10262
\l_ior_map_inline_loop ...... 10262
\l_ior_map_variable:N ...... 10288
\l_ior_map_variable_loop:N ...... 10288
\l_ior_new:N .......... 614, 10120, 10137
\l_ior_new_aux:N .......... 10124, 10128
\l_ior_open_stream:N .......... 10131
\l_ior_shell_open:N .......... 36905
\l_ior_show:NN ................. 10164
\l_ior_str_get:NN .......... 10222, 10246, 10265
\l_ior_stream_tl ............ 10086, 10134, 10138, 10145
\g_ior_streams_prop .......... 615, 10088, 10146, 10157, 10172, 10187
\g_ior Streams_seq ............. 10086, 10134, 10158, 10159
\c_ior_term:iorn .......... 10085, 10102, 10154, 10160, 10196, 10253
\c_ior_term:noprompt:iorn .......... 10242, 10252

iow commands:
\iow_allow_break: .................. 92, 302, 10474, 10516, 10521
\iow_allow_break:n ................. 625
\iow_char:N .... 86, 91, 3697, 3700, 3701, 3725, 3726, 3733, 3734, 4732, 4733, 4740, 4742, 4744, 4746, 4748, 4750, 5393, 5394, 6110, 6117, 6118, 6119, 6243, 8061, 8064, 8065, 8070
\iow_char:V .... 86, 91, 3697, 3700, 3701, 3725, 3726, 3733, 3734, 4732, 4733, 4740, 4742, 4744, 4746, 4748, 4750, 5393, 5394, 6110, 6117, 6118, 6119, 6243, 8061, 8064, 8065, 8070
\iow_char:W .... 86, 91, 3697, 3700, 3701, 3725, 3726, 3733, 3734, 4732, 4733, 4740, 4742, 4744, 4746, 4748, 4750, 5393, 5394, 6110, 6117, 6118, 6119, 6243, 8061, 8064, 8065, 8070

Index
\_iouw\_wrap\_do:  
\_iouw\_wrap\_end:  
\_iouw\_wrap\_end\_chunk:w  

\_iouw\_wrap\_fix\_newline:w  
c\_iouw\_wrap\_indent:n  
c\_iouw\_wrap\_indent\_marker\_tl  
c\_iouw\_wrap\_line\_nw  
c\_iouw\_wrap\_line\_aux:Nw  
c\_iouw\_wrap\_line\_end:N

c\_iouw\_wrap\_line\_loop:w  
c\_iouw\_wrap\_line\_seven:mmmmm

c\_iouw\_wrap\_line\_store\_do:n  
c\_iouw\_wrap\_next\_line:w  
c\_iouw\_wrap\_newline:n  
c\_iouw\_wrap\_newline\_marker\_tl  
c\_iouw\_wrap\_next\_nw  
c\_iouw\_wrap\_text\_n  
c\_iouw\_wrap\_trim:N  
c\_iouw\_wrap\_trim:w  
c\_iouw\_wrap\_trim\_aux:w  
c\_iouw\_wrap\_unindent:n  
c\_iouw\_wrap\_unindent\_marker\_tl  

\itshape  
\jshape  
\jcharwidowpenalty  
jfam  
jfont  
jis  
jobname  

\kern

\kanjiskip  
kansuji  
kansuujchar  
kcatecode  
kcatcode  
kchar  
khatchcode  
kchardef

kernel internal commands:

\Kernel\_backend\_align\_begin:  
\Kernel\_backend\_align\_end:  
\g\_kernel\_backend\_header\_bool  
\Kernel\_backend\_literal:n  
\Kernel\_backend\_literal\_pdf:n  
\Kernel\_backend\_literal\_postscript:n  
\Kernel\_backend\_literal\_svg:n  
\Kernel\_backend\_matrix:n  
\Kernel\_backend\_postscript:n  
\Kernel\_backend\_scope\_begin:  
\Kernel\_backend\_scope\_end:  
\Kernel\_chk\_cs\_exist:N

\Kernel\_chk\_defined:NTF  
\Kernel\_chk\_if\_free\_cs:N  
\Kernel\_chk\_parm\_from\_arg:-count:n

\Kernel\_dependency\_version\_check:n  
\Kernel\_dependency\_version\_check:nn

\Kernel\_deprecation\_code:nn

\Kernel\_deprecation\_error:nn  
\Kernel\_exp\_not:w  
\Kernel\_if\_free\_cs:n  
\Kernel\_item\_of\_the\_list:n  
\Kernel\_is\_odd:nn

\Index
Index

1465

308, 309, 310, 311, 312, 313, 314,
315, 316, 317, 318, 319, 320, 321,
322, 323, 324, 325, 326, 327, 328,
329, 330, 331, 332, 333, 334, 335,
336, 337, 338, 339, 340, 341, 342,
343, 344, 345, 346, 347, 348, 349,
350, 351, 352, 353, 354, 355, 356,
357, 358, 359, 360, 361, 362, 363,
364, 365, 366, 367, 368, 369, 370,
371, 372, 373, 374, 375, 376, 377,
378, 379, 380, 381, 382, 383, 384,
385, 386, 387, 388, 389, 390, 391,
392, 393, 394, 395, 396, 397, 398,
399, 400, 401, 402, 403, 404, 405,
406, 407, 408, 409, 410, 411, 412,
413, 414, 415, 416, 417, 418, 419,
420, 421, 422, 423, 424, 425, 426,
427, 428, 429, 430, 431, 432, 433,
434, 435, 436, 437, 438, 439, 440,
441, 442, 443, 444, 445, 446, 447,
448, 449, 450, 451, 452, 453, 454,
455, 456, 457, 458, 459, 460, 461,
462, 463, 464, 465, 466, 467, 468,
469, 470, 471, 472, 473, 474, 475,
476, 477, 478, 479, 480, 481, 482,
483, 484, 485, 486, 487, 488, 489,
490, 491, 492, 493, 494, 495, 496,
497, 498, 499, 500, 501, 502, 503,
504, 505, 506, 507, 508, 509, 510,
511, 512, 513, 514, 515, 516, 517,
518, 519, 520, 521, 522, 523, 524,
525, 526, 527, 528, 529, 530, 531,
532, 533, 534, 535, 536, 537, 538,
539, 540, 541, 542, 543, 544, 545,
546, 547, 548, 549, 550, 551, 552,
553, 554, 555, 556, 557, 558, 559,
560, 561, 562, 563, 564, 565, 566,
567, 568, 569, 570, 571, 572, 573,
574, 575, 576, 577, 578, 579, 580,
581, 582, 583, 584, 585, 586, 587,
588, 589, 590, 591, 592, 593, 594,
595, 596, 597, 598, 599, 600, 601,
602, 604, 605, 606, 607, 608, 609,
610, 612, 613, 614, 615, 616, 617,
618, 619, 620, 621, 622, 623, 624,
625, 626, 627, 628, 629, 630, 631,
632, 633, 634, 635, 636, 637, 638,
639, 640, 641, 642, 643, 644, 645,
646, 647, 648, 649, 650, 651, 652,
653, 654, 655, 656, 657, 658, 659,
660, 661, 662, 663, 664, 665, 666,
667, 668, 669, 670, 671, 672, 673,
674, 675, 676, 677, 678, 679, 680,
681, 682, 683, 684, 685, 686, 687,
Index

1101, 1102, 1103, 1104, 1105, 1106, 1107, 1108, 1109, 1110, 1111, 1112, 1113, 1114, 1115, 1116, 1117, 1118, 1119, 1120, 1121, 1122, 1123, 1124, 1125, 1126, 1127, 1128, 1129, 1130, 1131, 1132, 1133, 1134, 1135, 1136, 1137, 1138, 1139, 1140, 1141, 1142, 1143, 1144, 1145, 1146, 1147, 1148, 1149, 1150, 1151, 1152, 1153, 1154, 1155, 1156, 1158, 1160, 1161, 1162, 1163, 1164, 1166, 1167, 1168, 1169, 1170, 1171, 1172, 1173, 1174, 1175, 1176, 1177, 1178, 1179, 1180, 1181, 1182, 1183, 1184, 1185, 1186, 1187, 1188, 1189, 1190, 1191, 1192, 1193

\_kernel\_quark\_new conditional:Nn

\_kernel\_register\_log:Nn

\_kernel\_register\_show:Nn

\_kernel\_register\_show\_aux:Nn

\_kernel\_register\_str\_to\_other:n

\_kernel\_register\_str\_to\_other fast:n

\_kernel\_str\_to\_other fast:w

\_kernel\_sys\_configuration\_load:n

\_kernel\_tl\_set:Nn

\_kernel\_tl\_str:w

\_kernel\_tl\_to\_str:w

\_kernel\_tl\_set:Nn

\_kernel\_tl\_get:Nn

\_kernel\_quark new conditional:N

\_kernel\_quark new test:N

\_kernel\_register\_log:N

\_kernel\_register\_show:N

\_kernel\_register\_show\_aux:N

\_kernel\_str\_to\_other:n

\_kernel\_str\_to\_other fast:n

\_kernel\_str\_to\_other fast:w

\_kernel\_sys\_configuration\_load:n

\_keys\_choice\_int

\_keys\_choice\_tl

\_keys\_define:nn

\_keys\_if\_choice\_exist:nnn TF

\_keys\_if\_choice\_exist:nnnnn TF

\_keys\_path\_str

\_keys\_register\_show:N

\_keys\_if\_exist:nnn

\_keys\_if\_exist:nnn

\_keys\_if\_exist:nnnnn

\_keys\_if\_exist:nnnnn

\_keys\_if\_exist:nnnnnnn

\_keys\_if\_exist:nnnnnnn

\_keys\_if\_exist:nnnnnnnnn

\_keys\_if\_exist:nnnnnnnnn

\_keys\_path\_str

\_keys\_path\_str

keys commands:

\_keys\_choice\_int

\_keys\_choice\_tl

\_keys\_define:nn

\_keys\_if\_choice\_exist:nnn TF

\_keys\_if\_choice\_exist:nnnnn TF

\_keys\_if\_choice\_exist:nnnnnnn TF

\_keys\_if\_exist:nnn TF

\_keys\_if\_exist:nnnnn TF

\_keys\_if\_exist:nnnnnnn TF

\_keys\_path\_str
Index

21122, 21126, 21142, 21151, 21193, 21204, 21245, 21561, 21569, 21607, 21610, 21649, 21653, 21658, 21667, 21681, 21683, 21684, 21688, 21696, 21719, 21749, 21772, 21784, 21793 \l_keys_path_tl \ 20797, 20891, 20952 \keys_precompile:nnN \ 232, 21541 \keys_set:nnN \ 232, 225, 230–232, 21076, 21083, 21140, 21545 \keys_set_filter:nnN \ 232, 21470 \keys_set_filter:nnNN \ 232, 21470 \keys_set_groups:nn \ 232, 22174 \keys_set_known:nnN \ 231, 942, 21429 \keys_set_known:nnNN \ 231, 21429 \keys_show:nn \ 233, 21876 \l_keys_usage_load_prop \ 230, 20813, 21161, 21168, 21175 \l_keys_usage_preamble_prop \ 230, 20813, 21163, 21170, 21177 \l_value_tl \ 230, 20807, 21142, 21652, 21656, 21662, 21673, 21684, 21703, 21716, 21741, 21751, 21779 keys internal commands:
\_keys_bool_set:NNn \ 20916, 21219, 21221, 21223, 21225 \_keys_bool_set:nn \ 20916 \_keys_bool_set:inverse:NN \ 20916, 21227, 21229, 21231, 21233 \_keys_check_forbidden: \ 21109 \_keys_check_groups: \ 21611, 21619 \_keys_check_required: \ 21109 \c_keys_check_root_str \ 20782, 21116, 21222, 21226, 21683, 21702 \_keys_choice_find:n \ 20939, 21790 \_keys_choice_find:nn \ 21790 \_keys_choice_make: \ 20925, 20938, 20970, 21055, 21235 \_keys_choice_make:N \ 20938 \_keys_choice_make:aux:N \ 20938 \_keys_choice_make:nn \ 20969 \_keys_choice_make:NN \ 20969 \_keys_code_root_str \ \_keys_cmd_set:nn \ 20926, 20928, 20980, 20990, 21056, 21058, 21060, 21090, 21204, 21245 \_keys_cmd_set_direct:nn \ 20930, 20962, 20963, 20990, 21074, 21082 \c_keys_code_root_str \ 948, 20782, 20994, 20997, 21045, 21681, 21699, 21715, 21729, 21801, 21863, 21872, 21893 \_keys_default_set:NNpn \ \_keys_default_inherit: \ 20995, 21255, 21257, 21259, 21261, 21263, 21265, 21267, 21269 \c_keys_default_root_str \ \_keys_define: \ 20782, 21010, 21015, 21649, 21653, 21670, 21674 \_keys_default_set:n \ 20935, 21065, 21271, 21273, 21275, 21277 \_keys_define:n \ 20832, 20836 \_keys_define:nnn \ 20832, 20836 \_keys_define_aux:nn \ 20836 \_keys_define_code:n \ 20850, 20899 \_keys_define_code:w \ 20899 \_keys_execute: \ 21575, 21615, 21637, 21641, 21679 \_keys_execute:nn \ \_keys_execute:unknown \ 947, 21679 \l_keys_filtered_bool \ 20803, 21405, 21412, 21413, 21456, 21462, 21463, 21498, 21504, 21505, 21517, 21524, 21525, 21614, 21635, 21640 \_keys_find_key_module:nn \ 20799 \_keys_find_key_module:auxi:Nw \ 21429 \_keys_find_key_module:auxii:Nw \ 21429 \_keys_find_key_module:auxiii:Nw \ \_keys_find_key_module:auxiv:Nw \ \_keys_groups_clist \ 20791 \_keys_groups_set:n \ 21020 \c_keys_groups_root_str \ \_keys_initialise:n \ \_keys_inherit:n \ 21033 \_keys_inherit:NNn \ 21044, 21549 \_keys_inherit:auxi:Nw \ 21549 \_keys_inherit:auxii:Nw \ 21549 \_keys_inherit:auxiii:Nw \ \_keys_inherit:auxiv:Nw \ \_keys_cs_set:NNpNpn \ \_keys_define_code:w \ 21046, 21679, 21802, 21803 \_keys_execute:inherit: \ 21042, 21679 \_keys_execute:unknown: \ 947, 21679 \l_keys_filtered_bool \ 20803, 21405, 21412, 21413, 21456, 21462, 21463, 21498, 21504, 21505, 21517, 21524, 21525, 21614, 21635, 21640 \_keys_find_key_module:nn \ 20799 \_keys_find_key_module:auxi:Nw \ 21549 \_keys_find_key_module:auxii:Nw \ 21549 \_keys_find_key_module:auxiii:Nw \ \_keys_find_key_module:auxiv:Nw \ \_keys_groups_set:n \ 21020, 21549 \l_keys_groups_set:n \ 20791, 21022, 21023, 21030, 21609, 21624 \c_keys_groups_root_str \ \_keys_definitions \ 20782, 21025, 21029, 21607, 21610 \_keys_groups_set:n \ 21020, 21295 \_keys_inherit:n \ \_keys_inherit:root_str \ \_keys_initialise:n \ \_keys_legacy_if_inverse:nnn \ \_keys_legacy_if_inverse:nnn \ 20104, 21042, 21679, 21802, 21803 \_keys_execute:inherit: \ 21042, 21679 \_keys_execute:unknown: \ 947, 21679 \c_keys_groups_root_str \ \_keys_definitions \ 20782, 21025, 21029, 21607, 21610 \_keys_groups_set:n \ 21020, 21295 \_keys_inherit:n \ \_keys_inherit:root_str \ \_keys_initialise:n \
Index

\_\_keys Legacy_if_set:nnnn
................. 21050, 21052, 21053
\_\_keys Legacy_if_set_inverse:nn
................. 21051, 21319, 21321
\_\_keys Make:n ... 21072, 21323
\_\_keys Make:nn .. 21072, 21325
\_\_keys Module_str
.... 20794, 20828, 20831, 20833, 20875, 21077, 21190, 21197, 21422, 21425, 21427, 21552, 21557, 21567, 21570, 21576, 21715, 21716, 21719
\_\_keys Multichoice_find:n
................. 20941, 21790
\_\_keys Multichoice_make:
................. 20938, 20972, 21327
\_\_keys Multichoice_make:nn
... 20969, 21329, 21331, 21333, 21335
\_\_keys No_Value_Boolean
................. 20795, 20838, 20843, 20901, 21139, 21485, 21551, 21556, 21647, 21740, 21750, 21778
\_\_keys Only_Known_Bool
.... 20796, 21404, 21410, 21411, 21455, 21460, 21461, 21497, 21502, 21503, 21516, 21522, 21523, 21711
\_\_keys Parent:n
................. 20945, 20948, 20952, 21041, 21658, 21667, 21688, 21696, 21807
\_\_keys Parent Auxi:W
................. 21807
\_\_keys Parent Auxii:w
................. 21807
\_\_keys Parent Auxiii:n
................. 21807
\_\_keys Parent Auxiv:w
................. 21807
\_\_keys Precompile:n 20820, 20991, 20999, 20906, 21098, 21914, 21915
\_\_keys Precompile_Boolean
... 20811, 20822, 21543, 21546
\_\_keys Precompile_TL
... 20811, 20823, 21544, 21547
\_\_keys Prop:Put:NN
.... 21085, 21345, 21347, 21349, 21351
\_\_keys Property_find:n 20848, 20850
\_\_keys Property_find_auxi:w 20859
\_\_keys Property_find_auxii:w 20859
\_\_keys Property_find_auxiii:w 20859
\_\_keys Property_find_auxiv:w 20859
................. 20864, 20872, 20893, 20894
\_\_keys Property_str 20802, 20849, 20852, 20855, 20861, 20862, 20888, 20896, 20904, 20905, 20908, 20911
\_\_keys Props_root_str
20788, 20849, 20905, 20911, 21218, 21220, 21222, 21224, 21226, 21228, 21230, 21232, 21234, 21236, 21238, 21240, 21242, 21244, 21246, 21248, 21250, 21252, 21254, 21256, 21258, 21260, 21262, 21264, 21266, 21268, 21270, 21272, 21274, 21276, 21278, 21280, 21282, 21284, 21286, 21288, 21290, 21292, 21294, 21296, 21298, 21300, 21302, 21304, 21306, 21308, 21310, 21312, 21314, 21316, 21318, 21320, 21322, 21324, 21326, 21328, 21330, 21332, 21334, 21336, 21338, 21340, 21342, 21344, 21346, 21348, 21350, 21352, 21354, 21356, 21358, 21360, 21362, 21364, 21366, 21368, 21370, 21372, 21374, 21376, 21378, 21380, 21382, 21384, 21386, 21388, 21390, 21392, 21394, 21396, 21398
\_\_keys Quark_if_No_Value:NFF
................. 20810, 21735
\_\_keys Quark_if_No_Value:p:N 20819
\_\_keys Relative_tl
.... 20800, 21407, 21416, 21417, 21458, 21466, 21467, 21500, 21508, 21509, 21519, 21528, 21529, 21735, 21745, 21759, 21760, 21764, 21765, 21773, 21785
\_\_keys Selective_Boolean
.... 20803, 21406, 21414, 21415, 21457, 21464, 21465, 21499, 21506, 21507, 21518, 21526, 21527, 21573
\_\_keys Selective_Seq
.... 20805, 21534, 21537, 21539, 21622
\_\_keys Set:nn 21400, 21459, 21538
\_\_keys Set:nnn
.... 21400
\_\_keys Set:Filter:nnnn
... 21470
\_\_keys Set:Filter:nnnnN
... 21470
\_\_keys Set:Keyval:n
... 21426, 21549
\_\_keys Set:Keyval:nn
... 21426, 21549
\_\_keys Set:Keyval:nnn
... 21459
\_\_keys Set:Known:nn
... 21429
\_\_keys Set:Known:nnn
... 21429
\_\_keys Set:Selective:
... 21459
\_\_keys Set:Selective:nnn
... 21470
\_\_keys Show:n
... 21876
\_\_keys Show:Nn
... 21876
\_\_keys Show:Nw
... 21876
\_\_keys Show:w
... 21876
\_\_keys Store_Unused:
.... 21616, 21636, 21642, 21679
\_\_keys Store:Unused:w
.... 21763, 21784, 21789
\_\_keys Store:Unused_aux:
... 21679
\_\_keys Temp:w
... 21828, 21840
\_\_keys TempBool
.... 20808, 21621, 21628, 21633
\_keyval\_if\_empty:w .......................... 20701, 20710, 20722
\_keyval\_if\_recursion\_tail:w .......................... 20585, 20595, 20722
\_keyval\_key:nn ................................ 920, 20659, 20689, 20706, 20715, 20726, 20727, 20732
\_keyval\_misplaced\_equal\_after_-active\_error:w .......................... 918, 20600, 20610, 20661
\_keyval\_misplaced\_equal\_in_-split\_error:w .......................... 20617, 20622, 20626, 20630, 20646, 20651, 20661
\_keyval\_pair:nnnn ......................... 919, 20632, 20653, 20696
\_keyval\_split\_active:w .......................... 917, 918, 20599, 20607, 20616, 20681
\_keyval\_split\_active\_auxi:w .......................... 20598, 20602, 20634, 20684
\_keyval\_split\_active\_auxii:w .......................... 20598, 20602, 20634, 20684
\_keyval\_split\_active\_auxiii:w .......................... 20598, 20602, 20634, 20684
\_keyval\_split\_active\_auxiv:w .......................... 20598, 20602, 20634, 20684
\_keyval\_split\_other:w .......................... 20589, 20605, 20625, 20636, 20645
\_keyval\_split\_other\_auxi:w .......................... 919, 20637, 20640, 20655
\_keyval\_split\_other\_auxii:w .......................... 20640
\_keyval\_split\_other\_auxiii:w .......................... 20640
\_keyval\_split\_other\_auxiv:w .......................... 20640
\_keyval\_tmp:w .......................... 20558, 20694, 20697, 20719, 20720, 20741, 20779
\_keyval\_trim\_other:nn .......................... 20610, 20632, 20653, 20696
\_keyval\_trim\_auxi:w .......................... 20740
\_keyval\_trim\_auxii:w .......................... 20740
\_keyval\_trim\_auxiii:w .......................... 20740
\_keyval\_trim\_auxiv:w .......................... 20740
\kuten .................................................. 1142, 1180

L

\L .................................................. 29549, 31581, 31941
\l .................................................. 29549, 31581, 31941
\label ............................................. 29557, 29564, 31889
\language .......................................... 327
\LARGE .................................................. 31869
Index

\Large ................................. 31870 \lineskip .............................. 341
\large .................................. 31873 \lineskiplimit .......................... 342
\lastallocatedtoks ................. 3394 \linewidth ............................ 33089
\lastbox ................................ 328 \ln .................................. 27009, 27012
\lastkern ................................ 329 \lopenalty ................................ 252
\lastlinefit ............................. 545 \localbrokenpenalty .............. 840
\lastnamedcs ........................... 833 \localinterlinepenalty ........... 841
\lastnodechar ........................... 847 \localleftbox ......................... 846
\lastnodefont ......................... 1143 \localrightbox ...................... 847
\lastnodesubtype ...................... 1145 \loccount .............................. 10096, 10327
\lastnodetype ......................... 546 \loctoks ............................... 3366, 3367, 3393
\lastpenalty ......................... 330 \logb .................................. 253
\lastsavedboxresourceindex ........ 920 \long .................................. 182, 343, 19252, 19256
\lastsavedimageresourceindex ...... 922 \LongText .............................. 55, 97
\lastsavedimagereSOURCEpages ...... 924 \looseness ................................ 344
\lastskip ................................ 331 \lower .................................. 345
\lastxpos ................................ 926 \\lowercase ............................. 88, 346
\lastypos ................................ 927 \\lc ................................. 345
\latexu ............ 834 ltx.utils .......................... 99, 11652
\latexukesfunction ................. 835 ltx.utils.filtdump .................. 99, 11721
\lccode ............................... 85, 86, 332 ltx.utils.filemd5sum ........ 99, 11742
\leaders ................................ 333 ltx.utils.filemoddate ........... 99, 11751
\left .................................. 334 ltx.utils.filesize ................ 99, 11804
\leftghost ............................ 836 lua commands:
\lefthyphenmin ....................... 335 \luaescape:n ....................... 99, 11611, 11613, 36777
\leftmarginkern ...................... 687 \luaescape:x:n .................... 36776
\leftskip ................................ 336 \lua.now:n ......................... 98, 8734, 8743, 11612, 11613, 29086, 36779
\legacy commands:
\legacy_if:nTF .................. 100, 11889 \lua.now:x:n ................... 36778
\legacy_if_gset:n ............... 225, 21314 \lua_shipout:n ................. 98, 11613
\legacy_if_gset:nn ............... 100, 11906 \lua_shipout_e:n .............. 11613, 36781
\legacy_if_gset_false:n .............. 100, 11898, 11913 \lua_shipout_e:x:n ........... 36780
\legacy_if_gset_inverse:n ........ 225, 21314 \___lua_escape:n ............... 11608, 11618
\legacy_if_gset_true:n ........... 100, 11898, 11913 \___lua_now:n .................. 11608, 11613
\legacy_if_p:n .................. 100, 11889 \___lua_shipout:n ............. 11608, 11615
\legacy_if_set:n ............... 225, 21314 \luabytecode ....................... 842
\legacy_if_set:nn ............... 100, 11906 \luabytecodecall ................. 843
\legacy_if_set_false:n .......... 100, 11898, 11908 \luacopyinputnodes ............. 844
\legacy_if_set_inverse:n ........ 225, 21314 \luadef .......................... 845
\legacy_if_set_true:n ........... 100, 11898, 11908 \luadef ................................ 11814
\legno .................................. 337 \luacsapestring ..................... 848
\let .................................. 22, 179, 180, 338 \luafunc ......................... 849
\lecharcode ......................... 837 \luafuncall ......................... 850
\letterspacefont .................... 688 \luatex_if_engine:TF .......... 36784, 36786, 36788
\limits ................................ 1332, 339 \luatex_if_engine_p: .............. 36782
\LineBreak ............................ 58, 1250 \luatexalignmark ................. 1250
\linedir ................................ 838 \luatexaligntab .................. 1251
\linedirection ...................... 839 \luatexattribute .................. 1252
\linepenalty ......................... 340 \luatexattributedef .......... 1253
\linepenalty ......................... 340 \luatexbanner ....................... 851
\luatexbodydir ..................... 1289
Index

<table>
<thead>
<tr>
<th>Command</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>\luatexboxdir</td>
<td>1290</td>
</tr>
<tr>
<td>\luatexc当地etable</td>
<td>1254</td>
</tr>
<tr>
<td>\luatextclearmarks</td>
<td>1255</td>
</tr>
<tr>
<td>\luatexc当地edisplaystyle</td>
<td>1256</td>
</tr>
<tr>
<td>\luatexc当地edscriptstyle</td>
<td>1258</td>
</tr>
<tr>
<td>\luatexc当地edscriptstyle</td>
<td>1259</td>
</tr>
<tr>
<td>\luatexc当地edtextstyle</td>
<td>1260</td>
</tr>
<tr>
<td>\luatextfontid</td>
<td>1261</td>
</tr>
<tr>
<td>\luatextformatname</td>
<td>1262</td>
</tr>
<tr>
<td>\luatextleading</td>
<td>1263</td>
</tr>
<tr>
<td>\luatexitinitcatcodeable</td>
<td>1264</td>
</tr>
<tr>
<td>\luatextlateralua</td>
<td>1265</td>
</tr>
<tr>
<td>\luatextleftghost</td>
<td>1269</td>
</tr>
<tr>
<td>\luatextlocalbrokenpenalty</td>
<td>1272</td>
</tr>
<tr>
<td>\luatextlocalinterlinepenalty</td>
<td>1274</td>
</tr>
<tr>
<td>\luatextlocalleftbox</td>
<td>1280</td>
</tr>
<tr>
<td>\luatextlocalrightbox</td>
<td>1281</td>
</tr>
<tr>
<td>\luatextluasecapestring</td>
<td>1282</td>
</tr>
<tr>
<td>\luatextluafunctor</td>
<td>1297</td>
</tr>
<tr>
<td>\luatextmathdir</td>
<td>1297</td>
</tr>
<tr>
<td>\luatextmathstyle</td>
<td>1299</td>
</tr>
<tr>
<td>\luatextnomaders</td>
<td>1300</td>
</tr>
<tr>
<td>\luatextnormils</td>
<td>1301</td>
</tr>
<tr>
<td>\luatextoutputbox</td>
<td>1302</td>
</tr>
<tr>
<td>\luatexpagebottomoffset</td>
<td>1303</td>
</tr>
<tr>
<td>\luatexpagedir</td>
<td>1304</td>
</tr>
<tr>
<td>\luatexpageheight</td>
<td>1305</td>
</tr>
<tr>
<td>\luatexpagelleftoffset</td>
<td>1306</td>
</tr>
<tr>
<td>\luatexpagerightoffset</td>
<td>1307</td>
</tr>
<tr>
<td>\luatexpagetopoffset</td>
<td>1308</td>
</tr>
<tr>
<td>\luatexpagewidth</td>
<td>1309</td>
</tr>
<tr>
<td>\luatexpardir</td>
<td>1310</td>
</tr>
<tr>
<td>\luatexpostexhyphenchar</td>
<td>1311</td>
</tr>
<tr>
<td>\luatexpostexhyphenchar</td>
<td>1312</td>
</tr>
<tr>
<td>\luatexpreezhyphenchar</td>
<td>1313</td>
</tr>
<tr>
<td>\luatexpreezhyphenchar</td>
<td>1314</td>
</tr>
<tr>
<td>\luatexrevision</td>
<td>1315</td>
</tr>
<tr>
<td>\luatexrightghost</td>
<td>1316</td>
</tr>
<tr>
<td>\luatexsavecatcodeable</td>
<td>1317</td>
</tr>
<tr>
<td>\luatexscantextokens</td>
<td>1318</td>
</tr>
<tr>
<td>\luatexsuppressionsfontnotfounderror</td>
<td>1319</td>
</tr>
<tr>
<td>\luatexsuppressionsfontsnameerror</td>
<td>1320</td>
</tr>
<tr>
<td>\luatexsuppressionerror</td>
<td>1321</td>
</tr>
<tr>
<td>\luatexsuppressionmathparerror</td>
<td>1322</td>
</tr>
<tr>
<td>\luatexsuppressionoutererror</td>
<td>1323</td>
</tr>
<tr>
<td>\luatextextdir</td>
<td>1324</td>
</tr>
<tr>
<td>\luatextUchar</td>
<td>1325</td>
</tr>
<tr>
<td>\luatexversion</td>
<td>1326</td>
</tr>
</tbody>
</table>

M

<table>
<thead>
<tr>
<th>Command</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>\mag</td>
<td>347</td>
</tr>
<tr>
<td>\mark</td>
<td>348</td>
</tr>
<tr>
<td>\marks</td>
<td>547</td>
</tr>
</tbody>
</table>

mode commands:

<table>
<thead>
<tr>
<th>Command</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>\mode_if_horizontal:TF</td>
<td>70</td>
</tr>
<tr>
<td>\mode_if_horizontal_p:</td>
<td>70</td>
</tr>
</tbody>
</table>

meta commands:

<table>
<thead>
<tr>
<th>Command</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>\meta:n</td>
<td>225</td>
</tr>
<tr>
<td>\meta:nn</td>
<td>225</td>
</tr>
<tr>
<td>\midde</td>
<td>548</td>
</tr>
<tr>
<td>\min</td>
<td>253</td>
</tr>
<tr>
<td>\c_minus_one</td>
<td>36684</td>
</tr>
<tr>
<td>\mkrn</td>
<td>368</td>
</tr>
<tr>
<td>\mm</td>
<td>257</td>
</tr>
</tbody>
</table>
Index

\muskip_zero_new: N \quad 218, 20499
\g_tmap_muskip \quad 220, 20545
\l_tmap_muskip \quad 220, 20545
\g_tmbp_muskip \quad 220, 20545
\l_tmbp_muskip \quad 220, 20545
\c_zero_muskip \quad 220, 20494, 20496, 20543
\muskipdef \quad 375
\mutoglue \quad 550
my commands:
\l_my_int \quad 156

N
\n \quad 8916, 8918, 8920
nan \quad 256
nc \quad 257
nd \quad 257
\newbox \quad 810
\newcatcodetable \quad 28884
\newcount \quad 810
\newdimen \quad 810
\newlinechar \quad 80, 376
\newluabytecode \quad 36
\next \quad 53, 94, 102, 105, 108, 116
\NG \quad 29550, 31582, 31942
\ng \quad 29550, 31582, 31954
\c_nine \quad 36704
\noalign \quad 377
\noautosizing \quad 1146
\noautoxspace \quad 1147
\noboundary \quad 378
\nobreakspace \quad 31897
\noexpand \quad 81, 96, 99, 119, 379
\nolimits \quad 871
\noindent \quad 380
\notkerns \quad 872
\noligs \quad 873
\nolimits \quad 381
\nonscript \quad 382
\nonstopmode \quad 383
\normaldeviate \quad 928
\normalend \quad 1351, 1352, 10092, 10122, 10323
\normaleveryjob \quad 1353
\normalexpanded \quad 1362
\normalfont \quad 31857
\normalhoffset \quad 13857
\normalinput \quad 1365
\normalinput \quad 1354
\normalitaliccorrection \quad 1364, 1366
\normallanguage \quad 1355
\normalleft \quad 1372, 1373
\normalmathop \quad 1356
\normalmiddle \quad 1374
\normalmonth \quad 1357
\normalouter \quad 1358
\normalover \quad 1359
\normalright \quad 1375
\normalshowtokens \quad 1368
\normalsize \quad 31874
\normalunexpanded \quad 1361
\normalvcenter \quad 1360
\normalvoffset \quad 1367
\nospaces \quad 874
\notexpansion: \langle \text{token} \rangle \quad 196
\noutrule \quad 875
\nulldelimiter \quad 384
\nullfont \quad 385
\num \quad 239
\number \quad 386
\numexpr \quad 551

O
\O \quad 29551, 31583, 31943, 32233
\o \quad 29551, 31583, 31955, 3224
\odelcode \quad 1183
\odelimiter \quad 1184
\oE \quad 29552, 31584, 31944
\oe \quad 29552, 31584, 31956
\omathaccent \quad 1185
\omathchar \quad 1186
\omathchardef \quad 1187
\omathcode \quad 1188
\omit \quad 387
\c_one \quad 36688
\c_one_hundred \quad 36622
\openin \quad 388
\openout \quad 389
\or \quad 390
or commands:
2672, 2673, 2674, 3849, 3850, 4043,
4044, 4567, 4568, 4569, 4570, 4841,
4842, 4843, 4844, 4845, 6396, 6449,
7083, 7085, 7117, 7119, 7119, 7120,
7121, 7122, 7123, 7124, 7125, 7126,
7127, 7128, 7129, 7253, 7254, 10621,
10622, 10623, 10624, 10625, 10626,
10627, 13507, 13583, 13913, 13914,
13915, 13916, 13917, 14942, 14943,
17020, 17605, 17606, 17607, 17608,
17609, 17610, 17611, 17612, 17613,
17614, 17615, 17616, 17617, 17618,
17619, 17620, 17621, 17622, 17623,
17624, 17625, 17626, 17627, 17628,
17629, 17638, 17639, 17640, 17641,
17642, 17643, 17644, 17645, 17646,
17647, 17648, 17649, 17650, 17651,
17652, 17653, 17654, 17655, 17656,
The text is not legible due to the nature of the PDF content. It appears to be a list of variable names with some associated numbers.
\pdfstartthread ........................ 636
\pdfstrcmp .................................. 124, 705, 22, 708
\pdfsuppressstextinfo ................. 637
\pdftex commands:
  \pdftex_if_engine:TF ................... 36798, 36800, 36802
  \pdftex_if_engine_p: ................. 36796
  \pdftexbanner ......................... 682
  \pdftexrevision ....................... 683
  \pdftexversion ....................... 684
  \pdfthread ............................. 638
  \pdfthreadmargin ..................... 639
  \pdftracingfonts ..................... 680, 1240, 1241
  \pdftrailer ........................... 640
  \pdfuniformdeviate ................... 681
  \pdfuniquerename ..................... 641
  \pdfvariable .......................... 887
  \pdffontorig .......................... 642
  \pdfxform ............................. 643
  \pdfxformname ......................... 644
  \pdfimage .............................. 645
  \pdfimagebbox ......................... 646
\peek commands:
  \peek_after:Nw ........... 71, 191, 4219,
  \peek_analysis_map_break: 19447, 19460, 19485, 19526, 36578
  \peek_analysis_map_break:n ... 194, 4184, 4206
  \peek_analysis_map_break:n 194, 4184, 7936
  \peek_analysis_map_inline:n ... 45, 191, 194, 428, 551, 4196, 7929
  \peek_catcode:NTF ... 192, 19581, 34741
  \peek_catcode_collect_inline:Nn ... 309, 36558
  \peek_catcode_ignore_spaces:NTF . 36936
  \peek_catcode_remove:NTF ........... 192, 19581, 34801
  \peek_catcode_remove_ignore_-spaces:NTF 36936
  \peek_catcode_remove_ignore_-spaces:NTF 192, 195, 19581
  \peek_catcode_collect_inline:Nn .... 309, 36558
  \peek_catcode_ignore_spaces:NTF .... 36936
  \peek_catcode_remove:NTF ........... 192, 195, 19581
  \peek_catcode_remove_ignore_-spaces:NTF 36936
  \peek_gafter:Nw ........... 191, 19447
  \peek_meaning:NTF ........... 192, 19581
  \peek_meaning_collect_inline:Nn . . . 309, 36558
  \peek_meaningignore_spaces:NTF . . . 36936
  \peek_meaning_remove:NTF ... 192, 19581
  \peek_meaning_remove_ignore_-spaces:NTF 36936
  \peek_N_type:TF ...................... 193, 19559, 19593, 19634
  \peek_regex:NTF ...................... 194, 7873, 7888, 7890
  \peek_regex:nTF ................. 194, 502, 550, 552, 553, 7873, 7879, 7881
  \peek_regex_remove_once:n ........ 194
  \peek_regex_remove_once:NTF ....... 194, 7873, 7907, 7909
  \peek_regex_remove_once:NTF ....... 195, 552, 7873, 7897, 7899
  \peek_regex_replace_once:Nn 195, 7965
  \peek_regex_replace_once:nn ...... 195, 7965
  \peek_regex_replace_once:nnTF ..... 195, 7965
  \peek_regex_replace_once:n ...... 195, 7965
  \peek_regex_replace_once:nTF ....... 195, 524, 527, 550, 554, 7865, 7967, 7969
  \peek_remove_filler:n .............. 193, 19472, 34739, 34799
  \peek_remove_spaces:n .............. 192, 19456, 36945, 36948, 36951, 36954, 36957, 36960
  \\peek_token ........ 191, 19436, 19450
  \l_peek_token ........... 191, 194, 445–447, 879, 881, 882, 1332, 1379,
  \l_peek_token ........... 4224, 4225, 4226, 4269, 4329, 4332,
  \l_peek_token ........... 4377, 19436, 19448, 19465, 19490,
  \l_peek_token ........... 19493, 19504, 19542, 19554, 19574,
  \l_peek_token ........... 19601, 19602, 19603, 19606, 34748,
  \l_peek_token ........... 34755, 34769, 36584, 36585, 36586
\peek internal commands:
  \peek_collect:N .................. 1379, 36558
  \peek_collect:NN ................. 36558
  \peek_collect_remove:nw ........ 36558
  \l_peek_collect:tl ........... 1379,
  \l_peek_collect:tl ........... 36557, 36560, 36571, 36596, 36601
  \peek_collect_true:w .... 1379, 36558
  \peek_execute_branches: .... 1379
  \peek_execute_branches:_catcode: 882, 19548, 36559
  \peek_execute_branches:_catcode:aux: 19548
  \peek_execute_branches:_catcode:auxi: 19548
  \peek_execute_branches:_catcode:auxii: 19548
  \peek_execute_branches:_catcode:auxiii: 19548

Index

1478
Index

13156, 14041, 14981, 16264, 16265, 16401, 16408, 16694, 16696, 17916, 17922, 17930, 18085, 18298, 18307, 18370, 18381, 18456, 18468, 18522, 18526, 18533, 21724, 22905, 22939, 22965, 22973, 24517, 28365, 29039

\prg_new_conditional:Nnn

\prg_generate_conditional-

\prg_map_break:Nn

\prg_new_protected_conditional:Nnn

\prg_new_protected_conditional:Nn

\prg_replicate:nn

\prg_return_false:

\prg_new_conditional:Nnn

\prg_new_conditional:Nnn

\prg_new_equation:nn

\prg_new_equ:nn

20412, 20414, 20505, 20507, 24779, 24780, 29097, 29099, 32286, 32288

63, 1609, 8388

63, 1592, 5288, 7393, 7398, 7425, 7427, 8388, 8871, 10110, 10213, 10233, 10894, 11026, 11173, 11175, 11177, 11179, 11214, 11276, 12335, 12348, 12366, 12347, 12355, 14015, 14024, 16584, 16693, 16695, 16701, 16704, 16707, 16710, 18094, 18104, 18106, 18234, 18238, 19851, 19861, 19935, 29101, 30240, 36242, 36314, 36316

69, 111, 148, 549, 572, 963, 4417, 5046, 5801, 6403, 6429, 6575, 6583, 6746, 6912, 7024, 7664, 7672, 7732, 7848, 7850, 8632, 9294, 9477, 10452, 22308, 22468, 26209, 27061, 27369, 27625, 27671, 27708, 28231, 28239, 28698, 28801, 35312, 35320, 35388, 35422, 35434, 35437, 35517, 35518

63, 64, 360, 540, 794, 814, 843, 1586, 1652, 1660, 1814, 1819, 1832, 1837, 1845, 1862, 2147, 2873, 4847, 4858, 4861, 4866, 4870, 4871, 4879, 4882, 4887, 4890, 4927, 4930, 4951, 4954, 5295, 5300, 7503, 8388, 8435, 8478, 8538, 8554, 8564, 8580, 8590, 8681, 8683, 8685, 8687, 8875, 8883, 9183, 10117, 10200, 10216, 10236, 10903, 10931, 11056, 11186, 11206, 11220, 11239, 11248, 11259, 11273, 11280, 11893, 12284, 12294, 12309, 12318, 12328, 12344, 12358, 12372, 12385, 12402, 12417, 12606, 12714, 12732, 12740, 12750, 12766, 12789, 12800, 13221, 13226, 13231, 13238, 13243, 13251, 13259, 13879, 13887, 13903, 13919, 14022, 14031, 14582, 14585, 14588, 14615, 14618, 14635, 14638, 14641, 16074, 16082, 16093, 16103, 16250, 16532, 16598, 16617, 17224, 17256, 17261, 17284, 17322, 17330, 17876, 17883, 17948, 17951, 18097, 18111, 18223, 18258, 18264, 18901, 19006, 19101, 19106, 19113, 19120, 19125, 19130, 19135, 19140, 19145, 19150, 19155, 19160, 19177, 19182, 19187, 19192, 19226, 19229, 19241, 19341, 19366, 19383,
Index

19392, 19589, 19869, 19918, 19926,
19942, 20142, 20161, 20176, 20177,
20442, 20449, 21865, 21874, 22818,
22820, 23965, 23975, 24786, 24800,
24813, 29111, 29117, 29533, 29536,
30623, 30627, 30633, 30664, 30690,
30711, 30728, 32342, 32344, 32350,
32974, 32976, 34222, 34225, 35901,
35927, 35939, 35951, 36272, 36330
\prg_return_true: ........ 63, 64,
360, 537, 540, 638, 678, 691, 794,
892, 1372, 1586, 1652, 1660, 1817,
1834, 1842, 1847, 1860, 1865, 2147,
2865, 2873, 4836, 4850, 4858, 4861,
4866, 4870, 4882, 4887, 4890, 4925,
4949, 5291, 5297, 7501, 8388, 8433,
8476, 8536, 8552, 8562, 8578, 8588,
8681, 8683, 8685, 8857, 8896, 9183,
10115, 10198, 10203, 10206, 10219,
10239, 10901, 11032, 11070, 11178,
11221, 11237, 11246, 11257, 11279,
11895, 12282, 12292, 12307, 12316,
12326, 12342, 12356, 12372, 12383,
12400, 12415, 12694, 12712, 12730,
12748, 12764, 12787, 12798, 13221,
13226, 13231, 13238, 13243, 13251,
13259, 13897, 13901, 13909, 13922,
14022, 14031, 14582, 14588, 14620,
14635, 14641, 16072, 16080, 16091,
16101, 16250, 16530, 16602, 16620,
17256, 17282, 17320, 17328, 17876,
17881, 17944, 17947, 17953, 18100,
18114, 18224, 18254, 18264, 19091,
19096, 19101, 19106, 19113, 19120,
19125, 19130, 19135, 19140, 19145,
19150, 19155, 19160, 19176, 19182,
19190, 19240, 19364, 19390, 19857,
19867, 19918, 19931, 19940, 20142,
20177, 20441, 20450, 21864, 21873,
22811, 22816, 23960, 23981, 24784,
24802, 24811, 29107, 29534, 30631,
30636, 30640, 30647, 30650, 30653,
30656, 30659, 30662, 30676, 30679,
30682, 30685, 30688, 30700, 30703,
30706, 30709, 30720, 30723, 30726,
32342, 32344, 32350, 32973, 34223,
35900, 35926, 35939, 35950, 36292
\prg_set_conditional:Nnn ....... ........ 62, 1609, 8388
\prg_set_conditional:Npnn ....... ........ 62, 1609, 8388
\prg_set_protected_conditional:Nmm ................ 63, 1609, 8388
\prg_set_protected_conditional:Npnn ................ 63, 1592, 8388, 11039

prg internal commands:
\_\_\_prg_break_point:Nn ........ 379
\_\_\_prg_F_true:w .............. 1681
\_\_\_prg_generate_conditional:nNNnnNNnn ............. 1604, 1629, 1638
\_\_\_prg_generate_conditional:NNNNNNNNnN ............. 1638
\_\_\_prg_generate_conditional_—
  count:NNNNNN .......... 1609
\_\_\_prg_generate_conditional_—
  count:NNNNNNN .......... 1609
\_\_\_prg_generate_conditional_—
  fast:nw .......... 360, 361, 1638
\_\_\_prg_generate_F_form:NNNNnnNN ........ 1681
\_\_\_prg_generate_p_form:NNNNnnNN ........ 360, 1681
\_\_\_prg_generate_T_form:NNNNnnNN ........ 1681
\_\_\_prg_generate_TF_form:NNNNnnNN ........ 1681
\_\_\_prg_p_true:w .......... 1681
\_\_\_prg_replicate:N .......... 8632
\_\_\_prg_replicate .......... 8632
\_\_\_prg_replicate_0:n .......... 8632
\_\_\_prg_replicate_1:n .......... 8632
\_\_\_prg_replicate_2:n .......... 8632
\_\_\_prg_replicate_3:n .......... 8632
\_\_\_prg_replicate_4:n .......... 8632
\_\_\_prg_replicate_5:n .......... 8632
\_\_\_prg_replicate_6:n .......... 8632
\_\_\_prg_replicate_7:n .......... 8632
\_\_\_prg_replicate_8:n .......... 8632
\_\_\_prg_replicate_9:n .......... 8632
\_\_\_prg_replicate_first:N .......... 8632
\_\_\_prg_replicate_first:—:n .......... 8632
\_\_\_prg_replicate_first_0:n .......... 8632
\_\_\_prg_replicate_first_1:n .......... 8632
\_\_\_prg_replicate_first_2:n .......... 8632
\_\_\_prg_replicate_first_3:n .......... 8632
\_\_\_prg_replicate_first_4:n .......... 8632
\_\_\_prg_replicate_first_5:n .......... 8632
\_\_\_prg_replicate_first_6:n .......... 8632
\_\_\_prg_replicate_first_7:n .......... 8632
\_\_\_prg_replicate_first_8:n .......... 8632
\_\_\_prg_replicate_first_9:n .......... 8632
\_\_\_prg_set_eq_conditional:NNn ........ 64, 1728, 8388
\_\_prg\_set\_eq\_conditional:nnNnnNNw
..................... 1736, 1744
\_\_prg\_set\_eq\_conditional\_F:_
\_\_form:nnn ............... 1744
\_\_prg\_set\_eq\_conditional\_F:_
\_\_form:Wnnnn ............. 1781
\_\_prg\_set\_eq\_conditional:_
\_\_loop:nnnNw .............. 1744
\_\_prg\_set\_eq\_conditional\_p:_
\_\_form:nnn .............. 1744
\_\_prg\_set\_eq\_conditional\_T:_
\_\_form:nnn .............. 1744
\_\_prg\_set\_eq\_conditional\_T:_
\_\_form:Wnnnn ............ 1779
\_\_prg\_set\_eq\_conditional\_TF:_
\_\_form:nnn ............. 1744
\_\_prg\_set\_eq\_conditional\_TF:_
\_\_form:Wnnnn ........... 1777
\_\_prg\_T\_true:w .......... 1681
\_\_prg\_TF\_true:w ........ 361, 1681
\_\_prg\_use\_none\_delimit\_by\_q:\_recursion\_stop:w .......... 1590, 1667, 1749, 1754, 1761
\primitive .................. 780

prop commands:
\c\_empty\_prop .............. 205, 886, 19644, 19654, 19658, 19661, 19917
\prop\_clear:N .......... 198, 19657, 19664, 19696, 19710, 33756, 34660
\prop\_clear\_new:N ........ 198, 19663, 34853, 34884, 34925
\prop\_concat:NNN ......... 200, 886, 19682
\prop\_const\_from\_keyval:Nn .... 199, 19694, 32937, 32944, 35630
\prop\_count:N .......... 201, 19805, 36174
\prop\_gclear:N 198, 19657, 19667, 19702
\prop\_gclear\_new:N ........ 198, 1288, 19663, 33011, 33012
\prop\_gconcat:NNN .......... 200, 19682
\prop\_get:Nn .......... 71, 36804, 36806
\prop\_get:NN .......... 137, 138, 200, 201, 19761, 33996, 34000, 34069, 34073, 34232, 34341, 34429, 35427
\prop\_get:NNTF .. 200, 202, 9585, 9605, 9660, 10172, 10386, 14082, 19935, 21190, 28970, 33202, 34335, 34341, 34863, 35121, 35134, 35149, 35228, 35256, 35272, 35643, 35656
\prop\_gpop:NN ........ 201, 19772
\prop\_gpop:NNNTF .... 201, 203, 19851
\prop\_gput:N .......... 226, 21344
\prop\_gput:NN .......... 199, 3737, 3738, 8346, 8347, 9375, 9377, 10066, 10067, 10074, 10075, 10077, 10078, 10079, 10080, 10100, 10146, 10331, 10361, 13845, 13846, 13847, 13848, 13849, 13850, 13851, 13852, 13853, 13854, 13855, 13856, 13857, 13858, 13859, 13866, 13869, 14718, 14719, 19726, 19873, 20738, 20739, 21972, 21973, 22878, 22879, 28923, 29252, 29253, 32328, 32346, 32381, 33312, 35046, 35047, 35048, 35049, 35050, 35051, 35052, 35063, 35064, 35066, 35068, 35069, 35070, 35071, 35072, 35073, 35074, 35174, 35175, 35189, 35203, 35213, 35863, 35864
\prop\_gput\_from\_keyval:Nn 200, 19694
\prop\_gput\_if\_new:Nn .... 200, 19804
\prop\_gremove:Nn .......... 201, 10157, 10371, 19748, 28921
\prop\_gaet\_eq:NN 199, 19661, 19669, 19686, 33013, 33015, 33180, 33182, 33219, 33221, 33468, 33634, 33675
\prop\_gaet\_from\_keyval:Nn 199, 19694
\prop\_hput:Nn ........ 19873
\prop\_if\_empty:NTF ....... 202, 19820, 19840, 19915, 34221, 36171
\prop\_if\_empty:p:N ........ 202, 19915
\prop\_if\_exist:NTF ........ 201, 19664, 19667, 19911, 21087, 34219
\prop\_if\_exist:p:N ........ 201, 19911
\prop\_if\_in:NNTF ........... 202, 9380, 9386, 19922, 34865
\prop\_if\_in:p:N ........ 202, 19922
\prop\_item:NN 201, 203, 9381, 9387, 19794, 35490, 35529, 36805, 36807
\prop\_log:N ............. 205, 20013
\prop\_map\_break: ........ 204, 895, 896, 19952, 19953, 19954, 19955, 19956, 19978, 19990, 19991, 19992, 19993, 19994, 20009
\prop\_map\_break:n .......... 204, 19802, 19931, 20009
\prop\_map\_function:NN .... 203, 303, 896, 10187, 10401, 19810, 19828, 19843, 19946, 20028, 34144, 35718
\prop\_map\_inline:Nn ........ 203, 19691, 19971, 33478, 33480, 33483, 33501, 33503, 33577, 33594, 33655, 33657, 33661, 33663, 33843, 33862, 34043, 34052
\prop\_map\_tokens:Nn ........ 203, 800, 880, 894, 19796, 19924, 19986
\q_no_value  
\q_recursion_stop  
\q_if_recursion_tail_stop_-
\q_if_recursion_tail_stop:N  
\q_if_recursion_tail_break:n  
\q_recursion_tail  
\q_if_no_value:nTF  
\q_if_nil_p:n  
\q_if_nil:p  
\q_if_nil:NTF  
\q_if_recursion_tail_stop:n  
\q_no_value

\quark internal commands:
Index

\regex_replace:nnN            182
\regex_replace_all:NNn       57, 7421
\regex_replace_all:NN       
                          47, 57, 538, 7421
\regex_replace_all:NTF      57, 7421
\regex_replace_all:NTFT     57, 7421
\regex_replace_case_all:NN  
                          58, 7467, 7479
\regex_replace_case_all:NTF 
                          58, 7462, 7480, 7481, 7482, 7483, 7484
\regex_replace_case_once:NN 58, 7421
\regex_replace_once:nnNn    501, 6275, 6324, 6342, 6924
\regex_replace_once:NnNn         519
\l_regex_added_begin_int     7496, 7635, 7643, 7647, 7688, 7829, 7834, 7837, 7848, 7863
\l_regex_added_end_int       7496, 7637, 7643, 7648, 7689, 7831, 7834, 7838, 7850, 7864
\c_regex_all_catcodes_int    4905, 5027, 5131, 5729
\c_regex_ascii_lower_int     4187, 4549, 4555
\c_regex_ascii_max_control_int  4484, 4666
\c_regex_ascii_max_int       4484, 4659, 4667, 4857
\c_regex_ascii_min_int       4484, 4658, 4665
\_regex_assertion:nnN        466, 480, 512, 5307, 5389, 5998, 6111, 6602
\_regex_b_test:              
\l_regex_balance_int         454, 525, 548, 4483, 7001, 7033, 7292, 7309, 7308, 7521, 7523, 7524, 7776, 7806, 7830, 7832
\g_regex_balance_intarray    
                          451, 514, 60980, 60987, 7825, 7250
\g_regex_balance_tl          525, 529, 6942, 7002, 7032, 7058, 7075, 7085, 7160
__regex_begin               7486
\_regex_branch:n            466, 484, 508, 4680, 5032, 5107, 5539, 5592, 5777, 5887, 5895, 5979, 5981, 5984, 6093, 6485
\_regex_break_point:TF      455, 479, 505, 4496, 4502, 6391, 6392, 6608, 6625
\_regex_break_true:w        455, 456, 4496, 4502, 5007, 4514, 4521, 4525, 4532, 4538, 4587, 4599, 4615, 5342, 6632, 6638, 6644
\_regex_build:N             537, 6255, 7400, 7406, 7424, 7428
\_regex_build:n            503, 537, 6255, 7395, 7404, 7423, 7426
\_regex_build_aux:NNN       550, 6255, 7885, 7904, 7974
\_regex_build_aux:N         550, 6255, 7876, 7894, 7966
\_regex_build_for_cs:n      4610, 6331
\_regex_build_new_state:    6269, 6270

regex internal commands:
__regex_A_test:              466, 5371, 5393, 6009, 6012, 6018, 6118, 6602
__regex_action_cost:n       501, 505, 6391, 6392, 6400, 6850, 6876
__regex_action_free:n       501, 514, 6419, 6420, 6421, 6432, 6490, 6494, 6519, 6544, 6548, 6551, 6579, 6587, 6597, 6611, 6654, 6848, 6852
__regex_action_free_aux:nn  6852
__regex_action_free_group:n  501, 514, 6440, 6559, 6562, 6852
__regex_action_start_wildcard:N
                          501, 6272, 6292, 6845
__regex_action_submatch:n   
                          501, 6296, 6321, 6513, 6514, 6652, 6901, 6903
__regex_action_submatch_aux:w
                          6903
__regex_action_submatch_aux:w
                          6903
__regex_action_submatch_aux1:w
                          6903
__regex_action_submatch_aux2:w
                          6903
__regex_action_submatch_aux4:w
                          6903
6289, 6290, 6294, 6334, 6335, 6364, 
6373, 6405, 6439, 6443, 6487, 6502, 
6507, 6546, 6605, 6606, 6604, 6649 
\l_regex_build_tl ........ 484, 554, 
4477, 5024, 5031, 5049, 5054, 5057, 
5058, 5061, 5062, 5065, 5125, 5128, 
5168, 5182, 5186, 5311, 5325, 5366, 
5388, 5401, 5433, 5446, 5450, 5532, 
5535, 5538, 5544, 5545, 5548, 5591, 
5881, 5885, 5892, 5898, 5919, 5935, 
5953, 6092, 6149, 6152, 6163, 6193, 
6208, 6212, 6215, 6221, 7000, 7023, 
7034, 7037, 7088, 7157, 7214, 7217, 
7231, 7299, 8033, 8036, 8044, 8047 
\l_regex_build_transition_- 
left:NNN ........ 6300, 6548, 6562, 6579 
\l_regex_build_transition_- 
right:nN ................... 
... 6300, 6406, 6440, 6490, 6494, 
6519, 6544, 6551, 6559, 6587, 6597 
\l_regex_build_transitions_- 
lazyiness:NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN NN
\s_tl_nil \ldots 688, 12605, 12609, 12628, 12631, 12634, 13083
\s_tl_stop \ldots \ldots 673, 683, 687, 12183, 12185, 12399, 12405, 12441, 12451, 12456, 12457, 12466, 12470, 12472, 12474, 12476, 12485, 12486, 12496, 12497, 12506, 12511, 12513, 12515, 12572, 12574, 12579, 12581, 12611, 12634, 12656, 12658, 12676, 12691, 12705, 12729, 12754, 13048, 13058, 13083
\s_token_mark \ldots \ldots 877, 19004, 19423, 19424, 19433
\s_token_stop \ldots 871, 873, 19004, 19169, 19172, 19202, 19237, 19345, 19349, 19355, 19378, 19425, 19433
\scantokens \ldots \ldots 897
\scantokens \ldots \ldots 561
\scriptbaselineadjustfactor \ldots \ldots 1156
\scriptfont \ldots \ldots 429
\scriptscriptdelimiters \ldots \ldots 430
\scriptscriptfont \ldots \ldots 1158
\scrollmode \ldots \ldots 433
\sshape \ldots \ldots 31864
\selectfont \ldots \ldots 254
\selectfont \ldots \ldots 254
seq commands:
\selectfont \ldots \ldots 31836
\seq_gconcat:NNN \ldots \ldots 36894
\seq_gclear:N \ldots \ldots 143, 150, 152, 236, 7029, 11524, 16544, 16734, 16748, 16895, 16923, 36250, 36325
\seq_gclear_new:N \ldots \ldots 142, 16337
\seq_gcopy:N \ldots \ldots 142, 16337
\seq_gcopy_left:N \ldots \ldots 142, 16337
\seq_gcopy_right:N \ldots \ldots 142, 16337
\seq_gdel_item:NnN \ldots \ldots 142, 16337
\seq_gdup:NN \ldots \ldots 142, 16337
\seq_gempty:NNF \ldots \ldots 142, 16337
\seq_gempty_list:N \ldots \ldots 142, 16337
\seq_gempty_list:NN \ldots \ldots 142, 16337
\seq_gget:NN \ldots \ldots 142, 16337
\seq_gget_left:NNNTF \ldots \ldots 146, 16935
\seq_gget_right:NN \ldots \ldots 144, 16448
\seq_gget_right:NNTF \ldots \ldots 146, 16935
\seq_gget_item:NN \ldots \ldots 151, 11335, 16976, 29043, 34788
\seq_gget_item:NNTF \ldots \ldots 151, 11335, 16976, 29043, 34788
\seq_gget_left:NN \ldots \ldots 1151, 10315, 10318, 10326
\seq_gget_left:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NN \ldots \ldots 145, 16666
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NTF \ldots \ldots 146, 16701
\seq_gget_right:NTF \ldots \ldots 146, 16701
\seq_gget_right:NN \ldots \ldots 145, 16666
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NTF \ldots \ldots 146, 16701
\seq_gget_right:NTF \ldots \ldots 146, 16701
\seq_gget_right:NTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
\seq_gget_right:NNTF \ldots \ldots 146, 16701
Index

\seq_item:Nn .......................... 56, 145, 799, 9669, 9670, 9675, 16721, 16748
\seq_log:N .......................... 145, 16988
\seq_map_break: ...................... 144, 150, 305, 16751, 16764, 16805, 16815, 16838, 16845, 16854, 21629
\seq_map_break:n ............. 149, 799, 3449, 3452, 9608, 9622, 10945, 11045, 16751
\seq_map_function:NN ............. 6, 148, 303, 801, 6166, 6234, 9673, 11415, 16755, 17003, 17991
\seq_map_indexed_function:NN ... 148, 16842, 36896, 36897
\seq_map_indexed_init:NN ......... 149, 16842, 36894, 36895
\seq_map_indexed:Nn .............. 148, 153, 1369, 3449, 3452, 9003, 11044, 16474, 16801, 21622
\seq_map_tokens:N ................. 148, 10944, 11528, 16808
\seq_mapthread_function:NN ... 305, 36187
\seq_new:N ................. 6, 142, 3320, 4475, 4919, 6252, 6253, 6941, 9558, 9559, 10086, 10317, 10766, 10791, 10797, 10798, 16325, 16338, 16341, 16466, 16539, 17013, 17014, 17015, 17016, 18136, 18688, 18691, 20805, 28868, 28869, 28870, 34715
\seq_pop:NN ....................... 151, 6317, 6355, 6357, 7099, 16976
\seq_pop:NNTF ...................... 152, 16982
\seq_pop_item:NnN .................. 306, 36308
\seq_pop_item:NnNFT ............... 306, 36308
\seq_pop_left:NN ................. 144, 16634, 16978, 16979, 16984, 16985
\seq_pop_left:NFT ................. 146, 16701
\seq_pop_right:NN ................... 145, 6095, 6173, 16666
\seq_pop_right:NNTF .............. 146, 16701
\seq_push:Nn ............. 152, 6327, 6348, 6350, 7259, 16956, 16963
\seq_put_left:NN ................. 144, 9598, 16436, 16956, 16957, 16958, 16959, 16960, 16961, 16962, 16963, 16964, 16965
\seq_put_right:NN ................. 144, 152, 153, 6098, 6171, 9659, 11501, 16457, 16477, 36745
\seq_rand_item:N ............... 145, 16745
\seq_remove_all:N ............. 143, 147, 152, 153, 16833, 18168, 36747
\seq_remove_duplicates:N ... 147, 152, 153, 11413, 16467
\seq_reverse:N ........... 147, 792, 16509
\seq_set_eq:NN ............... 142, 153, 3449, 16332, 16433, 16468, 16540
\seq_set_eq:NN .............. 142, 305, 802, 6229, 36209
\seq_set_function:NN ........ 143, 16351, 18167
\seq_set_from_clist:NN ......... 143, 172, 788, 11409, 11423, 16351, 21537
\seq_set_from_clist:N ........... 143, 16352, 36229
\seq_set_from_function:NNN ... 305, 7652, 36229
\seq_set_inline_x:NNN .......... 305, 1370, 36219, 36230
\seq_set_item:N ........... 306, 1372, 36234
\seq_set_item:NnTF ............ 306, 36234
\seq_set_map:NN .............. 150, 16885
\seq_set_map:x:NNN .......... 150, 802, 6242, 16875
\seq_set_split:NN .......... 143, 6228, 6241, 16893, 18689, 18692
\seq_set_split:keep_spaces:NN . 143, 16383
\seq_show:N ............. 154, 598, 700, 16988
\seq_shuffle:N ............. 147, 16537
\seq_sort:N ................. 44, 147, 3448, 16527
\seq_use:N .................. 151, 6245, 16919
\seq_use:Nnnn ................. 150, 16919
\g__tapas_sea .............. 154, 17013
\l__tapas_seq .............. 154, 17013
\l_tapb_seq .............. 154, 17013
\l_tmb_seq .............. 154, 17013

seq internal commands:
\_seq_count:w ........... 803, 16895
\_seq_count_end:w ...... 802, 16895
\_seq_get_left:wnv ......... 16623
\_seq_get_right_end:NN .... 16648
\_seq_get_right_loop:nu . 796, 16648
\_seq_if_in: .......... 16584
\_seq_int_eval:w .... 36233, 36289, 36299
\l\_seq_internal_a_int ......... 16551, 16557, 16566, 16568, 16569
\l\_seq_internal_a_tl ...... 789, 1371, 1372, 16321, 16395, 16399, 16405, 16410, 16412, 16498, 16503, 16588, 16592, 36248, 36304, 36322
\l\_seq_internal_b_int ......... 16567, 16570, 16571
\l\_seq_internal_b_tl ....... 7372, 16321, 16494, 16498, 16591, 16592, 36323, 36331, 36338
\l\_seq_internal_seq ........ 16537
\l\_seq_item:n ............. 786, 791, 794, 795, 797–800, 803, 1368, 1369, 16316, 16440, 16448, 16458, 16460,
16465, 16515, 16516, 16518, 16523, 16553, 16589, 16628, 16631, 16641, 16656, 16659, 16672, 16673, 16684, 16728, 16737, 16762, 16767, 16768, 16769, 16770, 16782, 16787, 16793, 16797, 16813, 16819, 16820, 16821, 16822, 16865, 16867, 16881, 16891, 16902, 16903, 16904, 16905, 16906, 16907, 16908, 16909, 16914, 16915, 16930, 16945, 16948, 16951, 36225, 36226, 36248, 36294, 36297, 36335

Șeq_item::NW 16721
Șeq_item::nW 16721
Șeq_item::WnW 16721
Șeq_map_function::NW 799, 16555
Șeq_map_indexed::NN

..................... 16844, 16852, 16857
Șeq_map_indexed::nW 801, 16842
Șeq_map_tokens::nw 10808
Șeq_mapthread_function::NnWnn

..................... 36187
Șeq_mapthread_function::WWn 36187
Șeq_mapthread_function::WnW 36187
Șeq_pop::NNW

............. 16605, 16635, 16637, 16667, 16669
Șeq_pop_item::nm 1372, 36326, 36328
Șeq_pop_item::NoNN

............... 36315, 36317, 36320
Șeq_pop_item::NNNN
Șeq_pop_item::aux:w

............... 1372, 36326, 36335
Șeq_pop_item::def:

......... 786, 16505, 16555, 16779, 16805, 16838, 16883, 16898, 36217, 36227
Șeq_pop_item::def:

................. 16634, 16703, 16706
Șeq_pop_left::NNN
Șeq_pop_left::wvNN
Șeq_pop_right::NNN

.............. 791, 16666, 16709, 16712
Șeq_pop_right_loop::nn 16666
Șeq_pop_TF::NNNN

............... 797, 16605, 16694, 16696, 16703, 16706, 16709, 16712
Șeq_push_item::def: 16552, 16779
Șeq_push_item::def:n

.............. 786, 16489, 16779, 16803, 16832, 16881, 16891, 36215, 36225
Șeq_put_left::aux:w 790, 16436
Șeq_remove_all::aux:NNn 16483
Șeq_remove_duplicates::NN 16467
Șeq_remove_seq

16466, 16473, 16476, 16477, 16479
Șeq_reverse::NN 16509
Șeq_reverse_item::nw ....... 792, 793
Șeq_reverse_item::wnn ....... 16509
Șeq_set_filter::NNNN ....... 36209
Șeq_set_from_inline_x::NNnn 36219
Șeq_set_item::nnNnN .....
Șeq_set_item::nnNnNN

.......................... 1372, 36244, 36324
Șeq_set_item::nnNnNNN
Șeq_set_item::vn ......... 36234
Șeq_set_item::vnnNnN
Șeq_set_item::vnnNNN 1371, 36234
Șeq_set_item::vnnNNNN

.. 13684, 16386, 16388, 16390, 16391
Șeq_set_split::Nw
Șeq_set_split::w
Șeq_set_split::auxi:w
Șeq_set_split::auxii:w 789
Șeq_set_split_end: 789, 16383
Șeq_show::N
Șeq_show::validat:nn 16988
Șeq_shuffle::NN
Șeq_shuffle::item:n 16537
Șeq_tap::w

............. 16623, 16515, 16518, 16672, 16684
Șeq_use::NNNnN
Șeq_use::nwn
Șeq_use::nnwwnnn
Șeq_use::setup:w 16919
Șeq_wrap::item:n

............. 789, 16369, 16354, 16359, 16364, 16369, 16380, 16396, 16421, 16465, 16501, 17010, 36215

\setbox ................ 435
\setfontid ............. 898
\setlanguage .......... 436
\setrandomseed ........ 940
\c_seven ............. 36700
\sffamily ............ 437
\shellescape .......... 782
\Shipout ................ 1226
\Shipsout ............. 438, 1213, 1214
\ShortText .......... 54, 96
\show .................. 439
\showbox ............. 440
\showboxbreath ........ 441
\showboxdepth .......... 442
\showgroups .......... 562
\showifs ............. 563
\showlists 443  
\showmode 1160  
\showstream 1192  
\showthe 444  
\showtokens 564  
sign 253  
sin 254  
sind 254  
c_six 36698  
c_sixteen 36718  
\sjis 1161  
\skewchar 445  
\skip 446, 19261  
skip commands:  
\c_max_skip 217, 20475  
\skip_add:N 215, 20426  
\skip_const:Nn 215, 20476  
\skip_eval:n 216, 20396, 20475, 20476  
\skip_gadd:N 215, 20426  
\skip_ge:N 215, 21352  
\skip_ge:Nt 215, 909, 20416  
\skip_ge:NN 215, 20426  
\skip_gsub:N 215, 20426  
\skip_gzero:N 215, 20402, 20409  
\skip_gzero_new:N 215, 20406  
\skip_horizontal:N 217, 20459  
\skip_horizontal:n 217, 20459  
\skip_if_eq:nnTF 216, 20438  
\skip_if_eq:p:nn 216, 20438  
\skip_if_eq:nTF 216, 20444  
\skip_if_eq:n:TF 216, 20444  
\skip_log:N 217, 20471  
\skip_log:n 217, 20471  
\skip_new:N 215, 20407, 20409, 20412  
\skip_new:p:N 215, 20412  
\skip_new:NTF 216, 20444  
\skip_new:nTF 216, 20444  
\skip_set:N 226, 21352  
\skip_set:Nn 215, 20416  
\skip_set_eq:NN 215, 20422  
\skip_show:N 216, 20467  
\skip_show:n 216, 912, 20469  
\skip_sub:Nn 215, 20426  
\skip_use:N 216, 20449, 20456, 20457  
\skip_set:N 218, 20459  
\skip_set:NN 218, 20459  
\skip_set_eq:NN 215, 20402, 20407  
\skip_zero:N 215, 218, 899, 20402, 20407  
\l_tmpa_skip 217, 20477  
\l_tmpb_skip 217, 20477  
\c_zero_skip 217, 20477  
skip internal commands:  
\_skip_if_finite:w 20444  
\_skip_if_finite:TF 20444, 20454  
\skipdef 447  
\slshape 31865  
\small 31876  
sort commands:  
\sort_ordered: 36814  
\sort_return_same: 43, 44, 418, 3531, 36815  
\sort_return_swapped: 43, 44, 418, 3531, 36817  
\sort_reversed: 36816  
sort internal commands:  
\_sort:nNn 419, 420  
\l_sort_A_int 417, 3330, 3337, 3344, 3347, 3356, 3495, 3500, 3503, 3523, 3555, 3562, 3577, 3579, 3580  
\l_sort_B_int 417, 3330, 3500, 3504, 3512, 3514, 3515, 3576, 3568, 3577, 3578, 3587, 3588, 3590  
\l_sort_begin_int 415, 417, 3328, 3492, 3580, 3590  
\l_sort_block_int 411, 412, 416, 3227, 3339, 3344, 3348, 3351, 3356, 3357, 3422, 3483, 3486, 3493, 3496  
\l_sort_C_int 417, 3330, 3501, 3505, 3512, 3513, 3524, 3556, 3563, 3567, 3569, 3570, 3587, 3589  
\_sort_compare:n 414, 418, 3421, 3522  
\_sort_compute_range: 414, 413, 3361, 3409  
\_sort_copy_block: 416, 3502, 3510  
\_sort_disable_toksdef: 3408, 3687  
\_sort_disabled_toksdef:n 3687  
\_sort_end_int 412, 416, 417, 3328, 3484, 3492, 3493, 3494, 3495, 3496, 3497, 3498, 3515  
\_sort_error: 3681, 3693, 3711  
\_sort_i:nNNn 421  
\_sort_internal_seq 3320, 3470, 3477, 3478  
\_sort_internal_tl 3320, 3433, 3436, 3437  
\_sort_length_int 411, 412, 3322, 3419, 3483  
\_sort_level: 414, 424, 3423, 3481, 3685  
\_sort_loop:wNn 420, 421  
\_sort_main:NNNn 415, 3406, 3432, 3469
<table>
<thead>
<tr>
<th>Command</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text_aftergroup</td>
<td>1191, 198, 1464</td>
</tr>
<tr>
<td>\text_alignmark</td>
<td>783, 1250</td>
</tr>
<tr>
<td>\text_aligntab</td>
<td>784, 1251</td>
</tr>
<tr>
<td>\text_atop</td>
<td>199</td>
</tr>
<tr>
<td>\text_atopwithdelims</td>
<td>200</td>
</tr>
<tr>
<td>\text_attribute</td>
<td>785, 1252</td>
</tr>
<tr>
<td>\text_attributedef</td>
<td>786, 1253</td>
</tr>
<tr>
<td>\text_automaticdiscretionary</td>
<td>788</td>
</tr>
<tr>
<td>\text_automaticphenomenon</td>
<td>789</td>
</tr>
<tr>
<td>\text_automaticity</td>
<td>799, 1255</td>
</tr>
<tr>
<td>\text_clubpenalty</td>
<td>515</td>
</tr>
<tr>
<td>\text_clerical</td>
<td>805</td>
</tr>
<tr>
<td>\text_closeout</td>
<td>217, 10370</td>
</tr>
<tr>
<td>\text_closein</td>
<td>216, 10156</td>
</tr>
<tr>
<td>\text_clearmarks</td>
<td>799, 1255</td>
</tr>
<tr>
<td>\text_cleaders</td>
<td>215</td>
</tr>
<tr>
<td>\text_chardef</td>
<td>354, 214, 1452, 1481, 1483, 1783, 1784, 8394, 8416, 8421, 10145, 10360, 17155, 19334, 29599, 29601, 29603</td>
</tr>
<tr>
<td>\text_char</td>
<td>213</td>
</tr>
<tr>
<td>\text_charend</td>
<td>215</td>
</tr>
<tr>
<td>\text_cleanmark</td>
<td>799, 1255</td>
</tr>
<tr>
<td>\text_closelin</td>
<td>216, 10156</td>
</tr>
<tr>
<td>\text_closeout</td>
<td>217, 10370</td>
</tr>
<tr>
<td>\text_clubpenalties</td>
<td>515</td>
</tr>
<tr>
<td>\text_clubpenalty</td>
<td>218</td>
</tr>
<tr>
<td>\text_copy</td>
<td>219, 32275, 32277, 32301, 32310, 32319, 32327</td>
</tr>
<tr>
<td>\text_copyfont</td>
<td>651, 913</td>
</tr>
<tr>
<td>\text_count</td>
<td>220, 3363, 3379, 3387, 3388, 10093, 10095, 10324, 10326</td>
</tr>
<tr>
<td>\text_countdef</td>
<td>221</td>
</tr>
<tr>
<td>\text_cr</td>
<td>222</td>
</tr>
<tr>
<td>\text_crumpledstyle</td>
<td>800, 1256</td>
</tr>
<tr>
<td>\text_crumpledscriptstyle</td>
<td>802, 1257</td>
</tr>
<tr>
<td>\text_crumpledtextstyle</td>
<td>803, 1259</td>
</tr>
<tr>
<td>\text_curly</td>
<td>804, 1260</td>
</tr>
<tr>
<td>\text_curved</td>
<td>223</td>
</tr>
<tr>
<td>\text_creationdate</td>
<td>773</td>
</tr>
<tr>
<td>\text_currentcjktoken</td>
<td>1113, 1174</td>
</tr>
<tr>
<td>\text_currentgrouplevel</td>
<td>1191, 516, 28058, 29038, 29047, 29054, 35113</td>
</tr>
<tr>
<td>\text_currentgroup</td>
<td>517</td>
</tr>
<tr>
<td>\text_currentifbranch</td>
<td>518</td>
</tr>
<tr>
<td>\text_currentiflevel</td>
<td>519</td>
</tr>
<tr>
<td>\text_currentit</td>
<td>520</td>
</tr>
<tr>
<td>\text_currentspacingmode</td>
<td>1114</td>
</tr>
<tr>
<td>\text_currentxspacingmode</td>
<td>1115</td>
</tr>
<tr>
<td>\text_day</td>
<td>225, 1321, 1325</td>
</tr>
<tr>
<td>\text_def</td>
<td>226</td>
</tr>
<tr>
<td>\text_definewidth</td>
<td>227, 697, 698, 699, 1465, 1467, 1469, 1470, 1491, 1493, 1494, 1495, 1497, 1498, 1499, 1501, 1502, 1503</td>
</tr>
<tr>
<td>\text_defaulthyphenchar</td>
<td>228</td>
</tr>
<tr>
<td>\text_defaultskewchar</td>
<td>229</td>
</tr>
<tr>
<td>\text_defaultwidthchar</td>
<td>230</td>
</tr>
<tr>
<td>\text_delimiter</td>
<td>231</td>
</tr>
<tr>
<td>\text_delimiterfactor</td>
<td>232</td>
</tr>
<tr>
<td>\text_delimitershortfall</td>
<td>233</td>
</tr>
<tr>
<td>\text_detokenize</td>
<td>234</td>
</tr>
<tr>
<td>\text_dimension</td>
<td>235</td>
</tr>
<tr>
<td>\text_dimensions</td>
<td>236</td>
</tr>
<tr>
<td>\text_discretionary</td>
<td>807</td>
</tr>
<tr>
<td>\text_discretionaryligaturemode</td>
<td>1116</td>
</tr>
<tr>
<td>\text_disinhibitglue</td>
<td>1117</td>
</tr>
<tr>
<td>\text_displayskip</td>
<td>1118</td>
</tr>
<tr>
<td>\text_displayskip</td>
<td>1119</td>
</tr>
<tr>
<td>\text_displaywidth</td>
<td>1120</td>
</tr>
<tr>
<td>\text_displaylimits</td>
<td>1121</td>
</tr>
<tr>
<td>\text_displaystyle</td>
<td>1122</td>
</tr>
<tr>
<td>\text_displaystyle</td>
<td>1123</td>
</tr>
<tr>
<td>\text_displaywidth</td>
<td>1124</td>
</tr>
<tr>
<td>\text_divide</td>
<td>242, 3357, 5404</td>
</tr>
<tr>
<td>\text_doublehyphenpenalty</td>
<td>243</td>
</tr>
<tr>
<td>\text_doublehyphenpenalty</td>
<td>243</td>
</tr>
<tr>
<td>\text_doublehyphenpenalty</td>
<td>244, 32291</td>
</tr>
<tr>
<td>\text_draftmode</td>
<td>652, 914</td>
</tr>
<tr>
<td>\text_dumout</td>
<td>1177</td>
</tr>
<tr>
<td>\text_dump</td>
<td>245</td>
</tr>
<tr>
<td>\text_dviextension</td>
<td>809</td>
</tr>
</tbody>
</table>
Index

\text_string:D ........ 456, 1213, 1217, 1316, 1322, 1328, 1334, 1456
\text_string:D ........ 714, 1287
\text_string:D ........ 900, 1280
\text_string:D ........ 901, 1282
\text_string:D ........ 902, 1283
\text_string:D ........ 903, 1285
\text_string:D ........ 905
\text_string:D ........ 457
\text_string:D ........ 694
\text_string:D ........ 1162
\text_string:D ........ 1163
\text_string:D ........ 1165
\text_string:D ........ 906, 1305
\text_string:D ........ 907
\text_string:D ........ 458
\text_string:D ........ 459
\text_string:D ........ 568
\text_string:D ........ 1166
\text_string:D ........ 344, 384, 1008, 1014, 1015, 159, 460,
1901, 2184, 2333, 2337, 2666, 2708,
2799, 2818, 2834, 2838, 3443, 3475,
3523, 3524, 3555, 3556, 3562, 3563,
4040, 4154, 4291, 4409, 4428, 4434,
4437, 9015, 16558, 17033, 17034,
17209, 17210, 18597, 18667, 18673,
18679, 18685, 20273, 20274, 20275,
20345, 20346, 23480, 23968, 32374
\text_string:D ........ 461
\text_string:D ........ 462
\text_string:D ........ 463, 1315, 1319
\text_string:D ........ 464, 2809,
2838, 3416, 3443, 3475, 3512, 3523,
3524, 3555, 3556, 3562, 3563, 3567,
3577, 3587, 3864, 4040, 4409,
4412, 4414, 4419, 4427, 4428, 4433,
4434, 4437, 16558, 16569, 16570, 16571
\text_string:D ........ 908
\text_string:D ........ 909
\text_string:D ........ 465, 3694
\text_string:D ........ 466
\text_string:D ........ 467
\text_string:D ........ 569
\text_string:D ........ 468
\text_string:D ........ 1167
\text_string:D ........ 910
\text_string:D ........ 570
\text_string:D ........ 469
\text_string:D ........ 680, 941, 1239, 1241, 1245
\text_string:D ........ 573, 8888, 10909, 12117
\text_string:D ........ 472, 2212, 2218, 2219, 32389
\text_string:D ........ 473
\text_string:D ........ 474
\text_string:D ........ 475
\text_string:D ........ 476
\text_string:D ........ 574
\text_string:D ........ 1193
\text_string:D ........ 477
\text_string:D ........ 478, 18677, 18679
\text_string:D ........ 943, 1286, 29262
\text_string:D ........ 944, 1391, 18758, 29267
\text_string:D ........ 479
\text_string:D ........ 1168
\text_string:D ........ 945
\text_string:D ........ 946
\text_string:D ........ 947
\text_string:D ........ 948
\text_string:D ........ 949
\text_string:D ........ 950
\text_string:D ........ 951
\text_string:D ........ 952
\text_string:D ........ 953
\text_string:D ........ 954
\text_string:D ........ 955
\text_string:D ........ 956
\text_string:D ........ 957
\text_string:D ........ 958
\text_string:D ........ 959
\text_string:D ........ 960
\text_string:D ........ 961
\text_string:D ........ 962
\text_string:D ........ 963
\text_string:D ........ 964
\text_string:D ........ 965
\text_string:D ........ 966
\text_string:D ........ 967
\text_string:D ........ 968
\text_string:D ........ 969
\text_string:D ........ 970
\text_string:D ........ 971
\text_string:D ........ 972
\text_string:D ........ 973
\text_string:D ........ 974
\text_string:D ........ 975
\text_string:D ........ 976
\text_string:D ........ 977
\text_string:D ........ 978
\text_string:D ........ 979
\text_string:D ........ 980
\text_string:D ........ 982
\text_string:D ........ 1518
\text_string:D ........ 572
\text_string:D ........ 470
\text_string:D ........ 471
\text_string:D ........ 573
\text_string:D ........ 474
\text_string:D ........ 475
\text_string:D ........ 476
\text_string:D ........ 574
\text_string:D ........ 1193
\text_string:D ........ 477
\text_string:D ........ 478, 18677, 18679
\text_string:D ........ 943, 1286, 29262
\text_string:D ........ 944, 1391, 18758, 29267
\text_string:D ........ 479
\text_string:D ........ 1168
\text_string:D ........ 945
\text_string:D ........ 946
\text_string:D ........ 947
\text_string:D ........ 948
\text_string:D ........ 949
\text_string:D ........ 950
\text_string:D ........ 951
\text_string:D ........ 952
\text_string:D ........ 953
\text_string:D ........ 954
\text_string:D ........ 955
\text_string:D ........ 956
\text_string:D ........ 957
\text_string:D ........ 958
\text_string:D ........ 959
\text_string:D ........ 960
\text_string:D ........ 961
\text_string:D ........ 962
\text_string:D ........ 963
\text_string:D ........ 964
\text_string:D ........ 965
\text_string:D ........ 966
\text_string:D ........ 967
\text_string:D ........ 968
\text_string:D ........ 969
\text_string:D ........ 970
\text_string:D ........ 971
\text_string:D ........ 972
\text_string:D ........ 973
\text_string:D ........ 974
\text_string:D ........ 975
\text_string:D ........ 976
\text_string:D ........ 977
\text_string:D ........ 978
\text_string:D ........ 979
\text_string:D ........ 980
\text_string:D ........ 982
Index

\tl_map_tokens:nn .......... 111, 7073, 12493, 12518
\tl_map_variable:NNN ... 111, 12517
\tl_map_variable:Nn    111, 685, 12517
\tlmixed_case:n .......... 36902
\tlmixed_case:nn ........ 36902
\tlnew:N .................. 103, 107, 108, 109, 115
\tlput_left:NN .......... 103, 12903
\tlput_right:NN .......... 104, 1375, 4240, 4242, 4317, 4327, 4366, 4386, 4707, 10712, 10715, 10720, 11066, 12051, 16458, 18148, 18777, 18779, 18782, 18784, 18785, 18787, 18789, 18791, 18792, 18794, 18796, 18798, 18800, 18802, 29589, 35432, 35478, 36596, 36601
\tlrand_item:N .......... 114, 12919
\tlrand_item:n .......... 114, 12919
\tlrange:Nn ............ 115, 12926
\tlrange:nnn ........... 115, 128, 307, 12926
\tlrange:braced:Nnn ... 307, 36523
\tlrange:braced:nnn ... 115, 307, 36523
\tlrange:unbraced:Nnn ... 307, 36523
\tlrange:unbraced:nnn... 115, 307, 36523
\tlremove_all:N ........ 116, 117, 12273
\tlremove_once:N ....... 116, 12267
\tlreplace_all:Nn ....... 116, 789, 842, 12199, 12274, 16405
\tlreplace_once:Nn ....... 116, 12199, 12268, 18814
\tlrescan:nn 117, 118, 262, 672, 12101
\tlreverse:N ............ 108, 12890
\tlreverse:n ............ 108, 109, 12893
\tlreverse_items:n ... 108, 109, 12567
\tlset:N .................. 226, 21376
\tlset:Nn ................. 103, 117, 118, 144, 227, 308, 399, 600, 666, 668, 675, 926, 1377, 4213, 5065, 5828, 5905, 6152, 6215, 6745, 6766, 6776, 6792, 6797, 6840, 6872, 6910, 7261, 7922, 7923, 7983, 8869, 8904, 9309, 9539, 9567, 9650, 10210, 10223, 10256, 10529, 10566, 10892, 10927, 11023, 11051, 11166, 11168, 11170, 11172, 11212, 11446, 11449, 11450, 11451, 11452, 11459, 11460, 11461, 11463, 11467, 11983, 12110, 12338, 12351, 12352, 12520, 13048, 13068, 14083, 14233, 16395, 16399, 16427, 16494, 16503, 16588, 16591, 16608, 16616, 16635, 16644, 16703, 16834, 17449, 18060, 18066, 18075, 18082, 18336, 18775, 19514, 19751, 19767, 19768, 19776, 19777, 19779, 19875, 19878, 19855, 19856, 19865, 19877, 19900, 19939, 20315, 20801, 20983, 21192, 21407, 21416, 21446, 21458, 21466, 21488, 21500, 21508, 21519, 21528, 21539, 21626, 25056, 28996, 29006, 29539, 29542, 29557, 29561, 29563, 29594, 29971, 31824, 33207, 33851, 33852, 33973, 34213, 34241, 34311, 34338, 34345, 34362, 34370, 34373, 34411, 34426, 34467, 34463, 34654, 34658, 34702, 34710, 34714, 34843, 34851, 34868, 34871, 34912, 34928, 34952, 34961, 34981, 35015, 35017, 35019, 35022, 35039, 35233, 35421, 36248, 36323, 36568, 3672, 4091, 13024, 13824
\tlset_eq:NN ............... 103, 172, 564, 3425, 6934, 7358, 7841, 8409, 9613, 9621, 11069, 11951, 13121, 14012, 14021, 16343, 16344, 16345, 16346, 17971, 17972, 17973, 17974, 19669, 19670, 19671, 19672, 20891, 21547, 21571, 21651, 21672, 24713, 34231, 34340, 34425, 34914, 34934, 34957, 35276
\tlset_from:file:NN ........ 36852
\tlset_from:file:x:NN ...... 36854
\tlset_rescan:NN ........... 117, 118, 262, 637, 672, 12101
\tlset_x:NN ................ 227, 21376
\tlshow:N .................. 109, 110, 172, 440, 4091, 13024, 13824
\tlshow:n ................ 83, 109, 110, 377, 700, 1043, 1366, 2172, 2189, 8446, 10164, 10378, 13024, 13044, 13820, 17863, 18599, 18669,
Index

1532

\_tl_set_rescan_multi:nnN ...... 671, 673, 12101, 12152, 12174
\_tl_set_rescan_single:nnN ...... 673, 12144
\_tl_set_rescan_single:NNw ...... 673
\_tl_set_rescan_single_aux:nnNN ...... 12144
\_tl_set_rescan_single_aux:w ...... 673, 12144
\_tl_show:n ...... 13044
\_tl_show:NN ...... 13024
\_tl_tl_head:w ...... 12639, 12729
\_tl_tmp:w ...... 12843, 12848
\_tl_set_rescan_single:NNww ...... 12144
\_tl_set_rescan_single:nnNN ...... 12144
\_tl_tlp:w ...... 680, 687, 4263, 1246, 12370, 12371, 12377, 12390, 12599, 12638, 12809, 12824, 13090
\_tl_trimm: ...... 687, 688, 12586, 12591, 12599
\_tl_trimm_spaces:nn ...... 922, 12585, 12591, 12599
\_tl_trimm_spaces_auxi:w ...... 687, 12599
\_tl_trimm_spaces_auxii:w ...... 688, 12599
\_tl_trimm_spaces_auxiii:w ...... 689, 12599
\_tl_use_none_delimit_by_q_act_stop:w ....... 12809
\_tl_use_none_delimit_by_s_act_stop:w ....... 12843, 12848
\_tl_use_none_delimit_by_s_:stop:w ...... 683, 12453, 12505, 12513

token commands:
\c_alignment_tokenize ...... 186, 686, 3999, 18996, 19066, 19105, 29496, 34741
\c_catcode_letter_tokenize ...... 186, 869, 3995, 19008, 19066, 19134, 29508
\c_catcode_other_tokenize ...... 186, 869, 3993, 19011, 19066, 19139, 29511
\c_group_begin_tokenize ...... 186
\c_group_end_tokenize ...... 186
\c_math_subscript_tokenize ...... 186, 686, 4003, 19006, 19124, 29467, 29502, 34750
\c_math_superscript_tokenize ...... 186, 686, 4001, 18999, 19066, 19119, 29499, 34751
\c_math_toggle_tokenize ...... 186, 687, 3997, 18993, 19066, 19100, 29464, 29493
\c_parameter_tokenize ...... 186, 524, 686, 19066, 19109, 19112
\c_space_tokenize ...... 38, 107, 118, 186, 193, 309, 691, 686, 2766, 3833, 3863, 4065, 4191, 4226, 4329, 4775, 4816, 7111, 10636, 12709, 12747, 19005, 19066, 19129, 19465, 19490, 19603, 29468, 29505, 29665, 36586
\token_case_catcode:N ...... 191, 19396
\token_case_catcode:Nn ...... 191, 19396, 19398, 19400
\token_case_charcode:N ...... 191, 19396
\token_case_charcode:Nn ...... 191, 19396
\token_case_charcode:NnF ...... 191, 19396, 19406, 19408
\token_case_meaning:N ...... 191, 19396
\token_case_meaning:Nn ...... 191, 19396, 19414, 19416, 34748, 34755
\token_get_arg_spec:N ...... 36560
\token_get_prefix_spec:N ...... 36562
\token_get_replacement_spec:N ...... 188, 19142, 30337
\token_get_active:p:N ...... 188, 19142, 29919, 30293, 30312, 31776
\token_get_active:p:N ...... 188, 19103
\token_if_alignment:N ...... 188, 19103
\token_if_alignment:p:N ...... 188, 19103
\token_if_chardef:N ...... 189, 4146, 19215
\token_if_chardef:p:N ...... 189, 19215, 29436
\token_if_cs:N ...... 189, 19179, 29768, 30154, 30490, 30517, 30588, 31783
\token_if_cs:p:N ...... 189, 19179, 29918, 30610, 30748, 30850, 30903, 30931, 30976, 31775
\token_if_dim_register:N ...... 190, 4148, 19215
\token_if_dim_register:p:N ...... 190, 19215
\token_if_eq_catcode:N ...... 188, 191, 192, 309, 2766, 19152, 19397, 19399, 19401, 19403
\token_if_eq_catcode:Nn ...... 188, 19152
\token_if_eq_catcode:NnF ...... 188, 191, 192, 4816, 4821, 5456, 5565, 5669, 5761, 5709, 5847, 7097, 7111, 9849, 10636, 19157, 19405, 19407, 19409, 19411, 26736
\token_if_eq_charcode:N ...... 188, 19152
\token_if_eq_charcode:Nn ...... 189, 19157
\token_if_eq_charcode:NnF ...... 189, 19157
\token_if_eq_meaning:N ...... 189, 191
<table>
<thead>
<tr>
<th>Command</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>\tracinggifs</td>
<td>572</td>
</tr>
<tr>
<td>\tracinglostchars</td>
<td>470</td>
</tr>
<tr>
<td>\tracingmacros</td>
<td>471</td>
</tr>
<tr>
<td>\tracingnesting</td>
<td>573</td>
</tr>
<tr>
<td>\tracingonline</td>
<td>472</td>
</tr>
<tr>
<td>\tracingoutput</td>
<td>473</td>
</tr>
<tr>
<td>\tracingstacklevels</td>
<td>574</td>
</tr>
<tr>
<td>\tracingparagraphs</td>
<td>475</td>
</tr>
<tr>
<td>\tracingrestores</td>
<td>476</td>
</tr>
<tr>
<td>\tracingstacklevels</td>
<td>1193</td>
</tr>
<tr>
<td>\tracingstats</td>
<td>477</td>
</tr>
<tr>
<td>\ttfamily</td>
<td>31860</td>
</tr>
<tr>
<td>\c_{twelve}</td>
<td>36710</td>
</tr>
<tr>
<td>\c_{two}</td>
<td>36690</td>
</tr>
<tr>
<td>\c_{two_hundred_fifty_five}</td>
<td>36724</td>
</tr>
<tr>
<td>\c_{two_hundred_six}</td>
<td>36726</td>
</tr>
<tr>
<td>\mathcode</td>
<td>3678</td>
</tr>
<tr>
<td>\mathaccent</td>
<td>36710</td>
</tr>
<tr>
<td>\mathaccentuml</td>
<td>36724</td>
</tr>
<tr>
<td>\mathaccenthbox</td>
<td>36726</td>
</tr>
<tr>
<td>\mathaccentset</td>
<td>36726</td>
</tr>
<tr>
<td>\mathcharclass</td>
<td>36731</td>
</tr>
<tr>
<td>\mathchardef</td>
<td>36734</td>
</tr>
<tr>
<td>\mathcharfam</td>
<td>36736</td>
</tr>
<tr>
<td>\mathcharslot</td>
<td>36738</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36740</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36742</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36744</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36746</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36748</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36750</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36752</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36754</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36756</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36758</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36760</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36762</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36764</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36766</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36768</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36770</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36772</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36774</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36776</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36778</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36780</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36782</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36784</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36786</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36788</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36790</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36792</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36794</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36796</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36798</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36800</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36802</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36804</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36806</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36808</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36810</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36812</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36814</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36816</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36818</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36820</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36822</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36824</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36826</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36828</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36830</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36832</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36834</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36836</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36838</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36840</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36842</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36844</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36846</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36848</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36850</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36852</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36854</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36856</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36858</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36860</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36862</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36864</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36866</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36868</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36870</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36872</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36874</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36876</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36878</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36880</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36882</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36884</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36886</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36888</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36890</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36892</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36894</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36896</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36898</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36900</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36902</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36904</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36906</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36908</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36910</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36912</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36914</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36916</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36918</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36920</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36922</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36924</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36926</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36928</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36930</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36932</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36934</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36936</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36938</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36940</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36942</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36944</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36946</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36948</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36950</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36952</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36954</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36956</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36958</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36960</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36962</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36964</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36966</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36968</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36970</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36972</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36974</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36976</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36978</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36980</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36982</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36984</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36986</td>
</tr>
<tr>
<td>\mathcloseclosespacing</td>
<td>36988</td>
</tr>
<tr>
<td>\mathcloseinnerspacing</td>
<td>36990</td>
</tr>
<tr>
<td>\mathclosebinspacing</td>
<td>36992</td>
</tr>
</tbody>
</table>
\index{\protect\texttt{\textbackslash Umathcloseinnerspacing}}{971 \protect\texttt{\textbackslash Umathordpunctspacing}}{1033}
\index{\protect\texttt{\textbackslash Umathcloseopenspacing}}{973 \protect\texttt{\textbackslash Umathordrelspacing}}{1034}
\index{\protect\texttt{\textbackslash Umathcloseopspacing}}{974 \protect\texttt{\textbackslash Umathoverbarkern}}{1035}
\index{\protect\texttt{\textbackslash Umathcloseordspacing}}{975 \protect\texttt{\textbackslash Umathoverbarrule}}{1036}
\index{\protect\texttt{\textbackslash Umathclosepunctspacing}}{976 \protect\texttt{\textbackslash Umathoverbarygap}}{1037}
\index{\protect\texttt{\textbackslash Umathcloserelspacing}}{978 \protect\texttt{\textbackslash Umathoverdelimiterbgap}}{1038}
\index{\protect\texttt{\textbackslash Umathcode}}{979 \protect\texttt{\textbackslash Umathoverdelimitervgap}}{1040}
\index{\protect\texttt{\textbackslash Umathcodenum}}{980 \protect\texttt{\textbackslash Umathpunctbinspacing}}{1042}
\index{\protect\texttt{\textbackslash Umathconnectoroverlapmin}}{981 \protect\texttt{\textbackslash Umathpunctclosespacing}}{1043}
\index{\protect\texttt{\textbackslash Umathfractiondelsize}}{983 \protect\texttt{\textbackslash Umathpunctinnerspacing}}{1045}
\index{\protect\texttt{\textbackslash Umathfractiondenomdown}}{984 \protect\texttt{\textbackslash Umathpunctopenspacing}}{1047}
\index{\protect\texttt{\textbackslash Umathfractiondenomvgap}}{986 \protect\texttt{\textbackslash Umathpunctopsspacing}}{1048}
\index{\protect\texttt{\textbackslash Umathfractionnumup}}{988 \protect\texttt{\textbackslash Umathpunctordspacing}}{1049}
\index{\protect\texttt{\textbackslash Umathfractionnumvgap}}{989 \protect\texttt{\textbackslash Umathpunctpunctspacing}}{1050}
\index{\protect\texttt{\textbackslash Umathfractionrule}}{990 \protect\texttt{\textbackslash Umathpunctrelspacing}}{1052}
\index{\protect\texttt{\textbackslash Umathinnerbinspacing}}{991 \protect\texttt{\textbackslash Umathquad}}{1053}
\index{\protect\texttt{\textbackslash Umathinnerclosespacing}}{992 \protect\texttt{\textbackslash Umathradicaldegreeafter}}{1054}
\index{\protect\texttt{\textbackslash Umathinnerinnerspacing}}{994 \protect\texttt{\textbackslash Umathradicaldegreebefore}}{1056}
\index{\protect\texttt{\textbackslash Umathinneropenspacing}}{995 \protect\texttt{\textbackslash Umathradicaldegreeraise}}{1058}
\index{\protect\texttt{\textbackslash Umathinneropspacing}}{997 \protect\texttt{\textbackslash Umathradicalkern}}{1060}
\index{\protect\texttt{\textbackslash Umathinnerordspacing}}{998 \protect\texttt{\textbackslash Umathradicalrule}}{1061}
\index{\protect\texttt{\textbackslash Umathinnerpunctspacing}}{999 \protect\texttt{\textbackslash Umathradicalvgap}}{1062}
\index{\protect\texttt{\textbackslash Umathinnerrelspacing}}{1001 \protect\texttt{\textbackslash Umathrelbinspacing}}{1063}
\index{\protect\texttt{\textbackslash Umathlimitabovexgap}}{1002 \protect\texttt{\textbackslash Umathrelclosespacing}}{1064}
\index{\protect\texttt{\textbackslash Umathlimitabovexkern}}{1003 \protect\texttt{\textbackslash Umathrelinnerspacing}}{1065}
\index{\protect\texttt{\textbackslash Umathlimitabovexvgap}}{1004 \protect\texttt{\textbackslash Umathrelopenspacing}}{1066}
\index{\protect\texttt{\textbackslash Umathlimitbelowxgap}}{1005 \protect\texttt{\textbackslash Umathrelpexpacing}}{1067}
\index{\protect\texttt{\textbackslash Umathlimitbelowxkern}}{1006 \protect\texttt{\textbackslash Umathrelordspacing}}{1068}
\index{\protect\texttt{\textbackslash Umathlimitbelowxvgap}}{1007 \protect\texttt{\textbackslash Umathrelpuncspacing}}{1069}
\index{\protect\texttt{\textbackslash Umathlimitsubfactor}}{1008 \protect\texttt{\textbackslash Umathrelrelspacing}}{1070}
\index{\protect\texttt{\textbackslash Umathlimsupxgap}}{1009 \protect\texttt{\textbackslash Umathrelvspacing}}{1073}
\index{\protect\texttt{\textbackslash Umathrelvspacing}}{1010 \protect\texttt{\textbackslash Umathrelvspacing}}{1073}
\index{\protect\texttt{\textbackslash Umathopbinspacing}}{1011 \protect\texttt{\textbackslash Umathspaceafterscript}}{1075}
\index{\protect\texttt{\textbackslash Umathopenbinspacing}}{1012 \protect\texttt{\textbackslash Umathstackdenomdown}}{1076}
\index{\protect\texttt{\textbackslash Umathopeninnerspacing}}{1013 \protect\texttt{\textbackslash Umathstacknumup}}{1077}
\index{\protect\texttt{\textbackslash Umathopenpuncspacing}}{1014 \protect\texttt{\textbackslash Umathstackxvgap}}{1078}
\index{\protect\texttt{\textbackslash Umathopenpuncspacing}}{1015 \protect\texttt{\textbackslash Umathsubshiftdown}}{1079}
\index{\protect\texttt{\textbackslash Umathopenpuncspacing}}{1016 \protect\texttt{\textbackslash Umathsubshiftdrop}}{1080}
\index{\protect\texttt{\textbackslash Umathopenordspacing}}{1017 \protect\texttt{\textbackslash Umathsubupshiftdown}}{1081}
\index{\protect\texttt{\textbackslash Umathopenpuncspacing}}{1018 \protect\texttt{\textbackslash Umathsubupvgap}}{1082}
\index{\protect\texttt{\textbackslash Umathopenrelspacing}}{1019 \protect\texttt{\textbackslash Umathsubupmax}}{1083}
\index{\protect\texttt{\textbackslash Umathoperatorsize}}{1020 \protect\texttt{\textbackslash Umathsupbottommin}}{1084}
\index{\protect\texttt{\textbackslash Umathopinnerspacing}}{1021 \protect\texttt{\textbackslash Umathsupshiftdrop}}{1085}
\index{\protect\texttt{\textbackslash Umathopinnerspacing}}{1022 \protect\texttt{\textbackslash Umathsupshiftup}}{1086}
\index{\protect\texttt{\textbackslash Umathopinnerspacing}}{1023 \protect\texttt{\textbackslash Umathsupsubbottommax}}{1087}
\index{\protect\texttt{\textbackslash Umathopinnerspacing}}{1024 \protect\texttt{\textbackslash Umathunderbarkern}}{1088}
\index{\protect\texttt{\textbackslash Umathpunctspacing}}{1025 \protect\texttt{\textbackslash Umathunderbarrule}}{1089}
\index{\protect\texttt{\textbackslash Umathpunctspacing}}{1026 \protect\texttt{\textbackslash Umathunderbarygap}}{1090}
\index{\protect\texttt{\textbackslash Umathpunctspacing}}{1027 \protect\texttt{\textbackslash Umathunderdelimiterbgap}}{1091}
\index{\protect\texttt{\textbackslash Umathpunctspacing}}{1028 \protect\texttt{\textbackslash Umathunderdelimitervgap}}{1093}
\index{\protect\texttt{\textbackslash Umathpunctspacing}}{1029 \protect\texttt{\textbackslash undefine}}{227, 21392}
\index{\protect\texttt{\textbackslash Umathpunctspacing}}{1030 \protect\texttt{\textbackslash underline}}{480}
\index{\protect\texttt{\textbackslash Umathpunctspacing}}{1031 \protect\texttt{\textbackslash unexpanded}}{575, 2750, 2774}
Index 1538

vbox commands:
\vbox:n \hfill 271, 276, 32480
\vbox_get:Nn \hfill 276, 32494, 33055
\vbox_get:Nw \hfill 276, 32530, 33130
\vbox_get_end: \hfill 276, 32530, 33132
\vbox_get_split_to_ht:Nnn 277, 32569
\vbox_get_to_ht:Nnn \hfill 276, 32518
\vbox_get_to_ht:Nnw \hfill 276, 32551
\vbox_get_top:Nn \hfill 276, 32506
\vbox_set:Nn \hfill 276, 32494, 33049
\vbox_set:Nw \hfill 276, 32530, 33123
\vbox_set_end: \hfill 276, 32530, 33125
\vbox_set_split_to_ht:Nnn \hfill 277, 32569
\vbox_set_to_ht:Nnn \hfill 276, 32518
\vbox_set_to_ht:Nnw \hfill 276, 32551
\vbox_set_top:Nn \hfill ........ 276, 32506, 33069, 33146
\vbox_to_ht:n \hfill 276, 32484
\vbox_to_zero:n \hfill 276, 32484
\vbox_top:n \hfill 276, 32480
\vbox_unpack:N \hfill 277, 32565, 33069, 33146
\vbox_unpack_clear:N \hfill 36868
\vbox_unpack_drop:N \hfill 276, 32565, 33069, 33146
\vbox:n \hfill ........ 271, 276, 32480

vcenter \hfill 493

vcoffin commands:
\vcoffin_get:Nn \hfill 283, 33046
\vcoffin_get:Nw \hfill 283, 33121
\vcoffin_get_end: \hfill 283, 33121
\vcoffin_set:Nn \hfill 283, 33046
\vcoffin_set:Nw \hfill 283, 33121
\vcoffin_set_end: \hfill 283, 33121
\vfi \hfill 1173
\vfil \hfill 494
\vfill \hfill 495
\vfilneg \hfill 496
\v fuzz \hfill 497
\voffset \hfill 498
\vpack \hfill 911
\vrule \hfill 499
\vs csize \hfill 500
\vskip \hfill 501
\v split \hfill 502
\v ss \hfill 503
\vtop \hfill 504

W
\wd \hfill 505
\widowpenalties \hfill 577
\widowpenalty \hfill 506
\write \hfill 89, 507

X
\xdef \hfill 508

xetex commands:
\xetex_if_engine:TF \hfill ........ 36872, 36874, 36876
\xetex_if_engine:p \hfill 36870
\xeTeXcharclass \hfill 715
\xeTeXcharglyph \hfill 716
\xeTeXcountfeatures \hfill 717
\xeTeXcountglyphs \hfill 718
\xeTeXcountselectors \hfill 719
\xeTeXcountvariants \hfill 720
\xeTeXdashbreakstate \hfill 722
\xeTeXdefaultencoding \hfill 721
\xeTeXfeaturecode \hfill 723
\xeTeXfeatureset \hfill 724
\xeTeXfindfeaturebyname \hfill 725
\xeTeXfindselectorbyname \hfill 727
\xeTeXfindvariationbyname \hfill 729
\xeTeXfirstfontchar \hfill 731
\xeTeXfonttype \hfill 732
\xeTeXgenerateactualtext \hfill 733
\xeTeXglyph \hfill 735
\xeTeXglyphbounds \hfill 736
\xeTeXglyphindex \hfill 737
\xeTeXglyphname \hfill 738
\xeTeXinputencoding \hfill 739
\xeTeXinputnormalization \hfill 740
\xeTeXinterchartokenstate \hfill 742
\xeTeXinterchartoks \hfill 744
\xeTeXisdefaultselector \hfill 745
\xeTeXisexclusivefeature \hfill 747
\xeTeXlastfontchar \hfill 749
\xeTeXlinbreaklocale \hfill 751
\xeTeXlinbreakpenalty \hfill 752
\xeTeXlinbreakskip \hfill 750
\xeTeXXOTcountfeatures \hfill 753
\xeTeXXOTcountlanguages \hfill 754
\xeTeXXOTcountscripts \hfill 755
\xeTeXXOTfeaturetag \hfill 756
\xeTeXXOTlanguagesetag \hfill 757
\xeTeXXOTscripttag \hfill 758
\xeTeXXpdfpagecount \hfill 759
\xeTeXXpdfpagetitle \hfill 760
\xeTeXXpicfile \hfill 761
\xeTeXXrevision \hfill 762
\xeTeXXsectionsname \hfill 763
\xeTeXXtextfont \hfill 764
\xeTeXXupwardsmode \hfill 765
\xeTeXXuseglyphmetrics \hfill 766
\xeTeXXvariation \hfill 767
\xeTeXXvariationdefault \hfill 768
\xeTeXXvariationmax \hfill 769
\xeTeXXvariationmin \hfill 770
\xeTeXXversionname \hfill 771
\xeTeXXversion \hfill 772
Index

\kernskip \hspace \kernleaders \hspace \kernspaceskip \hspace \kernspcode \hspace \kernybaselineshift \hspace \kernyear \hspace \kernyoko \hspace \kernY \hspace \kernZ \hspace \kernc_zero

1169 511, 1334, 9009
509 1172
510
1170
1171
36686