Glossary of \TeX terms used to describe \LaTeX3 functions

The \LaTeX Project*

Released 2022-08-30

This file describes aspects of \TeX programming that are relevant in a \LaTeX3 context.

1 Reading a file

Tokenization.

Treatment of spaces, such as the trap that \texttt{\textbackslash --a} is equivalent to \texttt{\textbackslash -a} in expl syntax, or that \texttt{\textasciitilde} fails to give a space at the beginning of a line.

2 Structure of tokens

We refer to the documentation of \texttt{l3token} for a complete description of all \TeX tokens. 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. At any given time two tokens of the same shape automatically have the same meaning, but the converse does not hold, and the meaning associated with a given shape change when doing assignments.

Apart from a few exceptions, 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 Lua\TeX and X\TeX and less for other engines) and category code 13.
- A character token such as \texttt{A} or \texttt{#}, 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).

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 \texttt{\meaning}, together with their \LaTeX3 names and most common example:

1. \texttt{\textbackslash begin:group} character (\texttt{\textbackslash group_begin}, often {),
2. \texttt{\textbackslash end:group} character (\texttt{\textbackslash group_end}, often )}

*E-mail: latex-team@latex-project.org
3 math shift character (\texttt{math\_toggle}, often $),
4 alignment tab character (\texttt{alignment}, often &),
6 macro parameter character (\texttt{parameter}, often #),
7 superscript character (\texttt{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 (active) is discussed below. Input characters can also have several other category codes which do not lead to character tokens for later processing: 0 (escape), 5 (end line), 9 (ignore), 14 (comment), and 15 (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 implicit character tokens. The meaning is otherwise in the following list:

- a macro, used in \LaTeX{} for most functions and some variables (\texttt{tl, fp, seq}, ...),
- a primitive such as \texttt{\textbackslash def} or \texttt{\textbackslash topmark}, used in \LaTeX{} for some functions,
- a register such as \texttt{\count123}, used in \LaTeX{} for the implementation of some variables (int, dim, ...),
- a constant integer such as \texttt{\char"56} or \texttt{\mathchar"121}, used when defining a constant using \texttt{\textbackslash int\_const:Nn},
- a font selection command,
- undefined.

Macros can be \texttt{\textbackslash protected} or not, \texttt{\textbackslash long} or not (the opposite of what \LaTeX{} calls \texttt{nopar}), and \texttt{\textbackslash outer} or not (unused in \LaTeX{}). Their \texttt{\textbackslash meaning} takes the form

\texttt{(prefix) macro:(argument)-->(replacement)}

where (prefix) is among \texttt{\textbackslash protected,long\textbackslash outer}, (argument) describes parameters that the macro expects, such as \texttt{#1#2#3}, and (replacement) describes how the parameters are manipulated, such as \texttt{\textbackslash int\_eval:N\{#2+#1#3\}}. This information can be accessed by \texttt{\textbackslash cs\_prefix\_spec:N, \textbackslash cs\_parameter\_spec:N, \textbackslash cs\_replacement\_spec:N}.

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 (begin-group) and an arbitrary character code, then \TeX{} scans ahead to obtain an equal number of explicit character tokens with category code 1 (begin-group) and 2 (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 “\texttt{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).
3  Quantities and expressions

Integer denotations, dimensions, glue (including fill and true pt and the like).
Syntax of integer expressions (including the trap that -(1+2) is invalid).

4  \LaTeX3 terms

Terms like “intexpr” or “seq var” used in syntax blocks.