The *textalpha* package

Günter Milde

2020/10/30

Abstract

The *textalpha* package enables the use of Greek characters in text independent of font encoding or TeX engine.[1] Input is possible via text commands (`\textalpha` ...`\textOmega`) or Unicode literals[2].

Contents

1 Usage

Load this package in the preamble of your document with

```
\usepackage[<options>]{textalpha}
```

to get a unified interface for Greek letters and symbols that works in all text font encodings, with both, Unicode fonts and the *fontspec* package or 8-bit fonts and the *fontenc* package.

[1] This document was compiled using Unicode fonts (font encoding TU). For a version using 8-bit fonts, see textalpha-doc.pdf.

[2] Requires *greek-inputenc* or XeTeX/LuaTeX.

See the source of this document `textalpha-doc.tex` for a setup and usage example, the literate source of the package textalpha.sty for the implementation, and test-tuenc-greek.pdf for *Font setup for Greek with XeTeX/LuaTeX*.

1.1 Options

1.1.1 `normalize-symbols`

Mathematical notation uses variant shapes of some Greek letters as additional symbols. There are separate code points for the symbol variants in Unicode. TeX supports some of the variant shape symbols in mathematical mode

$$\theta|\vartheta,\ \phi|\varphi,\ \pi|\varpi,\ \rho|\varrho,\ \epsilon|\varepsilon$$

but not in the LGR font encoding used for Greek text in 8-bit TeX.

The variations have no syntactic meaning in Greek text and text fonts may use the variant shapes in place of the "regular" ones as a stylistic choice. However, some Greek texts use these GREEK ... SYMBOL Unicode literals in place of the corresponding GREEK LETTER ... characters.

The `normalize-symbols` option merges letters and symbols to Greek letters. This way, text copied from external sources can be compiled without errors even if it contains GREEK SYMBOL characters in place of GREEK LETTERS:

> The source of this text uses both variants for beta (β|ϐ), theta (θ|ϑ), phi (φ|ϕ), pi (π|ϖ), kappa (κ|ϰ), rho (ρ|ϱ), Theta (Θ|ϴ), and epsilon (ε|ϵ) in the LaTeX source.

This option is ignored with Unicode fonts.

Attention: Do not use this option in cases where the distinction between the symbol variants may be important (e.g. in a mathematical or scientific context). Try the alphabeta package with the respective characters in mathematical mode or use XeTeX/LuaTeX with Unicode fonts in these cases.

1.1.2 `keep-semicolon`

LGR is no standard text font encoding. Latin characters and some other ASCII symbols are mapped to Greek "equivalents" if LGR is the active font encoding. (See usage.pdf for a description of this Latin-Greek transliteration.)

Special care is required with the question mark characters: The LGR font encoding uses the Latin question mark as input for the *erotimatiko* and maps the semicolon to a middle dot (*ano teleia*). As a result, Unicode-encoded texts that use the semicolon as *erotimatiko* end up with an *ano teleia* in its place! Without special care, only the deprecated character 037E GREEK QUESTION MARK[3] works with both, Xe/LuaTeX and 8-bit TeX.

The `\textsemicolon` command inserts an *erotimatiko* in LGR and a semicolon else (i.e. always a character that looks like a semicolon):

> Latin (TU) a; b, Greek (TU) a; b

With the `keep-semicolon` option, character 003B SEMICOLON can be used for the *erotimatiko* also with LGR encoded fonts:

[3] The Unicode standard provides the code point 037E GREEK QUESTION MARK but says character 003B SEMICOLON and not 037E is the preferred character for a 'Greek question mark' (erotimatiko).

Latin (TU)	Greek (TU)	question mark character
Τί φῄς;	Τί φῄς;	037E GREEK QUESTION MARK
Τί φῄς;	Τί φῄς;	003B SEMICOLON
Τί φῄς?	Τί φῄς?	003F QUESTION MARK

This option is ignored with Unicode fonts (where the SEMICOLON literal always prints a semicolon character).

2 Limitations

Because the internal font encoding switch interferes with other work behind the scenes, kerning, diacritics and up/down-casing show problems if Greek letters are used without explicit change of the font encoding.

These problems can be avoided by use of *babel* and the correct language setting (greek), an explicit font encoding switch (e.g. wrapping the Greek text in `\ensuregreek`[4], or XeTeX/LuaTeX with Unicode fonts.

2.1 Diacritics

Accent macros do not work with Unicode literals as base character under 8-bit LaTeX engines. Use the Latin transcription or LICR commands.

Composition of diacritics (like `\accdasia\acctonos` or `\<\'`) fails in other font encodings. Long names (like `\accdasiaoxia`) work.

With LGR and TU, pre-composed glyphs are chosen if available. In other font encodings, accent macros do not select pre-composed characters. (The difference is a sub-optimal placement of the accent and becomes obvious if you drag-and-drop text from the PDF version of this document.):

ᾅ̈ᾅ̈ᾅ̈ᾅ̈ (TU) vs. ἅ̈ (TU).

According to Greek typographical tradition, diacritics (except the dialytika) are placed before capital letters in Titlecase and dropped in UPPERCASE:

ἁ ἐ ΐ ἣ ὅ ὗ ὦ   Ἁ Ἐ Ϊ Ἣ Ὅ Ὗ Ὦ   Α Ε Ϊ Η Ο Υ Ω

This fails for accent macros if the active font encoding does not support Greek. Pre-composed literal Unicode characters are handled correctly:

	LICR	Unicode
TU	Ά	Ά
TU	Ά	Ά

The dialytika marks a *hiatus* (break-up of a diphthong). It must be present in UPPERCASE even where it is redundant in lowercase (the hiatus can also be marked by an accent on the first character of a diphthong). The auto-hiatus feature works in LGR and TU font encodings only:

άυ, έι ↦ ΑΫ, ΕΪ (TU) vs. ΑΫ, ΕΪ (TU)

Currently, the auto-hiatus feature should works only reliably with LICR commands, not Unicode literals:

ἀυπνία ↦ ΑΫΠΝΙΑ (OK)

ἀυπνία ↦ ΑΥΠΝΙΑ (OK with LGR)

ἀυπνία ↦ ΑΥΠΝΙΑ (fail)

[4] The `\ensuregreek` macro ensures the argument is set in a font encoding supporting Greek.

2.2 Kerning

With 8-bit fonts, no kerning occurs between Greek characters in non-Greek text due to the internal font encoding switch: compare AYA (TU) to AYA (TU).

Compiling with LuaTeX provides kerning also on font encoding boundaries.

3 Test and Examples

3.1 Greek alphabet

Greek letters via Latin transcription (works only in LGR font encoding):
  a b g d e z h j i k l m n x o p r sv c t u f q y w
  A B G D E Z H J I K L M N X O P R S T U F Q Y W
Greek letters via default macros in the font for Latin script (here TU):
  α β γ δ ε ζ η θ ι κ λ μ ν ξ ο π ρ σ ς τ υ φ χ ψ ω
  Α Β Γ Δ Ε Ζ Η Θ Ι Κ Λ Μ Ν Ξ Ο Π Ρ Σ Τ Υ Φ Χ Ψ Ω
Archaic Greek letters and Greek punctuation
  ϝϜ ϟϞ ϙϘ ͳͲ ϛ⊘Ϛ · ; ʹ ͵
Diacritics
  Symbol macros:[5] ¨ ´ ` ~ ῾ ᾿ ˘ ¯ ῁ ΅ ῭ ῟ ῝ ῞ ῏ ῎ ῍
  Named macros: ¨ ´ ` ˜ ῾ ᾿ ͅ ι ῁ ΅ ῭ ῟ ῝ ῞ ῏ ῎ ῍ Only in LGR: ⊘ ⊘

Accent macros can start with `\a` instead of `\` when the short form is redefined, e. g. inside a *tabbing* environment. This also works for the new-defined Dasia and Psili shortcuts:

COL1	COL2	COL3	COL4
COL1		COL3	
Viele	Grüße	ἅ	ὢ

3.1.1 Sigma

The lower Sigma comes in two variants: `\textsigma` σ is used inside a word and `\textfinalsigma` ς (or `\textvarsigma` ς) at the end of words.

In LGR, the Latin letter `s` and the command `\textautosigma` print the "normal" sigma if followed by another letter and the final sigma if followed by space or punctuation. This is implemented via the font ligature mechanism and works only in LGR: σσ (TU) and σσ (TU).

The upper case of both sigma variants is `\textSigma`, the lower case of Sigma is `\textautosigma`.

Test: (Unicode literal and `\text...` commands in TU)
σ σ ς ς ς σ lowercase
Σ Σ Σ Σ Σ Σ MakeUppercase
ΣΣ ΣΣ uppercase
σσ σσ MakeLowercase

3.2 Greek Unicode characters in non-Greek text

With the *textalpha* package, greek-inputenc and input encoding `utf8`, Greek Unicode characters can be used in text with any font encoding. See Tables 1 and 2.

[5] Composite diacritics require wrapping in `\ensuregreek`.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
370	◦	◦	◦	◦	ʹ	͵	◦	◦				◦	◦	◦	;	
380					΄	΅	Ά	·	Έ	Ή	Ί		Ό		Ύ	Ώ
390	ΐ	Α	Β	Γ	Δ	Ε	Ζ	Η	Θ	Ι	Κ	Λ	Μ	Ν	Ξ	Ο
3A0	Π	Ρ		Σ	Τ	Υ	Φ	Χ	Ψ	Ω	Ϊ	Ϋ	ά	έ	ή	ί
3B0	ΰ	α	β	γ	δ	ε	ζ	η	θ	ι	κ	λ	μ	ν	ξ	ο
3C0	π	ρ	ς	σ	τ	υ	φ	χ	ψ	ω	ϊ	ϋ	ό	ύ	ώ	
3D0	◦	◦	◦	◦	◦	◦	◦	◦	Ϙ	ϙ	Ϛ	ϛ	Ϝ	ϝ	◦	ϟ
3E0	Ϡ	ϡ	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦
3F0	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦

Table 1: Greek and Coptic Unicode Block, input as literal Unicode characters in TU font encoding (legend: ◦ glyph missing in LGR).

Kerning is preserved if the active font encoding supports Greek: AŸA

Combined Diacritics work ᾷ, diacritics (except diaeresis) are dropped with MakeUppercase (μαΐστρος ↦ ΜΑΪΣΤΡΟΣ).

3.3 PDF strings

With *textalpha* and *greek-inputenc*, there are two options to get Greek letters in PDF strings: LICR macros and literal Unicode input.

3.3.1 λογος, λογος and logos

The subsection title above uses: LICR macros, Unicode input and the LGR transcription for the Greek word logos. Check the table of contents in the PDF viewer: LICR macros and Unicode literals work fine, the Latin transcription remains Latin in the PDF metadata.

3.3.2 α β γ δ ε ζ η θ ι κ λ μ ν ξ ο π ρ σ ς τ υ φ χ ψ ω

3.3.3 Α Β Γ Δ Ε Ζ Η Θ Ι Κ Λ Μ Ν Ξ Ο Π Ρ Σ Τ Υ Φ Χ Ψ Ω

3.3.4 ϝF ϟϟ ϙϘ ϡϠ ςϾς · ; ʹ ,

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
1F00	ἀ	ἁ	ἂ	ἃ	ἄ	ἅ	ἆ	ἇ	Ἀ	Ἁ	Ἂ	Ἃ	Ἄ	Ἅ	Ἆ	Ἇ
1F10	ἐ	ἑ	ἒ	ἓ	ἔ	ἕ			Ἐ	Ἑ	Ἒ	Ἓ	Ἔ	Ἕ		
1F20	ἠ	ἡ	ἢ	ἣ	ἤ	ἥ	ἦ	ἧ	Ἠ	Ἡ	Ἢ	Ἣ	Ἤ	Ἥ	Ἦ	Ἧ
1F30	ἰ	ἱ	ἲ	ἳ	ἴ	ἵ	ἶ	ἷ	Ἰ	Ἱ	Ἲ	Ἳ	Ἴ	Ἵ	Ἶ	Ἷ
1F40	ὀ	ὁ	ὂ	ὃ	ὄ	ὅ			Ὀ	Ὁ	Ὂ	Ὃ	Ὄ	Ὅ		
1F50	ὐ	ὑ	ὒ	ὓ	ὔ	ὕ	ὖ	ὗ		Ὑ		Ὓ		Ὕ		Ὗ
1F60	ὠ	ὡ	ὢ	ὣ	ὤ	ὥ	ὦ	ὧ	Ὠ	Ὡ	Ὢ	Ὣ	Ὤ	Ὥ	Ὦ	Ὧ
1F70	ὰ	ά	ὲ	έ	ὴ	ή	ὶ	ί	ὸ	ό	ὺ	ύ	ὼ	ώ		
1F80	ᾀ	ᾁ	ᾂ	ᾃ	ᾄ	ᾅ	ᾆ	ᾇ	ᾈ	ᾉ	ᾊ	ᾋ	ᾌ	ᾍ	ᾎ	ᾏ
1F90	ᾐ	ᾑ	ᾒ	ᾓ	ᾔ	ᾕ	ᾖ	ᾗ	ᾘ	ᾙ	ᾚ	ᾛ	ᾜ	ᾝ	ᾞ	ᾟ
1FA0	ᾠ	ᾡ	ᾢ	ᾣ	ᾤ	ᾥ	ᾦ	ᾧ	ᾨ	ᾩ	ᾪ	ᾫ	ᾬ	ᾭ	ᾮ	ᾯ
1FB0	ᾰ	ᾱ	ᾲ	ᾳ	ᾴ		ᾶ	ᾷ	Ᾰ	Ᾱ	Ὰ	Ά	ᾼ	᾽	ι	᾿
1FC0	῀	῁	ῂ	ῃ	ῄ		ῆ	ῇ	Ὲ	Έ	Ὴ	Ή	ῌ	῍	῎	῏
1FD0	ῐ	ῑ	ῒ	ΐ			ῖ	ῗ	Ῐ	Ῑ	Ὶ	Ί		῝	῞	῟
1FE0	ῠ	ῡ	ῢ	ΰ	ῤ	ῥ	ῦ	ῧ	Ῠ	Ῡ	Ὺ	Ύ	Ῥ	῭	΅	`
1FF0			ῲ	ῳ	ῴ		ῶ	ῷ	Ὸ	Ό	Ὼ	Ώ	ῼ	´	῾	

Table 2: Greek Extended Unicode Block, input as literal Unicode characters in TU font encoding.