The *textalpha* package

Günter Milde

2020/10/30

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}
```
Now you are ready to use literal Unicode characters** or the `\textalpha` `...\textOmega` macros anywhere in the text.[3] See the source of this document `textalpha-doc.tex` for a setup and usage example and greek-fontenc-doc for links to additional documentation.

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.

---

[1] This document was compiled using 8-bit fonts (font encoding T1). For a version using Unicode fonts, see textalpha-tu.pdf.

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

[3] Using the shorter `\alpha ...\Omega` macros (known from math mode) is possible with the *alphabeta* package.

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 a GREEK SYMBOL ... in place of a GREEK LETTER ...

> The source of this paragraph 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).

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). Use the respective characters in mathematical mode or XeTeX/LuaTeX with Unicode fonts.

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 babel-greek 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[4] 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 (T1) a; b, Greek (LGR) α; β

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

Latin (T1)	Greek (LGR)	question mark character
;	;	037E GREEK QUESTION MARK
;	;	003B SEMICOLON
?	;	003F QUESTION MARK

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

Test whether this works as expected in math mode: $ab; a\, b, (a;\alpha;2)$.

---

[4]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).

1.2 Symbol macros for Breathings

textalpha defines the macros `\<` and `\>` for the dasia (rough breathing) and psili (smooth breathing) diacritics.

2 Limitations

If Greek letters are used while the active font encoding does not support Greek, the internal font encoding switches interfere with other work behind the scenes. Kerning, diacritics and up/down-casing show problems that can be avoided by

- use of *babel* and the correct language setting,
- an explicit font encoding switch, e.g., wrapping in `\ensuregreek`[5], or
- XeTeX/LuaTeX with Unicode fonts.

2.1 Kerning

With pdfTeX and 8-bit fonts, no kerning occurs between Greek characters in non-Greek text due to the internal font encoding switch:

AΥA (T1) vs. AΥA (LGR).

Compiling with LuaTeX provides kerning also on font encoding boundaries.

2.2 Diacritics

With 8-bit TeX, accent macros do not work with Unicode literals as base character. Use the Latin transliteration or LICR commands.

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

�α vs. ἄ (LGR)

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.:

ᾰ (T1) vs. ᾰ (LGR).

In Greek typographical practice, diacritics (except the dialytika and sub-iota) are placed before capital letters in Titlecase (Ἀρχιμήδης) and dropped in uppercase (ΑΡΧΙΜΗΔΗΣ).

Diacritics input via standard accent macros are misplaced if the active font encoding does not support Greek.

---

[5]The `\ensuregreek` macro ensures the argument is set in a font encoding supporting Greek without adverse side-effects if the active font encoding is already LGR or TU.

With the `\MakeUppercase` implementation introduced 2022/06, Greek upcasing rules are only applied to literal characters if the text language is set to Greek with Babel and to standard accent macros if the documents loads Greek with Babel (i.e. not in this document).[6]

	named accent	standard accent	literal
LGR	´Α → Α	´Α → ´Α	´Α → ´Α
T1	Á → A	Á → Á	´Α → ´Α

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 or breathing on the first of two consecutive vowels). The auto-hiatus feature works in LGR and TU font encodings only:

άυ, έι → ΑΥ, ΕΙ (T1) vs. άυ, έι → ΑΫ, ΕΪ (LGR)

With the old implementation of `\MakeUppercase`, the auto-hiatus feature works with LICR macros but not Unicode literals. The new implementation works with Unicode literals, too, but only if the text language is Greek (i.e. not in this document).

ἀυπνία ↦ ΑΫΠΝΙΑ (LICR macros: OK with LGR or TU)

ἀυπνία ↦ ἈΥΠΝΊΑ (literal characters: fails without Babel)

3 Test and Examples

3.1 Greek alphabet

Greek literal characters in Latin text (font encoding T1):

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

Greek letters via default macros in Latin text (font encoding T1):

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

Greek letters via Latin transliteration (works only in LGR font encoding):

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 ϡϠ ϛϚϹ · ; ´ ‚

Diacritics

---

[6]With the pre-2022 `\MakeUppercase` implementation, the above rules were fully applied if the active font encoding is LGR or TU.

Short macros:[7] ¨ ´ ` ~ ῾ ᾿ ˘ ¯ ῀ ᾳ ᾀ ῂ ῄ ῆ ῇ ῞ ῍

Named macros: ¨ ´ ` ~ ῾ ᾿ , ͺ ῀ ᾳ ᾀ ῂ ῄ ῆ ῇ ῞ ῍

Only in LGR: ˷ ˬ

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:

col 1   col 2   col 3   col 4
Viele   Grüße   ἀ   ὠ

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 in LGR[8]:

σς (LGR) vs. ςς (T1).

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

Test Unicode literal and `\text...` commands:

no change:	σσ ςςςς ΣΣ ΣΣ
MakeUppercase:	ΣΣ ΣΣΣΣ ΣΣ ΣΣ
MakeLowercase (T1):	σσ ςςςς σς ςς
MakeLowercase (LGR):	σσ ςςςς σς σς

3.2 Greek literal characters in non-Greek text

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

Kerning is preserved if the active font encoding supports Greek. This can be secured by wrapping the Greek text part in `\ensuregreek` or setting the text language with Babel: AΫA

Combined Diacritics work for pre-composed characters: ᾱ̆.

Diacritics (except diaeresis) are dropped with MakeUppercase with LaTeX versions older than 2022/06 For other versions, set the language of to-be-upcased Greek text with Babel: μαΐστρος, δύο ↦ ΜΑΪΣΤΡΟΣ, ΔΎΟ.

---

[7]Composite diacritics require wrapping in `\ensuregreek`.
[8]TODO: Fix `\textautosigma` with Unicode fonts.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
370	◦	◦	◦	◦	ʹ	͵	◦	◦			ͺ	◦	◦	◦	;	
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	ϟ
3E0	Ϡ	ϡ	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦
3F0	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦	◦

Table 1: Greek and Coptic Unicode Block, input as literal Unicode characters in T1 font encoding (legend: ◦ glyph missing in LGR).

	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 T1 font encoding.

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 λογος

The subsection title above uses: LICR macros, Unicode input and the LGR transliteration for the Greek word λογος.

LICR macros and Unicode literals work fine everywhere, the Latin transliteration remains Latin in the PDF metadata (sidebar table of contents in the PDF viewer) and with Xe/LuaTeX.

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 ϝϜ ϟϞ ϙϘ ϡϠ ϛϚϙ · ; ´ ,

.