The *textalpha* package

Günter Milde

September 4, 2015

Abstract

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

Contents

1 Usage

Load this package in the preamble of your document with

```
\usepackage{textalpha}
```

eventually with options `normalize-symbols` and/or `keep-semicolon` .

If *textalpha* is loaded after the setup of Unicode fonts with the *fontspec* package under LuaTeX and XeTeX, it provides a compatible interface for Greek in text mode.

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 greek-euenc-doc.pdf for *Font setup for Greek with XeTeX/LuaTeX*.

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

1.1 option `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 but not in the LGR font encoding used for Greek in 8-bit TeX. TeX supports some of the variant shape symbols in mathematical mode ($\theta|\vartheta, \phi|\varphi, \pi|\varpi, \rho|\varrho, \epsilon|\varepsilon$).

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 in Unicode encoding use these GREEK ... SYMBOL characters 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:

> 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.
>
> Τηις τεξτ υσες βοτη αριαντς φορ βετα (β̦|ϐ), τηετα (ϑ̦|ϑ), πηι (φ̦|φ), πι (π̦|π), καππα (κ̦|ϰ), ρηο (ρ̦|ϱ), Τηετα (Θ̦|ϴ), ανδ επσιλον (ε̦|ε) ιν τηε ΛαΤεΞ σουρςε.

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 math mode or use XeTeX/LuaTeX with Unicode fonts in these cases.

This option is ignored with Unicode fonts.

1.2 `keep-semicolon`

LGR is no "standard 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*).
- 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),

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 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
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, too.

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`[2], or XeTeX/LuaTeX with Unicode fonts.

2.1 Diacritics

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

With LGR, 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.):

ἅ ἅ ἅ ἅ (LGR) vs. ἅ (T1).

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

ἁ ἐ ἵ ἣ ὅ ὗ ὧ   Ἁ Ἐ Ϊ Ἣ Ὅ Ὗ Ὧ   Α Ε Ϊ Η Ο Υ Ω

This fails for accent macros if the active font encoding is not LGR or Unicode. Pre-composed literal Unicode characters are handled correctly:

Ά/Ά (LICR/Unicode in LGR) vs. Á/Ά (LICR/Unicode in T1).

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 Unicode font encodings only:

άυ, έι ↦ ΑΫ, ΕΪ (LGR) vs. ΑΥ, ΕΙ (T1)

Currently, the second vowel of the diphthong must be given as macro, not Unicode literal if the auto-hiatus feature should work:

ἀυπνία ↦ ΑΫΠΝΙΑ (LGR) vs. ΑΥΠΝΙΑ (T1).

2.2 Kerning

No kerning occurs between Greek characters in non-Greek text due to the internal font encoding switch: compare ΑΥΑ (LGR) to ΑΥΑ(T1).

Compiling with LuaTeX provides kerning also over font encoding boundaries.

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

	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	A	B	Γ	Δ	E	Z	H	Θ	I	K	Λ	M	N	Ξ	O
3A0	Π	P		Σ	T	Υ	Φ	X	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	F	ϝ	*	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).

3 Test and Examples

3.1 Greek alphabet

Greek letters via Latin transcription in LGR font encoding:

A B Γ Δ E Z H Θ I K Λ M N Ξ O Π P Σ T Υ Φ X Ψ Ω

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

Greek letters via default macros in other font encoding (here T1):

A B Γ Δ E Z H Θ I K Λ M N Ξ O Π P Σ T Υ Φ X Ψ Ω

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

Archaic Greek letters and Greek punctuation

F·ϘϺϠ;

ϜϟϘϡϞ

Diacritics

Symbol macros:[3] ¨ ´ ` ~ ˚ ˇ ˘ ¯ ˝ ˛ ˙ ˜ ˚ ˇ ˘ ˝ ˛

Named macros: ¨ ´ ` ~ ˚ ˇ ˘ ¯ ˝ ˛ ˙ ˜ ˚ ˇ ˘ ˝ ˛

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	ἁ	ὠ

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.

Kerning is preserved if the font encoding is LGR: AΫA

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

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

	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 transcription for the Greek word λογος. 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.