Font setup for Greek with XeTeX/LuaTeX

Günter Milde

2020/10/30

The file tuenc-greek.def provides support for Greek LICR macros and upcasing of text with XeTeX and LuaTeX. It is loaded automatically by *textalpha*, *alphabeta*, and *babel-greek* when used with Unicode fonts (LuaTeX or XeTeX with *fontspec*).

Contents

LaTeX sets up the TU Unicode text font encoding if it detects the XeTeX or LuaTeX engines. The user must ensure that the selected font contains Greek glyphs (the default Latin Modern fonts have only capital Greek letters). **There are no errors for missing glyphs**, just warnings in the log file (but not in the console output) and empty spaces in the output document.

The *fontspec* package is the standard tool to select fonts in XeTeX/LuaTeX. *babel* provides a front-end to set up language-specific fonts with the `\babelfont` command.

`tuenc-greek.def` is loaded automatically by *textalpha*, *alphabeta*, or *Babel* (with the language option greek) if these packages are used with Unicode-aware TeX engines (XeTeX or LuaTeX). This provides support for Greek LICR input and fixes for MakeUppercase (cf. section **??**).

The LaTeX internal character representation (LICR) can be used as a verbose, fail-safe 7-bit ASCII encoding. It works under both, 8-bit TeX and XeTeX/LuaTeX. Use cases are macro definitions and generated text.

See the source of this document, `test-tuenc-greek.tex` for the input used in the examples below.

3.1 Greek alphabet

Greek letters via LICR macros:

Α Β Γ Δ Ε Ζ Η Θ Ι Κ Λ Μ Ν Ξ Ο Π Ρ 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

The small sigma is set with a different glyph if it ends a word:

σ `\textsigma`
ς `\textfinalsigma` or `\textvarsigma`

The `\textautosigma` currently works only with 8-bit LGR fonts: σι vs. ισ.

3.2 Diacritics

Greek diacritics can be input by named macro or symbol macro:

ά́ ή́ ὰ́ ὰ ῒ̈ ῒ ῗ ῗ ᾶ̓ ᾶ ἄ ἀ ἄ́ ἄ

XeTeX and LuaTeX with the Harfbuzz renderer normalise combining diacritics with the base letter to the corresponding pre-composed Unicode character (if such a character exists).[1]

ά ά ὰ ὰ ΐ ΐ ᾶ ᾶ ἀ ἀ ἄ ἄ

3.2.1 perispomeni vs. tilde

The Greek *perispomeni* has the look of a tilde but the semantic of a circumflex accent. The Unicode standard provides distinct code points. Only <Greek letter> + COMBINING GREEK PERISPOMENI is considered equivalent to the pre-composed character <Greek letter> WITH PERISPOMENI.

The `\accperispomeni` macro uses COMBINING GREEK PERISPOMENI, while the standard tilde-accent macro `\~` uses the COMBINING TILDE. Composite definitions for `\~` select the pre-composed character:

accent + base char / literal: ᾶ/ᾶ, ῆ/ῆ, ῖ/ῖ, ῦ/ῦ, ῶ/ῶ

There are no pre-composed capital letters with perispomeni: Ἆ͂, Ἦ͂, Ἶ͂, Ὗ͂, Ὦ͂ **??**

3.2.2 composite diacritics

Composite diacritics are defined using two combining diacritical characters.**??**

῍ ῎ ῏ ῝ ῞ ῟ ῭ ΅ ` ´ ῁ ῍ ῎ ῏

῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾

῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾ ῾

[1] Composite diacritics may overlap when they are not normalised to a pre-composed character. However, this is no problem in normal use as there are pre-composed Unicode characters for all letters/diacritic combinations that are actually used in the Greek language.

3.2.3 sub-iota

The sub-iota is input after the base letter.

- `\ypogegrammeni` sets a COMBINING GREEK YPOGEGRAMMENI: ᾳ ᾗ.

  A Greek capital letter followed by COMBINING GREEK YPOGEGRAMMENI is normalised to the corresponding Greek capital letter WITH [... AND] PROSGEGRAMMENI, if a mapping exists in the Unicode standard.

- `\prosgegrammeni` sets a spacing GREEK PROSGEGRAMMENI: A᾽ K᾽.

  Spacing is better with the pre-composed characters for Greek capital letters … WITH PROSGEGRAMMENI.

  Compare Aι (small letter iota) vs. A᾽ (spacing prosgegrammeni) vs. Aᾼ (pre-composed).

Test letters with ypogegrammeni and prosgegrammeni (literal/LICR):

unchanged	make lowercase	make uppercase.
ᾳα᾽/ᾳα᾽	ᾳα᾽/ᾳᾳ	AIAI/AᾼA᾽
AᾼA᾽/AᾼA᾽	ᾳα᾽/ᾳᾳ	AIAI/AᾼA᾽
ᾋΛ᾽/ᾋΛ᾽	ᾗλ᾽/ᾗᾗ	ΛIΛI/ᾋΛ᾽

3.3 Additional Greek symbols

3.3.1 symbols for Greek numbers

ϗ textkoppa
Ϟ textKoppa
ϙ textqoppa (archaic koppa)
Ϙ textQoppa (archaic Koppa)
ϛ textstigma
Ϛ textStigma (Sigma-Tau-Ligature in CB-fonts)[2]
ϡ textsampi
Ϡ textSampi
ϝ textdigamma
F textDigamma
ʹ textdexiakeraia
͵ textaristerikeraia

3.3.2 symbol variants

Mathematical notation uses variant shapes of some Greek letters as additional symbols. 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.

[2] the name “stigma” originally applied to a medieval sigma-tau ligature, whose shape was confusingly similar to the cursive digamma

TeX math		Unicode	
symbol	var symbol	"letter"	"symbol"
π	ϖ	π	ϖ
ρ	ϱ	ρ	ϱ
θ	ϑ	θ	ϑ
ϵ	ε	ε	ϵ
ϕ	φ	φ	ϕ
β	*missing*	β	ϐ
κ	*missing*	κ	ϰ
Θ	*missing*	Θ	ϴ

Table 1: Greek symbol variants in TeX and Unicode

Unicode defines separate code points for the symbol variants. TeX supports some of the variant shape symbols in mathematical mode, but its concept of "standard" vs. "variant" symbols differs from the distinction between "GREEK LETTER ..." vs. "GREEK ... SYMBOL" in the Unicode standard (see Table **??**).

`tuenc-greek.def` defines three TextCommands for each of these letters:

`\text<name>` selects the Unicode GREEK LETTER ... variant,

`\text<name>symbol` selects the Unicode GREEK ... SYMBOL variant,

`\textvar<name>` selects the variant shape according to TeX's mathematical mode

See Table **??** for the full list. The *alphabeta* package defines short macros that work in text and math mode.

3.3.3 Ancient Greek Numbers

Ancient Greek Numbers are missing in most fonts (including Libertine and Deja Vu). The "FreeSerif" font works fine:

𐅁𐅂𐅃𐅄

If the LGR font encoding is loaded via «fontenc» in the document preamble, Ancient Greek Numbers (as well as any other character) from LGR encoded 8-bit TeX fonts can be used after a font-encoding switch. (This document defines the `\lgrfont` command for this purpose in the preamble.)

𐅁𐅂𐅃𐅄

3.3.4 generic text symbols

There are some LICR macros for symbols from the 8-bit font encoding LGR that are not confined to Greek but missing in `tuenc.def` [2021/04/29 v2.0v] in TeXLive 23.

text		mathematics	
macro	output	macro	output
`\textbeta`	ϐ	`\beta`	β
`\textvarbeta`	ß	*missing*	
`\textbetasymbol`	ß		
`\textepsilon`	ε	`\epsilon`	ϵ
`\textvarepsilon`	ε	`\varepsilon`	ε
`\textepsilonsymbol`	ϵ		
`\texttheta`	θ	`\theta`	θ
`\textvartheta`	ϑ	`\vartheta`	ϑ
`\textthetasymbol`	ϑ		
`\textTheta`	Θ	`\Theta`	Θ
`\textvarTheta`	Θ	*missing*	
`\textThetasymbol`	Θ		
`\textkappa`	κ	`\kappa`	κ
`\textvarkappa`	ϰ	`\varkappa`	$\varkappa$
`\textkappasymbol`	ϰ		
`\textpi`	π	`\pi`	π
`\textvarpi`	ϖ	`\varpi`	ϖ
`\textpisymbol`	ϖ		
`\textrho`	ρ	`\rho`	ρ
`\textvarrho`	ϱ	`\varrho`	ϱ
`\textrhosymbol`	ϱ		
`\textphi`	φ	`\phi`	ϕ
`\textvarphi`	φ	`\varphi`	φ
`\textphisymbol`	ϕ		

Table 2: Macros for Greek symbol variants

; textsemicolon
µ textmicro
ə textschwa

The SI unit prefix MICRO SIGN is not upcased with MakeUppercase:

textmu: µ ↦ M but textmicro: µ ↦ µ.

4 Latin transliteration

The Latin transliteration known from LGR encoded 8-bit fonts[3] does not work with Unicode fonts.

For LuaTeX, there is a `transliteration.omega` *transform* that applies the transliteration system devised by Yannis Haralambous for the Omega system (cf. the Babel documentation for the Greek locale).

It is possible to set up LGR encoded fonts parallel to Unicode fonts (see the preamble of the source file `test-tuenc-greek.tex` for an example). After switching the font encoding to LGR, Greek text can be input via a *Latin transliteration*, e.g. «logos» becomes «λογος» and «`>aupn'ia`» becomes «ἀυπνία».

Mark that you cannot use Unicode input with LGR encoded fonts except when running in 8-bit compatibility mode. LICR macros work in both, TU and LGR: compare Ἰανουαρίου (TU) vs. Ἰανουαρίου (LGR).

5 UPPERCASE and lowercase

According to Greek typesetting conventions, Greek diacritics (except the dialytika and sub-iota) are placed to the left (instead of above) capital letters and dropped if text is set in UPPERCASE, e.g. μαΐστρος ↦ ΜΑΪΣΤΡΟΣ.

The 2022 `\MakeUppercase` implementation (cf. ltnews35.pdf) relies on Unicode data. It upcases literal characters according to the Greek conventions **if the text language is set to Greek** with *babel* or *polyglossia*. For the pre-2022/06 implementation, `tuenc-greek.def` contains uccode/lccode corrections (taken from Apostolos Syropoulos *xgreek* package) to get the same effect (independent of the text language).

In addition, `tuenc-greek.def` contains code to drop Greek diacritics input as accent macros. However, when using the symbol macros (`\' \` \~`) for tonos, varia, and perispomeni, special definitions from *babel-greek* ≥ 1.13.3 are required to distinguish them from acute, grave, and tilde accents that must be kept on Latin letters.

Named Greek accents and symbol accents on Greek vs. Latin letter:

ύύ úú ὺὺ ùù→ ΥΥ UÚ ΥΥ UÙ

Standard symbol accents on Latin letters are kept (OK). With pre-2023 LaTeX versions or babel-greek < 1.13.3, standard accents are also kept on Greek letters (sic!).

[3] See the *teubner* package or *babel-greek* for a description.

5.1 hiatus

Tonos and psili mark a *hiatus* (break-up of a diphthong) if placed on the first of two consecutive vowels (άι, άυ, έι, ὰυ, ἄι, ἄυ, ἔι). A dialytika must be placed on the second vowel if they are dropped:

άυλος ↦ ΑΫΛΟΣ, ἄυλος ↦ ΑΫΛΟΣ,

mάινα ↦ ΜΑΪΝΑ, κέικ↦ ΚΕΪΚ, ὰυπνία ↦ ΑΫΠΝΊΑ. νεράιδα → ΝΕΡΆΙΔΑ

Since 2023/02/10, this works with all literals

άι, άυ, έι, ἄι, ἄυ, ἔι, ὰυ ↦ ΑΪ, ΑΫ, ΕΪ, ΑΪ, ΑΫ, ΕΪ, ΑΫ,

with named accent-macro + Unicode literals:

άι, άυ, έι, ἄι, ἄυ, ἔι, ὰυ ↦ ΑΪ, ΑΫ, ΕΪ, ΑΪ, ΑΫ, ΕΪ, ΑΫ

and, with babel-greek 1.13.3 with short accent-macro + Unicode literals:

άι, άυ, έι, ἄι, ἄυ, ἔι, ὰυ ↦ ΑΪ, ΑΫ, ΕΪ, ΑΪ, ΑΫ, ΕΪ, ΑΫ

6 Character Tables

The file `char-list.tex` contains a listing of all Greek Unicode characters supported by the *greek-fontenc* package. See `char-list-tu.tex` for the output with XeTeX/LuaTeX.

7 TODO

Compilation error with MakeUppercase and combining ypogegrammeni in Greek locale.