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

1 Requirements

1.1 *fontspec* and suitable Unicode fonts

LaTeX sets up the TU Unicode text font encoding if it detects the XeTeX or LuaTeX engiges. 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 Xe/LuaLaTeX. Examples for suitable fonts are given in the greek-fontenc documentation.

2 Usage

`tuenc-greek.def` is usually not loaded directly, but by one of *textalpha*, *alphabeta*, or *Babel* (with the language option greek). If these packages are used with Unicode-aware TeX engines (XeTeX or LuaTeX), Unicode font setup is amended for use of the Greek script.

3 LICR input

The LaTeX internal character representation (LICR) is a verbose, fail-safe 7-bit ASCII encoding that can be used unaltered 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` macro, which automatically chooses the glyph according to the position, is not implemented for Unicode fonts.

3.2 Diacritics

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

ά́ά́ ᾽́᾽́ ὰ́ὰ́ ῝῝ ΐ̈ΐ̈ ῾̈῾̈ ᾶ̃ᾶ̃ ῾̃῾̃ ἄἄ ἂἂ ἀ̃ἀ̃ ἀ̓́ἀ̓́

XeTeX normalizes base letter and combining diacritics to the corresponding pre-composed character if such a mapping is defined in the Unicode standard.

ά ά ὰ ὰ ΐ ΐ ᾶ ᾶ ἀ ἀ ἀ ἀ

3.2.1 perispomeni vs. tilde

The Greek *perispomeni* has the look of a tilde but the semantic of a circumflex accent. The "named" `\accperispomeni` macro uses COMBINING GREEK PERISPOMENI, while the standard tilde-accent macro `\~` uses the COMBINING TILDE which is not normalized to GREEK LETTER ... WITH PERISPOMENI characters.

Composite definitions for `\~` select the pre-composed character:

α̃ = ᾶ, η̃ = ῆ, ι̃ = ῖ, ῦ = ῦ, ω̃ = ῶ

3.2.2 combined diacritics

Combined accents are defined using combining diacritical characters.

ἶ ἷ ἵ ᾰ̓́ ᾰ̓̀ ᾰ̓͂ ἲ ἳ ᾰ̔́ ᾰ̔̀ ᾰ̔͂ ἴ ἵ ᾰ̀ ᾰ́

ἷ ἷ ἷ ᾰ̓́ ᾰ̓̀ ἷ ἷ ᾰ̔́ ᾰ̔̀ ἷ ἷ ᾰ̀ ᾰ́

ἷ ἷ ἷ ᾰ̓́ ᾰ̓̀ ἷ ἷ ᾰ̔́ ᾰ̔̀ ἷ ἷ ᾰ̀ ᾰ́

Composite diacritics overlap when they are not normalized to a pre-composed character. However, this is not a major problem in normal use as pre-composed characters exist in Unicode for all letters that are used with diacritics in (ancient, polytonic or monotonic) Greek.

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 normalized to the corresponding Greek capital letter WITH [... AND] PROSGEGRAMMENI, if a mapping exists in the Unicode standard (by XeTeX but not by LuaTeX)

- `\prosgegrammeni` sets a spacing GREEK PROSGEGRAMMENI: Aι Kι.

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

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

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

unchanged	make lowercase	make uppercase.
ᾳαι/ᾳαι	ᾳαι/ᾳᾳ	AιAI/AιAι
AιAι/AιAι	ᾳαι/ᾳᾳ	AιAI/AιAι
ᾋΛι/ᾋΛι	ᾳλι/ᾳᾳ	ᾋΛI/ᾋΛι

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

3.3 Additional Greek symbols

3.3.1 symbols for Greek numbers

ϟ textkoppa
Ϟ textKoppa
ϙ textqoppa (archaic koppa)
Ϙ textQoppa (archaic Koppa)
ϛ textstigma
Ϛ textStigma (Sigma-Tau-Ligature in CB-fonts)[1]
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.

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 1).

`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' mathematical mode

See Table 2 for the full list. The *alphabeta* package defines short macros that work in text and math mode.

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

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

Table 2: Macros for Greek symbol variants

3.3.3 Ancient Greek Numbers

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

𐅀𐅁𐅂𐅃

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. babel-greek defines the `\textgreek` command for this purpose.

𝛢𝛨𝛸𝛭

3.3.4 generic text symbols

There are some LICR macros for some symbols from the 8-bit font encoding LGR that are not confined to Greek but not defined in `tuenc.def` [2018/08/11 v2.0j].

; 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 transcription

The Latin transcription known from LGR encoded 8-bit fonts[2] does not work with Unicode fonts.

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). The `\textgreek` macro can then be used for the input of Greek letters via the *Latin transcription*, 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, Unicode font encoding and LGR: compare Ἰανουαρίου (Unicode font set up via fontspec) vs. Ἰανουαρίου (LGR-encoded 8-bit font set up via NFSS commands).

5 UPPERCASE and lowercase

Capital Greek letters have Greek diacritics (except the dialytika and sub-iota) to the left (instead of above) and drop them if text is set in UPPERCASE, e.g. μαΐστρος ↦ ΜΑΪΣΤΡΟΣ.

The uccode/lccode corrections (taken from Apostolos Syropoulos xgreek package) ensure dropping of accents with `\MakeUppercase` for literal Unicode characters.

@uclclist additions ensure that upcasing also drops Greek diacritics. However, when the tonos, varia, and perispomeni accents are input using the symbol macros (`\' \` \~`), this does not work, as they cannot be distinguished from Latin acute,

[2] See the teubner package or the file usage.pdf from the *babel-greek* package for a description.

grave, and tilde accents.[3] If these accents should be dropped by MakeUppercase, they must be input as named macro:

ά\'α á\'a à\`α à\`a ↦ A\'A A\'A A\`A A\`A

5.0.1 hiatus

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

The «hiatus» feature works with macro input:

άυλος ↦ AΫΛΟΣ, ἄυλος ↦ AΫΛΟΣ,

μάινα ↦ MAΪNA, κέικ↦ KEΪK, ἀυπνία ↦ AΫΠΝΙA.

It does not work with Unicode literals:

άι, άυ, έι, ἄι, ἄυ, ἔι ↦ AI, AY, EI, AI, AY, EI

or accent-macro + Unicode literals (yet?):

άι, άυ, έι, ά̓ι, ά̓υ, ἕι ↦ AI, AY, EI, AI, AY, EI

6 Character Tables

The following tables list the Greek Unicode characters. In the input, the LICR macro is followed by the corresponding literal Unicode character.

6.1 Greek and Coptic Unicode block

Seldom used characters that are not part of LGR encoded TeX fonts have no LICR definition.

ʹʹ
ʹ ʹ ʺ ʺ ΆΑ ·· ΈΕ ΉΗ ΊΙ ΌΟ ΎΥ ΏΩ
ΐΐ AA BB ΓΓ ΔΔ EE ZZ HH ΘΘ II KK ΛΛ MM NN ΞΞ OO
ΠΠ PP ΣΣ TT YY ΦΦ XX ΨΨ ΩΩ ΪΪ ΫΫ άά έέ ήή ίί
ΰΰ αα ββ γγ δδ εε ζζ ηη θθ ιι κκ λλ μμ νν ξξ οο
ππ ρρ ςς σσ ττ υυ φφ χχ ψψ ωω ϊϊ ϋϋ όό ύύ ώώ
ϐϐ ϑϑ ϒ ϒ Ϋ φϕ ϖϖ
ϘϘ ϙϙ ϚϚ ϛϛ FF ﬀ ﬁﬁ ﬂﬂ
ϞϞ ϟϟ
ϰϰ ϱϱ ϴϴ ϵϵ

MakeUppercase (note: standard accents not dropped):

ʹʹ
ʹ ʹ ʺ ʺ ΑΑ ·· ΕΕ ΗΗ ΙΙ ΟΟ ΥΥ ΩΩ
ΪΪ AA BB ΓΓ ΔΔ EE ZZ HH ΘΘ II KK ΛΛ MM NN ΞΞ OO
ΠΠ PP ΣΣ TT YY ΦΦ XX ΨΨ ΩΩ ΪΪ ΫΫ ΑΑ ΕΕ ΗΗ ΙΙ
ΫΫ AA BB ΓΓ ΔΔ EE ZZ HH ΘΘ II KK ΛΛ MM NN ΞΞ OO
ΠΠ PP ΣΣ ΣΣ TT YY ΦΦ XX ΨΨ ΩΩ ΪΪ ΫΫ ΟΟ ΥΥ ΩΩ

[3] This might be fixed with \accACUTE, \accGRAVE, and \accTILDE definitions with corresponding @uclclist entries and composite definitions.

ϐB ϑΘ Y Y Ÿ ϕΦ ϖΠ
ϘϘ ϘϘ ϚϚ ϚϚ FF FF ϟϟ ϟϟ
ϡϡ ϡϡ
ϰK ϱP ΘΘ ϵE
MakeLowercase:
〃
’ ’ “ “ ;;
ά ·· έέ ήά ίί όό ύύ ώώ
ΐΐ αα ββ γγ δδ εε ζζ ηη θθ ιι κκ λλ μμ νν ξξ οο
ππ ρρ σσ ττ υυ φφ χχ ψψ ωω ϊϊ ϋϋ άά έέ ήή ίί
ΰΰ αα ββ γγ δδ εε ζζ ηη θθ ιι κκ λλ μμ νν ξξ οο
ππ ρρ ϛϛ σσ ττ υυ φφ χχ ψψ ωω ϊϊ ϋϋ όό ύύ ώώ
ϐϐ ϑϑ υ ύ ϋ φφ ϖπ
ϙϙ ϙϙ ϛϛ ϛϛ ϝϝ ϝϝ ϟϟ ϟϟ
ϡϡ ϡϡ
ϰκ ϱρ Θθ ϵϵ

The lowercase of Σ is σ (GREEK SMALL LETTER SIGMA).[4] The lowercase of Ϛ (GREEK LETTER STIGMA) is ϛ (GREEK SMALL LETTER STIGMA).

6.2 Greek Extended Unicode block

Note: There are no LICR definitions for spacing diacritical characters. Use the corresponding accent macro with an empty argument or a space.

ἀἀ ἁἁ ἂἂ ἃἃ ἄἄ ἅἅ ἆἆ ἇἇ ἈἈ ἉἉ ἊἊ ἋἋ ἌἌ ἍἍ ἎἎ ἏἏ
ἐἐ ἑἑ ἒἒ ἓἓ ἔἔ ἕἕ ἘἘ ἙἙ ἚἚ ἛἛ ἜἜ ἝἝ
ἠἠ ἡἡ ἢἢ ἣἣ ἤἤ ἥἥ ἦἦ ἧἧ ἨἨ ἩἩ ἪἪ ἫἫ ἬἬ ἭἭ ἮἮ ἯἯ
ἰἰ ἱἱ ἲἲ ἳἳ ἴἴ ἵἵ ἶἶ ἷἷ ἸἸ ἹἹ ἺἺ ἻἻ ἼἼ ἽἽ ἾἾ ἿἿ
ὀὀ ὁὁ ὂὂ ὃὃ ὄὄ ὅὅ ὈὈ ὉὉ ὊὊ ὋὋ ὌὌ ὍὍ ὎὎ ὏὏
ὐὐ ὑὑ ὒὒ ὓὓ ὔὔ ὕὕ ὖὖ ὗὗ ὙὙ ὛὛ ὝὝ ὟὟ
ὠὠ ὡὡ ὢὢ ὣὣ ὤὤ ὥὥ ὦὦ ὧὧ ὨὨ ὩὩ ὪὪ ὫὫ ὬὬ ὭὭ ὮὮ ὯὯ
ὰὰ άά ὲὲ έέ ὴὴ ήή ὶὶ ίί ὸὸ όό ὺὺ ύύ ὼὼ ώώ
ᾀᾀ ᾁᾁ ᾂᾂ ᾃᾃ ᾄᾄ ᾅᾅ ᾆᾆ ᾇᾇ ᾈᾈ ᾉᾉ ᾊᾊ ᾋᾋ ᾌᾌ ᾍᾍ ᾎᾎ ᾏᾏ
ᾐᾐ ᾑᾑ ᾒᾒ ᾓᾓ ᾔᾔ ᾕᾕ ᾖᾖ ᾗᾗ ᾘᾘ ᾙᾙ ᾚᾚ ᾛᾛ ᾜᾜ ᾝᾝ ᾞᾞ ᾟᾟ
ᾠᾠ ᾡᾡ ᾢᾢ ᾣᾣ ᾤᾤ ᾥᾥ ᾦᾦ ᾧᾧ ᾨᾨ ᾩᾩ ᾪᾪ ᾫᾫ ᾬᾬ ᾭᾭ ᾮᾮ ᾯᾯ
ᾰᾰ ᾱᾱ ᾲᾲ ᾳᾳ ᾴᾴ ᾵᾵ ᾶᾶ ᾷᾷ ΑᾹ ᾺΑ ΆΑ ΑιΑι ’ ’ ι ’ ’
῁ ῁ ῂῂ ῃῃ ῄῄ ῅῅ ῆῆ ῇῇ ῈΕ ΈΕ ῊΗ ΉΗ ΗιΗι ” ” ” ” ‚ ‚
ῐῐ ῑῑ ῒῒ ΐΐ ῖῖ ῗῗ ΙΙ ΪΙ ΊΙ ” ” ” ” ‚ ‚
ῠῠ ῡῡ ῢῢ ΰΰ ῤῤ ῥῥ ῦῦ ῧῧ ΥΥ ΫΫ ῪΥ ΎΥ ῬΡ ῭ ΅ ` ` ` `
ῲῲ ῳῳ ῴῴ ῶῶ ῷῷ ῸΟ ΌΟ ῺΩ ΏΩ ΩιΩι ’ ’ ‘ ‘

MakeUppercase:
AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA AA
EE EE EE EE EE EE EE EE EE EE EE EE
HH HH HH HH HH HH HH HH HH HH HH HH HH HH HH HH
II II II II II II II II II II II II II II II II
OO OO OO OO OO OO OO OO OO OO OO OO
YY YY YY YY YY YY YY YY YY YY YY YY
ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ
AA AA EE EE HH HH II II OO OO YY YY ΩΩ ΩΩ
ΑιΑι ΑιΑι ΑιΑι ΑιΑι ΑιΑι ΑιΑι ΑιΑι ΑιΑι ΑιΑι ΑιΑι ΑιΑι ΑιΑι ΑιΑι ΑιΑι ΑιΑι ΑιΑι

[4] With LICRs, it is `\textautosigma`.

HıHı HıHı HıHı HıHı HıHı HıHı HıHı HıHı HıHı HıHı HıHı HıHı HıHı HıHı HıHı HıHı
ΩıΩı ΩıΩı ΩıΩı ΩıΩı ΩıΩı ΩıΩı ΩıΩı ΩıΩı ΩıΩı ΩıΩı ΩıΩı ΩıΩı ΩıΩı ΩıΩı ΩıΩı
ĂĂ ĀĀ AıAı AıAı AıAı AA AıAı ĂĂ ĀĀ AA AA AıAı ’ ıI ’
˜¨ ˝ HıHı HıHı HıHı HH HıHı EE EE HH HH HıHı ” ” ˇ
ЇЇ ЇЇ ЇЇ ЇЇ II ”IЇ ЇЇ ЇЇ II II ˚ ˝ ˇ
ŸŸ ŸŸ ŸŸ ŸŸ PP PP YY ”YŸ ŸŸ ŸŸ YY YY PP¨ ˚¨ ´ `
ΩıΩı ΩıΩı ΩıΩı ΩΩ ΩıΩı OO OO ΩΩ ΩΩ ΩıΩı ´ ´ ‘

MakeLowercase:

ȧȧ ǡǡ ȃȃ ȃȃ ăă ăă ǎǎ ǎǎ ȧȧ ǡǡ ȃȃ ȃȃ ăă ăă ǎǎ ǎǎ
ȅȅ ȇȇ ȇȇ ȇȇ ȇȇ ȇȇ ȅȅ ȇȇ ȇȇ ȇȇ ȇȇ ȇȇ
ṅṅ ṅṅ ṅṅ ṅṅ ṅṅ ṅṅ ṅṅ ṅṅ ṅṅ ṅṅ ṅṅ ṅṅ ṅṅ ṅṅ ṅṅ
ìì íí ȉȉ ȉȉ ĭĭ ĭĭ ìì íí ȉȉ ȉȉ ĭĭ ĭĭ
ȯȯ ȯȯ ȍȍ ȍȍ ŏŏ ŏŏ ȯȯ ȯȯ ȍȍ ȍȍ ŏŏ ŏŏ
u̇u̇ úú ȕȕ ȕȕ ŭŭ ŭŭ ů̇ů̇ ů̇ů̇ úú ȕȕ ŭŭ ů̇ů̇
ẇẇ ẃẃ ẅẅ ẅẅ ẅẅ ẅẅ ẘẘ ẘẘ ẇẇ ẃẃ ẅẅ ẅẅ ẅẅ ẘẘ
àà áá èè éé ȅȅ ȇȇ ìì íí òò óó ȕȕ úú ẇẇ ẃẃ
ậậ ậậ ậậ ậậ ậậ ậậ ậậ ậậ ậậ ậậ ậậ ậậ ậậ ậậ
ṉṉ ṉṉ n̈n̈ n̈n̈ n̈n̈ n̈n̈ n̂n̂ n̂n̂ ṉṉ ṉṉ n̈n̈ n̈n̈ n̈n̈ n̈n̈ n̂n̂ n̂n̂
ộộ ộộ ộộ ộộ ộộ ộộ ộộ ộộ ộộ ộộ ộộ ộộ ộộ ộộ
ăă āā ậậ ąą ąą ãã ąą ăă āā àà áá ąą ’ ’ ˛ı ’ ’
˜ ˜ ˝ ˝ ṉṉ n̈n̈ ṉṉ ññ n̈n̈ èè éé ȅȅ ȇȇ n̈n̈ ” ” ” ” ˇ ˇ
ūū ūū ūū ūū ūū ūū ūū ìì íí ” ” ” ” ˇ ˇ
ŭŭ ōō ȕȕ ŭŭ ṗṗ ṗṗ ōō ȭȭ ŭŭ ōō ȕȕ úú ṗṗ ¨ ¨ ´ ´ ` `
ộộ ɷɷ ộộ ȭȭ ɷɷ òò óó ẇẇ ẃẃ ɷɷ ´ ´ ‘ ‘