

Font setup for Greek with XeTeX/LuaTeX

Günter Milde

September 26, 2020

The file `tuenc-greek.def` provides support for Greek LICR macros and upcasing of text with XeTeX and LuaTeX. It is automatically loaded by the `textalpha` and `alphabetalpha` packages as well as `babel-greek` when used with Unicode fonts (LuaTeX or XeTeX with `fontspec`).

Contents

1	Requirements	1
1.1	<code>fontspec</code>	1
1.2	suitable Unicode fonts	2
2	Usage	2
3	LICR input	2
3.1	Greek alphabet	2
3.2	Diacritics	2
3.2.1	perispomeni vs. tilde	3
3.2.2	combined diacritics	3
3.2.3	sub-iota	3
3.3	Additional Greek symbols	3
3.3.1	symbols for Greek numbers	3
3.3.2	symbol variants	4
3.3.3	Ancient Greek Numbers	4
4	Latin transcription	6
5	UPPERCASE and lowercase	6
5.0.1	hiatus	6
6	Character Tables	7
6.1	Greek and Coptic Unicode block	7
6.2	Greek Extended Unicode block	7

1 Requirements

1.1 fontspec

The `fontspec` package sets up the TU Unicode font encoding for XeTeX and LuaTeX. However, it does not provide Greek character names (LICR macros).

1.2 suitable Unicode fonts

With 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 the console output) and empty spaces in the output document.

Examples for suitable fonts are given in the [greek-fontenc README](#).

2 Usage

`tuenc-greek.def` is usually not loaded directly, but by one of [textalpha](#), [alphabetalpha](#), or [Babel](#) (with the language option `greek`). If these packages are loaded after [fontspec](#), 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, `tuenc-greek-doc.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:

```
ά ά ξξ àà ξξ ü ü ãã ää åå ðð
```

Up-to-date Xe/LuaTeX normalizes base letter and combining diacritics to the corresponding pre-composed character if such a mapping is defined in the Unicode standard:

```
ά ά à à ì ì ã ã ä ä å å
```

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:

$$\tilde{\alpha} = \tilde{\alpha}, \tilde{\eta} = \tilde{\eta}, \tilde{\iota} = \tilde{\iota}, \tilde{\upsilon} = \tilde{\upsilon}, \tilde{\omega} = \tilde{\omega}$$

3.2.2 combined diacritics

Combined accents are defined using combining diacritical characters.

í î ï ÿ
ï ÿ
ï ÿ

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 monotonoc) Greek.

3.2.3 sub-iota

The sub-iota (ypogegrammeni/prosgegrammeni) is input after the base letter.

- `\prosgegrammeni` sets a spacing GREEK PROSGEGRAMMENI: Αι Κι.
- `\ypogegrammeni` sets a COMBINING GREEK YPOGEGRAMMENI: α̣ κ̣.

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: Αι but Κ̣.

ΑιΑια̣α̣ ΑιΑι
Λ̣Λ̣λ̣λ̣ Λ̣Λ̣
α̣α̣α̣α̣ ΑιΑι

3.3 Additional Greek symbols

3.3.1 symbols for Greek numbers

ζ̣ textkoppa
Ζ̣ textKoppa
Ϟ̣ textqoppa (archaic koppa)
Ϛ̣ textQoppa (archaic Koppa)
ς̣ textstigma
Ϛ̣ textStigma (Sigma-Tau-Ligature in CB-fonts)¹
ϣ̣ textsampi

¹the 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	<i>missing</i>	β	β
κ	<i>missing</i>	κ	κ
Θ	<i>missing</i>	Θ	Θ

Table 1: Greek symbol variants in TeX and Unicode

ϑ textSampi
 ε textdigamma
 \textF textDigamma
 \text' textdexiakeraia
 \text, 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' mathemat-
ical mode

```

See Table 2 for the full list. The *alphabet* 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:

⊠⊡⊢⊣

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.

⊠⊡⊢⊣

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

Table 2: Macros for Greek symbol variants

4 Latin transcription

The Latin transcription known from LGR encoded 8-bit fonts² 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 `tuenc-greek-doc.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 « $\lambda\omicron\gamma\omicron\varsigma$ » and « $\lambda\upsilon\alpha\upsilon\pi\nu\acute{\alpha}$ » becomes « $\acute{\alpha}\upsilon\pi\nu\acute{\alpha}$ ».

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 $\text{\I}\alpha\nu\omicron\upsilon\alpha\rho\acute{\iota}\omicron\upsilon$ (Unicode font set up via `fontspec`) vs. $\text{\I}\alpha\nu\omicron\upsilon\alpha\rho\acute{\iota}\omicron\upsilon$ (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. $\mu\acute{\alpha}\iota\sigma\tau\rho\varsigma \mapsto \text{ΜΑΪΣΤΡΟΣ}$.

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

$\acute{\alpha}\acute{\alpha}\acute{\alpha}\grave{\alpha}\grave{\alpha} \mapsto \text{A}\acute{\text{A}}\text{A}\acute{\text{A}}\text{A}\grave{\text{A}}\text{A}\grave{\text{A}}$

5.0.1 hiatus

Tonos and dasia mark a *hiatus* (break-up of a diphthong) if placed on the first vowel of a diphtong ($\acute{\alpha}\iota$, $\acute{\alpha}\upsilon$, $\acute{\epsilon}\iota$, $\acute{\alpha}\iota$, $\acute{\alpha}\upsilon$, $\acute{\epsilon}\iota$). A dialytika must be placed on the second vowel if they are dropped.

The «hiatus» feature works with macro input:

$\acute{\alpha}\upsilon\lambda\omicron\varsigma \mapsto \text{A}\acute{\text{Y}}\text{Λ}\text{O}\Sigma$, $\acute{\alpha}\upsilon\lambda\omicron\varsigma \mapsto \text{A}\acute{\text{Y}}\text{Λ}\text{O}\Sigma$,
 $\mu\acute{\alpha}\iota\nu\alpha \mapsto \text{ΜΑ}\acute{\text{I}}\text{ΝΑ}$, $\acute{\kappa}\acute{\epsilon}\iota\kappa \mapsto \text{ΚΕ}\acute{\text{I}}\text{Κ}$, $\acute{\alpha}\upsilon\pi\nu\acute{\alpha} \mapsto \text{Α}\acute{\text{Y}}\text{Π}\text{I}\text{N}\text{A}$.

It does not work with Unicode literals:

$\acute{\alpha}\iota$, $\acute{\alpha}\upsilon$, $\acute{\epsilon}\iota$, $\acute{\alpha}\iota$, $\acute{\alpha}\upsilon$, $\acute{\epsilon}\iota \mapsto \text{AI}$, AY , EI , AI , AY , EI

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

$\acute{\alpha}\iota$, $\acute{\alpha}\upsilon$, $\acute{\epsilon}\iota$, $\acute{\alpha}\iota$, $\acute{\alpha}\upsilon$, $\acute{\epsilon}\iota \mapsto \text{AI}$, AY , EI , AI , AY , EI

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

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

