

Greek LICR definitions with XeTeX/LuaTeX

Günter Milde

September 8, 2014

The file `greek-euenc.def` provides support for Greek LICR macros with XeTeX and LuaTeX. It is automatically loaded by the `textalpha` and `alphabeta` packages as well as `babel-greek`, if the font encoding is set to one of the Unicode font encodings EU1 or EU2 (usually via the `fontspec` package for font setup with LuaTeX and XeTeX).

With the XeTeX or LuaTeX engines, the user should ensure that the selected font contains Greek glyphs (the default Latin Modern fonts have only capital Greek letters). Examples for suitable fonts are the “Deja Vu” or the “Linux Libertine” OpenType fonts.

See the source of this document `greek-euenc-doc.tex` for a setup and usage example.

1 LICR input

LICR macros are a verbose but failsafe 7-bit ASCII encoding that works unaltered under both, 8-bit TeX and XeTeX/LuaTeX. Use cases are macro definitions and generated text.

The example from `usage.tex` in `babel-greek` input using LICR macros:

```
Τί φήις; Ἰδὼν ἐνθέδε παῖδ' ἐλευθέραν  
τὰς πλησίον Νύμφας στεφανοῦσαν, Σώστρατε,  
ἔρῶν ἀπῆλθες εὐθύς;
```

1.1 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`

In the Latin transcription defined by the 8-bit LGR font encoding, the letter ‘s’ stands for `\textautosigma` which automatically chooses the glyph according to the position. This is not implemented for Unicode fonts.

1.2 Additional Greek symbols

ζ textKoppa
 ς textkoppa
 \wp textqoppa (archaic koppa)
 $\text{\textcircled{Q}}$ textQoppa (archaic Koppa)
 ζ textstigma
 $\text{\textcircled{S}}$ textStigma (Sigma-Tau-Ligature in CB-fonts)¹
 $\text{\textcircled{r}}$ textsampi
 $\text{\textcircled{R}}$ textSampi
 $\text{\textcircled{f}}$ textdigamma
 $\text{\textcircled{F}}$ textDigamma
 $\text{\textcircled{’}}$ textdexiakeraia
 $\text{\textcircled{,}}$ textaristerikeraia

Mathematical notation uses variant shapes for pi, kappa, rho, theta, and phi as additional symbols. These variations have no syntactic meaning in Greek text and are not given separate code-points in the LGR text encoding. Some text fonts use the variant shapes in place of the “regular” ones as a stylistic choice.

TeX’s concept of “standard” vs. “variant” symbols does not map to the distinction between “Greek Letter ...” vs. “Greek Symbol ...” in the Unicode standard:

math symbol	Unicode “letter”	var math symbol	Unicode “symbol”
π	π	ϖ	$\text{\textcircled{\omega}}$
ρ	ρ	ϱ	$\text{\textcircled{\varrho}}$
θ	θ	ϑ	$\text{\textcircled{\vartheta}}$
ϵ	ϵ	ε	$\text{\textcircled{\epsilon}}$
ϕ	ϕ	φ	$\text{\textcircled{\phi}}$
β	β	<i>missing</i>	$\text{\textcircled{\beta}}$
κ	κ	<i>missing</i>	$\text{\textcircled{\kappa}}$
Θ	Θ	<i>missing</i>	$\text{\textcircled{\Theta}}$

1.3 Diacritics

Greek diacritics can be input by named macro or symbol macro. Named macros are defined in xunicode-greek.sty for `\LastDeclaredEncoding` (here LGR).

$\acute{\alpha}$ $\acute{\chi}$ $\grave{\alpha}$ $\grave{\chi}$ $\text{\textcircled{\i}}$ $\text{\textcircled{\x}}$ $\text{\textcircled{\alpha}}$ $\text{\textcircled{\chi}}$ $\text{\textcircled{\acute{\alpha}}}$ $\text{\textcircled{\acute{\chi}}}$

There are currently no definitions for accent macros combined with literal unicode characters. Combining diacritical characters are used in the output.

$\acute{\alpha}$ $\acute{\alpha}$ $\grave{\alpha}$ $\grave{\alpha}$ $\text{\textcircled{\i}}$ $\text{\textcircled{\alpha}}$ $\text{\textcircled{\acute{\alpha}}}$ $\text{\textcircled{\acute{\alpha}}}$

Please mail a feature request if there is demand for definitions selecting precomposed characters. As proof of concept, this document defines a composite command for alpha with psili and perispomeni. Compare $\text{\textcircled{\acute{\alpha}}}$ with $\text{\textcircled{\acute{\alpha}}}$.

¹the name “stigma” originally applied to a medieval sigma-tau ligature, whose shape was confusably similar to the cursive digamma

Composite commands for combined diacritics are defined in “greek-fontenc.def”, also for the `\LastDeclaredEncoding`. There are composite definitions for all precomposed characters, the fallback definition uses combining characters:

í î ï ð ñ ò ó ô õ ö ÷ ù ú û ü ý ÿ

ï ï ð ð ñ ñ ò ò ó ó ô ô õ õ ö ö ÷ ÷ ù ù ú ú û û ü ü ý ý ÿ ÿ

ï ï ð ð ñ ñ ò ò ó ó ô ô õ õ ö ö ÷ ÷ ù ù ú ú û û ü ü ý ý ÿ ÿ

Problems: Composite diacritics with breathings overlap when set using two combining characters (in the tested fonts).

Upcasing with `\MakeUppercase` drops diacritics (except dialytika and sub-iota):

Ï X X Ï X X Ï X X

II X X II X X II X X

II X X II X X II X X

However, when the tonos, varia, and perispomeni accents are input using the symbol macros (`\' \` \~`), they behave like standard accents with XeTeX/LuaTeX. If these accents should be dropped by `MakeUppercase`, they must be input as named macro:

AA X \acute{X} AA X \grave{X} AA X \tilde{X}

The sub-iota is input after the base letter. Ligature definitions in the 8-bit LGR fonts ensure that precomposed characters are used. This does not work with Unicode fonts:²

α x Aι Xι, Aι Xι, see also the *Greek extended* Unicode block below.

Both, the standard LaTeX and the “xunicode” composite mechanism require the first part of a composite definition to be a macro requiring one argument. I.e. selecting pre-defined characters with sub-iota would require prefix `\accypogegrammeni` and `\accprosgegrammeni` accent macros and composite definitions for all combination of letters and letters with diacritic and sub-iota.

1.4 Greek and Coptic Unicode block

Greek symbols from the Greek and Coptic Unicode block are fully supported:

‘

;

‘ ‘ A · E H I O Y Ω í

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

Ï Ÿ á é ή í ú

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

ï ü ó ú ó

²When copy/pasting generated output from the “evince” PDF viewer, the COMBINING GREEK YPOGEGRAMMENI is merged with the base character. A capital letter + combining ypogegrammeni results in GREEK CAPITAL LETTER ... WITH [.. AND] PROSGEGRAMMENI). This indicates that the combining *ypogegrammeni* should also be used together with capital base letters.

Q o S F f 7 7 7

1.5 Greek Extended Unicode block

Characters from the Greek Extended Unicode block are only partially supported, accented characters with sub-iota currently use a combining prosgegrammeni or ypoge-grammeni character.

For comparision, the LICR macro is followed by the corresponding literal Unicode character:

άά άά άά άά άά άά άά άά ΆΑ ΆΑ ΆΑ ΆΑ ΆΑ ΆΑ ΆΑ ΆΑ ΆΑ ΆΑ
 έέ έέ έέ έέ έέ έέ έέ έέ ΈΕ ΈΕ ΈΕ ΈΕ ΈΕ ΈΕ ΈΕ ΈΕ
 ήή ήή ήή ήή ήή ήή ήή ήή ΗΗ ΗΗ ΗΗ ΗΗ ΗΗ ΗΗ ΗΗ ΗΗ ΗΗ ΗΗ
 ιι ιι ιι ιι ιι ιι ιι ιι ΙΙ ΙΙ ΙΙ ΙΙ ΙΙ ΙΙ ΙΙ ΙΙ ΙΙ ΙΙ
 όό όό όό όό όό όό όό ΌΟ ΌΟ ΌΟ ΌΟ ΌΟ ΌΟ ΌΟ ΌΟ ΌΟ
 ύύ ύύ ύύ ύύ ύύ ύύ ύύ ΎΥ ΎΥ ΎΥ ΎΥ ΎΥ ΎΥ ΎΥ ΎΥ
 ώώ ώώ ώώ ώώ ώώ ώώ ώώ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ
 άά άά έέ έέ ήή ήή ιι ιι όό όό ύύ ύύ ώώ ώώ
 άά άά άά άά άά άά άά ΆΑ ΆΑ ΆΑ ΆΑ ΆΑ ΆΑ ΆΑ ΆΑ ΆΑ
 ήή ήή ήή ήή ήή ήή ήή Ηι Ηι Ηι Ηι Ηι Ηι Ηι Ηι Ηι Ηι Ηι Ηι Ηι Ηι Ηι Ηι
 φφ φφ φφ φφ φφ φφ φφ Ωι Ωι Ωι Ωι Ωι Ωι Ωι Ωι Ωι Ωι Ωι Ωι Ωι Ωι Ωι Ωι Ωι
 άά άά άά άά άά άά ΆΑ ΆΑ ΆΑ ΆΑ Αι Αι " ιι "
 ~ ~ ήή ήή ήή ήή ΈΕ ΈΕ ΗΗ ΗΗ Ηι Ηι ~ ~ ~ ~
 ιι ιι ιι ιι ιι ΙΙ ΙΙ ΙΙ ΙΙ ~ ~ ~ ~
 υύ υύ υύ υύ ρρ ρρ υύ υύ ΎΎ ΎΎ ΎΎ ΎΎ ΡΡ ~ ~ ~ ~
 φφ φφ φφ φφ φφ ΌΟ ΌΟ ΩΩ ΩΩ Ωι Ωι " "

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:



2 Latin transcription

The `\textgreek` macro and LGR encoded fonts can also be used for the input of Greek letters via the *Latin transcription* provided by this font encoding, e.g. «logos» becomes « $\lambda\omicron\gamma\omicron\varsigma$ » and «`\>aupn\ 'ia`» becomes « $\acute{\alpha}\upsilon\pi\nu\acute{\iota}\alpha$ » (capitalized « $\acute{\text{A}}\acute{\text{T}}\text{I}\text{N}\text{I}\text{A}$ »). For details of the Latin transliteration see the `teubner` package or the file `usage.pdf` from the `babel-greek` package.

Mark, that you cannot use Unicode input with LGR encoded fonts except when running in 8-bit compatibility mode because of the incompatibility between Xe/LuaTeX and the `inputenc` package.

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, see *The LaTeX font guide* [`fntguide.pdf`]).