

Greek LICR definitions with XeTeX/LuaTeX

Typesetting Greek texts requires a font containing Greek letters. With the XeTeX or LuaTeX engines, the user should ensure that the chosen font contains the required 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.

1 LICR input

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

Τί φηις; Ἰδὼν ἐνθὲδε παῖδ' ἐλευθέραν
τάς πλησίον Νύμφας στεφανοῦσαν, Σώστρατε,
ἔρων ἀπῆλθες εὐθύς;

1.1 Greek alphabet

Greek letters via LICR macros:

A B Γ Δ E Z H Θ I K Λ M N Ξ O Π P Σ T Y Φ X 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, the letter ‘s’ stands for `\textautosigma` which automatically chooses the glyph according to the position.

1.2 Additional Greek symbols

Ϝ `textkoppa`

ϝ `textqoppa` (archaic koppa)

Ϟ `textQoppa` (archaic Koppa)

Ϸ `textstigma`

ϸ `textStigma` (Sigma-Tau-Ligature in CB-fonts)¹

Ϡ `textsampi`

ϡ `textSampi`

Ϣ `textdigamma`

ϣ `textDigamma`

ϥ `textdexiakeraia`

Ϧ `textaristerikeraia`

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

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	ϖ
ρ	ρ	ϱ	ϱ
θ	θ	ϑ	ϑ
ϵ	ϵ	ε	ϵ
ϕ	ϕ	φ	ϕ
β	β	<i>missing</i>	\mathfrak{b}
κ	κ	<i>missing</i>	\mathfrak{k}
Θ	Θ	<i>missing</i>	\mathfrak{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 EU1).

άά χχ òò ùù ιι ξξ ãã õõ áá χχ άά χχ

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

$$\acute{\imath} \acute{\imath} \acute{\imath} \acute{x} \acute{x} \acute{x} \grave{\imath} \grave{\imath} \grave{\imath} \grave{x} \grave{x} \grave{x} \tilde{\imath} \tilde{\imath} \tilde{\imath} \tilde{x} \tilde{x} \tilde{x}$$
$$\begin{array}{cccccccccccccccccccc} \ell & \ell & \ell & x & x & x & \ell & \ell & \ell & x & x & x & \ell & \ell & \ell & x & x & x \end{array}$$
$$\begin{array}{cccccccccccccccccccc} \text{''} & \text{''} & \text{''} & \text{' } & \text{' } & \text{' } & \text{''} & \text{''} & \text{''} & \text{' } & \text{' } & \text{' } & \text{~} & \text{~} & \text{~} & \text{~} & \text{~} & \text{~} \\ \text{L} & \text{L} & \text{L} & \text{X} & \text{X} & \text{X} & \text{L} & \text{L} & \text{L} & \text{X} & \text{X} & \text{X} & \text{L} & \text{L} & \text{L} & \text{X} & \text{X} & \text{X} \end{array}$$

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

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

$$\ddot{\Pi} \ddot{X} \ddot{X} \ddot{\Pi} \ddot{X} \ddot{X} \ddot{\Pi} \ddot{X} \ddot{X}$$

Π Χ Χ Π Χ Χ Π Χ Χ

ΠΧΧΠΧΧΠΧΧ

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

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

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 for a suffix (in contrast to post-fix) *prosgegrammeni*, the combining *ypogegrammeni* could be used together with a capital base letter.

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 ι
Α Β Γ Δ Ε Ζ Η Θ Ι Κ Λ Μ Ν Ξ Ο Π Ρ Σ Τ Υ Φ Χ Ψ Ω
İ Ÿ á é ή í ú
α β γ δ ε ζ η θ ι κ λ μ ν ξ ο π ρ σ τ υ φ χ ψ ω
ĩ ü ó ú ő
Q q S s F f ʒ ʔ ʀ

```

1.5 Greek Extended Unicode block

Characters from the Greek Extended Unicode block are only partially supported (multi-accented characters with sub-iota currently use a combining prosgegrammeni or ypogegrammeni character).

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

```

άά άά άά άά άά άά ΆΆ ΆΆ ΆΆ ΆΆ ΆΆ ΆΆ ΆΆ ΆΆ
έέ έέ έέ έέ έέ έέ ΈΈ ΈΈ ΈΈ ΈΈ ΈΈ ΈΈ ΈΈ
ήή ήή ήή ήή ήή ήή ήή ήή ήή ήή ήή ήή ήή ήή
ιι ιι ιι ιι ιι ιι ιι ιι ιι ιι ιι ιι ιι ιι ιι
οο οο οο οο οο οο ΟΟ ΟΟ ΟΟ ΟΟ ΟΟ ΟΟ ΟΟ ΟΟ
ύύ ύύ ύύ ύύ ύύ ύύ ύύ ύύ ύύ ύύ ύύ ύύ ύύ
ώώ ώώ ώώ ώώ ώώ ώώ ώώ ώώ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ
άά άά έέ έέ ήή ήή ù ù ò ò ù ù ώώ ώώ
άά άά άά άά άά άά άά άά ΆΆ ΆΆ ΆΆ ΆΆ ΆΆ ΆΆ ΆΆ
ήή ήή ήή ήή ήή ήή ήή ήή ήή ήή ήή ήή ήή ήή
ώώ ώώ ώώ ώώ ώώ ώώ ώώ ώώ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ ΩΩ
άά άά άά άά ΆΆ ΆΆ ΆΆ ΆΆ ΆΆ ΆΆ ΆΆ

```

ᐃᑦᐃ ᐃᑦᐃ ᐃᑦᐃ ᐃᐃ ᐃᑦᐃ ᐅᐅ ᐅᐅ ᐅᐅ ᐅᐅ ᐅᐅ ᐅᐅ ᐅᐅ

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