

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 miss them). Examples for suitable fonts are the “Deja Vu” or the “Linux Libertine” OpenType fonts.

1 LICR input

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

Ϸ `textkoppa`

Ϡ `textqoppa` (archaic koppa)

ϡ `textQoppa` (archaic Koppa)

Ϸ `textstigma`

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

ϳ `textsampi`

ϴ `textSampi`

ϕ `textdigamma`

Ϟ `textDigamma`

ϟ `textnumeralsigngreek` (Dexia keraia)

Ϡ `textnumeralsignlowergreek` (Aristeri keraia)

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.

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

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

math symbol	Unicode «letter»	var math symbol	Unicode «symbol»
ϵ	ε	ε	€
κ	κ	<i>missing</i>	κ
π	π	ϖ	ω
ρ	ρ	ϱ	ϱ
θ	θ	ϑ	ϑ
ϕ	φ	φ	ϕ

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 ξξ àà ùù ùù ãã õõ áá ùù áá ùù

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:

ùù ùù ùù ùù ùù ùù

ùù ùù ùù ùù ùù ùù

ùù ùù ùù ùù ùù ùù

Problems:

The COMBINING GREEK DIALYTIKA TONOS character is set behind instead of above the base letter.

Composite diacritics defined with `xunicode`’s `\DeclareEncodedCompositeAccents` drop one of the constituents.

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

ùù ùù ùù ùù ùù ùù

ùù ùù ùù ùù ùù ùù

ùù ùù ùù ùù ùù ùù

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 ξ Α ξ, 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 requires prefix `\accypogegrammeni { }` and `\accprosgegrammeni` accent macros.

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

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 ü

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 ħ í ü

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

ĩ ü ó ú ő

Q q S s F f L l T t

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 ypoge-grammeni character).

For comparison, the LICR macro is followed by the same character as literal Unicode char in this list:

$\grave{\alpha}\acute{\alpha} \quad \acute{\alpha}\acute{\alpha} \quad \grave{\alpha}\grave{\alpha} \quad \acute{\alpha}\acute{\alpha} \quad \acute{\alpha}\acute{\alpha} \quad \acute{\alpha}\acute{\alpha} \quad \tilde{\alpha}\tilde{\alpha} \quad \tilde{\alpha}\tilde{\alpha} \quad \grave{A}\grave{A} \quad \grave{A}\grave{A} \quad \textit{\'A}\textit{'A} \quad \textit{\'A}\textit{'A} \quad \textit{\'A}\textit{'A} \quad \textit{\'A}\textit{'A} \quad \textit{\'A}\textit{'A} \quad \tilde{A}\tilde{A} \quad \tilde{A}\tilde{A}$

$\mathbb{E}\mathbb{E}$ $\mathbb{E}\mathbb{E}$ $\mathbb{E}\mathbb{E}$ $\mathbb{E}\mathbb{E}$ $\mathbb{E}\mathbb{E}$ $\mathbb{E}\mathbb{E}$ $\mathbb{E}'\mathbb{E}$ $\mathbb{E}'\mathbb{E}$ $\mathbb{E}''\mathbb{E}$ $\mathbb{E}''\mathbb{E}$ $\mathbb{E}''\mathbb{E}$ $\mathbb{E}''\mathbb{E}$

[illegible][illegible]

00 00 00 00 00 00 '00 '00 "0"0 "0"0 '0"0 "0"0

ǔǔ ǘǘ ǚǚ ǜǜ ǖǖ ǘǘ ǚǚ ǜǜ ‘Y’Y “Y”Y ”Y’Y ȳȳȳȳ

[illegible]

àà áá èè éé ήή ήή ìí ίί òò óó ùú ύύ ώώ ώώ

[illegible]

ṛḥ ṛḡ ṛḥ ṛḡ ṛḥ ṛḡ ṛḥ ṛḡ ṚḤ ṚḢ ṚḤ ṚḢ ṚḤ ṚḢ ṚḤ ṚḢ ṚḤ ṚḢ ṚḤ ṚḢ ṚḤ ṚḢ ṚḤ ṚḢ

[illegible]

ǎāàαάăăĂĀÀ'AAıı'

~ òòóò~E'E'H'H, " " " "

ũũũũũĩĩĩĩĩ

ŭ ū ü ǔ ǖ Ǘ Ǚ ǚ ǜ Ỳ Ỳ Ỳ Ỳ Ỳ

$\omega_1\omega_1\tilde{\omega}\tilde{\omega}'O'O'\Omega'\Omega\Omega_i'$

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

⌘⌘⌘⌘