Test Greek LaTeX internal character representations (LICR macros)

Günter Milde

2023/03/08

This document tests the compatibility of "luainputenc" and the Greek font setup for TU and PU. It uses only ASCII input.

See the source `test-inputenc.tex` for the input used in the examples.

TODO: Compiling with LuaTeX fails: `tuenc-greek.def` uses literal Unicode characters which is incompatible with legacy 8-bit encodings defined via "luainputenc".

Contents

1 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 (with any ASCII-compatible input encoding) and XeTeX/LuaTeX. Use cases are macro definitions and generated text.

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 Υ 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`

With Unicode fonts (Xe/LuaTeX, font encoding TU), the `\textautosigma` macro (which automatically chooses the glyph according to the position) does not work with LaTeX versions older than 2022/06 (requires the new `\MakeLowercase` implementation).

1.2 Diacritics

Greek accents are tonos = oxia, varia, psili, dasia, dialytika, and perispomeni.
    Greek diacritics can be input by named macro or symbol macro:

άά άά ὰὰ ὰὰ ϊ ϊ ϋϋ ᾶᾶ ᾶᾶ ἀἀ ἀἀ ἁἁ ἁἁ

Diacritics as spacing characters:
    with empty argument: ´ ´ ` ᾿ ῾ ¨ ˜ ᾽ ᾿ ῾ ῀ ῀ ῀ ῀ ῀
    with space as argument: ´ ´ ` ᾿ ῾ ¨ ˜ ᾽ ᾿ ῾ ῀ ῀ ῀ ῀ ῀
    with protected space: ´ ´ ` ᾿ ῾ ¨ ˜ ᾽ ᾿ ῾ ῀ ῀ ῀ ῀ ῀
    with nobreakspace: ´ ´ ` ᾿ ῾ ¨ ˜ ᾽ ᾿ ῾ ῀ ῀ ῀ ῀ ῀

1.2.1 mute iota

The mute iota is input after the base letter.

- `\ypogegrammeni` following a Greek letter sets a sub-iota (corresponding to COMBINING GREEK YPOGEGRAMMENI), e.g. ᾳ.

  In Unicode, a GREEK CAPITAL LETTER ... followed by COMBINING GREEK YPOGEGRAMMENI is normalized to GREEK CAPITAL LETTER ... WITH [... AND] PROSGEGRAMMENI, if a corresponding letter exists in the Unicode standard. In LGR fonts, this is implemented via a

ligature definition (set the Babel language or wrap in `\ensuregreek`): Aͺ but Λͺ.

The shape and position of the mute iota with pre-composed capital letters depends on the selected font, both sub-iota and adscript iota are possible.

- `\prosgegrammeni` sets an adscript iota (GREEK PROSGEGRAMMENI), e.g. Aι. In Unicode fonts the prosgegrammeni is spaced similar to the letter iota. In the CB Greek fonts, the only visible difference to the pre-composed characters is a slightly increased spacing.

  Copy/Paste may convert the adscript iota to a small letter iota!

Ypogegrammeni and prosgegrammeni following matching/not-matching base character (unchanged, lowercase, uppercase):

AιAι ᾳᾳ AιAι

ΛͺΛι λͅλͅ ΛͺΛι

ᾳαι ᾳᾳ AιAι

Using `\ypogegrammeni` for the mute iota with both, small and capital letters usually gives better results.

1.3 Additional Greek symbols

1.3.1 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
Ϝ textDigamma
ʹ textdexiakeraia
͵ textaristerikeraia

1.3.2 generic text symbols

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: $\mu \mapsto M$ but textmicro: $\mu \mapsto \mu$.

---

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

Table 1: Macros for Greek symbol variants

2 Greek in section headings

The packages *textalpha* and *alphabeta* as well as *babel-greek*, add hyperref support for LICR input with non-standard accents or combined diacritics.

2.1 0374 ʹ ͵ ; ΅ ᾿Α·ΈΉΊΌΎΏϊ

2.2 0391 ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΤΥΦΧΨΩ

2.3 03AA ΪΫ́άέήίΰ

2.4 03B1 αβγδεζηϑικλμνξοπρςστυφχψω

2.5 03CA ϊϋόύώβϑφπϞϙϠϡϜϝϟϟϏϗκϱΘε

2.6 1F00 ἀἁἂἃἄἅἆἇἈἉἊἋἌἍἎἏ

2.7 1F10 ἐἑἒἓἔἕἘἙἚἛἜἝ

2.8 1F20 ἠἡἢἣἤἥἦἧἨἩἪἫἬἭἮἯ

2.9 1F30 ἰἱἲἳἴἵἶἷἸἹἺἻἼἽἾἿ

2.10 1F40 ὀὁὂὃὄὅὈὉὊὋὌὍ

2.11 1F50 ὐὑὒὓὔὕὖὗὙὛὝὟ

2.12 1F60 ὠὡὢὣὤὥὦὧὨὩὪὫὬὭὮὯ

2.13 1F70 ὰάὲέὴήὶίὸόὺύὼώ

2.14 1F80 ᾀᾁᾂᾃᾄᾅᾆᾇᾈᾉᾊᾋᾌᾍᾎᾏ

2.15 1F90 ᾐᾑᾒᾓᾔᾕᾖᾗᾘᾙᾚᾛᾜᾝᾞᾟ

2.16 1FA0 ᾠᾡᾢᾣᾤᾥᾦᾧᾨᾩᾪᾫᾬᾭᾮᾯ

2.17 1FB0 ᾰᾱᾲᾳᾴᾶᾷᾸᾹᾺΆᾼ᾽ι᾿

2.18 1FC0 ῀῁ῂῃῄῆῇῈΈῊΉῌ῍῎῏

2.19 1FD0 ῐῑῒΐῖῗῘῙῚΊ῝῞῟

2.20 1FE0 ῠῡῢΰῤῥῦῧῨῩῪΎῬ῭΅`

2.21 1FF0 ῲῳῴῶῷῸΌῺΏῼ´῾

5