Test LICR macros in LuaTeX's 8-bit compatibility mode

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.

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:

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) does not work with Unicode fonts (Xe/LuaTeX, font encoding TU).

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:

ά᾽ ᾽ύ ὰ᾽ ᾽ὺ ῗ ΐ ῧ ΰ ᾶ᾽ ᾽ῦ ὰ᾽ ᾽ὺ ά᾽ ᾽ύ

Combining diacritics were misplaced with the 2021 version of LuaTeX in 8-bit compatibility mode. This can be solved by selecting the "Harfbuzz" renderer (see the source).

ε̈ ε̆ ῠ̆ α̣ α̣

Diacritics as spacing characters:
with empty argument: ´ ˊ ` ˋ ¨ ¨ ~ ~ ¯ ¯ ˘ ˘ ˚ ˚ ˝ ˝ ¸ ¸
with space as argument: ´ ˊ ` ˋ ¨ ¨ ~ ~ ¯ ¯ ˘ ˘ ˚ ˚ ˝ ˝ ¸ ¸
with protected space: ´ ˊ ` ˋ ¨ ¨ ~ ~ ¯ ¯ ˘ ˘ ˚ ˚ ˝ ˝
with nobreakspace: ´ ˊ ` ˋ ¨ ¨ ~ ~ ¯ ¯ ˘ ˘ ˚ ˚ ˝ ˝

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. α̱.

  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`): Αι but Α̱.

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

ᾋᾉ ᾳᾳ ᾋᾉ

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
F textDigamma
′ 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: µ ↦ M but textmicro: µ ↦ µ.

[1] the name "stigma" originally applied to a medieval sigma-tau ligature, whose shape was confusingly similar to the cursive digamma

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 Greek and Coptic

2.1.1 ΄ , ; ΄ ΅ Ά·ΈΉΊΌΎΏΐ

2.1.2 ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΤΥΦΧΨΩ

2.1.3 ΪΫάέήίΰ

2.1.4 αβγδεζηθικλμνξοπρςστυφχψω

2.1.5 ϊϋόύώϐϑϕϖϘϙϚϛϜϝϞϟϠϡϴϵ

2.2 Greek Extended

2.2.1 ἀἁἂἃἄἅἆἇἈἉἊἋἌἍἎἏ

2.2.2 ἐἑἒἓἔἕἘἙἚἛἜἝ

2.2.3 ἠἡἢἣἤἥἦἧἨἩἪἫἬἭἮἯ

2.2.4 ἰἱἲἳἴἵἶἷἸἹἺἻἼἽἾἿ

2.2.5 ὀὁὂὃὄὅὈὉὊὋὌὍ

2.2.6 ὐὑὒὓὔὕὖὗὙὛὝὟ

2.2.7 ὠὡὢὣὤὥὦὧὨὩὪὫὬὭὮὯ

2.2.8 ὰάὲέὴήὶίὸόὺύὼώ

2.2.9 ᾀᾁᾂᾃᾄᾅᾆᾇᾈᾉᾊᾋᾌᾍᾎᾏ

2.2.10 ᾐᾑᾒᾓᾔᾕᾖᾗᾘᾙᾚᾛᾜᾝᾞᾟ

2.2.11 ᾠᾡᾢᾣᾤᾥᾦᾧᾨᾩᾪᾫᾬᾭᾮᾯ

2.2.12 ᾰᾱᾲᾳᾴᾶᾷᾸᾹᾺΆᾼι᾽ ι῀

2.2.13 ῁ ῂῃῄῆῇῈΈῊΉῌ῍ ῎ ῏

2.2.14 ῐῑῒΐῖῗῘῙῚΊ῝ ῞ ῟

2.2.15 ῠῡῢΰῤῥῦῧῨΥῪΎῬ῭ ΅ `

2.2.16 ῲῳῴῶῷῸΌῺΏῼ´ ῾

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

Table 1: Macros for Greek symbol variants