Test LGR font encoding definitions

Günter Milde

2020/10/30

The file lgrenc.def provides a comprehensive set of macros to typeset Greek with LGR encoded fonts. It works for both, monotonic and polytonic Greek, independent of the *Babel* package.

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

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

1 Symbols

See the source file lgrenc-test.tex for the macros used to access the symbols.

1.1 Generic text symbols

Latin: + - = < > – — [()] { } \ | ‰ ‱ ␣

LGR: + - = – — [()]

< `\textless`
> `\textgreater`
{ `\textbraceleft`
} `\textbraceright`
\ `\textbackslash`
| `\textbar`
‰ `\textperthousand` (Per-mille symbol is missing in LGR.)
␣ `\textvisiblespace`

Quotes:[1] «a» «α», ‘a’ ‘α’, “a” “α” (double quotes wrong with Kerkis fonts)

Single guillemets and base-quotes (‹a› „a” ‚a’) are missing in LGR.

Ligature break up with `\textcompwordmark`: AY fi AϒΥ ϊ ↦ AY fi AϒΥ ¨ι

[1] Single quotes need special attention to prevent conversion to accents. Test the input conventions: ‘α’ ‘α’ ‘α’ ‘α’ but not ὰ́ ἔ ΐ̓

Spacing accent chars: ⁀a ˆα ˆι ∼a ˜α ˜ι ˘a ˘α ˘ι ¯a ¯α ¯ι ¨a ¨α ¨ι ´a ´α ´ι `a `α `ι

Letter schwa and Euro symbol: ə `\textschwa`, € `\texteuro`

Some ASCII symbols are replaced by different symbols in LGR encoding other symbols are composed from Latin letters and show Greek letters in LGR. *babel-greek* redefines some affected macros to use a standard font encoding, however this cannot be done in a font encoding definition file.

Beware that `" # & ' ; < > ?  @` becomes ’ ῎ · ΄ · ῾ ᾿ ; ῏

The *textcomp* package[2] provides pre-composed coyright ©, registered ® and trademark ™ symbols that work in all font encodings. In LGR they come out as: © `\textcopyright`, ® `\textregistered`, ™ `\texttrademark`.

textcomp also provides the upright MICRO SIGN and OHM SIGN for SI units: R = 5 µΩ

In LGR, `\textmicro` and `\textohm` are aliases to `\textmu` and `\textOmega` that do not change case: Αντίσταση = 5 µΩ, ΑΝΤΙΣΤΑΣΗ = 5 µΩ, αντίσταση = 5 µΩ.

1.2 Greek alphabet

Greek letters via Latin transliteration and 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 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

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

The small sigma is set with a different glyph if it ends a word:

σ `\textsigma`
ς `\textfinalsigma`

In the Latin transliteration, the letter ‘s’ stands for `\textautosigma` which automatically chooses the glyph according to the position.

1.3 additional Greek symbols

ϟ `\textkoppa` (numeral koppa = 90)
ϟ `\textKoppa` (numeral Koppa = 90)[3]
ϙ `\textqoppa` (archaic koppa)
Ϙ `\textQoppa` (archaic Koppa)
ϛ `\textstigma`
ϛ `\textvarstigma`

[2] loaded by default in not too old LaTeX
[3] In LGR, there is no separate code point for uppercase koppa.

Ϛ \textStigma (Sigma-Tau-Ligature in CB-fonts)[4]
ϡ \textsampi
Ϡ \textSampi
ϝ \textdigamma
Ϝ \textDigamma
ʹ \textdexiakeraia (dexia keraia)
͵ \textaristerikeraia (aristeri keraia)

Up/Downcasing of the additional Symbols from the Greek And Coptic Unicode block:

ʹ ͵ ΄ ; ΅ ΆA · ΈΉΊΌΎΏΐΪΫάέήίΰϊϋόύώϘϙϚϛϜϝϞϟϠϡ

MakeUppercase:
ʹ ͵ ΄ ; ¨ A · E H I O Υ Ω Ϊ Ϋ A E H I Ϋ Ϊ Ϋ O Υ Ω Ϙ Ϙ Ϛ Ϛ F F Ϟ Ϟ Ϡ Ϡ

MakeLowercase:
ʹ ͵ ΄ ; ΅ ά · έ ή ί ό ύ ώ ΐ ϊ ϋ ά έ ή ί ΰ ϊ ϋ ό ύ ώ ϙ ϙ ϛ ϛ ϝ ϝ Ϟ Ϟ ϡ ϡ

1.4 aliases

Aliases are defined in the included file greek-fontenc.def.

Names matching mathematical variant symbols:

ε \textvarepsilon = ε \textepsilon
φ \textvarphi = φ \textphi
ς \textvarsigma = ς \textfinalsigma

Compatibility aliases for hyperref's puenc.def:

μ \textmugreek = μ \textmu
Ϟ \textkoppagreek = Ϟ \textkoppa
Ϟ \textKoppagreek = Ϟ \textKoppa
Ϛ \textStigmagreek = Ϛ \textStigma
ϛ \textstigmagreek = ϛ \textstigma
Ϡ \textSampigreek = Ϡ \textSampi
ϡ \textsampigreek = ϡ \textsampi
ϝ \textdigammagreek = ϝ \textdigamma
Ϝ \textDigammagreek = Ϝ \textDigamma
ʹ \textnumeralsigngreek = ʹ \textdexiakeraia
͵ \textnumeralsignlowergreek = ͵ \textaristerikeraia

Two Unicode code points and names for one character:

ʹ \accoxia = ʹ \acctonos
᾽ \acckoronis = ᾽ \accpsili

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

1.5 symbol variants

Mathematical notation distinguishes variant shapes for pi ($\pi|\varpi$), rho ($\rho|\varrho$), theta ($\theta|\vartheta$), beta ($\beta|\varbeta$), and kappa ($\kappa|\varkappa$) (characters for the last two variant symbols are not included in TeX's standard math fonts). These variations have no syntactic meaning in Greek text and are not given code-points in the LGR encoding. Greek text fonts use the shape variants interchangeably.

2 Diacritics

Capital Greek letters have diacritics (except dialytika, breve, macron, and sub-iota) to the left (instead of above) and drop them if text is set in UPPERCASE. This is implemented for all combinations that are used in Greek texts (i.e. for which pre-composed Unicode character exist), but not for, e.g., $\tilde{\Omega}$).

Different conventions exist for the treatment of the sub-iota with uppercase letters. The CB-Fonts use a capital Iota "index" (A_I, H_I, Ω_I).

LaTeX standard accents[5] (Latin, Greek, Greek Capitals ↦ UPPERCASE)

à á ã ä â ā ă ȧ å ǎ ă a̲ ą ạ ą ↦ À Á Ã Ä Â Ā Ă Ȧ Å Ǎ Ă A̲ Ą Ạ Ą

ὰ ά ᾶ ϊ ᾶ ᾱ ᾰ ὰ ᾰ ᾰ ᾳ ᾳ ᾳ ↦ A A A Ä Â Ā Ă Ȧ Å Ǎ Ă A̲ Ą Ạ

᾽Α᾽ΑÃÄ ÂĀĂȦÅǍĂ A̲ Ą Ạ ↦ AAAÄ ÂĀĂȦÅǍĂ A̲ Ą Ạ

Additional Greek diacritics (Greek, Greek Capitals[6] ↦ UPPERCASE)

ὰ ὲ ῖ ΐ ῒ ῗ ῆ ὅ ὅ ὗ ῶ ᾳ ↦ A E I Ϊ Ϊ Ϊ H O O ϒ Ω A_I

῾A ῾E ῾I ῾H ῾O ῾O ῾ϒ ῾Ω A_I ↦ A E I H O O ϒ Ω A_I

Input variants and their conversion with MakeUppercase:

ᾰ ᾰ ᾰ, ᾰ ᾰ ᾰ ᾰ ᾰ, ῆ ῆ ῆ ῆ ῆ ῆ, ῆ ῆ, ῖ ῖ, ῖ ῖ, ῖ ῖ ῖ,
ῦ ῦ, ῦ ῦ, ῦ ῦ, ὼ, ὼ, ῶ ῶ, ῶ ῶ, ᾳ ᾳ

A A A, A A A A A, H H H H H H$_I$, H H, I I, I I, Ϊ Ϊ
ϒ ϒ, ϒ ϒ, Ϋ Ϋ, Ω, Ω, Ω Ω, Ω Ω, A_I A_I

ὰ ὲ ῖ ῆ ῆ ὅ ὗ ῶ ῾A ῾E Ϊ ῾H ῾O ῾ϒ ῾Ω, ᾳ ᾳ
A E Ϊ H H O ϒ Ω A E Ϊ H O ϒ Ω, A_I A_I

῾A ῾A ῾A ῾A ↦ A A A A.

Input variants and their conversion with MakeLowercase:

῾A ῾A ῾A, ῾A ῾A ῾A ῾A, ῾H ῾H ῾H ῾H ῾H$_I$, ῾H ῾H, ῾I ῾I, ῾I ῾I
ᾰ ᾰ ᾰ, ᾰ ᾰ ᾰ ᾰ, ῆ ῆ ῆ ῆ ῆ ῆ, ῆ ῆ, ῖ ῖ, ῖ ῖ

[5] The ogonek (*little hook*) accent ˛ (\k) is not defined in LGR.
[6] The dialytika is not used on Initial letters.

᾽Υ ᾽Υ, ᾽Ω, ᾽Ω ᾽Ω, ᾽Ω ᾽Ω, A_I A_I A_I.
ὖ ὖ, ὠ, ὡ, ὦ ὦ, ὦ ὦ, ᾳ ᾳ ᾳ.

ά ὲ ή ῆ ό ὖ ὦ ᾽A ᾽E Ϊ ᾽H ᾽O ᾽Υ ᾽Ω A_I A_I A_I
ά ὲ ή ῆ ό ὖ ὦ ά ὲ ϊ ῆ ό ὖ ὦ ᾳ ᾳ ᾳ

᾽A ᾽A ᾽A ᾽A ↦ ᾴ ᾴ ᾴ ᾴ

The tilde character can be used in combined accents. However, in documents not defining the Babel language *greek* or *polutonikogreek*, better use the tilde-accent macro, as the tilde produces a no-break space if converted with `\MakeUppercase` or `\MakeLowercase`:

combined accent with tilde character:
῀ ῁ ῝ ῞ ῟ ↦ ¨ I I ¨ Υ Υ Υ
Ῐ ᾽I Ῡ ᾽Υ Ῠ ↦ ¨ ι ᾽ ι ¨ υ ᾽ υ ῟

combined accent with tilde-accent macro:
῁ ῟ ↦ Ϊ Ῡ
Ῐ Ῡ ↦ ῁ ῟

Accents input via the Latin transliteration are not dropped with MakeUppercase, unless Babel is loaded and the current language is Greek (because the required local re-definitions of the `uccode` are done in `greek.ldf` from the *babel-greek* package).

ά ϊ ὰ ὰ ά ᾳ ↦ ᾽A Ϊ ᾽A ᾽A ᾽A A_I

Accent macros can start with `\a` instead of `\` when the short form is redefined, e. g. inside a *tabbing* environment. This also works for the locally defined dasia and psili shortcuts `\<` and `\>`:

COL1	COL2	COL3	COL4
COL1		COL3	
Viele	Grüße	ά	ὠ

Combinations with named accents: ᾴ ᾇ ᾴ.

The dialytika must be kept in UPPERCASE, e. g.

μαΐστρος ↦ ΜΑΪΣΤΡΟΣ or εὐζωΐα ↦ ΕΥΖΩΪΑ.

This is implemented for all input variants of diacritics with dialytika:

ϊ ΐ ΐ ῗ ϋ ΰ ὺ ῦ ↦ Ϊ Ϊ Ϊ Ϊ Ῡ Ῡ Ῡ Ῡ,

Tonos and dasia mark a *hiatus* (break-up of a diphthong) if placed on the first vowel of a diphthong (άι, άυ, έι). A dialytika must be placed on the second vowel if they are dropped: (ΑΪ, ΑΫ, ΕΪ).

άυλος ↦ ΑΫΛΟΣ, ἄυλος ↦ ΑΫΛΟΣ, μάινα ↦ ΜΑΪΝΑ, κέικ ↦ ΚΕΪΚ,
ἀυπνία ↦ ΑΫΠΝΙΑ, ρωμέικα ↦ ΡΩΜΕΪΚΑ

Test the auto-hiatus feature for side-effects:

A B (must keep space after A).

Kerning (see the input):

AO AΨ AI Aϒ PA OA ϒA Δϒ [
̓AO AΨ AI Aϒ PA OA ϒA Δϒ [
̔AO AΨ AÏ Aϔ PA OA ϒA Δϒ [
̓́AO AΨ AÏ Aϔ PA OA ϒA Δϒ [
́AO AΨ AÏ Aϔ PA OA ϒA Δϒ [
̔́AO AΨ AÏ Aϔ PA OA ϒA Δϒ [
̓̀AO AΨ AÏ Aϔ PA OA ϒA Δϒ [
̀AO AΨ AI Aϒ PA OA ϒA Δϒ [
̔̀AO AΨ AI Aϒ PA OA ϒA Δϒ [
̃AO AΨ AI Aϒ PA OA ϒA Δϒ [
̓̃AO AΨ AI Aϒ PA OA ϒA Δϒ [
̔̃AO AΨ AI Aϒ PA OA ϒA Δϒ [
̓̔̃AO AΨ AI Aϒ PA OA ϒA Δϒ [
̈AO AΨ AÏ Aϔ PA OA ϔA Δϒ [
̈ÄO ÄΨ ÄI Äϒ P̈A ÖA ϒÄ Δϔ [
̓̈ÄO ÄΨ ÄI Äϒ P̈A ÖA ϒÄ Δϔ [

Rows 3 . . . 7: Look-ahead (to check for a hiatus) breaks kerning before ‘A’ with tonos or psili.

Rows 15 and 16: Like in any font encoding, there is no kerning for non-defined accent-letter-combinations (dialytica on A O Δ).

Downcasing should keep diacritics (of course, it cannot regenerate “manually" dropped ones): ΄A Ϊ ϔ ΅A ↦ ά ΐ ΰ ᾴ