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

Ligature break up with `\textcompwordmark`: AY fi AϒϊAϒ ϊ ↦ AY fi AϒϊAϒ ϊ

[1] Single quotes need special attention to prevent conversion to accents. Test the input conventions: ‘α’ ‘α’ ‘α’ ‘α’ but not ὰ᾽ ἐ᾽ ῒ᾽

Spacing accent chars: ^a ^α ^ι ~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 with `\latintext`, however this cannot be done in a font encoding definition file.

Beware that `"#&';<>?@` becomes ¨ˆ˙´·‘’;˜.

The *textcomp* package provides pre-composed coyright ©, registered ® and trademark ™ symbols that work in all font encodings. In LGR (with textcomp), 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 transcription and 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 ς τ υ φ χ ψ ω

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 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 transcription, 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)[2]
ϙ `\textqoppa` (archaic koppa)
Ϙ `\textQoppa` (archaic Koppa)
ϛ `\textstigma`
ϝ `\textvarstigma`

[2]Modern typographical practice normally does not observe a contrast between uppercase and lowercase forms for numeric koppa. In LGR, there is no separate code point for Koppa.

Ͳ \textStigma (Sigma-Tau-Ligature in CB-fonts)[3]
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 ͵ ΄ ; ΄ ΅ Ά · Έ Ή Ί Ό Ύ Ώ ΐ Ϊ Ϋ ά έ ή ί ΰ ϊ ϋ ό ύ ώ Ϙ ϙ Ͳ ϛ F ϝ ϟ ϟ Ϡ ϡ

MakeUppercase:
ʹ ͵ ΄ ; ¨ Α · Ε Η Ι Ο Υ Ω Ϊ Ϊ Ϋ Α Ε Η Ι Ϋ Ϊ Ϋ Ο Υ Ω Ϙ Ϙ Ͳ Ͳ 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
Ͳ \textStigmagreek = Ͳ \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:

΄ \accoxia = ΄ \acctonos
᾽ \acckoronis = ᾽ \accpsili

[3] 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, and kappa (characters for the last two 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 interchangeabely.

2 Diacritics

Capital Greek letters have Greek diacritics (except the dialytika 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[4] (Latin, Greek, Greek Capitals $\mapsto$ 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 α̲ ᾳ ᾳ ↦ A A A Ä Â Ā Ă Ȧ Å Ǎ Ă A̲ Ą Ạ

᾽Α΄ΑÃÄ ÂĀĂȦÅǍĂ Α̲ Ą Ạ ↦ AAAÄ ÂĀĂȦÅǍĂ Α̲ Ą Ạ

Additional Greek diacritics (Greek, Greek Capitals[5] $\mapsto$ UPPERCASE)

ὰ ὲ ῖ ΐ ῒ ΐ ῇ ὅ ὅ ῦ ῶ ᾳ ↦ A E I Ϊ Ϊ Ϊ H O O ϒ Ω A_I

΄Α ΅Ε ΅Ι ΅Η ΅Ο ΅Ο ΅ϒ ΅Ω A_I ↦ A E I H O O ϒ Ω A_I

Input variants and their conversion with MakeUppercase:

ᾴ ᾷ ᾶ, ᾷ ᾷ ᾷ ᾷ ᾷ, ῄ ῇ ῇ ῇ ῇ ῇ, ῇ ῇ, ῒ ΐ, ΐ ΐ, ῒ ΐ ΐ,
ῦ ῦ, ῦ ῦ, ῦ ῦ, ὼ, ώ, ῶ ῶ, ῶ ῶ, ᾳ ᾳ

A A A, A A A A A, H 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 ΐ ῇ ῇ ὅ ῦ ῶ ΄Α ΅Ε Ϊ ΅Η ΅Ο ΅ϒ ΅Ω, ᾳ ᾳ
A E Ϊ H H O ϒ Ω A E Ϊ H O ϒ Ω, A_I A_I

΅Α ΅Α ΅Α ΅Α ↦ A A A A.

Input variants and their conversion with MakeLowercase:

΅Α ΅Α ΅Α, ΅Α ΅Α ΅Α ΅Α, ΅Η ΅Η ΅Η ΅Η ΅H_I, ΅Η ΅Η, ΅Ι ΅Ι, ΅Ι ΅Ι
ᾴ ᾷ ᾶ, ᾷ ᾷ ᾷ ᾷ, ῄ ῇ ῇ ῇ ῇ ῇ, ῇ ῇ, ῒ ΐ, ΐ ΐ

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

῞Υ ῞Υ, ᾽Ω, ῾Ω, ῞Ω ῞Ω, ῞Ω ῞Ω, A_I A_I A_I.
῟ ῟, ὼ, ώ, ῶ ῶ, ῶ ῶ, ᾳ ᾳ ᾳ.

ά ὲ ί ῆ ό ῦ ῶ ῾A ᾽E Ϊ ῾H ῞O ῞Υ ῞Ω A_I A_I A_I
ά ὲ ί ῆ ό ῦ ῶ ά ὲ ϊ ῆ ό ῦ ῶ ᾳ ᾳ ᾳ

῞A ῞A ῞A ῞A ↦ ᾰ ᾰ ᾰ ᾰ

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 Ῡ ↦ ¨ 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	ὼ

Combinations with named accents: ᾰ ᾱ ᾶ.

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 ῒ ῗ ϋ ΰ ῢ ῧ ↦ 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, ᾰυλος ↦ ΑΫΛΟΣ, μάινα ↦ ΜΑΪΝΑ, κέικ, ↦ ΚΕΪΚ
ἀυπνία ↦ ΑΫΠΝΙΑ

Test the auto-hiatus feature for side-effects:

A B (must keep space after A).

Kerning (see the input):

AO AΨ AI AϒU PA OA ϒA ΔϒU [
̓AO AΨ AI AϒU PA OA ϒA ΔϒU [
̓AO AΨ AÏ AϒU̓ PA OA ϒA ΔϒU [
̔AO AΨ AÏ AϒU̓ PA OA ϒA ΔϒU [
́AO AΨ AÏ AϒU̓ PA OA ϒA ΔϒU [
̓́AO AΨ AÏ AϒU̓ PA OA ϒA ΔϒU [
̔́AO AΨ AÏ AϒU̓ PA OA ϒA ΔϒU [
̀AO AΨ AI AϒU PA OA ϒA ΔϒU [
̓̀AO AΨ AI AϒU PA OA ϒA ΔϒU [
̃AO AΨ AI AϒU PA OA ϒA ΔϒU [
̔̃AO AΨ AI AϒU PA OA ϒA ΔϒU [
̈AO AΨ AI AϒU PA OA ϒA ΔϒU [
̓̈AO AΨ AI AϒU PA OA ϒA ΔϒU [
̔̈AO AΨ AÏ AϒU̓ PA OA ϒU̓A ΔϒU [
̈ÄO ÄΨ ÄI ÄϒU ̈PA ÖA ϒÄ ΔϒU̓ [
̃ÄO ÄΨ ÄI ÄϒU ̈PA ÖA ϒÄ ΔϒU̓ [

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 Ï ϔ ̓́A ↦ ά ϊ ϋ ᾶ