

Greek Unicode with *inputenc*

Günter Milde

July 16, 2013

Abstract

The definitions in `lgrenc.dfu` provide UTF-8 support for Greek based on `inputenc` and the LICRs (LaTeX internal character representations) defined in `lgrenc.def` from the `greek-fontec` package.

1 Usage

If the *greek-fontenc* package is installed, the line

```
\usepackage[utf8]{inputenc}
```

in the document preamble allows the use of Greek Unicode literals in the document body, if the current font encoding is LGR.¹ (The font encoding is automatically set to LGR by babel for the *greek* language.)

In combination with the *textalpha* package from `lgrx`, Greek Unicode literals can be used in text with any font encoding just like the symbols provided by package “textcomp” (i.e. with some limitations to kerning, MakeUppercase/MakeLowercase and multiple accents, described in `textalpha-test` from the “lgrx” package):

```
\usepackage[utf8]{inputenc}
\usepackage{textalpha}
```

In combination with the *alphabet* package (also from `lgrx`), Greek Unicode literals can also be used in math mode:

```
\usepackage[utf8]{inputenc}
\usepackage{alphabet}
```

2 Supported Characters

LGR is no “standard font encoding”. Latin characters are mapped to Greek if LGR is the active font encoding. This means you need an explicit font-encoding switch for Latin words and abbreviations in Greek text.

Unicode definitions exist for all non-ASCII characters that can be rendered with an LGR-encoded font.

¹All *inputenc* Unicode definitions only work for text in the correct font encoding.

2.1 Greek and Coptic

legend: * glyph missing in LGR, [space] Unicode point not defined

[illegible]

2.2 Greek Extended

[illegible]

2.3 Other Unicode Blocks

Latin-1 Supplement: “ « - ' . »

IPA Extensions: \emptyset

Spacing Modifier Letters: \sim_{α} (here followed by letter alpha)

General Punctuation: – — ‘ ’ ‰ ZWNJ (zero width no joiner, prevents kerning and ligatures, e.g. ΑΥ vs. ΑΥ and ‘α vs. α)

Currency Symbols: €

Letterlike Symbols: Ω

Ancient Greek Numbers: Δ H X M

3 MakeUppercase, MakeLowercase

Capital Greek letters have diacritics (except the dialytika) to the left (instead of above) and drop them in uppercase, e.g. $\mu\acute{\alpha}\iota\sigma\tau\rho\omicron\varsigma \mapsto \text{ΜΑΪΣΤΡΟΣ}$.

Tonos and dasia on the first vowel of a diphthong (αι, αυ, ει) imply a *hiatus*. A dialytika must be placed on the second vowel if they are dropped (Αἶ, Αῖ, Εἶ).

The auto-hiatus feature in lgrxenc.def works nice with the Latin transcription and with character-macros (Αἶ, Αῖ, Εἶ) and also if the first character is wrapped in `\ensuregreek` (as done by the lgrenc.dfu definition for accented characters) or a literal Unicode character (Αἶ, Αῖ, Εἶ) but not if the second character of the diphthong is a Unicode literal (ΑΙ, ΑΥ, ΕΙ).

Therefore, the diaeresis is missing in the following examples: αυλος \mapsto ΑΥΛΟΣ, αυλος \mapsto ΑΥΛΟΣ, μαινα \mapsto MAINA, κεικ, \mapsto KEIK, αυπνια \mapsto ΑΥΠΝΙΑ.

Fixing this shortcoming requires knowledge of what the `\LGR@ifnextchar` “sees” when the next character is an upcased Unicode literal.

As an ugly workaround, use `\textiota` resp. `\textupsilon` for the character that should get the diaeresis: αυπνια \mapsto ΑῖΠΝΙΑ.

The following subsections test MakeUppercase and MakeLowercase with all characters defined in lgrenc.dfu:

3.1 Greek and Coptic

Characters of the Greek and Coptic Unicode Block:

’; ’’Α·ΕΗΤΟΥΩΑΒΓΔΕΖΗΘΙΚΑΜΝΞΟΠΡΣΤΥΦΧΨΩΪΫϞϠϡ
άέήίύαβγδεζηθικλμνξοπρςστυφχψωϊϋόώϟϠϡ

MakeUppercase:

’; ·Α·ΕΗΙΟΥΩΑΒΓΔΕΖΗΘΙΚΑΜΝΞΟΠΡΣΤΥΦΧΨΩΪΫϞϠϡ
ΑΕΗΙΥΑΒΓΔΕΖΗΘΙΚΑΜΝΞΟΠΡΣΣΤΥΦΧΨΩΪΫΟΥΩϞϠϡ

Letters and ypogegrammeni upcased, tonos dropped, dialytika kept.

There is no capital Koppa in LGR, therefore Ϟ is left unchanged with MakeUppercase.

MakeLowercase:

’; ’’ά·έήίούώαβγδεζηθικλμνξοπρςστυφχψωϊϋόώϟϠϡ
άέήίύαβγδεζηθικλμνξοπρςστυφχψωϊϋόώϟϠϡ

The lowercase of Σ is the «auto-sigma» (`\textautosigma`): ΣΣ \mapsto σς. Add a ZWNJ or use the `\noboundary` macro to prevent conversion to final sigma: σσ. The lowercase of GREEK LETTER STIGMA Ϟ is ϟ.

3.2 Greek extended

MakeUppercase:

Α Α Α Α Α Α Α Α Α Α Α Α Α Α Α
Ε Ε Ε Ε Ε Ε Ε Ε Ε Ε Ε
Η Η Η Η Η Η Η Η Η Η Η Η Η Η Η
Ι Ι Ι Ι Ι Ι Ι Ι Ι Ι Ι Ι Ι Ι Ι
Ο Ο Ο Ο Ο Ο Ο Ο Ο Ο Ο
Υ Υ Υ Υ Υ Υ Υ Υ Υ Υ Υ Υ
Ω Ω Ω Ω Ω Ω Ω Ω Ω Ω Ω Ω Ω Ω Ω
Α Α Ε Ε Η Η Ι Ι Ο Ο Υ Υ Ω Ω
Α₁ Α₁ Α₁ Α₁ Α₁ Α₁ Α₁ Α₁ Α₁ Α₁ Α₁ Α₁ Α₁ Α₁ Α₁
Η₁ Η₁ Η₁ Η₁ Η₁ Η₁ Η₁ Η₁ Η₁ Η₁ Η₁ Η₁ Η₁ Η₁ Η₁
Ω₁ Ω₁ Ω₁ Ω₁ Ω₁ Ω₁ Ω₁ Ω₁ Ω₁ Ω₁ Ω₁ Ω₁ Ω₁ Ω₁ Ω₁
Ά Ά Α₁ Α₁ Α₁ Α Α₁ Ά Ά Α Α Α₁ ,
“ Η₁ Η₁ Η₁ Η Η₁ Ε Ε Η Η Η₁
Ϊ Ϊ Ϊ Ϊ Ϊ Ϊ Ϊ Ϊ Ϊ Ϊ Ϊ
Υ̂ Υ̂ Υ̂ Υ̂ Ρ Ρ Υ̂ Υ̂ Υ̂ Υ̂ Υ̂ Ρ ...
Ω₁ Ω₁ Ω₁ Ω₁ Ω₁ Ο Ο Ω Ω₁

MakeLowercase:

ά ά ἀ ἀ ἀ ἀ ἀ ἀ ἀ ἀ ἀ ἀ ἀ
έ έ ε ε ε ε ε ε ε ε ε ε ε
ή ή η η η η η η η η η η η
ι ι ι ι ι ι ι ι ι ι ι ι ι ι ι
ο ό ô õ ö ö ö ö ö ö ö ö ö
ύ ύ υ υ υ υ υ υ υ υ υ υ υ υ υ
ώ ώ ð ð ð ð ð ð ð ð ð ð ð ð ð
ὰ ᾶ ἐ ἐ ἡ ἱ ἰ ὀ ὁ ὐ ὗ ὠ
ᾱ ῶ ῶ ῶ ῶ ῶ ῶ ῶ ῶ ῶ ῶ ῶ ῶ ῶ ῶ
ἥ ἥ ἥ ἥ ἥ ἥ ἥ ἥ ἥ ἥ ἥ ἥ ἥ ἥ ἥ
ὀ ὀ ὀ ὀ ὀ ὀ ὀ ὀ ὀ ὀ ὀ ὀ ὀ ὀ ὀ
ᾳ ᾳ ᾳ ᾳ ᾳ ᾳ ᾳ ᾳ ᾳ ᾳ ᾳ ᾳ ᾳ ᾳ
ῖ ῖ ῖ ῖ ῖ ῖ ῖ ῖ ῖ ῖ ῖ ῖ ῖ ῖ ῖ
ϐ ϐ ϐ ϐ ϐ ϐ ϐ ϐ ϐ ϐ ϐ ϐ ϐ ϐ
ϑ ϑ ϑ ϑ ϑ ϑ ϑ ϑ ϑ ϑ ϑ ϑ ϑ ϑ

3.3 Other Unicode Blocks

MakeUppercase works without effect on non-letter symbols and the letter shwa:

“ « - ‘ . » ə ˘ A — ‘ ’ %₀ AΥ € ☒ ☐ ☒ ☐

MakeLowercase works without effect on non-letter symbols, too:

“ « - ‘ . » ə ˘ α — ‘ ’ %₀ αυ € ☒ ☐ ☒ ☐