

Greek Unicode with `lgrenc.dfu`

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

Like other *inputenc* Unicode definitions, they only work for text in the correct font encoding (in this case LGR). The font encoding is automatically set to LGR by babel for the languages *greek* and *polutonikogreek*.

With the combination

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

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).

1 Supported Characters

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

1.1 Greek and Coptic

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

	0	1	2	3	4	5	6	7	8	9	A	B	C	Δ	E	Φ
370	*	*	*	*	'		*	*				*	*	*		
380					'	´	Α	·	Ε	Η	Ι		Ο		Υ	Ω
390	ι	Α	Β	Γ	Δ	Ε	Ζ	Η	Θ	Ι	Κ	Λ	Μ	Ν	Ξ	Ο
3A0	Π	Ρ		Σ	Τ	Υ	Φ	Χ	Ψ	Ω	Ϊ	Ϋ	ά	έ	ή	ί
3B0	ύ	α	β	γ	δ	ε	ζ	η	θ	ι	κ	λ	μ	ν	ξ	ο
3C0	π	ρ	ς	σ	τ	υ	φ	χ	ψ	ω	ϊ	ϋ	ό	ύ	ώ	
3D0	*	*	*	*	*	*	*	*	*	Ω	ϙ	ϐ	ϑ	ϒ	*	ι
3E0	Λ	λ	*	*	*	*	*	*	*	*	*	*	*	*	*	*
3F0	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

1.2 Greek Extended

	0	1	2	3	4	5	6	7	8	9	A	B	C	Δ	E	Φ
1F00	ᾀ	ᾁ	ᾂ	ᾃ	ᾄ	ᾅ	ᾆ	ᾇ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ
1F10	ἔ	ἕ	἖	἗	Ἐ	Ἑ			Ἐ	Ἐ	Ἐ	Ἐ	Ἐ	Ἐ		
1F20	ῥ	ῑ	ῒ	ΐ	῔	῕	ῖ	ῗ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ
1F30	ἰ	ἱ	ἲ	ἳ	ἴ	ἵ	ἶ	ἷ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ
1F40	ὀ	ὁ	ὂ	ὃ	ὄ	ὅ			Ὀ	Ὀ	Ὀ	Ὀ	Ὀ	Ὀ		
1F50	ὠ	ὡ	ὢ	ὣ	ὤ	ὥ	ὦ	ὧ		Υ		Υ		Υ		Υ
1F60	ὠ	ὡ	ὢ	ὣ	ὤ	ὥ	ὦ	ὧ	Ω	Ω	Ω	Ω	Ω	Ω	Ω	Ω
1F70	ᾀ	ᾁ	ᾂ	ᾃ	ᾄ	ᾅ	ᾆ	ᾇ	ὀ	ὁ	ὂ	ὃ	ὄ	ὅ		
1F80	ᾀ	ᾁ	ᾂ	ᾃ	ᾄ	ᾅ	ᾆ	ᾇ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ
1F90	ῥ	ῑ	ῒ	ΐ	῔	῕	ῖ	ῗ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ	Ἡ
1FA0	ῥ	ῑ	ῒ	ΐ	῔	῕	ῖ	ῗ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ
1FB0	ᾀ	ᾁ	ᾂ	ᾃ	ᾄ		ᾆ	ᾇ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ			
1FC0	˜	˜	ῥ	ῑ	ΐ		ῖ	ῗ	Ἐ	Ἐ	Ἡ	Ἡ	Ἡ			
1FD0	ῖ	ῗ	Ῐ	Ῑ			Ὶ	Ί	Ἰ	Ἰ	Ἰ	Ἰ				
1FE0	ῤ	ῥ	ῗ	Ῐ	Ὶ	Ί	ῤ	ῥ	Υ	Υ	Υ	Υ	Υ			
1FF0			ὦ	ὧ	Ὠ		ὠ	ὡ	Ὀ	Ὀ	Ὀ	Ὀ	Ὀ			

1.3 Other Unicode Blocks

Latin-1 Supplement: “ « ’ · »

IPA Extensions: ə

Spacing Modifier Letters: Ϳ (here followed by letter alpha)

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

Currency Symbols: €

Letterlike Symbols: Ω

Ancient Greek Numbers: ͵Ͷͷ͸͹

2 MakeUppercase, MakeLowercase

Capital Greek letters have diacritics (except the dialytika) to the left (instead of above) and drop them in uppercase, e.g. μαῖστρος ↦ ΜΑῖΣΤΡΟΣ.

Tonos and dasia on the first vowel of a diphthong (αῖ, αὖ, εῖ) imply a *hiatus*. A dialytika must be placed on the second vowel if they are dropped (Aῖ, Aῦ, Eῖ).

The auto-hiatus feature in lgrxenc.def works nice with the Latin transcription and with character-macros (Aῖ, Aῦ, Eῖ) 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 (Aῖ, Aῦ, Aῖ) but not if the second character of the diphthong is a Unicode literal (AI, AY, EI).

Therefore, the diaeresis is missing in the following examples: $\acute{\alpha}\upsilon\lambda\omicron\varsigma \mapsto \text{AY}\Lambda\text{O}\Sigma$, $\acute{\alpha}\upsilon\lambda\omicron\varsigma \mapsto \text{AY}\Lambda\text{O}\Sigma$, $\mu\acute{\alpha}\iota\nu\alpha \mapsto \text{MAIN}\alpha$, $\kappa\acute{\epsilon}\iota\kappa, \mapsto \text{KEIK}$, $\acute{\alpha}\upsilon\pi\nu\acute{\iota}\alpha \mapsto \text{AY}\Pi\text{NIA}$.

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: $\acute{\alpha}\upsilon\pi\nu\acute{\iota}\alpha \mapsto \text{A}\acute{\text{Y}}\Pi\text{NIA}$.

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

2.1 Greek and Coptic

Characters of the Greek and Coptic Unicode Block:

```
';; 'A·EHTO·T·Ź·ABΓΔΕΖΗΘΙΚΑΜΝΞΟΠΡΣΤΥΦΧΨΩΪΫϞϠϡ
άέήίύαβγδεζηθικλμνξοπρςστυφχψωϊϋόύώϟϠϡϣ
```

`MakeUppercase`:

```
';; "·A·EHIOYŹ·ABΓΔΕΖΗΘΙΚΑΜΝΞΟΠΡΣΤΥΦΧΨΩΪΫϞϠϡ
ΑΕΗΙΥΑΒΓΔΕΖΗΘΙΚΑΜΝΞΟΠΡΣΤΥΦΧΨΩΪΫΟΥΩϞϠϡϣ
```

Letters and ypogegrammeni upcased, tonos dropped, dialytika kept.

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

`MakeLowercase`:

```
';; 'A·έ·ή·ί·ό·ύ·ά·β·γ·δ·ε·ζ·η·θ·ι·κ·λ·μ·ν·ξ·ο·π·ρ·ς·σ·τ·υ·φ·χ·ψ·ω·ϊ·ϋ·ό·ύ·ώ·ϟ·Ϡ·ϡ·ϣ
άέήίύαβγδεζηθικλμνξοπρςστυφχψωϊϋόύώϟϠϡϣ
```

The lowercase of Σ is the «auto-sigma» (`\textautosigma`): $\Sigma\Sigma \mapsto \sigma\varsigma$. Add a ZWNJ or use the `\noboundary` macro to prevent conversion to final sigma: $\sigma\sigma$. The lowercase of GREEK LETTER STIGMA $\var�$ is τ .

2.2 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
Υ̂ Υ̂ Υ̂ Υ̂ Ρ Ρ Υ̂ Υ̂ Υ̂ Υ̂ Υ̂ Ρ ...
Ω₁ Ω₁ Ω₁ Ω Ω Ω Ο Ο Ω Ω Ω

MakeLowercase:

ά ά ἀ ἀ ᾶ ᾶ ᾶ ᾶ ᾶ ἀ ἀ ἀ ᾶ ᾶ ᾶ
έ έ ε̇ ε̇ ε̇ ε̇ ε̇ ε̇ ε̇ ε̇ ε̇
ή ή η̇ η̇ η̇ η̇ η̇ η̇ η̇ η̇ η̇ η̇ η̇
ι̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇
ο̇ ο̇ ο̇ ο̇ ο̇ ο̇ ο̇ ο̇ ο̇ ο̇ ο̇ ο̇ ο̇ ο̇ ο̇
ύ ύ υ̇ υ̇ υ̇ υ̇ υ̇ υ̇ υ̇ υ̇ υ̇ υ̇ υ̇ υ̇ υ̇ υ̇
ώ ώ ω̇ ω̇ ω̇ ω̇ ω̇ ω̇ ω̇ ω̇ ω̇ ω̇ ω̇ ω̇ ω̇ ω̇
ὰ ἀ ἐ ἐ ἡ ή ι̇ ι̇ ο̇ ο̇ υ̇ υ̇ ώ ώ
ᾶ ᾶ ᾷ ᾷ ᾷ ᾷ ᾷ ᾶ ᾷ ᾷ ᾷ ᾷ ᾷ ᾷ ᾷ
ή ή η̇ η̇ η̇ η̇ η̇ η̇ η̇ η̇ η̇ η̇ η̇ η̇ η̇
ὀ ὀ ὂ ὂ ὃ ὃ ὃ ὃ ὃ ὃ ὃ ὃ ὃ ὃ ὃ ὃ
ᾱ ᾱ ῥ ῥ ᾶ ᾶ ᾶ ᾶ ᾶ ᾶ ᾶ ᾶ ᾶ ᾶ
~ή η η̇ η̇ η̇ ε̇ ε̇ ἡ η̇ ~
ϊ̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇ ι̇
ϋ ϋ ϋ̇ ϋ̇ ϋ̇ ϋ̇ ϋ̇ ϋ̇ ϋ̇ ϋ̇ ϋ̇ ϋ̇ ϋ̇ ϋ̇ ϋ̇
ϖ ϖ ϖ̇ ϖ̇ ϖ̇ ϖ̇ ϖ̇ ϖ̇ ϖ̇ ϖ̇ ϖ̇ ϖ̇ ϖ̇ ϖ̇ ϖ̇

2.3 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 € ☒ ☒ ☒ ☒