

Greek Unicode with 8-bit TeX and *inputenc*

Günter Milde

February 22, 2023

Abstract

The definitions in `lgrenc.dfu` provide UTF-8 support for the Greek script based on *inputenc* and the *LaTeX internal character representation* macros (LICRs) defined in the *greek-fontenc* package.

1 Requirements

The *inputenc* standard package enables the use of non-ASCII characters with 8-bit TeX. However, it misses definitions for Greek characters. The *greek-inputenc* package extends *inputenc* to allow the use of Greek literals in the document source.

As with all *inputenc* definitions, this only works if the active font encoding supports the characters. For the Greek script, this is usually the non-standard *LGR* font encoding set up by *greek-fontenc*.

2 Usage

There are several alternatives to activate Greek Unicode input for 8-bit TeX¹ (see also the source document `greek-utf8.tex`):

- Define the LGR font encoding and the UTF8 input encoding (the order does not matter), e.g.,

```
\usepackage[T1,LGR]{fontenc}
\usepackage[utf8]{inputenc}
```

Ensure that LGR is the active font encoding whenever a Greek character is used in the text (see below).

- For text in the Greek language, it is recommended to use the *Babel* package with the Greek language definitions in *babel-greek*. Babel sets the font encoding automatically to LGR and Greek Unicode characters work as expected. Write in the preamble, e.g.,

¹The XeTeX and LuaTeX engines use utf8 as native input encoding. They do not require (and, except in 8-bit compatibility mode, do not work with) the *inputenc* and *greek-inputenc* packages.

```

\usepackage[utf8]{inputenc}
\usepackage[LGR,T1]{fontenc}
\usepackage[english,greek,german]{babel}

```

and use `\foreignlanguage` or `\selectlanguage` to set the text language to Greek (see the *babel-greek* documentation for detailed examples).

Τί φηίς; Ἰδὼν ἐνθῆδε παῖδ' ἔλευθέραν τὰς πλησίον Νύμφας στεφανοῦσαν, Σώστρατε, ἔρῶν ἀπῆλθες εὐθύς;

- Greek literal characters can also be used in PDF-strings (bookmarks and ToC entries with *hyperref*). See *greek-fontenc* for a *hyperref test and usage example*.
- In combination with the *textalpha* package from *greek-fontenc*, Greek Unicode characters can be used in text with any font encoding – just like the symbols provided by the “textcomp” package (i.e. with some limitations described in *textalpha-doc*). With the preamble lines

```

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

```

it is straightforward to write about π -mesons, γ -radiation, or a 50 k Ω resistor.²

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

```

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

```

$$\tan \beta = \frac{\sin \beta}{\cos \beta}.$$

3 Warning: unsafe ASCII input

LGR is no “standard font encoding”. Latin characters and some other ASCII symbols are mapped to Greek equivalents if LGR is the active font encoding. (See [usage.pdf](#) for a description of this Latin-Greek transliteration.)

This means you need an explicit language and/or font-encoding switch for Latin words and abbreviations in Greek text, e.g., not « $\eta\acute{\iota}\alpha$ αντίσταση 750- $\kappa\Omega$ » but « $\eta\acute{\iota}\alpha$ αντίσταση 750-k Ω »

Special care is also required with the question mark characters:

- The Unicode standard says character 003B SEMICOLON and not 037E GREEK QUESTION MARK, is the preferred character for a ‘Greek question mark’ (erotimatiko),

²The MIKRO SIGN and OHM SIGN characters are set up by *inputenc* with *textcomp* characters for any font encoding while GREEK CAPITAL LETTER OMEGA works only with the LGR font encoding.

- The LGR font encoding maps a SEMICOLON to a middle dot (ano teleia), while the Latin question mark “?” is mapped to the erotimatiko.

As a result, only the deprecated character 037E GREEK QUESTION MARK works with both, Xe/LuaTeX and 8-bit TeX. Compare the source [greek-utf8.tex](#) and the PDF output:

Latin (T1)	Greek (LGR)	question mark character
Τί φήις;	Τί φήις;	037E GREEK QUESTION MARK
Τί φήις;	Τί φήις·	003B SEMICOLON
Τί φήις?	Τί φήις;	003F QUESTION MARK

With the *textalpha* package’s “keep-semicolon” option, the SEMICOLON character can be used for the erotimatiko also with LGR encoded fonts.

4 Supported Characters

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

4.1 Greek and Coptic

	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	*	*	*	*	*	*	*	*	Ϡ	ϡ	Ϣ	ϣ	Ϥ	ϥ	*	ι
3E0	Δ	λ	*	*	*	*	*	*	*	*	*	*	*	*	*	*
3F0	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

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

4.2 Greek Extended

	0	1	2	3	4	5	6	7	8	9	A	B	C	Δ	E	Φ
1F00	ά	ά̂	ά̃	ά̄	ά̅	ά̆	ά̇	ά̈	Ά	Ά̂	Ά̃	Ά̄	Ά̅	Ά̆	Ά̇	Ά̈
1F10	έ	έ̂	έ̃	έ̄	έ̅	έ̆	έ̇	έ̈	Έ	Έ̂	Έ̃	Έ̄	Έ̅	Έ̆	Έ̇	Έ̈
1F20	ή	ή̂	ή̃	ή̄	ή̅	ή̆	ή̇	ή̈	Ή	Ή̂	Ή̃	Ή̄	Ή̅	Ή̆	Ή̇	Ή̈
1F30	ι	ι̂	ι̃	ῑ	ι̅	ῐ	ι̇	ϊ	Ί	Ί̂	Ί̃	Ί̄	Ί̅	Ί̆	Ί̇	Ί̈
1F40	ο	ο̂	ο̃	ο̄	ο̅	ο̆	ο̇	ο̈	Ό	Ό̂	Ό̃	Ό̄	Ό̅	Ό̆	Ό̇	Ό̈
1F50	ύ	ύ̂	ύ̃	ύ̄	ύ̅	ύ̆	ύ̇	ύ̈	Υ	Υ̂	Υ̃	Ῡ	Υ̅	Ῠ	Υ̇	Ϋ
1F60	ώ	ώ̂	ώ̃	ώ̄	ώ̅	ώ̆	ώ̇	ώ̈	Ω	Ω̂	Ω̃	Ω̄	Ω̅	Ω̆	Ω̇	Ω̈
1F70	ά	ά̂	έ̅	έ̆	ή̅	ή̆	ι̅	ῐ	ο̅	ο̆	ύ̅	ύ̆	ώ̅	ώ̆		
1F80	ά̂	ά̃	έ̃	έ̄	ή̃	ή̄	ι̃	ῑ	ο̃	ο̄	ύ̃	ύ̄	ώ̃	ώ̄	Ά̂	Ά̃
1F90	ή̂	ή̃	ή̃	ή̄	ή̃	ή̄	ή̃	ή̄	Ή̂	Ή̃	Ή̃	Ή̄	Ή̃	Ή̄	Ή̂	Ή̃
1FA0	φ̂	φ̃	φ̃	φ̄	φ̃	φ̄	φ̃	φ̄	Ό̂	Ό̃	Ό̃	Ό̄	Ό̃	Ό̄	Ό̂	Ό̃
1FB0	ά̂	ά̃	έ̃	έ̄	ή̃	ή̄	ι̃	ῑ	ο̃	ο̄	ύ̃	ύ̄	ώ̃	ώ̄	Ά̂	Ά̃
1FC0	~	~	ή̃	ή̄	ή̃	ή̄	ή̃	ή̄	Έ̂	Έ̃	Ή̂	Ή̃	Ή̃	Ή̄	Ή̂	Ή̃
1FD0	ι̂	ι̃	ι̃	ῑ	ι̃	ῑ	ι̃	ῑ	Ί̂	Ί̃	Ί̃	Ί̄	Ί̃	Ί̄	Ί̂	Ί̃
1FE0	υ̂	υ̃	υ̃	ῡ	υ̃	ῡ	υ̃	ῡ	Υ̂	Υ̃	Υ̃	Ῡ	Υ̃	Ῡ	Υ̂	Υ̃
1FF0			φ̃	φ̄	φ̃	φ̄	φ̃	φ̄	Ό̂	Ό̃	Ό̃	Ό̄	Ό̃	Ό̄	Ό̂	Ό̃

4.3 Other Unicode Blocks

Latin-1 Supplement : “ « ’ · »

IPA Extensions : ə LATIN SMALL LETTER SCHWA

Spacing Modifier Letters : ˘α (here followed by letter alpha)

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

Currency Symbols : €

Letter-like Symbols : Ω

Ancient Greek Numbers : ͵Ͷ ͵ͷ ͵͸ ͵͹ ͵ͺ ͵ͻ ͵ͼ ͵ͽ

5 Test up/downcasing

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

Tonos and dasia on the first vowel of a diphthong (ά, ά̂, έι) imply a *hiatus*. A dialytika must be placed on the second vowel if they are dropped (AĪ, Aÿ̂, EĪ).

The implementation of `\MakeUppercase` changed in the 2022/06 LaTeX release (cf. LaTeX News 35). Since then, Greek uppercase rules are only applied if the text language is set to “greek” with Babel.

With the “old” `\MakeUppercase`, the auto-hiatus feature in LGR works with the Latin transcription and LICR macros (AĪ, Aÿ̂, EĪ) but not with literal Unicode characters.

It works also if only the first character of the diphthong is wrapped in `\ensuregreek` (as done by the `lgrenc.dfu` definition for accented characters)) or input as literal Unicode character (`AĪ`, `AŶ`, `Aİ`) but not if the second character of the diphthong is a Unicode literal (`AI`, `AΥ`, `EI`).

As an ugly workaround, use `\textiota` resp. `\textupsilon` for the character that should get the diaeresis: $\acute{\alpha}\pi\nu\acute{\iota}\alpha \mapsto A\check{\Upsilon}P\check{\Upsilon}NIA$.

With the “new” (post 2022/06) `\MakeUppercase`, the auto-hiatus feature in LGR works with literals from the “Greek and Coptic” Unicode block and some from “Greek extended”. See the following examples: $\acute{\alpha}\lambda\omicron\varsigma \mapsto A\Upsilon\Lambda O\Sigma$, $\acute{\alpha}\lambda\omicron\varsigma \mapsto A\Upsilon\Lambda O\Sigma$, $\acute{\alpha}\lambda\omicron\varsigma \mapsto A\Upsilon\Lambda O\Sigma$, $\acute{\alpha}\lambda\omicron\varsigma \mapsto A\Upsilon\Lambda O\Sigma$, $\acute{\alpha}\lambda\omicron\varsigma \mapsto A\Upsilon\Lambda O\Sigma$.

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

5.1 Greek and Coptic

Characters of the Greek and Coptic Unicode Block:

```

';; ' "A·E·HTO·Y·ΩiABΓΔΕΖΗΘΙΚΑΜΝΞΟΠΡΣΤΥΦΧΨΩİŶŸŲTFλ
άέήίύάβγδεζηθικλμνξοπρςστυφχψωϊϋόύώϑϕϗλ

```

`MakeUppercase`:

```

';; "A·E·HIOYŶiABΓΔΕΖΗΘΙΚΑΜΝΞΟΠΡΣΤΥΦΧΨΩİŶŸŲTFλ
ΑΕΗΙŶΑΒΓΔΕΖΗΘΙΚΑΜΝΞΟΠΡΣΣΤΥΦΧΨΩİŶŸŲŲTFλ

```

Letters and ypogegrammeni upcased, tonos dropped, dialytika kept.

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

`MakeLowercase`:

```

';; ' "ά·έ·ή·ί·ύ·ά·ί·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·ϑ·ϕ·ϗ
άέήίύάβγδεζηθικλμνξοπρςστυφχψωϊϋόύώϑϕϗ

```

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

5.2 Greek extended

`MakeUppercase`:

```

A A A A A A A A A A A A A A A
E E E E E E E E E E
H H H H H H H H H H H H H H H
I I I I I I I I I I I I I I I
O O O O O O O O O O O O
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
A A E E H H I I O O Υ Υ Ω Ω
A1 A1 A1 A1 A1 A1 A1 A1 A1 A1 A1 A1 A1 A1 A1 A1
H1 H1 H1 H1 H1 H1 H1 H1 H1 H1 H1 H1 H1 H1 H1 H1
Ω1 Ω1 Ω1 Ω1 Ω1 Ω1 Ω1 Ω1 Ω1 Ω1 Ω1 Ω1 Ω1 Ω1 Ω1 Ω1

```

Ā Ā A₁ A₁ A₁ A₁ A₁ Ā Ā A A A₁ ı
 Ħ Ħ Ħ Ħ Ħ Ħ Ħ Ħ Ħ Ħ Ħ Ħ Ħ Ħ Ħ Ħ
 Ĩ Ĩ Ĩ Ĩ Ĩ Ĩ Ĩ Ĩ Ĩ Ĩ Ĩ Ĩ Ĩ Ĩ Ĩ Ĩ
 Ŷ Ŷ Ŷ Ŷ P P Ŷ Ŷ Ŷ Ŷ P P ...
 Ω₁ Ω₁ Ω₁ Ω Ω O O Ω Ω Ω₁

MakeLowercase:

ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ
 ê ê ê ê ê ê ê ê ê ê ê ê ê ê ê
 ħ ħ ħ ħ ħ ħ ħ ħ ħ ħ ħ ħ ħ ħ ħ
 ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ ĩ
 ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ
 ư ư ư ư ư ư ư ư ư ư ư ư ư ư ư
 ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ
 ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ
 ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ
 ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ ǎ
 ư ư ư ư ư ư ư ư ư ư ư ư ư ư ư
 ư ư ư ư ư ư ư ư ư ư ư ư ư ư ư
 ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ ỏ

5.2.1 Iota subscript vs. iota adscript

Pre-composed capital letters with *mute iota* decompose to the base letter and COMBINING GREEK YPOGEGRAMMENI (U+0345) even if they are named ... WITH [... AND] PROSGEGRAMMENI for “historic reasons”.³

Accordingly, the “canonical” LICR for capital letters with mute iota is the base character LICR followed by `\ypogegrammeni`.

Compare letters followed by `\prosgegrammeni` and `\ypogegrammeni` with the pre-composed characters and with character + literal GREEK YPOGEGRAMMENI.

prosgegrammeni: α₁α₁ αα₁ / Α₁Α₁ Α₁Α₁ / Α₁Α₁ Α₁Α₁
 MakeUppercase Α₁Α₁ Α₁Α₁ / Α₁Α₁ Α₁Α₁ / Α₁Α₁ Α₁Α₁
 MakeLowercase αα αα / αα αα / α̣α̣ α̣α̣
 ypogegrammeni: αα αα₁ / Α₁Α₁ Α₁Α₁ / Α₁Α₁ Α₁Α₁
 MakeUppercase Α₁Α₁ Α₁Α₁ / Α₁Α₁ Α₁Α₁ / Α₁Α₁ Α₁Α₁
 MakeLowercase αα αα / αα αα / α̣α̣ α̣α̣

5.3 Other Unicode Blocks

MakeUppercase does not change non-letter symbols and the letter shwa (there is a capital Cyrillic schwa in T2A encoded fonts):

« - ' · » ə ˘A — ‘ ’ %₀ ΑΥ € ☒ ☒ ☒ ☒

³cf. Nick Nicholas ‘Titlecase and Adscripts’

MakeLowercase does not change non-letter symbols either:

.. « - ' · » ə ~ α - — ‘ ’ %₀₀ αυ € ☒ ☒ ☒ ☒

6 Test kerning/ligatures

check for kerning and unwanted ligatures:

Aàα Aáα Aâα Aãα Aäα Aåα Aǎα Aãα Aäα Aåα A'Αα A'Αα A"Αα A"Αα A'Αα
A'Αα A'Αα A'Αα
Aèα Aéα Aêα Aëα Aξα Aζα A'Εα A'Εα A"Εα A"Εα A'Εα A'Εα
Aήα Aήα Aῆα Aῆα Aῆα Aῆα Aῆα Aῆα A'Ηα A'Ηα A"Ηα A"Ηα A'Ηα
A'Ηα A'Ηα A'Ηα
Aία Aία Aϊα Aϊα Aϊα Aϊα Aϊα Aϊα A'Ια A'Ια A'Ια A'Ια A'Ια A'Ια
A'Ια
Aòα Aóα Aδα Aδα Aδ̄α Aδ̄α A'Οα A'Οα A"Οα A"Οα A'Οα A'Οα
Aύα Aύα Aῦα Aῦα Aῦα Aῦα Aῦα Aῦα A'Υα A'Υα A"Υα A"Υα
Aώα Aώα Aῶα Aῶα Aῶα Aῶα Aῶα Aῶα A'Ωα A'Ωα A"Ωα A"Ωα A'Ωα
A'Ωα A'Ωα A'Ωα
Aàα Aáα Aèα Aéα Aήα Aήα Aìα Aία Aòα Aόα Aύα Aύα Aώα Aώα
Aḗα Aḗα Aḡα Aḡα Aḗα Aḗα Aḗα Aḗα A'Α₁α A'Α₁α A"Α₁α A"Α₁α A'Α₁α
A'Α₁α A'Α₁α A'Α₁α
Aḥα Aḥα Aḥ̄α Aḥ̄α Aḥ̄α Aḥ̄α Aḥ̄α Aḥ̄α A'Η₁α A'Η₁α A"Η₁α A"Η₁α A'Η₁α
A'Η₁α A'Η₁α A'Η₁α
Aḟα Aḟα Aḟ̄α Aḟ̄α Aḟ̄α Aḟ̄α Aḟ̄α Aḟ̄α A'Ω₁α A'Ω₁α A"Ω₁α A"Ω₁α
A'Ω₁α A'Ω₁α A'Ω₁α A'Ω₁α
Aǎα Aāα Aḗα Aḗα Aḗα Aḗα Aḗα Aḗα A'Αα A'Αα AA₁α A'α A₁α
A'α
A~α A~α Aḡα Aḡα Aḡα Aḡα Aḡα A'Εα A'Εα A'Ηα A'Ηα AH₁α A~α A~α
A~α
Aïα Aïα Aïα Aïα Aïα Aïα Aïα Aïα A'Ια A'Ια A"α A"α A"α
Aῦα Aῦα Aῦα Aῦα Aῖα Aῖα Aῖα Aῖα A'Υα A'Υα A"Υα A"Υα A'Υα A'Υα
A'α A'α A'α
Aῶα Aῶα Aῶα Aῶα Aῶα Aῶα Aῶα Aῶα A'Ωα A'Ωα AΩ₁α A'α A'α