

Greek Unicode with 8-bit TeX and *inputenc*

Günter Milde

September 14, 2014

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 Motivation

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 (see `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, e.g.,

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

Ensure, LGR is the active font encoding whenever a Greek character is used in the text.

- 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.,

```
\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).

¹ The 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.

Τί φής; Ἴδὼν ἐνθ' ἔδε παῖδ' ἐλευθέραν τὰς πλησίον Νύμφας
στεφανοῦσαν, Σώστρατε, ἔρῳ ἀπῆλθε εὐθύς;

- 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 package “textcomp” (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.

- 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 «ἡία αντίσταση 750-κΩ» but «ἡία αντίσταση 750-kΩ»

Special care is also required with the question mark characters: While the Unicode standard says character 003B SEMICOLON and not 037E, is the preferred character for ‘Greek question mark’, the LGR font encoding maps it 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 φής;	Τί φής;	037E GREEK QUESTION MARK
Τί φής;	Τί φής;	003B SEMICOLON
Τί φής?	Τί φής;	003F QUESTION MARK

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

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

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
370	*	*	*	*	´	´	*	*				*	*	*	;	
380					´	´	Á	.	É	Η	Í		Ó		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	*	*	*	*	*	*	*	*	*	ϱ	ϲ	Ϝ	ϝ	Ϟ	ϟ	*
3E0	λ	λ	*	*	*	*	*	*	*	*	*	*	*	*	*	*
3F0	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

4.2 Greek Extended

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
1F00	ά	ά	ᾀ	ᾀ	ᾀ	ᾀ	ᾀ	ᾀ	Ά	Ά	Ἀ	Ἀ	Ἄ	Ἄ	Ἄ	Ἄ
1F10	έ	έ	ἔ	ἔ	ἔ	ἔ			Έ	Έ	Ἐ	Ἐ	Ἐ	Ἐ	Ἐ	Ἐ
1F20	ή	ή	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	Ή	Ή	Ἠ	Ἠ	Ἠ	Ἠ	Ἠ	Ἠ
1F30	ι	ι	ἰ	ἰ	ἰ	ἰ	ἰ	ἰ	Ϊ	Ϊ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ
1F40	ό	ό	ὀ	ὀ	ὀ	ὀ			Ό	Ό	Ὢ	Ὢ	Ὢ	Ὢ	Ὢ	Ὢ
1F50	ύ	ύ	ὺ	ὺ	ὺ	ὺ	ὺ	ὺ		Υ		Υ		Υ		Υ
1F60	ώ	ώ	ῶ	ῶ	ῶ	ῶ	ῶ	ῶ	Ώ	Ώ	Ὤ	Ὤ	Ὤ	Ὤ	Ὤ	Ὤ
1F70	ὰ	ὰ	ἔ	έ	ή	ή	ι	ι	ὀ	ό	ὸ	ὸ	ὸ	ὸ	ὸ	ὸ
1F80	ᾀ	ᾀ	ᾀ	ᾀ	ᾀ	ᾀ	ᾀ	ᾀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ
1F90	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	ἦ	Ἠ	Ἠ	Ἠ	Ἠ	Ἠ	Ἠ	Ἠ	Ἠ
1FA0	ἰ	ἰ	ἰ	ἰ	ἰ	ἰ	ἰ	ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ
1FB0	ᾀ	ᾀ	ᾀ	ᾀ	ᾀ		ᾀ	ᾀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ	Ἀ		Ἰ
1FC0	ῶ	ῶ	ῶ	ῶ	ῶ		ῶ	ῶ	Ἐ	Ἐ	Ἐ	Ἐ	Ἐ	Ἐ		Ἰ
1FD0	ἰ	ἰ	ἰ	ἰ			ἰ	ἰ	Ἰ	Ἰ	Ἰ	Ἰ	Ἰ			Ἰ
1FE0	ὺ	ὺ	ὺ	ὺ	ρ	ρ	ὺ	ὺ	Υ	Υ	Υ	Υ	Υ	Ρ		Ἰ
1FF0			ῶ	ῶ	ῶ		ῶ	ῶ	Ό	Ό	Ὤ	Ὤ	Ὤ	Ὤ		Ἰ

4.3 Other Unicode Blocks

Latin-1 Supplement: “ ‹ ′ ‰ »

IPA Extensions: ə

Spacing Modifier Letters: ˘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. ά)

Currency Symbols: €

Letter-like Symbols: Ω

Ancient Greek Numbers: ͵ Ͷ ͷ ͸

5 Test up/downcasing

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

Tonos and dasia on the first vowel of a diphthong (ά, ά, έ) imply a *hiatus*. A dialytika must be placed on the second vowel if they are dropped (ΑΪ, ΑΥ, ΕΪ).

