Greek LICR definitions with XeTeX/LuaTeX

The textalpha package from the greek-fontenc bundle is a convenient way to load Greek LICR macros for use anywhere in the document and with any TeX engine supporting the e-TeX extensions (pdfTeX, XeTeX, or LuaTeX). With XeTeX or LuaTeX, it loads the file greek-euenc.def.

Typesetting Greek texts requires a font containing Greek letters. With the XeTeX or LuaTeX engines, the user should ensure that the choosen font contains the required glyphs (the default Latin Modern fonts have only capital Greek letters). Examples for suitable fonts are the "Deja Vu" or the "Linux Libertine" OpenType fonts.

1 LICR input

The example from usage.tex in babel-greek input using the LICR macros:

```
Τί φήις; Ίδὼν ἐνθέδε παῖδ' ἐλευθέραν τὰς πλησίον Νύμφας στεφανοῦσαν, Σώστρατε, ἐρῶν ἀπῆλθες εὐθύς;
```

1.1 Greek alphabet

Greek letters via LICR macros:

```
ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΤΥΦΧΨΩ
```

```
αβγδεζηθικλμνξοπρστυφχψω
```

The small sigma is set with a different glyph if it ends a word:

- σ textsigma
- ς textfinalsigma or textvarsigma

In the Latin transcription, the letter `s' stands for \textautosigma which automatically chooses the glyph according to the position.

1.2 Additional Greek symbols

- 4 textkoppa
- o textqoppa (archaic koppa)
- Q textQoppa (archaic Koppa)
- ς textstigma
- ς textStigma (Sigma-Tau-Ligature in CB-fonts)¹
- ₹ textsampi
- ን textSampi
- ε text
digamma
- F textDigamma

 $^{^{1}\}mathrm{the}$ name "stigma" originally applied to a medieval sigma-tau ligature, whose shape was confusably similar to the cursive digamma

' textdexiakeraia

, textaristerikeraia

Mathematical notation uses variant shapes for pi, kappa, rho, theta, and phi as additional symbols. These variations have no syntactic meaning in Greek text and are not given separate code-points in the LGR text encoding. Some text fonts use the variant shapes in place of the "regular" ones.

TeX's concept of "standard" vs. "variant" symbols does not map to the distinction between "Greek Letter ..." vs. "Greek Symbol ..." in the Unicode standard:

math symbol	Unicode «letter»	var math symbol	Unicode «symbol»
π	π	ϖ	\odot
ho	ρ	ϱ	Q
heta	θ	ϑ	θ
ϵ	ε	arepsilon	€
ϕ	φ	arphi	ф
β	β	missing	в
κ	κ	missing	и
Θ	Θ	missing	θ

1.3 Diacritics

Greek diacritcs can be input by named macro or symbol macro. Named macros are defined in xunicode-greek.sty for \LastDeclaredEncoding (here LGR).

There are currently no defintions for accent macros combined with literal unicode characters. Combining diacritical characters are used in the output.

άάὰὰϊῖᾶᾶἀἀάἁ

Please mail a feature request if there is demand for definitions selecting precomposed characters. As proof of concept, this document defines a composite command for alpha with psili and perispomeni. Compare $\tilde{\alpha}$ with $\tilde{\alpha}$.

Composite commands for combined diacritics are defined in "greek-fontenc.def", also for the \LastDeclaredEncoding. There are composite definitions for all precomposed characters, the fallback definition uses combining characters:

Problems: Composite diacritics with breathings overlap when set using two combining characters (in the tested fonts).

Upcasing with \MakeUppercase drops diacritics (except dialytika and sub-iota):

ΪΙΧΧΙΙΧΧΙ

 $\amalg X X \amalg X X \amalg X X$

II X X II X X II X X

However, when the tonos, varia, and perispomeni accents are input using the symbol macros (\'\`\~), they behave like standard accents with XeTeX/LuaTeX. If these accents should be dropped by MakeUppercase, they must be input as named macro:

```
A'A XX A'A XX AA XX
```

The sub-iota is input after the base letter. Ligature definitions in the 8-bit LGR fonts ensure that precomposed characters are used. This does not work with Unicode fonts:²

α x Aι Xι, see also the *Greek extended* Unicode block below.

Both, the standard LaTeX and the "xunicode" composite mechanism require the first part of a composite definition to be a macro requiring one argument. I.e. selecting pre-defined characters with sub-iota would require prefix \accypogegrammeni and \accprosgegrammeni accent macros and composite definitions for all combination of letters and letters with diacritic and sub-iota.

1.4 Greek and Coptic Unicode block

Greek symbols from the Greek and Coptic Unicode block are fully supported:

```
΄,
΄,
΄*Α·Έ Ή ΊΟΥΩ ΐ

ΑΒΓΔΕ ΖΗ ΘΙΚΛΜΝΞΟΠΡΣΤΥΦΧΨΩ
ΪΫά ἐ ἡ ἱ ΰ

αβγδεζηθικλμνξοπρςστυφχψω
ϊΰό ὑ ώ

Ο ο ς ς F ε ΄ ʹ ງ »
```

1.5 Greek Extended Unicode block

Characters from the Greek Extended Unicode block are only partially supported, accented characters with sub-iota currently use a combining prosgegrammeni or ypogegrammeni character.

For comparision, the LICR macro is followed by the corresponding literal Unicode character:

ἀά ἀὰ αὰ αὰ ακ κά ἄἄ ἄἄ ἄἄ ἀΑ ΑΑ ΑΑ "Α" Α"Α "Α" Α"Α "Α" Α"Α "Α" Α"Α

éé éé ëë ëë ëë ëë E'E'E'E'E'E'E'E'E

²When copy/pasting generated output from the "evince" PDF viewer, the COMBINING GREEK YPOGE-GRAMMENI is merged with the base character. A capital letter + combining ypogegrammeni results in GREEK CAPITAL LETTER ... WITH [.. AND] PROSGEGRAMMENI). This indicates that for a subfix (in contrast to post-fix) prosgegrammeni, the combining *ypogegrammeni* could be used together with a capital base letter.

Ancient Greek Numbers are missing in most fonts (including Libertine and Deja Vu). The "FreeSerif" font works fine.

XIXIXIX

If the LGR font encoding is loaded via «fontenc» in the document preamble, Ancient Greek Numbers (as well as any other character) from LGR encoded 8-bit TeX fonts can be used after a font-encoding switch. babel-greek defines the \textgreek command for this purpose:

ΔΗΧΙΜ

2 Latin transcription

The \textgreek macro and LGR encoded fonts can also be used for the input of Greek letters via the *Latin transcription* provided by this font encoding, e.g. «logos» becomes « λ 0 γ 0 ς » and «\>aupn\'ia» becomes « α 0 π 0 π 0) (capitalized «A Π 1 π 1). For details of the Latin transliteration see the teubner package or the file usage.pdf from the babel-greek package.

Mark, that you cannot use Unicode input with LGR encoded fonts except when running in 8-bit compatibility mode because of the incompatibility between Xe/LuaTeX and the inputenc package.

LICR macros work in both, Unicode font encoding and LGR: compare Ἰανουαρίου (Unicode font set up via fontspec) vs. Ἰανουαρίου (LGR-encoded 8-bit font set up via NFSS commands, see *The LaTeX font guide* [fntguide.pdf]).