Greek and hyperref

Günter Milde

2020/10/30

Contents

On 2010-11-05, Heiko Oberdiek wrote in comp.text.tex:

> `\pdfstringdef` (converting TeX code to PDF strings for bookmarks) supports NFSS2 and needs active characters. Encoding stuff based on the internal font machinery of TeX (letters with catcode 11 or 12, ligatures) does not work, because the strings don't reach TeX's stomach.

The *greek-fontenc* package allows input of Greek characters in a way that "reaches TeX's stomach" and hence works in both, the main document as well as in PDF strings (e.g. bookmarks). Hyperref's "puenc.def" font encoding file defines LICR macros for monotonic Greek (Greek characters of the "Greek and Coptic" unicode block).

All utf8-encoded literal Unicode characters work in PDF strings. With *greek-fontenc* and *greek-inputenc*, this enables use of all Greek character in text and PDF strings.

With the *textalpha* package, Greek letters can be used without explicit change of the font encoding or Babel language (with some caveats, see textalpha-doc.pdf). For correct hyphenation and other fixes, mark Greek text parts with the Babel language `greek`. There should be no space around a language switch: |λογος|.

1 Transcription: λογος, λογος

Text input using the Latin transliteration: In the PDF-bookmark are Latin letters instead of Greek ones.

2 Macros: λογος, λογος, λογος, λογος,

Babel-Greek or *textalpha* package with `\textalpha` ... `\textomega` macros; *alphabeta* package with `\alpha` ... `\Omega` macros. With 8-bit TeX (pdflatex), literal Greek Unicode characters are converted to LICR Macros, too.

Works, if the `unicode` or `pdfencoding=auto` option is given to *hyperref*.[1]

3 Kerning: ΑΥΛ ΑΥΛ ΑΥΛ

Kerning is impeded if the font encoding is switched for every single character. To fix this, wrap the Greek part in a command switching to a font encoding supporting Greek, either `\ensuregreek{...}` (with package *textalpha* or *Babel*) or `\foreignlanguage{greek}{...}` (with *Babel*).

4 Literal Unicode input

The following subsection headings contain all characters from the “Greek and Coptic” and “Greek Extended” Unicode Blocks that are supported by the LGR font encoding.

4.1 ʹ͵; ΄ ΅ Ά·ΈΉΊΌΎΏΐΑΒΓΔΕΖΗΘΙΚΛΜΝ

4.2 ΞΟΠΡΣΤΥΦΧΨΩΪΫϘϚϜϠ

Greek and Coptic Unicode block: punctuation and uppercase letters

4.3 άέήίΰαβγδεζηθικλμν

4.4 ξοπρςστυφχψωϊϋόύώϐϑϕϖϘϙϛϝϟϡ

Greek and Coptic Unicode block: lowercase letters

[1] With the “xpdf” viewer, Greek letters are not shown in PDF bookmarks.

4.5 ἀἁἂἃἄἅἆἇἈἉἊἋἌἍἎἏ

4.6 ἐἑἒἓἔἕἘἙἚἛἜἝ

4.7 ἠἡἢἣἤἥἦἧἨἩἪἫἬἭἮἯ

4.8 ἰἱἲἳἴἵἶἷἸἹἺἻἼἽἾἿ

4.9 ὀὁὂὃὄὅὈὉὊὋὌὍ

4.10 ὐὑὒὓὔὕὖὗὙὛὝὟ

4.11 ὠὡὢὣὤὥὦὧὨὩὪὫὬὭὮὯ

4.12 ὰάὲέὴήὶίὸόὺύὼώ

4.13 ᾀᾁᾂᾃᾄᾅᾆᾇᾈᾉᾊᾋᾌᾍᾎᾏ

4.14 ᾐᾑᾒᾓᾔᾕᾖᾗᾘᾙᾚᾛᾜᾝᾞᾟ

4.15 ᾠᾡᾢᾣᾤᾥᾦᾧᾨᾩᾪᾫᾬᾭᾮᾯ

4.16 ᾰᾱᾲᾳᾴᾶᾷᾸᾹᾺΆᾼ᾽ι᾿

4.17 ῀῁ῂῃῄῆῇῈΈῊΉῌ῍῎῏

4.18 ῐῑῒΐῖῗῘῙῚΊ῝῞῟

4.19 ῠῡῢΰῤῥῦῧῨῩῪΎῬ῭΅`

4.20 ῲῳῴῶῷῸΌῺΏῼ´῾

Greek Extended Unicode block: Input as literal precomposed Unicode character works fine.

5 LICR command input

textalpha loads definitions for LICR input with non-standard accents or combined diacritics characters also for PU (hyperref).

5.1 Greek and Coptic

5.1.1 ʹ͵ͺ;΄΅Ά·ΈΉΊΌΎΏΐΑΒΓΔΕΖΗΘΙΚΛΜΝ

5.1.2 ΞΟΠΡΣΤΥΦΧΨΩΪΫ

5.1.3 άέήίΰαβγδεζηθικλμν

5.1.4 ξοπρςστυφχψωϊϋόύώϐϑϕϖϘϙϚϛϜϝϞϟϠϡ

5.2 Greek Extended

5.2.1 ἀἁἂἃἄἅἆἇἈἉἊἋἌἍἎἏ

5.2.2 ἐἑἒἓἔἕἘἙἚἛἜἝ

5.2.3 ἠἡἢἣἤἥἦἧἨἩἪἫἬἭἮἯ

5.2.4 ἰἱἲἳἴἵἶἷἸἹἺἻἼἽἾἿ

5.2.5 ὀὁὂὃὄὅὈὉὊὋὌὍ

5.2.6 ὐὑὒὓὔὕὖὗὙὛὝὟ

5.2.7 ὠὡὢὣὤὥὦὧὨὩὪὫὬὭὮὯ

5.2.8 ὰάὲέὴήὶίὸόὺύὼώ

5.2.9 ᾀᾁᾂᾃᾄᾅᾆᾇᾈᾉᾊᾋᾌᾍᾎᾏ

5.2.10 ᾐᾑᾒᾓᾔᾕᾖᾗᾘᾙᾚᾛᾜᾝᾞᾟ

5.2.11 ᾠᾡᾢᾣᾤᾥᾦᾧᾨᾩᾪᾫᾬᾭᾮᾯ

5.2.12 ᾰᾱᾲᾳᾴᾶᾷᾸᾹᾺΆᾼ᾽ι᾿

5.2.13 ῀῁ῂῃῄῆῇῈΈῊΉῌ῍῎῏

5.2.14 ῐῑῒΐῖῗῘῙῚΊ῝῞῟

5.2.15 ῠῡῢΰῤῥῦῧῨῩῪΎῬ῭΅`

5.2.16 ῲῳῴῶῷῸΌῺΏῼ´῾

6 Alias character names

6.1 Θεφςβϑπϰρʹ͵µϚϜϝϠϡ

7 Makeuppercase

According to Greek typesetting conventions, diacritics (except the dialytika) are dropped in UPPERCASE.

The LaTeX `\Makeuppercase` implementation changed fundamentally with the release in 06/2022.[2] Since the change, `\Makeuppercase` is also supported

[2] This change cannot be reverted with the rollback mechanism “for technical reasons”.

in PDF-strings. Greek typesetting rules are ony applied if the text language is set to `greek` with Babel's `\setlanguage` or `\foreignlanguage`.

The changes broke the support for upcasing of polutonic Greek (combined accents) with LICR input and of Greek with the LGR Latin transliteration. Support is restored as of 2023-09-07, some issues remain with PDF strings.

7.1 Literal Unicode input

7.1.1 ʹ„; ʹ ̈ ΆΈΉΊΌΎΏΪΑΒΓΔΕΖΗΘΙΚΛΜΝ

Text language English — diacritics not dropped.

7.1.2 ʹ„; ʹ ̈ A·EHIOΥΩΪΑΒΓΔΕΖΗΘΙΚΛΜΝ

Language set to Greek inside the `\subsection` command – diacritics dropped in the text and ToC but not in the PDF sidebar (sic!).

7.1.3 ʹ„; ʹ ̈ A·EHIOΥΩΪΑΒΓΔΕΖΗΘΙΚΛΜΝ

Language switched before the `\subsection` command – diacritics dropped in text, ToC, and PDF sidebar.

7.1.4 ΞΟΠΡΣΤΥΦΧΨΩΪΫϚϜFⅬ

7.1.5 ΑΕΗΙΫΑΒΓΔΕΖΗΘΙΚΛΜΝ

7.1.6 ΞΟΠΡΣΣΤΥΦΧΨΩΪΫΟΥΩΒΘΦΠϚϚϜFⁱⅬ

7.1.7 AAAAAAAAAAAAAAA

7.1.8 EEEEEEEEEEE

7.1.9 HHHHHHHHHHHHHHH

7.1.10 I I I I I I I I I I I I I I I

7.1.11 OOOOOOOOOOO

7.1.12 ΥΥΥΥΥΥΥΥΥΥΥ

7.1.13 ΩΩΩΩΩΩΩΩΩΩΩΩΩΩΩ

7.1.14 AAEEHHIIOOΥΥΩΩ

7.1.15 $A_IA_IA_IA_IA_IA_IA_IA_IA_IA_IA_IA_IA_IA_IA_I$

7.1.16 $H_IH_IH_IH_IH_IH_IH_IH_IH_IH_IH_IH_IH_IH_IH_I$

7.1.17 $Ω_IΩ_IΩ_IΩ_IΩ_IΩ_IΩ_IΩ_IΩ_IΩ_IΩ_IΩ_IΩ_IΩ_IΩ_IΩ_I$

7.1.18 ĂĀ$A_I$$A_I$$A_IAA_I$ĂĀAAA$_I$'I'

7.1.19 ῀῁ῌῌῌΗῌΕΕΗΗῌ῝ ῞ ῀

7.1.20 Ῐ Ῑ Ϊ Ϊ Ι Ϊ Ῐ Ῑ Ι Ι ῾῾῀

7.1.21 ῨῩΫΫΡΡΥΫῨῩΥΥΡ῍῍

7.1.22 ῼῼῼΩῼΟΟΩΩῼ῎

7.2 LICR input

7.2.1 ΄ ͵ ; ¨ Α·ΕΗΙΟΥΩΪΑΒΓΔΕΖΗΘΙΚΛΜΝ

7.2.2 ΞΟΠΡΣΤΥΦΧΨΩΪΫ

7.2.3 ΑΕΗΙΫΑΒΓΔΕΖΗΘΙΚΛΜΝ

7.2.4 ΞΟΠΡΣΣΤΥΦΧΨΩΪΫΟΥΩΒΘΦΠϘϘϠϠϜϜϟϟϡϡ

7.2.5 ΑΑΑΑΑΑΑΑ ΑΑΑΑΑΑΑΑ

7.2.6 ΕΕΕΕΕΕ ΕΕΕΕΕΕ

7.2.7 ΗΗΗΗΗΗΗΗ ΗΗΗΗΗΗΗΗ

7.2.8 Ι Ι Ι Ι Ι Ι Ι Ι Ι Ι Ι Ι Ι Ι Ι Ι

7.2.9 ΟΟΟΟΟΟ ΟΟΟΟΟΟ

7.2.10 ΥΥΥΥΥΥΥΥ ΥΥΥΥ

7.2.11 ΩΩΩΩΩΩΩΩ ΩΩΩΩΩΩΩΩ

7.2.12 ΑΑΕΕΗΗΙΙΟΟΥΥΩΩ

7.2.13 ᾼᾼᾼᾼᾼᾼᾼᾼᾼᾼᾼᾼᾼᾼᾼᾼ

7.2.14 ῌῌῌῌῌῌῌῌῌῌῌῌῌῌῌῌ

7.2.15 ῼῼῼῼῼῼῼῼῼῼῼῼῼῼῼῼ

7.2.16 ᾸᾹᾼᾼᾼΑᾼᾸᾹΑΑᾼ ͺ

7.2.17 ¨ ῌῌῌΗῌΕΕΗΗῌ

7.2.18 Ῐ Ῑ Ϊ Ϊ Ι Ϊ Ῐ Ῑ Ι Ι

7.2.19 ῨῩΫΫΡΡΥΫῨῩΥΥΡ¨ ¨

7.2.20 ῼῼῼΩῼΟΟΩΩῼ

7.3 Ηιατυς

7.3.1 άι άυ έι ἅι ἅυ ἕι ἀυ → ΑΪ ΑΫ ΕΪ ΑΪ ΑΫ ΕΪ ΑΫ

7.3.2 άι άυ έι ἅι ἅυ ἕι ἀυ → ΑΪ ΑΫ ΕΪ ΑΪ ΑΫ ΕΪ ΑΫ

In PDF strings, the hiatus feature only works with literal input.

8 Conclusion

For Greek text parts in section headers use either literal Unicode characters[3] or macros. For proper kerning und upcasing in the main document, set the text language of Greek text parts to `greek`. If you use polytonic Greek, set the `polutoniko` language attribute.

[3]Combining Unicode characters do not work with inputenc and 8-bit LaTeX. (This is a general restriction.) Use pre-composed Unicode characters or accent macros for letters with diacritics.