

**Notes on the naming of some characters proposed in the FCD of ISO/IEC 10646:2012 (3<sup>rd</sup> ed.) (JTC1/SC2/WG2 N3945, JTC1/SC2 N4168)**

Karl Pentzlin — 2011-02-02

This paper addresses the naming of the following characters proposed for inclusion:

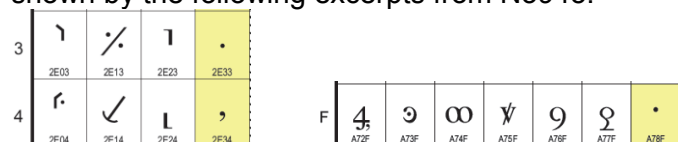
U+2E33 RAISED DOT

U+2E34 RAISED COMMA (both proposed in JTC1/SC2/WG2 N3912)

U+A78F LATIN LETTER GLOTTAL DOT

(proposed in JTC1/SC2/WG2 N3567 as LATIN LETTER MIDDLE DOT)

shown by the following excerpts from N3945:



**1. Dots, Points, and "middle" and "raised" characters encoded in Unicode 6.0**

(without script-specific ones and fullwidth/small forms; of similar characters within a block, a single character is selected as representative):

*Dots and Points:*

- U+002E FULL STOP
- U+00B7 MIDDLE DOT
- U+02D9 DOT ABOVE
- U+0387 GREEK ANO TELEIA
- U+2024 ONE DOT LEADER
- U+2027 HYPHENATION POINT
- U+22C5 DOT OPERATOR
- U+2E31 WORD SEPARATOR MIDDLE DOT

*U+02D9 has the property Sk (Symbol, modifier), U+22C5 has Sm (Symbol, math), while all others have Po (Punctuation, other). U+A78F is proposed to have Lo (Letter, other).*

*"Middle" and "raised" characters:*

- U+02F4 MODIFIER LETTER MIDDLE GRAVE ACCENT
- U+2E0C LEFT RAISED OMISSION BRACKET (*representing several similar characters in the same block*)
- U+02F8 MODIFIER LETTER RAISED COLON
- U+A71B MODIFIER LETTER RAISED UP ARROW

The following table will show these characters using some different widespread fonts.

Font	002E	00B7	02D9	0387	2024	2027	22C5	2E31	02F4	2E0C	02F8	A71B
Andron Mega CorpusX.E	x.E	x·E	x`E	α·E	x.E	x·E	x·E	---	x·E	x`E	x·E	---
Arial	x.E	x·E	x`E	α·E	x.E	x·E	x·E	---	x·E	---	x·E	---
Times New Roman	x.E	x·E	x`E	α·E	x.E	x·E	x·E	---	x·E	---	x·E	---

Cambria	x.E	x·E	x`E	α·E	x.E	x·E	x⊠E	---	x⊠E	---	x⊠E	---
Cardo	x.E	x·E	x`E	α·E	x.E	x·E	x·E	---	x·E	x`E	---	---
Charis SIL	x.E	x·E	x`E	---	x.E	x·E	---	---	x·E	---	x·E	x↑E
Code 2000	x.E	x·E	x`E	α·E	x.E	x·E	x·E	---	x·E	x`E	x·E	x↑E
Consolas	x.E	x·E	x`E	α·E	x.E	x·E	x·E	---	x⊠E	---	x⊠E	---
Courier New	x.E	x·E	x`E	α·E	x.E	x·E	x·E	---	x·E	---	x·E	---
DejaVu Sans	x.E	x·E	x`E	α·E	x.E	x·E	x·E	---	x·E	---	x·E	---
DejaVu Serif	x.E	x·E	x`E	α·E	x.E	x·E	x·E	---	x·E	---	x·E	---
Everson Mono	x.E	x·E	x`E	α·E	x.E	x·E	x·E	x·E	x·E	x`E	x·E	---
Galatia SIL	x.E	x·E	---	α·E	---	---	---	---	---	---	---	---
Garamond (Adobe)	x.E	x·E	x`E	α·E	x.E	x·E	x·E	---	x·E	---	x·E	---
Gentium	x.E	x·E	x`E	α·E	x.E	x·E	---	---	---	---	---	---
Junicode	x.E	x·E	x`E	α·E	x.E	x·E	---	---	---	x`E	---	---
KadmosU	x.E	x·E	x`E	α·E	---	x·E	---	---	---	x`E	---	---
Linux Libertine	x.E	x·E	x`E	α·E	x.E	x·E	x·E	---	x·E	---	x·E	---
Lucida Sans Unicode	x.E	x·E	x`E	α·E	x.E	x·E	x·E	---	---	---	---	---
Minion Pro	x.E	x·E	x`E	α·E	---	x·E	x·E	---	x·E	---	x·E	---
MS Gothic	x.E	x·E	x`E	α·E	x.E	x·E	x·E	---	---	---	---	---
Roman Cyrillic	x.E	x·E	x`E	α·E	x.E	x·E	---	x·E	x·E	x`E	x·E	x↑E
Segoe UI	x.E	x·E	x`E	α·E	x.E	x·E	x·E	---	x·E	---	x·E	---

A script specific example for a "raised dot" is (copied from Unicode 6.0):



## 2. The term "raised" and the naming of U+2E33/U+2E34

As shown in the proposal N3912, the proposed characters U+2E33/U+2E34 are centered at the middle between the baseline and the x-height line

As the previous table shows, all characters name "raised" so far are in fact centered in the middle of the x-height line and the cap-height line (or, in the case of the Canadian Syllabics example where these height terms do not apply, slightly below the letter-height line).

Thus, to call the proposed characters U+2E33/U+2E34 "raised", would introduce an unsystematic and arbitrary use of this term into the Unicode character naming.

Moreover, to call U+2E33 "RAISED DOT" while U+18DF is also basically called "... RAISED DOT", seems contradictory, as the dots are definitively placed differently.

Thus, a naming like this would be more exact:

U+2E33 DOT AT HALF X-HEIGHT  
U+2E34 COMMA AT HALF X-HEIGHT

However, such naming is more strongly referring to a specific glyph design than usual for Unicode character naming.

Thus, the following naming seems more appropriate, introducing the term "lifted" (which until now does not occur in any name of a Unicode character) to denote "slightly above of the baseline":

U+2E33 LIFTED DOT  
U+2E34 LIFTED COMMA

## 3. Use of the proposed U+A78F LATIN LETTER GLOTTAL DOT vs. U+00B7 MIDDLE DOT

In the proposal N3567, the necessity of introducing U+A78F is based on its use as a true letter, which has to be recognized as such by its "Lo" property by searching algorithms, etc.

This paper provides another argument for encoding the character, complete independent of the reasons listed in N3567 and in addition to them.

In fact, it seems inappropriate to unify the proposed letter with U+00B7 as it has a different appearance, thus establishing a character identity independent of its properties.

"Middle", as used in the name of U+00B7 MIDDLE DOT, is a well-established typographic term. It denotes the middle between the baseline and the cap-height line.

The table provided above shows that almost all fonts listed there supply an appropriate glyph, showing the dot in fact at such a vertical position. This position usually is considerably lower than the x-height line, as also can be seen in the table.

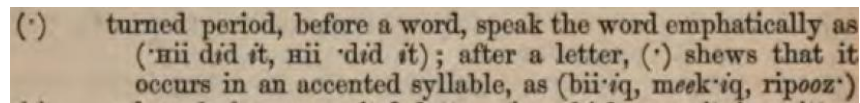
The proposed U+A78F LATIN LETTER GLOTTAL DOT, however, consistently appears in the examples provided in N3567 at the x-height line, thus being optically "in line" with small Latin letters (which also supports the position that it *is* a Latin letter).

See e.g. the following excerpt of fig. 5 of N3567:

'Phags-pa:  [Line 1]  
bh·rum

Moreover, while N3567 only relies on a special use of this letter (in translations of 'Phags-pa), this letter in fact is found in several other extensions of the Latin script, where the use of U+00B7 MIDDLE DOT would be typographically unacceptable.

For instance, it is used in Paleotype:



(Ellis, Alexander J: On Early English Pronunciation, Part I; London 1869; p. 12)

Based on this use, a more unspecific name than "LATIN LETTER GLOTTAL DOT" seems to be appropriate, like:

U+A78F LATIN LETTER DOT AT X-HEIGHT

(while the name "LATIN LETTER MIDDLE DOT" used in N3567 is inapplicable anyway).

However, such naming is more strongly referring to a specific glyph design than usual for Unicode character naming.

Thus, the following naming seems more appropriate, introducing the term "elevated" (which until now does not occur in any name of a Unicode character) to denote "usually at x-height":

U+A78F LATIN LETTER ELEVATED DOT

Also, instead to the informative note now given in N3945:

- used for transliteration for Phags-Pa and phonetic transcription for Tangut

the following set of alternative name and informative notes seem appropriate:

- = turned full stop (referring to metal type)
- usually displayed at x-height
- used as glottal dot in transliteration for Phags-Pa and phonetic transcription for Tangut