

Comments on L2/18-020 *Proposal to define Standardized Variation Sequences for BOPOMOFO LETTER I*

Liang Hai (梁海)

lianghai@gmail.com

24 January 2018

L2/18-020 *Proposal to define Standardized Variation Sequences for BOPOMOFO LETTER I* is unclear and possibly misleading to readers that are not familiar with Bopomofo and Unicode–OpenType.

Note my terminology:

- "Lying stroke (—)" vs "standing stroke (|)"
- "Vertical text" vs "horizontal text"

Background documents about Taiwan's behavior reform of the Bopomofo letter I:

- [WG2 N3246](#) *Proposal to encode two Bopomofo characters in UCS*
- [WG2 N4609](#) *"Reply to your company's inquiry to usage of Bopomofo Letter I, as description, please check."*

1 Traditional and reformed behaviors of the Bopomofo letter I

- The *traditional* behavior: Letter I is written with a lying stroke (—) in vertical text, while with a standing stroke (|) in horizontal text, ie, letter I changes appearance depending on text orientation so its stroke direction is always perpendicular to the direction text flows.
- Due to technical limitations, in 2008 Taiwan introduced the *reformed* behavior: Letter I should now be written with a lying stroke (—) in both horizontal and vertical text, while the traditional behavior is still allowed. This reform effectively has only changed the appearance in horizontal text. Also note, other regions (eg, mainland China) are not affected by this reform.

Representative glyphs in Unicode

Principally, Unicode uses the glyphs in horizontal text as representative glyphs for characters:

- Originally the representative glyph for letter I in Unicode conformed with the traditional behavior, ie, a standing stroke (I) in horizontal text, while the appearance of a lying stroke (—) is considered a variant for vertical text.
- In Unicode 8.0, the representative glyph was changed to conform with Taiwan's reformed behavior, ie, a lying stroke (—) for both horizontal and vertical text. The annotation "the vertical stroke form is considered a rendering variant" refers to the expected appearance in horizontal text when the traditional behavior is implemented.

Implementation with OpenType

In terms of implementation:

- With the traditional behavior, fonts and text rendering environments are required to enable the appearance-changing ability of letter I: show the glyph of standing stroke (I) by default for horizontal text, and change to the glyph of lying stroke (—) for vertical text. The OpenType technology is typically used to achieve such a behavior.
- The reformed behavior practically requires the default and only glyph of letter I should be of a lying stroke (—) for both horizontal and vertical text, while relieves fonts and text rendering environments from the required appearance-changing ability described above.
- In Taiwan, while the reformed behavior is officially preferred now, the traditional behavior is explicitly stated to still be allowed.
- Thus the distinction between traditional and reformed behaviors can and should be achieved with appropriate fonts.

Confusion and mistakes in the proposal

- Figure 3 and 4 simply show the expected appearance (—) in vertical text as always. Letter I's expected appearance in vertical text has never been modified.
- Figure 5 shows the traditional appearance (I) in horizontal text. Nothing said for clarifying if letter I in this specific writing system for the Taiwanese language is also affected by the reformed behavior (which was presumably decided mostly for Mandarin).
- "TCA had explained that these two forms are needed for Taiwan at the same time on WG2 N3246 & N4609." — It's still not clear these two appearances are needed to contrast in encoding.
- Figure 6 is a glyph list of printed type (possibly a set of movable metal type) for the traditional behavior, not really relevant to whether the two appearances need to be differentiated in encoding.
- Figure 7 and 8 (as well as figure 12), similarly to figure 5, show the traditional appearance (I) in horizontal text.
- Table 1, figure 10, and figure 11:
 - The analysis of distinction is not reliable.
 - The source of figure 10 and 11 is actually a comparison list of two other (drafted) transcription systems, with Bopomofo serving as the keys.
 - The duplication of letter I in the sections "Initials" and "Rhymes" (not really "Finals" here because the medial part is only included in the following "Combined Finals" section) is meant for duplicated keys, because the corresponding spellings in the other two transcription systems actually distinguish an /i/ sound in initials/medials and an /i/ sound in rhymes.
 - Without running text examples, it's not clear if the single instance of a lying stroke (—) is an intended contrast or some mistake (eg, typesetter misunderstood text orientation).
- "CNS 11643, Taiwan's official standard, has included two forms at the same time." — It's not clear how the two characters are intended to behave. It's not clear what relationship the two characters have, especially, how the newly added character affects the original one.
- "We think it is necessary to use two forms in plain text at the same time." — It's still not clear the two appearances need to contrast in the same text orientation.

- It's not clear what behavior the proposed standardized variation sequence is intended to have. Does it always show as a standing stroke (|)? Does it follow the traditional behavior of letter I instead? What's the relationship between it and the characters in CNS 11643?

Note on related characters

There is a nasalized version of the letter I, (at least traditionally) written with the same traditional behavior of letter I. The two appearances of letter *nasalized I* is encoded as separate characters (similar to the two characters of letter I in CNS 11643):

- U+31AA BOPOMOFO LETTER INN: Lying stroke (—) with a small circle, used in vertical text.
- U+31B3 BOPOMOFO LETTER INNN: Standing stroke (|) with a small circle, used in horizontal text.

How are these closely related characters intended to behave? And should they have their behavior synced to letter I's?

Conclusion

Given the weak evidence supplied in the proposal, I don't see necessity to introduce encoding-level distinction for the two appearances.

The traditional and reformed behaviors of letter I are already supported by fonts, although there're too many badly produced fonts, which is the actual issue we need to work on.

A standardized variation sequence needs to be supported by fonts to be valuable. The introduction of a standardized variation sequence will only complicate the situation.