

## **Universal Multiple-Octet Coded Character Set International Organization for Standardization**

**Doc Type:** Working Group Document

**Title:** Proposal to add standardized variation sequences for fullwidth East Asian punctuation

**Author:** Ken Lunde (Adobe Systems Incorporated)

**Status:** Corporate Full Member Contribution

**Action:** For consideration by the UTC

**Date:** 2018-01-21

### **Background**

This proposal is a split version of [L2/17-056](#) that was originally discussed during UTC #138 in early 2014 as [L2/14-006](#). L2/17-056 itself was discussed during UTC #153 in late 2017 for the purpose of soliciting feedback. The scope of this particular proposal is limited to eight fullwidth East Asian punctuation characters.

Regional conventions affect how particular fullwidth punctuation should display, with the general difference being justified to a corner of the em-box (lower-left for horizontal layout, and upper-right for vertical) versus centered within the em-box. The fundamental issue is that the glyphs for these characters share the same Unicode code point, meaning that an explicit font change or layout feature invocation (such as the OpenType 'locl' GSUB feature) must be used to specify or distinguish them, which is generally not possible in “plain text” environments.

Although “rich text” environments are becoming more common, including those that support language-tagging and the OpenType 'locl' GSUB feature, “plain text” environments persist, and are likely to continue to persist for a long time due to their robust nature. In addition, environments that support variation sequences outnumber those that support language-tagging.

### **Proposal Summary**

This document is a proposal for adding standardized variation sequences for eight (8) fullwidth punctuation characters that use VS1 and VS2 (aka U+FE00 and U+FE01) to distinguish between *corner-justified* and *centered* forms, respectively, whose usage varies according to well-established East Asian regional conventions.

### **Fullwidth Characters With Ambiguous Alignment**

This proposal covers eight East Asian punctuation characters that are unambiguously fullwidth, because they have the *East Asian Width* (see [UAX #11](#)) property W (*East Asian Wide*) or F (*East Asian Fullwidth*), yet remain ambiguous in terms of whether their glyphs are corner-justified or centered. While single-language or single-region East Asian fonts need to include only one of the forms, Pan-CJK fonts that support multiple East Asian languages and regions, such as those of the open source *Source Han* and *Noto CJK* typeface families, need to include both forms to accommodate multiple—and conflicting—typographic conventions.

To summarize the regional conventions for these eight characters, China (PRC) uses corner-justified forms, Taiwan (ROC) and Hong Kong SAR use centered forms, and Japan uses corner-justified periods and commas, but uses centered exclamation points, question marks, colons, and semicolons. The Koreas—ROK and DPRK—typically do not use fullwidth punctuation except for vertical writing, which is relatively rare.

### **Standardized Variation Sequences**

Standardized variation sequences offer a solution to this glyph-level alignment ambiguity by using variation selectors to support these conventions. A font with appropriate entries in its Format 14 (*Unicode Variation*

Sequences) 'cmap' subtable can enable these distinctions to be shown and preserved in “plain text” environments. Below is a complete list of the proposed standardized variation sequences as they would appear in the UCD’s *StandardizedVariants.txt* file:

```
# Corner-justified form and centered form variation sequences

3001 FE00; corner-justified form;      # IDEOGRAPHIC COMMA
3001 FE01; centered form;              # IDEOGRAPHIC COMMA
3002 FE00; corner-justified form;      # IDEOGRAPHIC FULL STOP
3002 FE01; centered form;              # IDEOGRAPHIC FULL STOP
FF01 FE00; corner-justified form;      # FULLWIDTH EXCLAMATION MARK
FF01 FE01; centered form;              # FULLWIDTH EXCLAMATION MARK
FF0C FE00; corner-justified form;      # FULLWIDTH COMMA
FF0C FE01; centered form;              # FULLWIDTH COMMA
FF0E FE00; corner-justified form;      # FULLWIDTH FULL STOP
FF0E FE01; centered form;              # FULLWIDTH FULL STOP
FF1A FE00; corner-justified form;      # FULLWIDTH COLON
FF1A FE01; centered form;              # FULLWIDTH COLON
FF1B FE00; corner-justified form;      # FULLWIDTH SEMICOLON
FF1B FE01; centered form;              # FULLWIDTH SEMICOLON
FF1F FE00; corner-justified form;      # FULLWIDTH QUESTION MARK
FF1F FE01; centered form;              # FULLWIDTH QUESTION MARK
```

The table below demonstrates an actual implementation—using a fully-functional OpenType/CFF font with an appropriately-built Format 14 'cmap' subtable that specifies the UVSes (*Unicode Variation Sequences*) that correspond to the proposed standardized variation sequences. This OpenType/CFF font is also attached to this proposal, and can be extracted and used. Although not shown in this document, vertical forms of fullwidth glyphs, if any, are supported via the 'vert' GSUB feature. The table uses VS1 and VS2 as described in this proposal. Red registration marks are used to draw attention to how their glyphs are typically aligned within the em-box, with prototypical characters surrounding them: 汉 to indicate China, 永 to indicate Taiwan and Hong Kong SAR, and あ to indicate Japan.

Unicode	VS1—Corner-Justified	VS2—Centered
U+3001	あ漢、 漢あ	永、 永
U+3002	あ漢。 漢あ	永。 永
U+FF01	漢! 漢	あ永! 永あ
U+FF0C	あ漢, 漢あ	永, 永
U+FF0E	あ漢. 漢あ	永・永
U+FF1A	漢: 漢	あ永: 永あ
U+FF1B	漢; 漢	あ永; 永あ

Unicode	VS1—Corner-Justified	VS2—Centered
U+FF1F	𠄎? 𠄎	あ永? 永あ

The table below is in response to the first comment in Section 24 of the Script Ad Hoc report (aka [L2/18-039](#)) that starts on the top of page 28, and shows the eight characters in a vertical context, which illustrates that the corner-justified forms are uniformly justified to the upper-right corner, and that the centered forms remain centered (green-shaded) except for Japan whose convention is to additionally rotate the still-centered form:

Unicode	U+3001	U+3002	U+FF01	U+FF0C	U+FF0E	U+FF1A	U+FF1B	U+FF1F
VS1—Corner-Justified	あ 𠄎、 𠄎あ	あ 𠄎。 𠄎あ	𠄎! 𠄎	あ 𠄎、 𠄎あ	あ 𠄎。 𠄎あ	𠄎: 𠄎	𠄎; 𠄎	𠄎? 𠄎
VS2—Centered	永、 永	永。 永	あ永! 永あ	永、 永	永。 永	永: 永 あ あ	あ永; 永あ	あ永? 永あ

While this does not affect the proposal—because modern apps use the OpenType 'vert' (*Vertical Alternates*) GSUB feature to access vertical forms—seven of the eight characters happen to have compatibility forms in the [Vertical Forms](#) block whose forms are justified to the upper-right corner: U+FE10 ’, U+FE11 ` , U+FE12 °, U+FE13 ∴, U+FE14 ∵, U+FE15 !, and U+FE16 ?. The only outlier is U+FF0E . FULLWIDTH FULL STOP whose cell is red-shaded in the table above.

In response to the second comment in the same Script Ad Hoc report, modern CJK font resources include only fullwidth glyphs for these characters, and it is the responsibility of the layout engine to make them halfwidth when appropriate, either by applying the 'halt' (*Alternate Half Widths*) or 'vhal' (*Alternate Vertical Half Metrics*) GPOS features (preferred), or to do so mechanically. The important point is that the glyphs themselves remain corner-justified or centered.

## **Rationale & Conclusion**

This proposal addresses the varying regional conventions for fullwidth East Asian punctuation, which is a real-world issue for Pan-CJK fonts that support multiple East Asian languages and regions, especially in “plain text” environments with limited font-selection capability, or in environments that lack support for per-character language-tagging.

That is all.