

Title: Suggested edits to the documentation of U+2060 WORD JOINER in ISO/IEC 10646
Source: Laurențiu Iancu and Ken Whistler
Status: Individual contribution
Action: For consideration by the Unicode Technical Committee
Date: 2015-07-29

Background

This document proposes a few edits to the documentation of U+2060 WORD JOINER in ISO/IEC 10646, in alignment with the updates currently being done in the Unicode Standard 8.0 core specification to clarify the segmentation effects of U+2060. The suggested edits are proposed to be considered by the Unicode Technical Committee for ballot comments on the Committee Draft of ISO/IEC 10646 Fifth Edition.

In Section 23.2 *Layout Controls* of the Unicode Standard Version 7.0 core specification, U+2060 WORD JOINER is incorrectly described as marking the absence of word boundaries [TUS70]. That section is being corrected in the Unicode Standard Version 8.0 core specification to indicate that U+2060 marks the absence of line breaking opportunities. The change is only editorial; the actual line breaking and word segmentation behavior of U+2060 stays unchanged as it has been defined for many years in UAX #14 and UAX #29 [UAX14, UAX29].

In ISO/IEC 10646, U+2060 WORD JOINER is described in Section F.1.2 *Word boundary indicators*, based on the corresponding documentation in the Unicode Standard core specification. The proposed edits to Section F.1.2 aim to bring the documentation of U+2060 in the two standards in synchrony, to avoid any confusion for implementers who may refer to one standard or the other.

Proposal

The suggested edits are in the title and body of Section F.1.2 as highlighted below in the text reproduced from p. 2414 of ISO/IEC 10646:2014 Fourth Edition [10646].

F.1.2 Word and line boundary indicators

ZERO WIDTH SPACE (200B): This character behaves like a SPACE in that it indicates a word boundary, but unlike SPACE it has no presentational width. For example, this character could be used to indicate word boundaries in Thai, which does not use visible gaps to separate words.

WORD JOINER (2060) and **ZERO WIDTH NO-BREAK SPACE** (FEFF): These characters behave like a NO-BREAK SPACE in that they indicate the absence of word boundaries line breaking opportunities, but unlike NO-BREAK SPACE they have no presentational width. For example, these characters could be inserted after the fourth character in the text "base+delta" to indicate that there is to be no word-break line breaking opportunity between the "e" and the "+".

NOTE – For additional usages of the ZERO WIDTH NO-BREAK SPACE for "signature", see Clause 10.

References

- [10646] International Organization for Standardization, *ISO/IEC 10646:2014: Information Technology – Universal Coded Character Set (UCS), Fourth Edition*. Available from the ISO/IEC ITTF website, <http://standards.iso.org/ittf/PubliclyAvailableStandards/>.
- [TUS70] The Unicode Consortium, *The Unicode Standard, Version 7.0.0*, October 2014, <http://www.unicode.org/versions/Unicode7.0.0/>. Description of U+2060 WORD JOINER at <http://www.unicode.org/versions/Unicode7.0.0/ch23.pdf#G20199>.
- [UAX14] Andy Heninger, *Unicode Standard Annex #14: Unicode Line Breaking Algorithm*, Version 8.0, <http://www.unicode.org/reports/tr14/tr14-35.html>.
- [UAX29] Mark Davis, Laurențiu Iancu, *Unicode Standard Annex #29: Unicode Text Segmentation*, Version 8.0, <http://www.unicode.org/reports/tr29/tr29-27.html>.