

TITLE: Details on Kaithi punctuation (based on N3389)

SOURCE: Deborah Anderson, Script Encoding Initiative, UC Berkeley

STATUS: Liaison contribution

ACTION: For consideration by WG2

DISTRIBUTION: ISO/IEC JTC1/SC2/WG2

DATE: 2 February 2009

This document responds to a request by the UK for details on which characters should be used for Kaithi punctuation, which were discussed in the Kaithi proposal N3389.

1. (5.7) Word boundaries

When the need arises to encode written Kaithi as is found in a particular manuscript, then U+2E31 WORD SEPARATOR MIDDLE DOT can be used if the original character is dot-like. U+2010 HYPHEN can be used if the word boundary is more like a dash.

2. (5.8) Swash marks

The proposal N3389 recommended U+2028 LINE SEPARATOR, U+2029 PARAGRAPH SEPARATOR, U+10FB GEORGIAN PARAGRAPH SEPARATOR, or U+1368 ETHIOPIC PARAGRAPH SEPARATOR be used for swash marks. However, the LINE SEPARATOR and PARAGRAPH SEPARATOR do not have visible glyphs (except in a show-hidden editing mode), and the GEORGIAN and ETHIOPIC PARAGRAPH SEPARATORS have distinctive shape, quite different from the swash marks of Kaithi. Hence, we will propose two characters for the swashes, KAITHI SECTION MARK and KAITHI DOUBLE SECTION MARK. (Note: These were first called KATHI WORD SEPARATOR [single horizontal bar] and KAITHI SENTENCE SEPARATOR [double horizontal bar], but the new proposed name better reflects their actual use.) These characters will be suggested in ballot comments for the April 2009 WG2 meeting.

3. (5.8) Other marks

To indicate the cross, U+002B PLUS SIGN should be used.

4. (5.9) Hyphenation

For Kaithi hyphenation, users should follow whatever is the recommended practice found in similar Indic script traditions (which might be U+2010, U+002D, etc.).

5. (5.10) Abbreviation

U+110BB KAITHI ABBREVIATION SIGN should be used for the circle indicating abbreviations. In order to capture other scribal abbreviations that appear in manuscripts, mark-up or a script-specific font should be used. These other scribal traditions are considered to be idiosyncratic and out of scope for encoding.

6. (5.11) Enumeration

U+100BC KAITHI ENUMERATOR SIGN, as contained in the proposal, will cover inline usage (such as example E on page 34). A separate enumerator sign, U+110BD KAITHI NUMBER SIGN should be used to cover the use of an enumerator that interacts with digits (similar to U+0600 ARABIC NUMBER SIGN).