

To: South Asian Subcommittee and UTC

From: Deborah Anderson, Rick McGowan, and Ken Whistler

Title: Review of Indic-related L2 documents and Recommendations to the UTC

Date: 6 August 2010

We reviewed several Vedic and Indic-related script documents and make the following recommendations:

## **I. Grantha-related topics**

### **1. GRANTHA OM**

**Documents: L2/10-263 “Comments on Grantha OM” (Rajan)**

**L2/10-267 “Comments on L2/09-345, 10-062 and 10-154” (Sharma)**

**L2/10-274 “Usage Evidence for Grantha OM glyph” (Ganesan)**

**L2/10-286 “Comments on L2/10-274” (Sharma)**

Based on the above set of documents, it is clear there is disagreement about encoding the Grantha OM symbol. The glyph provided as evidence by N. Ganesan on page 2 in L2/10-274 differs from the glyph that was originally proposed (and appears in the revised Grantha chart L2/10-265). Also, the description in L2/10-274 does not accord with shapes of long o and right-hand side “anusvara,” as noted by Sharma in L2/10-267.

Hence, we recommend the removal of 11350 GRANTHA OM from L2/10-265 (the revised Grantha chart), leaving a hole at 11350. We would welcome a well-formed and well-argued proposal for a Grantha OM symbol, drawing upon evidence such as that shown in L2/10-274 and not conflated with what has been claimed to be an Oriya OM symbol. If the evidence of an Oriya OM symbol warrants it, we would welcome a proposal for such a symbol on its own merits.

### **2. Name Change for GRANTHA SIGN ANUNASIKA**

**Document: L2/10-259 “Follow-up #2 to my Grantha proposal” L2/09-372 (Sharma), §1**

We agree that the name 11301 GRANTHA SIGN ANUNASIKA is incorrect, and recommend the name for 11301 be changed to GRANTHA SIGN CANDRABINDU in the revised Grantha chart and names list, document L2/10-265.

### **3. On the use of Variation Selector Sequences in Grantha**

**Document: L2/10-259 “Follow-up #2 to my Grantha proposal” L2/09-372 (Sharma), §2 and §3**

We recommend the UTC not accept the variation sequences as proposed in sections 2 and 3. The use of variation sequences for handling sub-base RA (§2.1), sub-base YA (§2.2), sub-base LA (§2.3), and old-style vowel signs for vocal I and II (§3) is not appropriate. Variation sequences are characters that specify a particular glyph for the preceding single base character. They cannot be used to specify the presentation for a sequence of characters. (See chapter 16 section 4 of *The Unicode Standard 5.2*.)

Sharma also has suggested (§2.4) using variation selectors as a way to represent the 3 presentation forms of viramas in Grantha. This is also inappropriate, since viramas are combining marks, and variation selectors cannot be used with combining marks. See below #4 **Chillu**, in which an alternative possibility is put forward, involving the encoding of a ligating virama.

All these issues are interrelated, and an overall approach needs to come up with ways of indicating variant forms for display, if the variants are considered important for users.

We recommend that sections 2 and 3 be re-worked to suggest alternative possibilities for the representation of these variant forms (if required to be distinguished in plain text), without the use of variation sequences.

(Note: The discussion in section 2 of L2/10-259 refers to “PR 37”, which is PRI 37. The actual reference should be to the *background* document for PRI 37, document L2/04-279, “Proposal on Clarification and Consolidation of the Function of ZERO WIDTH JOINER in Indic Scripts” by Peter Constable.)

#### 4. Chillu

**Documents: L2/10-267 “Comments on L2/09-345, 10-062 and 10-154” (Sharma)**

**L2/10-232 “Official submission of a request regarding the Grantha Chillu Marker” (Sharma)**

**L2/10-284 “Comments on L2/10-194” (Sharma)**

Sharma states in L2/10-267, p. 12:

If Ganesan had been reading my documents as well as I have been reading his, he should understand that the reason is that the semantic equality between the different virama forms in Grantha should be maintained. If the same chillu marker were to be renamed as a ligating virama and provided a decomposition to the regular virama, then I would have no objection to there being such a character that CONSONANT + THAT CHARACTER produces a virama ligature.

Based on the above comment, we recommend that 1134F GRANTHA SIGN CHILLU (which is contained in the chart in L2/10-265) be renamed to GRANTHA SIGN LIGATING VIRAMA.

This provides a way to represent the consonant-virama ligations for Sanskrit. If it also provides

a way to handle Malayalam transliteration, that is another point in its favor, but the proposed solution focuses on addressing the Sanskrit situation, a primary concern raised in the documents above (and in the feedback documents in #5 **Grantha User-feedback**, below).

We note that if this recommendation is taken for the Grantha encoding, Grantha would have two viramas encoded -- a situation which is not unprecedented. There are other instances in the Unicode Standard in which because of presentation issues for a Brahmi-based script, both a virama and a killer (or two types of viramas) have been encoded.

## **5. Grantha User-feedback**

**Documents: L2/10-233 "Request from scholars of the Grantha user community" (Sharma)**

**L2/10-283 "Request from scholars of the Grantha user community #2" (Sharma)**

**L2/10-285 "Comments on Grantha encoding" (Shastri)**

We recommend the UTC note this feedback.

## **6. Vedic Characters for Grantha**

**Document: L2/10-235 "Proposal to encode Vedic characters for the Grantha script" (Sharma)**

We recommend the approval of this proposal, with the following two name changes:

a. 11373 COMBINING GRANTHA SYLLABLE VI > COMBINING GRANTHA LETTER VI

b. 11374 COMBINING TAMIL LETTER PA > COMBINING GRANTHA LETTER PA

The change in (a) makes the name consistent with how these "combining letters" are handled in the Devanagari Extended block (cf. A8FO COMBINING DEVANAGARI LETTER VI).

The change requested in (b) will bring the name in line with the other characters in the Grantha block, all of which will share the Grantha script value. An annotation can be added that identifies this character as having the form of a PA associated with Tamil, rather than the form of a PA specific to Grantha. Handling this by annotation is better than introducing a character name which would call into question character naming rules and the script identity for the character.

## **II. Tamil**

**Document: L2/10-256 "Proposal to encode characters for Extended Tamil" (Sharma)**

We recommend no action be taken at this point, as this document warrants further study.

However, "EXTENDED" should be taken out of the names, they should only contain "TAMIL," since the text covered by these characters will effectively be written in one script, Tamil.

### **III. Vedic proposals**

The UTC should quickly review the following two proposals:

**1. Two Vedic characters: VEDIC SIGN JIHVAMULIYA and VEDIC SIGN UPADHMANIYA**  
**Doc: L2/10-257 “Request to add two characters to the Vedic Extensions block” (Anderson and Sharma)**

[The UTC in May 2010 requested a proper proposal be created, and the general category should be Lo for the two characters; this proposal fulfills this request.]

**2. Vedic Rotated Ardhavisarga**

**Doc: L2/10-234 “Request for encoding 1CF3 VEDIC SIGN ROTATED ARDHAVISARGA (replaces L2/09-343; WG2 N3861)” (Sharma)**

[The UTC in May 2010 had approved this character, but requested general category be changed to gc=Mc. This is done in the above proposal.]

Note: L2/09-344 “Request for encoding 1CF4 VEDIC TONE CANDRA ABOVE” (Sharma) was already approved by the UTC. No changes were requested at the last UTC, and it has already been forwarded to WG2.

### **IV. Other Indic character and script proposals:**

**1. RUPEE currency sign**

**Documents: L2/10-258 “Proposal to encode the Indian Rupee Symbol” (GoI)**

**L2/10-251 “Proposal to Encode India’s National Currency Symbol” (Deka)**

**L2/10-249 “Proposal to encode the INDIAN RUPEE SIGN” (Everson)**

We recommend the encoding of the INDIAN RUPEE SIGN, based on L2/10-258 (revised), which has an appropriate name, glyph, and character properties for this character.

**2. Sindhi script proposal**

**Document: L2/10-271 “Proposal to Encode the Sindhi Script (WG2 N3871)” (Pandey)**

We recommend the approval of this proposal.

We recommend a few minor editorial corrections be made in a revised version (i.e., section 4.3, modify the range to “..1164F” and remove second sentence; section 4.10, describe use of nukta more fully; section 4.12 add “decimal” before the word “digits”; remove “preliminary” throughout the proposal, if referring to an earlier version of this document).

## **V. Other topics**

### **1. Abbreviation sign**

**Document: L2/10-266 “Comments on L2/10-083” (Sharma)**

We recommend the UTC respond by saying the comments in this document are noted. The justification provided does not, however, warrant changing the UTC position, which supported encoding script-specific abbreviation signs for North Indic scripts.

### **2. Brahmi glyph change**

**Document: L2/10-260 “Proposal to change the glyph for Brahmi Number One Thousand (U+11065)” (Glass and Baums)**

We agree with this request to change the glyph for U+11065 BRAHMI NUMBER ONE THOUSAND.

Because Brahmi will appear in Unicode 6.0, the UTC needs to decide on how to handle this glyph change. The options are: (a) approve the glyph change now for Unicode 6.0 or (b) wait, issue a glyph erratum, and fix the glyph after Unicode 6.0.