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: 5 May 2010

We reviewed several Vedic and Indic scripts documents in the L2 document registry and provide the following recommendations to the UTC:

1. L2/09-342 "Misrepresentation in Unicode of characters related to the Sanskrit sounds Jihvamuliya and Upadhmaniya" [Sharma]; comment in L2/09-404 "Feedback on Vedic Docs" [Scharf]

a. We recommend the general category property for 0CF1 KANNADA SIGN JIHVAMULIYA and 0CF2 KANNADA SIGN UPADHMANIYA be changed from So to Lo for 6.0. As noted in L2/09-342, these are not used as symbols, so a change to Lo is deemed appropriate.

b. We recommend the header above the two characters 0CF1 and 0CF2 be changed from "Vedic signs" to wording that indicates the characters are used in general Sanskrit usage.

c. We recommend the annotation for these two characters be changed to wording similar to that provided in this document, viz.:
• Denotes the sounds jihvamuliya and upadhmaniya (velar and bilabial voiceless fricatives) in Sanskrit.
• Despite its name and inclusion in this block, this character is not limited to Vedic.

d. In our view, the question remains open as to whether a pair of "generic" characters for jihvamuliya and upadhmaniya should be separately encoded in the Vedic Extensions block for use by Indic scripts writing Sanskrit, or whether a separate pair should be encoded for each Indic script used to write Sanskrit. Sharma recommends the encoding of the Devanagari shapes and counsels the UTC to approve other script-specific characters, as evidence is provided. (Scharf does not come down one way or the other on this question in his comment in L2/09-404.)

If the UTC decides to encode two "generic" characters for use with Indic scripts, then:
• a proposal summary form is needed.
• the script property for the two Kannada characters should be changed from Common back to Kannada.

2. L2/09-343 "Request for encoding 1CF3 ROTATED ARDHAVISARGA" [Sharma]; comment in L2/09-404 "Feedback on Vedic Docs" [Scharf]

a. The contrastive evidence provided in L2/09-343 supports the need for this character. (As noted by Scharf in L2/09-404, however, this character is actually a glyph variant). We recommend the character 1CF3 VEDIC SIGN ROTATED ARDHAVISARGA be accepted, but with the gc=Mc (see below, 2.c). We recommend the name be modified from "ROTATED ARDHAVISARGA," as proposed in L2/09-343, to include "VEDIC SIGN" in its name, in order to bring it in line with 1CF2 VEDIC SIGN ARDHAVISARGA.

b. We recommend an annotation be added, noting that this character is used in general Sanskrit text and not only specifically Vedic text, although it appears in the Vedic Extensions block.

c. We recommend not accepting the request to change the general category from Mc to Lo for VEDIC SIGN ARDHAVISARGA, as well as for visarga, anusvara, and other characters. This change could significantly affect existing data and implementations. However, a well-formed proposal that argues in favor of such a change, with a discussion of the implications for current data and implementations, would be welcome for consideration by the UTC.

3. L2/09-344 Request for encoding 1CF4 VEDIC TONE CANDRA ABOVE [Sharma]; comment in L2/09-404 "Feedback on Vedic Docs" [Scharf]

a. We recommend the acceptance of the 1CF4 VEDIC TONE CANDRA ABOVE, based on the evidence provided in L/209-344. (Peter Scharf also supported the encoding of this character in L2/09-404.) The name matches that for 1CD8 VEDIC TONE CANDRA BELOW. The appropriate properties are given in L2/09-344.

4. L2/09-423 Request for an annotation for the generic Indic dandas [Sharma]

a. Based on the evidence provided in L2/09-423, we recommend an annotation be added to 0965 DEVANAGARI DOUBLE DANDA, along the lines of that suggested in L2/09-423: "These characters are also used as abbreviation signs in some South Indian scripts."

5. L2/09-339 for normative aliases for some generic Indic characters [Sharma]

a. We recommend not accepting the request to make the current informative aliases be changed to normative aliases. Normative aliases are used when there is an error in a name (The Unicode Standard 5.2, Chapter 4, p. 129), which is not the case here.

b. We recommend informative aliases be added to 0964 ("Indic Danda") and 0965 ("Indic Double Danda").

Note: The request to add name aliases to 0951 DEVANAGARI STRESS SIGN UDATTA and 0952 DEVANAGARI STRESS SIGN ANUDATTA has already been accommodated in the draft

Unicode 6.0 names list file, but they were added as informative name aliases, for the reason given above (5.a.).

6. L2/10-062 "A note on Grantha OM Sign and only two kinds of virama in Grantha" [Ganesan]

a. The evidence for GRANTHA OM, provided in L2/10-062, is sufficient to recommend including 11350 GRANTHA OM in the repertoire of Grantha characters. (It is noted here that the Summary Grantha Report, L2/10-053, inadvertently omitted the fact that a GoI expert also had requested this character. Ganesan pointed this out in L2/10-062.)

b. As a result of the above action, we recommend the glyph for GRANTHA OM be added to the Grantha chart in L2/10-071 and its name be included in the names list in the same document.

Regarding comments on "chillu", see below, 8.d.

7. L2/10-153 "Proposal to encode Grantha chillu marker sign" [Ganesan]
See below, 8.d.

8. L2/10-085 "Feedback on Dr. Anderson's Grantha Summary dt 2010-Jan-28 [Sharma]

a. We recommend not accepting the request to change the name for anusvara and visarga from "GRANTHA SIGN __" to "GRANTHA LETTER __." The Grantha script should follow the pattern of naming and properties found in the other Indic scripts. (Regarding the request to change the general category property for anusvara and visarga, see above, 2.c.)

b. The comment about the general category property for GRANTHA VOWEL SIGN VOCALIC L OLD and GRANTHA VOWEL SIGN VOCALIC LL OLD is noted. This topic is postponed for later discussion.

c. The names GRANTHA VOWEL SIGN AU and GRANTHA AU LENGTH MARK seem appropriate, given the similar situation for Malayalam. However, we recommend that an annotation can be added to cover the usage of these characters, as suggested by Sharma in his proposal L2/09-372, pages 62 and 63.

d. Regarding the question of a ligating virama and the "chillu marker", we note the following comments in Sharma's proposal L2/09-372, page 22:

> *Thus we conclude that despite Grantha having these distinct forms of presenting the virama, there is no semantic difference between forms.*
>
> *We also draw the UTC's attention to the unambiguous statement, appended to this document in paragraph 17, of Dr. R. Krishnamurti Shastri, well known Sanskrit scholar and chief trustee of*

3

the major Grantha publisher of Tamil Nadu, Heritage India Educational Trust, that using these "special vowelless forms" as he terms them are in no way semantically distinct from adding the "vowel absence marker" that is the virama. Hence there should be no doubt about this matter.

Based on the above statements, we recommend the use of the GRANTHA SIGN VIRAMA and, for chillu forms, fonts (and ligature tables) be employed. That is, we recommend against encoding any chillus, chillu marker, or any special means of indicating the presence or absence of chillu forms.

If the representations are truly equivalent and there is truly no semantic difference between chillu and non-chillu representation of texts, then the rendering as chillu or conjunct is a freely variable choice that can be worked out in font software. No underlying difference in encoding is warranted. The encoding model is therefore "pure" in terms of using the virama, which fits with the "script agnostic" nature of the Sanskrit text as documented. Input from engineers implementing rendering of Indic scripts is encouraged, so they may work out any details or uncover any difficulties that may arise as a result of this encoding model.

e. Regarding the request for a name change to GRANTHA AU LENGTH MARK, see above, 8.c.

f. The repertoire in the Revised Chart and Names List, L2/10-071, only included characters that reflected consensus between the various proposals, so Vedic characters were not included. Currently encoded Vedic characters should cover many of the needs of Vedic scholars, but an explicit list of those characters that are still missing is needed. The Samavedic characters, in particular, are an issue. We recommend a specific document be submitted proposing any missing characters.

g. The request for annotations to be added for three characters, RRA, NNNA, and LLLLA, has already been accommodated in the Revised Chart L2/10-071, so no action is required.

h. While the discussion on how to deal with "Extended Tamil" is useful, we recommend postponing this discussion until Grantha has been approved.