

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: 27 October 2010

I. GRANTHA

1. Document: L2/10-409 Unicode Standard for Grantha Script, Summary of the Meeting Sept 6th, New Delhi - Gov't of India / Manoj Jain

We recommend the UTC endorse this meeting summary, which reflects consensus between the Government of India and experts, with the proviso that a few items approved by consensus in the summary still require discussion by the UTC. These are listed below, under **a-c**.

[Note: The set of characters in L2/10-409, pp. 6-14, is the same as that contained in L2/10-265R Revised Chart and Names List for Grantha, except for the items **a-c** below and a few annotations.]

a. 1135D GRANTHA SIGN PLUTA

Document: L2/10-331 Request to encode 1135D GRANTHA SIGN PLUTA - Sharma

We recommend encoding this character.

b. 1137E GRANTHA VOWEL SIGN SUB-BASE VOCALIC L and 1137F GRANTHA VOWEL SIGN SUB-BASE VOCALIC LL

Document: L2/10-341 Encoding sub-base Grantha vowel signs for Vocalic L and LL – Sharma

We do not make a specific recommendation on these characters, but defer to the UTC. In our opinion, the right-side spacing character should be encoded as the fundamental form, with the sub-base form handled by fonts. However, those members who implement Indic scripts should voice their views.

c. 1134F GRANTHA SIGN LIGATING VIRAMA

Documents:

L2/10-404 Finalizing the Grantha virama model - Sharma

L2/10-303 Proposal for a Grantha Character Name - Ligating Prepausal Virama at U+1134F - Naga Ganesan

L2/10-332 Comments on [L2/09-405, L2/10-297, L2/10-298 and] L2/10-303 - Sharma

L2/10-424 Letter regarding Prepausal Virama (Grantha) - Ulrike Niklas

L2/10-414 Letter regarding Grantha virama - Krishnaswamy Nachimuthu

RESPONSE: L2/10-423 Response to L2/10-414 on the GRANTHA "PREPAUSAL" VIRAMA - Sharma

L2/10-406 Letter regarding Grantha vowelless consonants - T. Ganesan

L2/10-297 Comment on L2/10-285: Handwritten styles in Tamil, Grantha, and English scripts - Naga Ganesan

We recommend the UTC agree with the decision by the Government of India to remove 1134F GRANTHA SIGN LIGATING VIRAMA, and direct implementers to use L2/10-404 to inform their implementations.

d. General Category of GRANTHA SIGN VIRAMA

Document: L2/10-405 On the GC of 1134D Grantha Sign Virama – Sharma

We recommend the general category for 1134D Grantha Sign Virama be Mc, based on the case made in L2/10-405. (The gc-Mc is reflected in the Grantha script proposal, L2/10-426, e below.)

e. Proposal for Grantha

Document: L2/10-426 Proposal to encode the Grantha script in Unicode - Government of India

We recommend the characters in this proposal be approved.

We recommend the UTC approve the three canonical equivalences in this document (1134A GRANTHA VOWEL SIGN, 1134B GRANTHA VOWEL SIGN OO, and 1134C GRANTHA VOWEL SIGN AU).

An updated proposal, with the addition of those characters in (a) and (b), if approved by the UTC, should be sent to WG2.

We recommend the chart in L2/10-265 be updated with changes based on the decisions by the UTC and suggest the authors word annotations as reflected in the proposal (L2/10-426) and the meeting report (L2/10-409). This new, auxiliary Grantha document with the updated chart and nameslist should be sent to WG2.

II. TAMIL

1. "Extended Tamil"

Documents:

[Core Document] L2/10-256 Proposal to encode characters for Extended Tamil - Sharma

L2/10-379 Follow-up to Extended Tamil proposal - Sharma

L2/10-407 Follow-up #2 to Extended Tamil proposal - Sharma

L2/10-363 Naming of Characters in Extended Tamil Block in SMP -INFITT

RESPONSE: L2/10-380 Response to L2/10-363 from INFITT- Sharma

L2/10-298 Tamil Letters with Numbered Superscripts – Comment on L2/10-256 - Naga Ganesan

RESPONSE: L2/10-332 Comments on L2/09-405, L2/10-297, L2/10-298 and L2/10-303 - Sharma

L2/10-425 Comments on L2/10-256R and L2/10-363 - Vinodh Rajan

We do not recommend encoding characters for Extended Tamil, as these can already be represented with sequences. However, we agree that some changes may need to be made to fonts and rendering engines in order to represent the text as desired. Indic rendering engines, for example, will need to

know that the superscript numbers should be treated as diacritics (that is, in the nukta class). As with other Indic script topics, members who have implementations should comment.

2. Tamil Fractions

Documents:

L2/10-334 Request to encode Tamil fractions (replaces L2/09-376 ; revision 2) - Sharma

L2/10-408 Encoding of Tamil Fractions in Unicode - INFITT / Naga Ganesan

L2/10-428 Revised Tamil Fractions Proposal - Sharma

We recommend a decision on Tamil fractions be postponed, as there are still inconsistencies between the proposals. In the interim, we recommend a consensus document be created. For the consensus document, we recommend the non-Tamil names be used as character names (as in L2/10-334), 1/16 not receive a canonical decomposition, and 1/32 be included, based on the ligated form as document on page 9 of L2/10-408.

We recommend the Roadmap allocation be deferred pending further research, as several symbols are reported to be proposed in the near future, but the number of characters is not known.

III. OTHER PROPOSALS AND DOCUMENTS

1. Vedic Tone Asterisk Above

Documents: L2/10-349 Request to encode 1CF7 VEDIC TONE ASTERISK ABOVE - Sharma

L2/10-398 Follow-up to Asterisk Above Proposal L2/10-349 - Sharma

We recommend not accepting this character, and suggest instead the use of 20F0 COMBINING ASTERISK ABOVE. However, Indic rendering implementers should comment on this.

2. Deprecate Telugu Length Mark

Document: L2/10-339 Request to deprecate 0C55 TELUGU LENGTH MARK - Sharma

We recommend against deprecating 0C55 TELUGU LENGTH MARK, and instead remand the topic to the Ed Committee, which can consider adding sequences to the TELUGU VOWEL LETTER table. (The table would clarify which characters to use (and not use) when representing vowel letters in text.)

3. Brahmi Numbers

Document: L2/10-340 Using ZWJ in the encoded representation of Brahmi numbers - Sharma

We recommend no change be made now, but remand this topic to the Ed Committee to make sure no text is included in 6.0, pending more discussion by interested parties.

4. Jihvamuliya and Upadhmaniya characters in Brahmi and Sharada

Document: L2/10-330 Concerning the Brahmi and Sharada characters for jihvamuliya and upadhmaniya (revised) - Sharma

We do not make a recommendation on this document, but instead defer to the UTC for discussion, and in particular for the views of those members with Indic scripts implementations.

The UTC needs to decide if the standard should retain the current handling of jihvamuliya and upadhmaniya (as jihva./upadh. + virama + consonant), or if a change should be made, as reflected in L2/10-330. This topic affects the Brahmi block description for 6.0 (and will affect the future block description for Sharada, which is in the third edition of 10646).

5. South Indian Candrabindus

Document: L2/10-392 Request to encode South Indian candrabindu-s – Sharma

We recommend encoding three characters as proposed and adding an annotation to 0C01 TELUGU SIGN CANDRABINDU along the lines suggested in L2/10-392.

The three characters are:

0C00 TELUGU SIGN NON-SPACING CANDRABINDU

0C81 KANNADA SIGN CANDRABINDU

0D01 MALAYALAM SIGN CANDRABINDU

6. Khojki

Document: L2/10-32 Revised Proposal to Encode the Khojki Script – Pandey

We recommend the approval of the characters as proposed in L2/10-32. The author of the proposal needs to verify whether the block could be packed into four columns, which could allow two other historic scripts to be located in the same row (which begins with U+11200), or if it is likely five columns will be needed for additional characters.

7. Mahajani

Document: L2/10-377 Preliminary proposal to encode Mahajani – Pandey

As this is a preliminary proposal, we recommend no action be taken at this time, but members should review the document.

A few comments: Rather than encode a virama solely for the SHRII ligature, we would recommend encoding the SHRII ligature itself. We also suggest the author respond in the proposal to the comment made by John Cowan, who recommended dropping “A” in the letter names, since the script is an alphabet and not an abugida. (In our view, the current set of letter names – with “A” – is fine.)

8. Script Name Change for Sindhi

Document: L2/10-420 Request to Rename ‘Sindhi’ to ‘Khudawadi’ – Pandey

We recommend the UTC respond that this requested change is too late to be accommodated, since “Sindhi” has been assigned a script code in ISO 15924 and various dependent registries are already

affected. A name change at this point would create a permanent mismatch between the Unicode/ISO 10646 block name and the ISO 15924 script code.

At this point, clarification on the name of the script should be handled as an annotation for Sindhi in the ISO 15924 code tables, such as "Sindhi (Khudawadi)". Or, if it seems appropriate to call out a specific Khudawadi variety of Sindhi as a separate entity, it could be registered in 15924 as a variant of Sindhi.