

To: UTC

L2/12-267

From: Deborah Anderson, Rick McGowan, and Ken Whistler

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

Date: 21 July 2012

This set of recommendations covers documents on Indic-related topics submitted as of 19 July 2012 (§I-§IX, pp. 1-5), as well a copy of recommendations and comments on documents that were not covered at the May 2012 UTC meeting (§X-§XVII, pp. 5-8).

I. SIDDHAM

Documents: L2/12-234 Proposal to Encode the Siddham Script - Anshu Pandey

L2/12-221 Comments on naming the "Siddham" encoding - Vinodh Rajan, Shriramana Sharma

Discussion: We reviewed L2/12-221 and L2/12-234.

The script proposal itself, L2/12-234, now includes decompositions for two vowel signs, which were requested at the May 2012 meeting. Earlier versions of the script proposal have been seen by the UTC, and the current proposal appears to be very mature.

In L2/12-221, two members of the Indic community recommend the name "Siddham" for the script.

Recommendation: We recommend the UTC approve the script with the name "Siddham."

II. TAMIL

Documents: L2/12-231 Proposal to encode Tamil fractions and symbols - Shriramana Sharma

L2/12-232 Regarding my Tamil fractions/symbols proposal - Shriramana Sharma

Discussion: We reviewed L2/12-232, and only quickly reviewed the proposal itself, L2/12-231, which is long (66 pages) and complicated.

Two questions for the UTC are raised in §1 and §2 of L2/12-232:

In §1, the author asks whether the unification of 0BF3 TAMIL DAY SIGN with the measure of grain "nāḷi/paṭi" is appropriate. While the glyph and character properties (gc=So) are shared between the two, the meaning is different, and collation would also be different.

In §2, the author asks whether the numeric property he has assigned for 11FD2 TAMIL FRACTION DOWNSCALING FACTOR KIIZH is correct.

In §3 of L2/12-232, the author requests the UTC take no action on the Tamil fractions and symbols proposal, as he is soliciting feedback from relevant parties in India and Sri Lanka.

Recommendation: Concerning L2/12-232, we recommend:

For §1, the UTC should collectively review the unification of TAMIL DAY SIGN for "nāḷi/paṭi" and make a decision, preferably at the August 2012 meeting.

For §2, the numeric values proposed for TAMIL FRACTION DOWNSCALING FACTOR KIIZH appear to be appropriate if this character is to be encoded.

In view of §3, we recommend the UTC members review the main proposal, L2/12-231, and if they have any expertise or knowledge of details of Tamil fractions and symbols, they should relay feedback to the document's author.

III. MALAYALAM

Documents: L2/12-225 Proposal to encode 0D5F MALAYALAM LETTER ARCHAIC II

Shriramana Sharma

L2/12-233 Comments on Encoding "Duplicate" Indic Characters - Vinodh Rajan

L2/12-236 Update on Brahmi and other Indic unification issues - Shriramana Sharma

Discussion: We reviewed the three documents that discuss MALAYALAM ARCHAIC LETTER II: the main proposal, L2/12-225, and comments in L2/12-233 and L2/12-236 (specifically §3).

While the comments in L2/12-233 and L2/12-236 deal in part with proposed Brahmi characters—now moot because the Brahmi proposal has been withdrawn—they also touch on arguments raised about encoding a new MALAYALAM ARCHAIC LETTER II.

The rationale for encoding a new MALAYALAM ARCHAIC LETTER II is based on the fact that the written forms of the proposed MALAYALAM ARCHAIC LETTER II and 0D08 MALAYALAM LETTER II are distinct and have no common glyphic skeleton, although their sound values are the same. (According to Sharma in §3 of L2/12-236, this same argument does not apply to Brahmi, a paleographic script with many variations and inconsistencies.)

Rajan's document, L2/12-233, argues the proposed character is a variant glyph and should be handled at the font level.

Of particular interest to the UTC is Rajan's concern ("Font Level Handling", page 3) that there is at least one font that already treats "archaic letter ii" as an alternate glyph. NOTE: The link to the font URL in L2/12-233 should have an ASCII dash. The correct URL is: <http://www.aai.uni-hamburg.de/indtib/INDOLIPi/Indolipi.htm>.

A core issue is whether a clear case has been made that the "archaic letter ii" is a distinction that should be maintained in plain text.

Recommendation: We recommend the encoding of a separate character MALAYALAM ARCHAIC LETTER II be discussed in UTC, taking into consideration that if the encoding for 0D08 MALAYALAM LETTER II were to be split, an existing implementation may be invalidated.

IV. BRAHMI

Documents: L2/12-236 Update on Brahmi and other Indic unification issues - Shriramana Sharma

(WITHDRAWN: L2/12-226 Proposal to add two characters for Brahmi - Shriramana Sharma]
L2/12- L2/12-233 Comments on Encoding "Duplicate" Indic Characters - Vinodh Rajan

Discussion: We reviewed L2/12-236 and L2/12-226. We note that in L2/12-236 the author has withdrawn the proposal for the separate encoding of BRAHMI SIGN PULLI and BRAHMI LETTER OLD TAMIL LLA.

Recommendation: We recommend no action be taken by the UTC.

V. AHOM

Document: L2/12-222 **Proposal to add the Ahom Script in the SMP - Martin Hosken, Stephen Morey**

Discussion: We reviewed this document, and have several comments to be relayed to the authors before submission of an Ahom proposal to WG2:

a. Location

The codepoints in the proposal need to be revised; the proposal uses codepoints that are now taken up by Khudawadi, which is on an advanced ballot (DAM1). The script should be moved to 11700-1173F, which is the current location on the Roadmap.

b. Digits

The section on Digits is problematical, since this is not a decimal radix system. The section should be re-written with more explanation and details, including:

- clarification that this is not a decimal radix system.
- text addressing the “open issue” referred to in the sentence: “Full details of what should be used specifically for 2, 3 and 4 are an open issue.”
- at least one additional attestation of numbers with accompanying explanation (besides ‘58’ in figure 3).

Also, the properties for these characters should be changed from “Nd” to “No”, with 0 to 9 getting numeric values in two columns, and 10 and 20 getting a numeric value in the third column.

c. Variant forms

The section on Variant forms also needs revision, as variation sequences need to be explicitly proposed, and VS are not allowed for ligatures. Suggested wording might be, “We have the following forms.... One way to analyze them is to use variants in fonts. Another approach is to define variation sequences. We ask the UTC to make recommendations on how to handle these forms.”

d. Sort order

The sort order for medials needs to be clarified.

The following sentence (page 4) is unclear: “For the purposes of default collation, vowels are ordered according to their codepoint value, likewise for the two medials. The relative block order is:

Consonants, Vowels, Medials.” The sentence suggests the order has the medials follow the vowels, but in the code chart, the medials are between the consonants and vowels.

Regarding the medial consonants, the proposal should recommend which of the following behaviors is preferred:

1. Medial consonants be given primary weights separate from the set of initial consonants, e.g., as in chart order, with the medial consonants following the entire main set of consonants.
2. Medial consonants be given primary weights separate from the initial consonants, but interfiled, so that medial *la*, for example would be immediately after *la* in sort order.
3. Medial consonants be given primary weights identical to the corresponding initial consonant, but be distinguished at a secondary level. (This would be roughly the same as the distinction between a letter and an accented letter, in the case of alphabets.)

Recommendation: We recommend the above comments be relayed to the authors, and that other UTC members carefully review the proposal and also send comments to the authors.

VI. KANNADA

Document: L2/12-203 **Supporting Tulu language written in the Kannada script - Shriramana Sharma**

Discussion: We reviewed this document, which is primarily a “FYI” document to UTC members regarding a potential revision to the orthography for the Tulu language in the Kannada script.

Recommendation: We recommend the UTC take no action, but members with Kannada rendering engines should be on notice that the orthography for the Tulu language in the Kannada script is complicated, and the representation for certain vowels may change.

VII. SHARADA

Document: L2/12-224 **Annotation for 111C4 SHARADA OM - Shriramana Sharma**

Discussion: We reviewed this proposal.

Recommendation: We recommend the UTC note this short document and suggest it go to the Editorial Committee as feedback on the names list.

VIII. GONDI

Document: L2/12-235 **Revised Preliminary Proposal to Encode the Gondi Script - Anshuman Pandey**

Discussion: We reviewed this document, which is a revision of an earlier preliminary proposal. We note the script is already on the Roadmap.

Recommendation: We recommend UTC members review this proposal and send the author any feedback.

IX. ROHINGYA

Document: L2/12-214 Preliminary Proposal to Encode the Rohingya Script - Anshuman Pandey
2012-06-25

Discussion: We reviewed this document, which is preliminary only. Note that the script proposal includes a TATWEEL. Additional information on the extent of its use should be included in the next version of the proposal, along with a discussion of whether a specific TATWEEL should be separately encoded.

Recommendation: We recommend UTC members review this proposal and send the author feedback.

X. COORGI-COX

Document: L2/12-217 Introducing the Coorgi-Cox Alphabet - Anshuman Pandey

Discussion: We reviewed this preliminary proposal. The script has not yet been put on the Roadmap, which seems appropriate since the extent of its use is not yet verified.

Recommendation: We recommend UTC members review this proposal and send the author feedback.

SCRIPT TOPICS CARRIED OVER FROM MAY 2012 UTC MEETING

XI. Generic Indic comments

Document: L2/12-106 Request for editorial updates to various Indic scripts, §1 - Shriramana Sharma

a. L2/12-147 Review of Indic-related L2 documents and Recommendations to the UTC

Discussion: We reviewed §1 of L2/12-106.

Recommendation: While the proposed text changes to the text seem reasonable, we recommend UTC members, particularly those with rendering engines for Devanagari, review this request and verify the proposed change in the text of Chapter 9 is acceptable.

b. L2/12-184 from GOI: The proposal is being referred to experts to give their views.

XII. TAMIL

Document: L2/12-106 Request for editorial updates to various Indic scripts, §2 Tamil Debit and Credit - Shriramana Sharma

a. L2/12-147 Review of Indic-related L2 documents and Recommendations to the UTC

Discussion: We reviewed §2 of L2/12-106.

Recommendation: We recommend the glyph for U+0BF6 TAMIL DEBIT SIGN be changed, based upon this document and personal email feedback from Uma Umamaheswaran, who was an author of the original proposal (with INFITT), L2/01-375R. Additional feedback from

INFITT would also be welcome. For U+0BF7 TAMIL CREDIT SIGN, we recommend additional feedback be sought from UTC members and other interested parties.

b. L2/12-180 from INFITT:

[...] Since the original Tamil Debit Sign glyph is based on the Government of Tamil Nadu 8-bit standards for Tamil and the proposal that INFITT co-authored with Dr. Uma Umamaheswaran, INFITT has an interest in reviewing Mr. Shriramana Sharma's proposal (L2/12-106). INFITT's WG02 members Mr. P. Chellappan (who provided the glyphs in the INFITT-Uma proposal) and Dr. Ramakrishnan are part of the user community that uses the debit sign in their accounting. They insist that the shape in the INFITT-Uma proposal is what they learned to use. Still, INFITT would like to review this further and offer up additional comments once we have done more research and have had an opportunity to consult additional experts in the user community. We expect to submit our comments before the August UTC meeting.

c. L2/12-184 from GOI:

The proposal is being referred to experts to give their views. Hence we request the UTC to defer the decision on the above proposals till next UTC meeting

XIII. MALAYALAM

Document: L2/12-106 Request for editorial updates to various Indic scripts, §3 Malayalam - Shriramana Sharma

a. L2/12-147 Review of Indic-related L2 documents and Recommendations to the UTC

Discussion: We reviewed §3 of L2/12-106.

Recommendation: We recommend the Malayalam block intro text be modified to acknowledge the glyphic form of the dot reph as a small vertical stroke, besides the dot shape. Before making changes to the text regarding C2 conjoining forms, we suggest the UTC solicit feedback from members with Malayalam rendering engines.

L2/12-184 from GOI:

The proposal is being referred to experts to give their views. Hence we request the UTC to defer the decision on the above proposals till next UTC meeting

XIV. ORIYA

Document: L2/12-106 Request for editorial updates to various Indic scripts, §4 Oriya - Shriramana Sharma

a. L2/12-147 Review of Indic-related L2 documents and Recommendations to the UTC

Discussion: We reviewed §4 of L2/12-106.

Recommendation: Because we have no expertise on current rendering engines, we defer to the UTC on the proposed wording change, specifically to get feedback from those members with working Oriya implementations.

b. L2/12-184 from GOI:

The proposal is being referred to experts to give their views. Hence we request the UTC to defer the decision on the above proposals till next UTC meeting

XV. BRAHMI

Document: L2/12-106 Request for editorial updates to various Indic scripts, §5 Brahmi - Shriramana Sharma

a. L2/12-147 Review of Indic-related L2 documents and Recommendations to the UTC

Discussion: We reviewed §5 of L2/12-106.2

Recommendation: We recommend the UTC approve this glyph change.

b. L2/12-184 from GOI:

The proposal is being referred to experts to give their views. Hence we request the UTC to defer the decision on the above proposals till next UTC meeting

c. L2/12-165 Brahmi Letter LLA glyph in TUS - Ganesan

d. L2/12-236 Update on Brahmi and other Indic unification issues - Shriramana Sharma

In §2, the author provides an update to L2/12-147 §5 and notes that additional feedback is being sought on which glyph is the oldest and most appropriate for the code chart.

XVI. BENGALI

Documents: L2/12-121 Proposal to Encode the Sign ANJI for Bengali - Anshuman Pandey (replaces L2/11-359)

a. L2/12-147 Review of Indic-related L2 documents and Recommendations to the UTC

Discussion: We reviewed this document and found it to be sound.

Recommendation: We recommend the UTC approve the character. We note that there will be ample opportunity for National Bodies to comment on it, once it is on a ballot.

b. L2/12-184 from GOI (concerning L2/11-359 and L2/12-121)

In the proposal document L2/11-359 the author has not provided any usage example of the proposed sign. Also in the revised document L2/12-121, the author has provided one example from the Devanagari document, which is to be examined by the experts. Experts have been requested to give feedback on the issue, so far no usage example could be found in the Bengali script.

XVII. DEVANAGARI

Document: L2/12-123 Proposal to Encode the Sign SIDDHAM for Devanagari -Anshuman Pandey

a. L2/12-147 Review of Indic-related L2 documents and Recommendations to the UTC

Discussion: We reviewed this document. Because the layout for Tibetan is quite different from Devanagari, the separate encoding of this character for Devanagari is warranted in our opinion.

Recommendation: We recommend the character be renamed to “DEVANAGARI SIDDHAM SIGN”, in order to help differentiate it from script “Siddham”, and that the UTC approve it.

b. L2/12-184 from GOI:

The proposal is being referred to experts to give their views. Hence we request the UTC to defer the decision on the above proposals till next UTC meeting

XVIII. SHARADA

Document: L2/12-124 Proposal to Encode Signs for Writing Kashmiri in Sharada -Anshuman Pandey

a. L2/12-147 Review of Indic-related L2 documents and Recommendations to the UTC

Discussion: We reviewed this proposal, and find it sound.

Recommendation: We recommend the UTC accept these characters.