

TO: UTC

FROM: Deborah Anderson, Ken Whistler, Rick McGowan, and Roozbeh Pournader

SUBJECT: Recommendations to UTC #137 November 2013 on Script Proposals

DATE: 31 October 2013

EUROPEAN

Latin

1. [L2/13-181](#) Precomposed Tilde Letters – Bodor

Discussion: We reviewed this short document, which asked for precomposed Latin letters with a tilde. (The author also asks for his “International Phonemic Alphabet” to be supported, as shown on <http://zhoroscop.wordpress.com/>.)

We consider the request for precomposed Latin letters unacceptable, because all of these letters are currently representable by sequences involving combining characters in Unicode.

Recommendation: We recommend the UTC go on record as not wanting this request to progress, but it may wish to direct the author to the relevant FAQs on the Unicode website, including http://www.unicode.org/faq/char_combmark.html, which discusses precomposed characters and display issues.

AFRICAN

2. [L2/13-191](#) Preliminary proposal to encode the Adlam script (WG2 N4488) – Everson/SEI

Discussion: We reviewed this preliminary proposal, which is for a relatively new script.

Our group had several comments that the author and user community should address in a future revision, including:

- More script samples are needed, and they should show consistency. The current set of figures suggests that the script is still in development and is not yet ready for encoding in a standard.
 - Specific comments:
 - Several glyph shapes in figures 1-4 differ from one another and from the code chart (where lowercase shapes are just smaller versions of uppercase), cf. ALIF and the glyph for the digit ‘8’
 - Case distinction does not appear to be shown in the publication (figure 5)
- The “Structure” section states that the script can be written with letters either separate or joined, and is “a question of font style.” Implementation of the script will require additional details.
- The re-use of existing punctuation marks in section 8 suggests these characters should have entries in Script_Extensions.txt, so a section of the proposal should mention this.
- The glyph for 1E948 ADLAM DOT BELOW does not match the glyph, which is a small line below.

Recommendation: We recommend the UTC review this proposal and send comments to the author.

AMERICAN

Cherokee

3. [L2/13-190](#) Proposal for the addition of Cherokee characters (WG2 N4487) – Everson and Feeling

Discussion: We reviewed this proposal, which requested the addition of an archaic letter MV in the current Cherokee block, a new LOW COLON punctuation mark in the Supplemental Punctuation block, and a new SMP Cherokee Supplement block with 86 uppercase letters.

A question was raised about the LOW COLON, since the samples provided (figs. 10 and 11) do not support interpretation as a normal-sized colon set low on the line, but seems rather to be a colon that is smaller in size and could even be a subscript colon.

While the CHEROKEE LETTER MV and LOW COLON appear somewhat less problematical, the new uppercase letters will require discussion between those members having Cherokee implementations and fonts and the user community. Although the addition of uppercase letters to current lowercase letters does not break casing stability, the transition will be problematical for several years and we feel the community should be informed of all the potential ramifications.

Recommendation: We recommend the UTC discuss this proposal with user community participation.

MIDDLE EASTERN

Arabic

4. [L2/13-178](#) Proposal to encode five Arabic script characters for the Bravanese – Banfunzi

Discussion: We reviewed this proposal for 5 Arabic characters for the Bravanese (Chimiini) orthographies:

ARABIC LETTER BEH WITH SMALL ARABIC LETTER MEEM ABOVE

ARABIC LETTER PEH WITH SMALL ARABIC LETTER MEEM ABOVE

ARABIC LETTER TEH WITH CURL

ARABIC LETTER REH WITH SMALL ARABIC LETTER NOON ABOVE

ARABIC LETTER YEH WITH TWO DOTS BELOW AND SMALL ARABIC LETTER NOON ABOVE

While it may appear that the characters could be represented with combining marks, it should be noted that the letters appearing above the base characters are not that same as the Koranic annotation marks.

We recommend more discussion take place on the following points:

- review the name and glyph for ARABIC LETTER TEH WITH CURL
- review the entries in ArabicShaping.txt for these characters, as they may need to be corrected
- remove “SMALL ARABIC LETTER” from the names of four characters.

We also suggest the the schematic names (in ArabicShaping.txt) be sent to the Editorial Committee for review.

Recommendation: We recommend the UTC encode 5 characters after discussion, taking into consideration the points raised above.

SOUTH AND CENTRAL ASIA

Siddham

5. Siddham variants

[L2/13-189](#) Comments on N4407R Proposal to Encode Variants for Siddham Script (WG2 N4486) – Glass

[L2/13-195](#) Practical approach to encoding Siddham variants – Pandey

Background docs:

[L2/13-110](#) Proposal to Encode Variants for Siddham Script – Kawabata

[L2/13-136](#) Additional Siddham Variants – Pandey

Discussion: We reviewed the two new documents. The document by Andrew Glass, which was based on discussion at the July/August 2013 UTC, recommended the variants proposed by Kawabata (et al.) be used only for logographs, but not as palaeographs for syllables. He also asked whether the dependent U forms occur only with the base HA. If they do, then it would make sense to encode the whole syllable. The “Practical Approach” document by Anshuman Pandey recommends taking the characters used as ideographs out of the Siddham block proper and placing them into a separate block. He also recommended a name change for the six characters, so they would be more descriptive.

We deem the “Practical Approach” advocated by Anshuman Pandey to be a good suggestion and worthy of more consideration. Having a separate area just for logographically used characters would make it less likely that the variants would be used in non-logographic contexts.

One consideration is the available space in the Siddham block and surrounding columns: currently, the six variants are in a separate column (115E0..115E5), which is followed by an empty column (115F0..115FF). There are two empty columns before the Siddham block (11560..1157F). What is the intended use of the open columns? One option would be to create a block as proposed by Pandey, physically separated from the main Siddham block. An alternative approach would be to keep the variants where they are, but have them set off with a clear identification in the headers.

Recommendation: We recommend the UTC discuss the suggestion proposed in [L2/13-195](#) by Anshuman Pandey, and consider making a ballot comment on 4th edition to delay consideration of the 6 Siddham letter variants, while possible alternative approaches are discussed and pending the outcome of discussions in the U.S. and in Japan with experts.

Tamil

6. Tamil font [L2/13-174](#) Request to change the Unicode chart font for Tamil (WG2 N4476) – Sharma

Discussion: We reviewed this document, which requested a change of the Tamil chart font.

In our judgment, the rationale for the font change seems reasonable and the Lohit Tamil typeface reflects a stylistic improvement to the current font. As noted in the document, the new Tamil fractions characters were designed in the same Lohit Tamil font style. Also, because the Lohit Tamil font is a fully functioning font, it could help in the production process of the Core Spec, annexes and technical reports.

Recommendation: We recommend the UTC accept the font change for Tamil as proposed, and support a ballot comment on the 4th edition to this effect, but caution the editors take care in converting the font.

7. Tamil fraction and symbol names

Documents:

[L2/13-193](#) Comment on Tamil Nadu Government Request L2/13-161, Use of Tamil Names for Tamil Fractions and Symbols – Ganesan

[L2/13-175](#) Response to L2/13-161 on naming Tamil fractions and symbols (WG2 N4477) – Sharma

Background docs:

[L2/13-161](#) Request to Use Tamil names in Tamil fractions and symbols – Tamil Virtual Academy’s request to use Tamil names in Tamil fractions and symbols)

[L2/13-047](#) Revised proposal to encode Tamil fractions and symbols– Sharma

Discussion: We reviewed the new documents. In [L2/13-193](#) Naga Ganesan reported that the Tamil Nadu Government set up a committee to review the Tamil character names, with the result that the committee all supported the changes described in [L2/13-161](#). Document [L2/13-175](#) is a response by S. Sharma that questioned the “Tamil transcriptions [sic] standard” referred to in [L2/13-175](#) and provided rationale for his spelling, which more closely reflects how the words are pronounced.

While we agree that it is not clear what Tamil transcription standard is being referred to, the Tamil Virtual Academy’s recommendation seems to be based on dictionary forms, such as those in the University of Madras Tamil Lexicon, where initial c is the preferred spelling (instead of initial s), initial p (instead of initial b), medial k (instead of medial g), medial t (instead of medial d), and the long vowels spelled out. By relying on the entries in a standardized reference (such as a dictionary), the spelling in the Unicode Standard will be as consistent as possible.

For the representation of certain liquids and nasals, such as ṛ, ḷ, ḹ, ṇ and ṇ̄, the spelling RR, LLL, LL, NN and NNN is the convention already present in the names list. The ASCII-based spelling of these letters (ṛ, ḷ, ḹ, ṇ, and ṇ̄) when they appear geminated would be, we agree with Sharma, rather awkward. As noted in [L2/13-175](#), pillai would be spelled PILLLLAI.

For GEJAM / KESAM, see §8 below.

Recommendation: We recommend the UTC adopt the current spelling described above (except for the liquids and nasals mentioned), and support a ballot comment for PDAM1 reflecting this. For the cited liquids and nasals, we recommend the UTC invite the TVA to prepare acceptable suggestions, particularly on how to deal with these letters when geminated. We also recommend the changes to the annotations be directed to the Editorial Committee for review.

8. Name Change to One Character

[L2/13-176](#) Request to change one character name from N4430 L2/13-047 (WG2 N4478) – Sharma change one character name in N4430 [L2/13-047](#))

Discussion: We reviewed this proposal, which requests a name change from GEJAM to GAJAM.

Recommendation: We recommend the UTC change the name to KEJAM, which is one of the spellings found in the University of Madras Tamil Lexicon. We suggest the UTC support a ballot comment for the PDAM1 to this effect. Because the word has alternate transliterations (as mentioned in [L2/13-176](#)), the UTC may consider entertaining a request to add the alias KAJAM.

Grantha

9. [L2/13-179](#) Request to reallocate one character within the Grantha block (WG2 N4479) – Sharma

Discussion: We reviewed this proposal, which requested moving one character, GRANTHA SIGN COMBINING ANUSVARA ABOVE, from U+1137D to U+11300. This character is currently in PDAM1. Moving the ANUSVARA to the beginning of the block would keep it in line with the location of the anusvara in other Indic script blocks. Since the character is only in the PDAM stage, it would be fairly easy to make the change at this point.

Recommendation: We recommend the UTC agree to move GRANTHA SIGN COMBINING ANUSVARA ABOVE from U+1137D to U+11300, and draft a ballot comment on PDAM1 in support of this change.