

TO: UTC **L2/15-312**
FROM: Deborah Anderson, Ken Whistler, Rick McGowan, Roozbeh Pournader, Andrew Glass, and Laurentiu Iancu
SUBJECT: Recommendations to UTC #145 November 2015 on Script Proposals
DATE: 1 November 2015

The recommendations below are based on documents available to the members of this group at the time they met, October 9, and do not include documents submitted later to the document registry. Some documents have been revised based on the recommendations.

SOUTH ASIA

Indic


1. Dogra

Document: [L2/15-234](#) Proposal to encode the Dogra script – Pandey

Discussion: We reviewed this document, which is an update of [L2/15-213](#)

The following comments were made:

- Mention in the introduction that a third Takri-spin-off, Sirmauri, could be separately encoded.
- Section 3.15 mentions that fraction signs and currency marks may be represented with the characters in the Common Indic Number Forms block. Add the code points for the block in addition to the reference to the proposal (L2/07-354).
- Add “Old” to Dogra in the captions of the relevant figures.
- The caption for figure 14 has an extra space in “qalamra o” [Note: In revised proposal, L2/15-234, the caption now reads “qalamrau”]
- Identify what characters can be used to represent Old Dogra I, II, U and UU
- In Section 3.14, adjust the text to reflect the decision to unify Dogra digits with Devanagari digits, and adjust Section 4.5 ScriptExtensions.txt accordingly
- In Section 3.12 in the middle of page 8, adjust the following example by removing the ZWJ (and massage the text around this example accordingly):

᳚᳚ ᳚᳚ <᳚ RA, Q VIRAMA,  ZERO WIDTH JOINER, *᳚᳚* NNA>

Remove the following repha example:

᳚᳚ ᳚᳚ <᳚ RA, Q VIRAMA, *᳚᳚* NNA>

- In section 4.3 Syllabic categories, the code point ranges and general categories need to be fixed (and made consistent with gc properties in Section 4.1).

For example,

Section 4.3

```
# Indic_Syllabic_Category=Vowel_Dependent
```

```
1182C..11835 ; Vowel_Dependent # Mc [4] DOGRA VOWEL SIGN E .. DOGRA VOWEL SIGN AU
```

Section 4.1

```
11832;DOGRA VOWEL SIGN E;Mn;0;NSM;;;;;N;;;;;
```

```
11833;DOGRA VOWEL SIGN AI;Mn;0;NSM;;;;;N;;;;;
```

```
11834;DOGRA VOWEL SIGN O;Mn;0;NSM;;;;;N;;;;;
```

Recommendations: We recommend the UTC members review this proposal and send feedback to the author, including the comments above.

2. *Gunjala Gondi*

Document: [L2/15-235](#) Proposal to encode the Gunjala Gondi script – Pandey

Discussion: We reviewed this document and recommend the following:

- Add script extensions information on the dandas
- Identify Telugu in figures 9 and 10
- Provide a comparison chart showing Gunjala Gondi and Modi
- Change the virama (U+11D97) to be like that in Masaram (i.e., it should appear with a plus sign in a box and have the same properties, note that the entry in `Indic_Positional_Category=Bottom` should be removed). If, at a later time, a halant is found, provide evidence for it and propose the character.
- Add information in the Background section on the usage of Masaram Gondi vs. Gunjala Gondi (i.e., geographical location, language use [is Gunjala is used for Southern Gondi and Masaram only for Northern Gondi (ISO 639-3: gno)?], etc.)
- Clarify whether font permissions have been obtained (if needed).

Recommendations: We recommend the UTC members review this proposal and send feedback to the author, including the comments above.

CENTRAL ASIA

3. *Tocharian*

Document: [L2/15-236](#) Proposal to Encode the Tocharian Script – Wilson

Discussion: We briefly reviewed this proposal and had the following comments:

- In Section 4.9, provide non-typeset examples of the number “1” by itself and in combination.
- Section 4.12 Virama needs additional work and clarification: although one virama is proposed, the “proposed implementation” mentions a *Fremdzeichen* with virama, a *Fremdzeichen* with two viramas, as well as a standard virama.
- In Section 5.1 Character Properties, remove “reserved for” and “this position shall not be used”
- In Section 5.2, change the code points beginning with “113--” to “11E--” and verify other code points in properties section (for example, VIRAMA should be 114E4).
- In figures 5 through 11, identify in the captions what the images contain, circling specific forms.
- In Section 4.13, subscript independent vowel letters, which are marked by a virama, are proposed. Some current font technologies don’t support this model. Further investigation is needed whether it would be better to extend such font technologies or propose separate dependent vowels to be encoded.

We encourage Andrew Glass, Roozbeh Pournader, Anshuman Pandey, and others to carefully review the proposal and get back to the author.

Recommendations: We recommend the UTC members review this proposal and send feedback to the author, including the comments above.

INDONESIA AND PHILIPPINES

4. *Makasar Script*

Documents: [L2/15-233](#) Proposal to Encode the Makasar Script – Pandey

Discussion: We briefly reviewed this proposal and had the following comments:

- Change “Makassarese” to “Makasar” in Section 4.6 (page 10)
- The note on the names list page (p. 17) says: “This script is known indigenously as Ukiri' Jangangjangang and in English as Old Makassarese and Makassarese Bird Script.” Based on feedback from the Ian Caldwell, “Makassarese” was used only by one scholar (Bill Cummings), and he no longer publishes on South Sulawesi; instead, “Makasar” is the term that is in wide use. As a result, it would be advisable to change the note to instead use “Old Makasar”.
- Section 3.2 mentions that “‘Makassarese’ is the most commonly known [spelling]”; this too should be aligned with the names list note and eschew “Makassarese”.

Recommendations: We recommend the UTC members review this proposal and send feedback to the author, including the comments above.

5. *Kulitan*

Document: [L2/15-232](#) Towards an encoding for Kulitan – Pandey

Discussion: We reviewed this introductory document on the Kulitan script, which has both vertical and horizontal orientations, although horizontal orientation is not currently used.

We have the following comments:

- Provide examples showing the text in horizontal orientation
- Take a section of text, and demonstrate how it can be unambiguously syllabified. This could involve either adding marks showing syllabification or a different encoding solution that encodes codas. This will enable support for different orientations of text. Compare Hangul jamo, in which each “chunk” becomes a syllable of its own.

Recommendations: We recommend the UTC members review this document and send feedback to the author, including the comments above.

AFRICA

6. *Medefaidrin*

Document: [L2/15-298](#) Proposal for encoding the Medefaidrin (Oberi Okaimé) script in the SMP of the UCS (Revised) – Rovenchak

Discussion:

The script ad hoc reviewed a revised version of the proposal. The following points were made:

- The revised proposal answers the question about the different forms of zero that occur in the examples (cf. the representation of 0 in 40, 60 and 80 in figure 3)
- The proposed names don't reflect the names in figure 1. Do the letters have names (for example is M called EM)?
- Is there a capital version of AIVA 'or'? If text were in all-caps, how would it appear? Is the symbol for AIVA a shorthand for the word, like “&”?

- In figure 1, what do the “-“ marks in the chart indicate (fourth row on the right-hand side)? The uppercase for SE is missing in figure 1, but figure 4 (1961) has complete set of upper and lower case. Explain the differences between figure 4 and figure 1.
- In the code chart, have the digits occur consecutively (i.e., the order should be: NINE 16E89, TEN 16E8A, ELEVEN 16E8B, etc.)
- The use of alternate forms for 1, 2 and 3 needs further explanation in order to understand if the variation is consistent. Are there rules to explain the usage of the alternate forms, or is their usage is totally arbitrary?
- Explain the differences between figures 3 and figure 5. For example, in Figure 3, numbers 1, 2, 3 have an extra serif.

Recommendations: We recommend the UTC review this proposal, and address the question whether Nd is the appropriate General Category for numbers in base 20. We further recommend members send comments to the the author.

7. Mandombe

Document: [L2/15-297](#) Preliminary proposal for encoding the Mandombe script in the SMP of the UCS (Revised) – Rovenchak

Discussion: We reviewed this revised proposal, in which the script is now encoded clearly a syllabary. The following comments and questions were raised during discussion:

- Show how bare consonants would be represented (if they are)
- Spell out the character names
- Put the characters in sort order. (The current grouping is in 16s, although this is probably not inherent in the script.)
- Remove PILUKA, unless a strong case can be made
- Remove ELLIPSIS, as it can be represented with three MANDOMBE DOT characters
- Provide examples of combining marks in running text
- Draw circles in examples to show identified characters in figures
- Provide more information on calendar symbols. What is the model for the calendar and dates? Provide examples showing how dates are represented.
- Explain the dot above “Janvier” in figure 9
- Number the pages

Recommendations: We recommend the UTC members review this preliminary proposal and send feedback to the author, including the comments above.

EAST ASIA

8. Hentaigana

Document: [L2/15-239](#) Proposal of Japanese HENTAIGANA – Japan NB

Discussion: We reviewed this proposal, which is a welcome contribution. We have the following comments:

- Provide the justification for how characters are distinguished.
- Provide a mapping between the different lists used to create the proposed repertoire of characters. (As noted on page 4, the official list had 168 HENTAIGANA, whereas an academic version had 275. Because the proposed list was 299, it appears that particular forms have been over-distinguished from others coming from the same source. Cf. 238 and 239, which are

calligraphic variants of the same character, and 244, 245 and 246, also variants of the same character. A comparison chart showing a mapping between the different lists would be helpful.)

- #53 HENTAIGANA LETTER KE VARIANT3 and #28 HENTAIGANA LETTER KA VARIANT6 are identical, but have different sounds. These two should be unified. [NOTE: In L2/15-300, the letter is #34 HENTAIGANA LETTER KA-KE]
- Identify whether any of the Hentaigana use the sound mark and if they do, which ones.

Note: *An updated revised list of HENTAIGANA characters appears in L2/15-300.*

Recommendation: We recommend the UTC review L2/15-300, taking the comments above into consideration.

SYMBOLS

Currency Symbols

9. Bitcoin

Document: [L2/15-229](#) Proposal for addition of bitcoin sign - Ken Shirriff

Discussion: We reviewed this proposal, which provided good examples and addressed the main concerns raised by an earlier proposal (L2/11-130). The proposed codepoint is acceptable.

Recommendation: We recommend the UTC review the proposal and accept 20BF BITCOIN SIGN for encoding.

10. Japanese TV symbols

Document: [L2/15-238](#) Proposal to include additional Japanese TV symbols – Japan NB

Discussion: This set of characters are part of a revision of the ARIB standard to accommodate Ultra High Definition moving pictures. It proposes 19 non-decomposable atomic symbols. All proposed characters can go in the Enclosed Alphanumeric Supplement block, except for SQUARED CJK UNIFIED IDEOGRAPH U+914D, which would go in the Enclosed Ideographic Supplement block.

Recommendations: We recommend the UTC review the proposal and after discussion, approve the 19 characters.

NUMBER SYSTEMS

11. Siyaq

a) Indic Siyaq

Document [L2/15-121R2](#) Proposal to Encode Indic Siyaq Numbers – Pandey

Discussion: We reviewed this proposal, which has been reviewed by various experts in India and elsewhere. The revised proposal incorporates revisions recommended by Roozbeh Pournader at a meeting in August.

The revised proposal includes a new Background section and new alternate forms for 1, 2, 10,000 and LAKH MARK. The alternate forms of primary numbers, used to represent primary units in compounds,

are now called “INDIC SIYAQ NUMBER PREFIXED ONE”, etc., instead of “INDIC SIYAQ ALTERNATE NUMBER ONE”. The INDIC SIYAQ NUMBER ONE HUNDRED THOUSAND has been removed. More details are provided on the alternate method of writing lakhs and crores in the Deccani style (Section 5.8), and on fractions (section 5.8) and currency (section 5.9)

Comments from the discussion:

- Add the ISO forms to the proposal.
- Identify the font source and get appropriate permissions (if needed).

Recommendation: We recommend the UTC members review this proposal and send feedback to the author, including the comments above.

b) Ottoman Siyaq

Document: [L2/15-072R](#) Proposal to Encode Ottoman Siyaq Numbers – Pandey

Discussion: We quickly reviewed the proposal for Ottoman Siyaq Numbers, which includes several modifications from the earlier proposal. The revised proposal incorporates changes recommended by Roozbeh Pournader at a meeting in August, and the proposal has had review by experts.

The revised proposal includes now figures from a key reference source (Fekete), glyph changes (for 6, 8, 20, and 10,000), a new alternate form for 10,000, and further information on glyphic variants and alternate forms. Fractions are called out as characters still under investigation. An extra column has been added (1ED40..1ED4F) to accommodate future characters (such as fractions).

Comments from the discussion:

- Add the ISO forms to the proposal.
- Include a Background section to the proposal giving approximate dates of usage and other information.

Recommendation: We recommend the UTC members review this proposal and send feedback to the author, including the comments above.

SHORTHAND SYSTEMS

12. Pitman Shorthand

Document: [L2/15-116](#) Encoding Pitman Shorthand Scripts – Ramachandran

Discussion: We reviewed this proposal. Our assessment is that currently the proposal reflects a fairly complete catalog of the glyphs used by Pitman. However, it is unclear in many cases which of these glyphs are simply contextual forms of underlying letters. A more complete assessment of the contextual forming of shapes is required in order to have a reliable listing of encodable characters for the writing system. For example, 84 PITMAN INITIAL LIGHT W HOOK FOR LETTER KAY and 85 PITMAN INITIAL HEAVY W HOOK FOR LETTER GAY appear to be contextual variants of the same character (with the weight of the hook affected by the neighboring letter).

The following specific comments were made:

- 158 and 159 have typos: correct INITITAL to INITIAL
- 156 should be: PITMAN SYMBOL LEFT PARENTHESSES

157 should be: PITMAN SYMBOL RIGHT PARENTHESSES

- The codepoint allocation issue mentioned on page 18 should be postponed until the repertoire is worked out further.

Recommendations: We recommend the UTC members review this proposal and send feedback to the author, including the comments above.