TO:	UTC
FROM:	Deborah Anderson, Ken Whistler, Roozbeh Pournader, Lisa Moore, Liang Hai, Richard Cook, Peter Constable, Manish Goregaokar, Ben Yang, and Marek Jeziorek
SUBJECT:	Recommendations to UTC #157 September 2018 on Script Proposals
DATE:	14 September 2018

The Script Ad Hoc group met on 27 August 2018 in order to review proposals. The following represents feedback on proposals that were posted in the Unicode document registry at the time the group met. A table of contents is provided below.

EUROPE
1. Latin
Latin Letter Thorn with Diagonal Stroke
Documents: L2/18-286 Feedback on N4936 (L2/17=236) Proposal to add LATIN LETTER THORN WITH DIAGONAL STROKE – Stokes et al.

Comments: We reviewed this document, which provided experts' feedback on the proposed upper- and lowercase forms of LATIN LETTER THORN WITH DIAGONAL STROKE, which are listed for ballot in the CD. This document adds feedback from Odd Einar Haugen, an expert whose opinion was not included in L2/18-242 (which was seen at the last UTC). Prof. Haugen, who started the MUFI project, provides examples of the thorn with diagonal stroke and horizontal stroke in Medieval Nordic texts, and states that there is no semantic distinction between the thorn with horizontal vs. thorn with diagonal stroke, to the best of his knowledge.

We reiterate the earlier recommendations (from the July Script Ad Hoc recommendations, L2/18-241):

The evidence provided in the document with expert feedback… strongly suggests that the user community considers these two shapes as glyph differences and is using the existing character for both shapes. Encoding a new character for one of the forms would be very disruptive to the user community, as a partial uptake of a new character could throw much of their searching and display processes into disarray.

Recommendations: We recommend the UTC review this document and consider making a ballot comment, citing the feedback contained in this document.

2. Palaeohispanic
Document: L2/18-283 Proposal to encode the Palaeohispanic script – Ferrer et al.

Comments: We reviewed this revised proposal, which made changes to the January 2017 version of the proposal (L2/18-030).

The points listed below arose during discussion. (The cited code points refer to the document under review; any major shift in the repertoire in future revisions will likely result in reassignment of code points.)

- In the list of properties on page 8, the code point U+10239 appears twice; the second one should be U+1023A.
- We recommend the authors include variants in the main chart of unification at the end of the document, so all letter variants -- instead of just a representative form -- in every writing system are listed. This may help justify certain unification cases. (Because the unification chart at the end summarizes the information on the different writing systems, figures 1-10 are not necessary.)
- Based on the charts on pp. 27-29:
  - The glyph for EC- ke/ge in U+10212 (p. 27) should be moved up to U+10211
  - Including the pronunciation under the glyph was a useful improvement.
  - We suggest a review be done that compares the earliest chart from the Palaeohispanic proposal against the chart here, to be sure letters are not being omitted.
  - Some unifications are still unclear, specifically where the shapes are dissimilar or where there is no apparent semantic basis for the unification (i.e., the Southern script glyphs for U+1020B, U+10228, U+10229).
  - There appears to be a three-way distinction that is missed by placing U+10237, U+10238, U+10239 and U+1023A at the end of the repertoire. U+10237 should be moved to a position after (or before) U+1020F/U+10210; U+10238 should be after (or before) U+10211/ U+10212; U+10239 should appear after (or before) U+1021D/ U+1021E; and U+1023A should be after (or before) U+1021F/ U+10220.
  - The unifications that are questionable are largely in the Southern scripts, especially the Southwestern script. Based on this observation, we recommend a chart with just the Northern script be put forward, as well as a separate one for Southern scripts. In this way, it will be possible to view the Northern scripts more clearly and see if the Southern scripts should be handled as additional characters, or as a separate script.

- We recommend the authors start using Unibook (http://unicode.org/unibook/) for generating the chart and names list. This will make review easier and will aid in the eventual publication of the chart and names list.

**Recommendations**: We recommend the UTC review this proposal at their leisure and send the authors comments.

AFRICA
3. Adlam
Document: L2/18-282 Proposal to encode ADLAM LETTER APOSTROPHE for ADLaM script – Barry et al.

Comments: We reviewed this document, which requested one alphabetic element for Adlam.

The following summarizes the comments made during discussion:

- Change the name of the character from ADLAM MODIFIER LETTER APOSTROPHE to ADLAM NASALIZATION MARK.
- Discuss what happens if the nasalization mark appears at the beginning of word (which would not normally occur) or the end of a word (which would occur during typing). Should the mark cause the neighboring character to join or not?
- Show the desired display when the nasalization mark when it occurs at the beginning or end of words.
- Provide an explanation for the examples shown on page 1 (which shows the nasalization mark hanging over tatweel, occupying its own portion of horizontal space) versus figure 1 (below, where the mark doesn't occupy extra space and can touch the preceding letter). Which is the desired typography?

 page 1   vs.   Figure 1:

- Modify the transcription/transliteration (i.e., hindu vs. hiⁿdu), or provide an extra column of phonetic transcription (e.g., in IPA, /hin.du vs. hi.ⁿdu/) in addition to the existing letter-to-letter transliteration and adjust the typography, if needed (as noted above) in the following:

 hindu   'ancient'

 hin'du   'it'

- Mention the implications of hindu vs hiⁿdu on syllabification or syllable boundaries.
- Provide transliterations or phonetic transcriptions for all highlighted examples in figures, especially for those cases other than /ⁿd/.

Based on the above information, the properties (such as general category Mc instead of Lm) can be determined.

Recommendations: We recommend the UTC review this proposal and send the authors comments, including those above.

MIDDLE EAST
4. Book Pahlavi
Document: L2/18-276 Preliminary proposal to encode Book Pahlavi in Unicode – Pandey

Comments: We reviewed this very detailed preliminary proposal for Book Pahlavi. Earlier proposals for Book Pahlavi were written by Michael Everson et al. (L2/07-234), Roozbeh Pournader (L2/13-141) and most recently Abe Meyers (L2/14-077). It was first noted in UTR #3.

Dr. Pandey reported that the repertoire in this proposal represents the characters and behavior of the *typical* usage of letters; more detailed explanation and examples will be included in the next revision of the proposal.

The following comments were raised during discussion:

- Unify the BOOK PAHLAVI END OF WORD MARK with BOOK PAHLAVI LETTER WAW-NUN-AYIN-RESH. According to Roozbeh Pournader, scholars can't consistently distinguish this mark from letters.
- Add in a names list after the code chart.
- For punctuation, review the Book Pahlavi original proposal by Everson (L2/07-234), and make sure the proposed Book Pahlavi punctuation is consistent with decisions made about Avestan punctuation. (Note that more information on punctuation is included on page 6 of N3193 = L2/07-004, a proposal for Medievalist and Iranianist punctuation.)
- The two fixed form letters, which are proposed to cover cases where the normal joining behavior isn't operating, will involve further review by the script ad hoc.
- Include a note that the joining model around numbers is complex.
- Discuss the alternate forms of ligatures (i.e., Ahreman ligatures) which were cited in the proposal by Abe Meyers (L2/14-077).
- After removing END OF WORD MARK, cross-check the chart, code points, and list of character properties. (For example, the last character in the chart is U+10BD9, but the list of character properties on page 37 ends in 10BD8.)
- In section 6, explain the color scheme and double-check the use of the colors in the examples. Also describe how the examples illustrate the so-called ligatures in actual words.
- In collation, are the different kaphs and lamedhs considered different letters or should they be considered equal in collation?

Recommendations: We recommend the UTC review this proposal and send the author comments. Unicode reviewers are especially encouraged to review section 6, in particular the two fixed forms in 6.2 and 6.4.2.

5. Yezidi
Document: L2/18-284 Preliminary proposal for encoding the Yezidi script – Karaca et al.

Comments: We reviewed this revised document for Yezidi, which replaces L2/18-238.
The following comments were made:

- In section 2, show ligated vs. non-ligated forms of ligatures.

- In section 3, the list of newly created letters includes PHE. However, the 1911 chart of letters in figure 8 includes PHE (but not PE). It appears section 3 should have PE instead of PHE.
- We recommend the authors use Unibook (http://unicode.org/unibook/) for generating the chart and names list. This will make review easier and will aid in the eventual publication of the chart and names list, and with appropriate configuration will also create the dotted circles.
- The chart and names list need to match: the chart on page 6 is missing a HAMZA and still contains ALEF-MADDA (which was decomposed in this revision). Also, dotted circles are missing in the chart glyphs for COMBINING HAMZA MARK and COMBINING MADDA MARK.
- Provide the character properties.
- Show the use of hyphenation in both modern and historic texts, with as many examples as possible. In modern texts, is it regularly placed above the last letter (which suggests a combining mark) or does it just closely follow the last letter? In old texts, does it have any combining mark behavior?
- For collation, do users want to sort the historical letters at the end? For example, if users were to look for the historical letters, would they be interleaved with other letters, or would users expect them at the end?
- Including information on how hamza would sort. (Hamza is often near aleph, at the beginning of the alphabet.)
- For digits, provide examples of the digits in in use, including their use within a line of text.

Recommendations: We recommend the UTC review this proposal at their leisure and send the authors comments.

SOUTH AND CENTRAL ASIA
6. Syloti Nagri

Document: L2/18-259 Syloti Nagri feedback – Srinidhi
Background documents:
L2/17-418 Encoding model to represent conjuncts in Syloti Nagri – Srinidhi and Sridatta
L2/05-130 Encoding Model for Syloti Nagri Conjoining Behaviour - Constable
L2/03-146 Alternate Encoding Models for Syloti Nagri - Constable

Comments: We reviewed the feedback in L2/18-259, which was a response to the Script Ad Hoc recommendations in L2/18-241.

The 2017 document from Srinidhi and Sridatta (in L2/17-418) made two requests of the UTC:

(1) discuss the plain-text representation of cross-cluster ligatures and false conjuncts and make changes if necessary to §15.1 Syloti Nagri in the Core Spec
(2) consider a change in the Indic Syllabic category for U+A806 SYLOTI NAGRI SIGN HASANTA, from Pure_Killer to Virama.

In the feedback document L2/18-259, Srinidhi mentions that cross cluster ligatures and false conjuncts are "highly rare and occur in some handwritten sources....[but] are not attested [in] modern printed sources." To verify this statement, we recommend those working on Syloti Nagri at SIL be engaged in the discussion, since some materials are printed in locations such as the UK and Bangladesh.

In the feedback document, Srinidhi also states that *hasanta* is used both as a visible killer and a consonant stacker, whence his request to change the Indic Syllabic category. To verify current usage, it is important to see how SIL implementations are currently using *hasanta*.[1] Until more information is received, we feel it is premature to change the property of *hasanta* to have the Indic Syllabic category of Virama. Depending upon the answer, the text in the Core Spec may need to be changed.

Finally, in his feedback, Srinidhi answered the two questions posed by the Script Ad Hoc, "Are all the cited conjoining forms orthographically significant, and are conjuncts more common than ligatures?" He replied that true consonant conjuncts are significant and common, and new fonts (such as Surma) use *hasanta* for conjunct formation. Again, information from existing implementations is needed before any change to the Core Spec should be considered.

Recommendations: We recommend the UTC assign an AI to Debbie Anderson to contact Lorna Evans of SIL to determine the usage of Syloti Nagri cross-cluster ligatures and false conjuncts in publications and to provide information on how SIL implementations are using Syloti Nagri *hasanta*.

7. Takri
Document: L2/18-247 Additional information on L2/18-084 TAKRI VOWEL SIGN VOCALIC R – Anderson
Background document: L2/18-084 Proposal to encode the TAKRI VOWEL SIGN VOCALIC R

Comments: We reviewed this document, which provided additional information on TAKRI VOWEL SIGN VOCALIC R by the proposal authors. The proposed character is based on the Mandeali form of Takri.

As noted in the original Takri proposal (L2/09-424), Takri "is the traditional writing system for Chambeali, Dogri, and several 'Pahari' languages, such as Jaunsari, Kulvi, and Mandeali." Takri as originally proposed also covered the Sirmauri variety of the script, but Sirmauri has now been separately proposed (L2/18-085). However, the authors Srinidhi and Sridatta don't believe Mandeali should be separately encoded. Instead, they suggest TAKRI VOWEL SIGN VOCALIC R should either be encoded as a Takri character, or unified with the Sharada vowel sign vocalic r (U+111B8 SHARADA VOWEL SIGN VOCALIC R), depending on whether Mandeali should be unified with Takri or Sharada, because the proposed character is only attested in Mandeali text thus far. It was noted that Takri was descended from Sharada.

We recommend asking the authors whether the specimens of the "Mandeali form of Takri" in figure 37 in L2/09-424 should be represented by Takri or Sharada? (They should also compare figure 36, a chart showing the Mandeali form of Takri.) A change to the text of the Core Spec may need to be made, depending upon the response from Srinidhi.

Recommendations: We recommend the UTC review this document and assign an AI to Debbie Anderson to ask Srinidhi and Sridatta how the text in figures 36 and 37 of L2/09-424 should be represented (i.e., by Takri or Sharada characters), and also ask Anshuman Pandey for his opinion.

---

[1] As summarized in the earlier Script Ad Hoc recommendations (L2/18-241), the UTC had agreed with Peter Constable in L2/05-130 that the model for the script is one with a virama, with no conjoining behavior (like Burmese). For special forms, such as rare cross-cluster ligatures or false conjuncts, L2/05-130 recommended OpenType features or ZWJ. The only way to be able display such special conjoining behavior today would be with joiners.

8. Bengali
Documents: L2/18-287 Further remarks on the encoding model of Vedic *gomukha* characters – Sharma
Background documents:
L2/18-035 Encoding model issues with the Vedic *gomukha* characters – Sharma
L2/17-098 Indic editorial updates – Srinidhi et al.

Comments: We reviewed the documents from Shriramana Sharma. We also reviewed the Script Ad Hoc report (L2/18-241), which recommended the UTC get "input from those members with Indic rendering engines and the Bengali user community [and] suggest the author work with Srinidhi and Sridatta, authors of L2/17-098, and consider why U+09FC BENGALI LETTER VEDIC ANUSVARA would not be suitable."

Sharma relayed his personal view that using Bengali Vedic anusvara would not be a problem, noting he had even shown use of the character in similar Sama Vedic contexts, but his preference would be to encode a new dot with script=inherited in Vedic Extensions. (This "dot above" would complement dots below U+1CDD..U+1CDF).

Recommendations: Based on feedback from Sharma, we recommend the UTC wait for a proposal before making any change.   We also suggest assigning an AI to Debbie Anderson to ask Srinidhi or Sridhatta to pull together the discussion and make a concrete proposal.

9. Khitan Small Script
a. Khitan Small Script Cluster Model
Document: L2/18-285 Further information on Khitan Small Script clusters – Anderson
Background docs:
L2/18-121R Cluster Formation Model for Khitan Small Script – West et al.
L2/18-213   KSS Ad Hoc Report – Anderson et al.

Comments: We reviewed the document L2/18-285, which contained two Khitan Small Script examples provided by Prof. Wu. The document also contains confirmation from two authors of the Khitan Small Script proposal that the encoding model covers the examples by Prof. Wu. (This document was prepared in response to the Script Ad Hoc recommendation in L2/18-241 [p. 10] that KSS experts in China provide examples of any "unexpected contexts" of KSS characters. Such documentation may be needed in the future Core Spec block introduction.)

Recommendations: Since the examples are covered by the proposed KSS encoding model, we recommend the UTC note this document.

Note: Regarding the Script Ad Hoc recommendation in L2/18-241 [p. 10] about the location of the ITERATION MARK in the main Khitan Small Script block, Andrew West has privately confirmed that "there is no requirement that the character [iteration mark] have any property other than Lo, and the character will require no special treatment in the nameslist".

b. Glyph for U+16FE4 KHITAN SMALL SCRIPT FILLER
Document: L2/18-214 Feedback on Additional repertoire for ISO/IEC 10646:2017 (5th ed.) beyond Amendment 2 (L2/18-211) – Marín Silva

Comments: We briefly reviewed the comment in this document which recommended changing the glyph for U+16FE4 KHITAN SMALL SCRIPT FILLER from KITSF to KSSF. We agree that KSSF is a better abbreviation.

Recommendations: We recommend the UTC add a ballot comment, asking for the glyph for U+16FE4 KHITAN SMALL SCRIPT FILLER from KITSF to KSSF.

SOUTHEAST ASIA
10. Wancho
Document: L2/18-264 **Error in three Wancho character names (Anderson)**

Comments: We reviewed this document, which identified errors in the names for three Wancho tone marks. The three new characters were added at the June 2018 ISO SC2 and WG2 meeting, appeared in DAM2, and were approved at the July 2018 UTC meeting. However, the original proposal author and the script's creator both confirm the names in DAM2 (and approved by the UTC) are incorrect for U+1E2ED, U+1E2EE, and U+1E2EF

Recommendation: We recommend the UTC change the following three character names
from:	to:
1E2ED WANCHO TONE TUP MANG	1E2ED WANCHO TONE TUPNI
1E2EE WANCHO TONE OKOI	1E2EE WANCHO TONE KOI
1E2EF WANCHO TONE OKOI MANG	1E2EF WANCHO TONE KOINI

We also recommend the UTC assign an AI to Debbie Anderson to check the DAM2 ballot and make a ballot comment if the names don't reflect those above.

SYMBOLS AND NUMERICAL NOTATION SYSTEMS
11. Legacy Computer and Teletext characters

Document: L2/18-275 Proposal to add characters from legacy computers and teletext

Comments: We reviewed this revised proposal and accompanying mapping files and spreadsheet. The following summarizes the comments.
Note: Responses to the comments from Doug Ewell were received after the Script Ad Hoc had met and are appended after each comment below.

- The proposed glyphs for U+1FBB5 and U+1FBB6 are different from the source glyphs. Is there a reason for the difference?

  Response from Doug Ewell:
  We have added a new paragraph to page 3:
  Graphic characters on text-oriented legacy platforms were designed for restricted resolution, typically an 8 × 8 cell. Many of these characters are shown with improved resolution in the code charts beginning on page 11. For example, two characters from the Apple MouseText set, LEFTWARDS and RIGHTWARDS ARROW AND UPPER AND LOWER ONE EIGHTH BLOCK, were displayed in the Apple II series with an incomplete upper line (Figure 1), but are shown in the

code charts with a complete (broken) line. The code chart glyphs are illustrative only and do not imply a change in character identity

- For U+IFBC1, the proposed glyph is quite different from the source glyph. Is there a reason for the difference?

  Response from Doug Ewell:
  We have withdrawn the character.

- For U+2088+U+2087, we believe the mappings to be incorrect (i.e., to legacy superscripts and subscripts). Instead, we recommend these two characters be moved to the section for legacy characters with no proposed mappings.

  Response from Doug Ewell:
  These have been removed from the revised version of the document.

- There was a difference of opinion on some of the mappings that involved shades/checkerboards (e.g., U+1FB95/U+1FB96, U+2425, U+1FB8C…U+1FB91).

Recommendation: We recommend the UTC discuss the proposal.