TO: UTC L2/16-037
FROM: Deborah Anderson, Ken Whistler, Rick McGowan, Roozbeh Pournader, Andrew Glass, and Laurentiu Iancu
SUBJECT: Recommendations to UTC #146 January 2016 on Script Proposals
DATE: 22 January 2016

The recommendations below are based on documents available to the members of this group at the time they met, January 19, 2016.

EUROPE
1. *Latin*
Document: L2/15-327 Proposal to add Medievalist punctuation characters – Everson

Discussion: We reviewed this document, which requested 21 characters. Many of the proposed characters require more detailed analysis, specifically providing examples that show contrasts in manuscripts, in old transcriptions, and how the marks are represented in text today.

Specific comments raised in the discussion:

- §1 Introduction. In the list of the proposed characters on pages 1 and 2, include dotted guide-lines, which show the placement of the characters in relation to the baseline, mid-line, and top line, and solid lines separating individual table cells.
- §2.2.3. *Punctus versus*. The text suggests that two glyphs for the same character are being proposed: PUNCTUS VERSUS MARK and LOW PUNCTUS VERSUS MARK.
- §2.4 *Distinctiones*. "Note too that ⁖ is the Georgian paragraph separator; no 'generic' punctuation mark for that has been encoded."
  Is this a request to unify the Latin ⁖ with U+10FB Georgian Paragraph Separator? If so, it can be added to ScriptExtensions.txt.
- §4 Linebreaking. The assignment of SY as the LB property for DOTTED SOLIDUS should be reviewed by the UTC, since the SY class currently has only one member and it would be prudent to be cautious about adding another member to SY.
- Comments on specific characters or figures:
  - The POSITURA MARK may be a candidate for encoding; it has a distinct function and is discussed. Can an example be providing contrasting it with a MIDDLE COMMA?
  - Are PUNCTUS EXCLAMATIVUS and PUNCTUS INTERROGATIVUS ever in contrast with the more usual "!" or "?" characters? They appear to be early forms, which are reflected by odd glyphs in the font, but there is no plain-text distinction demonstrated. Figure 17 looks to be a glyph variant of a "!".
  - Figures 5 and 9 show the TWO DOTS OVER COMMA only in transcribed form, and appear to be a sequence of DOT, COMMA, DOT, instead of a single atomic character. Can the original text be shown and further support provided for a single character?
  - In figures 6 and 8, COLON WITH SIDEWAYS REVERSED RAISED COMMA could be considered a semicolon and a swash or a dash, based on the manuscripts. It is not clear this is an atomic character.
  - For MIDDLE COMMA, the transcription in figure 25 has all dots as middle dots, but the original text shows dots above the baseline or on the baseline. This figure does not show the MIDDLE DOT in contrast to a baseline dot (nor MIDDLE COMMA contrasting with baseline comma).

- MEDIEVAL COMMA is only shown in transcription. Provide an example from an original manuscript text.
- Evidence for VERTICAL FIVE DOTS is restricted to figure 26. Provide additional evidence, particularly showing systematic representation of the five dots in Runic transcriptions where the four vertical dot and five vertical dot marks are distinguished.

Recommendations: We recommend the UTC accept TRIPLE DAGGER and DOTTED SOLIDUS for inclusion in the Supplemental Punctuation block, and discuss the LB property for DOTTED SOLIDUS.

SOUTH AND SOUTHEAST ASIA
Indic
2. Devanagari
Document: L2/15-335 Proposal to encode the Devanagari letter and vowel sign AY - Pandey

Discussion: We reviewed this proposal, which proposed two characters devised in the 19th century to represent sounds in languages in north India for which no Devanagari characters were available.

In 2009, a proposal asking for these characters (and others) by Anshuman Pandey was submitted (L2/09-320). At the time the response on how to represent the AY letter and vowel sign was to use two U+0946 DEVANAGARI VOWEL SIGN SHORT E characters. However, putting two vowel signs and getting them to look acceptable is difficult. Because these two characters complete a set alongside the already encoded DEVANAGARI VOWEL SIGN and VOWEL LETTER AW, it is reasonable to accept them for encoding.

Recommendations: We recommend the UTC accept the two characters U+A8FE DEVANAGARI LETTER AY and U+A8FF DEVANAGARI VOWEL SIGN AY.

We also suggest these two characters and U+094F DEVANAGARI VOWEL SIGN AW and U+0975 DEVANAGARI LETTER AW be annotated, indicating their relationship to Hoernle's conventions.

3. Nandinagari
Document: L2/16-002 Proposal to encode the Nandinagari script – Pandey

Discussion: We reviewed this proposal, which built significantly off an earlier version, L2/13-002. Specific comments raised during the discussion include:

- In §3.7, on page 7, the section on joiners and non-joiners is confusing. Why is the model not conforming to the usual pattern for Devanagari (etc.) where the order is VIRAMA ZWJ? What model is being proposed? Add a table showing how this model compares with that of Devanagari. If a joiner is being proposed, compare it against other Indic scripts, especially Devanagari, and make a case for it.
- §3.8 reads: "The ANUSVARA is used for marking nasalization. It is generally placed to the right of a base letter, but may also be placed above it." Show samples of both placements, since two may be required (cf. Grantha and Malayalam).
- Add a chart showing which characters should be used to represent the vowel letters, similar to Table 12-1 in TUS (i.e., for LETTER AI, use 11B9B, not 11B9A LETTER E, 11BCA VOWEL SIGN E, etc.)
- Should the headstroke be unified with Devanagari headstroke (U+A8FB), since it doesn't interact with anything? Or should it be separately encoded, as was done for Sharada (U+111DC)?

- Minor editorial comments:
    - p. 6: replace ZWJ with the glyph for ZWJ in the following: “This model also uses the generic control characters ZWJ U+200D ZERO WIDTH JOINER”
    - p. 8: Replace RA with YA in the following
      “The default is to use the post-base form of the second letter: yya …<RA, VIRAMA, YA>
      The alternative is to use a conjoined ligature: yya <RA, ZWJ, VIRAMA, YA>”

Recommendations: We recommend the UTC discuss this proposal, and specifically address whether a separate headstroke character should be encoded. We encourage UTC members to send the author feedback.

4. Hanifi Rohingya

Document: L2/15-278 Proposal to encode the Hanifi Rohingya script in Unicode (revised) - Pandey

Discussion: We reviewed this revised proposal. While it is getting close to being a mature proposal, a few issues remain.

The finals need to be explained more fully. Could they considered optional or required ligatures? Or are they contrastively used, and hence would be appropriate for separate encoding? Where did the information on them come from? Are there examples in the figures showing them? (If so, identify them in the images and in the captions.) Provide examples in running text.

Other comments:

- Is SAKIN dual-joining or right-joining? Does SAKIN occur in the middle of a word? If so, provide examples.
- Add FINAL DA/MA/LA to Table 1
- Names list (p. 14): correct the names for FINAL LETTER DA (LA, MA) to LETTER FINAL DA (LA, MA)
- §4.1 (p. 9): correct the character name from SUKUN to SAKIN
- §4.4 (p. 11): replace TONE-2 and TONE-3 with the appropriate character names
- §4.2 Arabic shaping: the last field should have the script name (otherwise it implies the script is Arabic or Syriac)

Recommendations: We recommend the UTC review this document and send comments to the author. We encourage those with expertise in joining scripts to review the Hanifi Rohingya shaping data proposed for ArabicShaping.txt.

5. Pau Cin Hau

Document: L2/16-014 Revised code chart for the Pau Cin Hau Syllabary – Pandey

Discussion: This document is a code chart for the Pau Cin Hau Syllabary, with code points modified from the earlier chart in L2/13-067. The current location in the document (3 rows, U+14100..U+1452F) bumps up against Anatolian Hieroglyphs (U+14400-).

Recommendations: We recommend the proposal author update the codepoints to match the updated Roadmap (14700-14BFF).

6. Lepcha
Document: L2/15-332 "Lepcha Unicode – Published Standard" – Daehler, SIBLAC

Discussion: We reviewed this document, which followed up on an earlier Error Report, in which the author stated that the syllabic structure in Table 13-4 of TUS was incorrect and the description on page 542 about RAN was also incorrect.

In §2, the author asks for a change to the block intro text for Lepcha. The problem text is:

> "The combining mark U+1C36 LEPCHA SIGN RAN occurs only after the inherent vowel *-a* or the dependent vowels *-aa* and *-i*. When it occurs together with a final consonant sign, the *ran* sign renders above the sign for that final consonant."

The author reports RAN with *–aa* is an error, and should be removed.

§3 is a comment on Lepcha Syllabic Structure, in which the author states that phonology, typing order, and backing store should be the same.

Recommendations: We recommend the first topic (§2) be remanded to the Editorial Committee.
We recommend the UTC respond to the author regarding §3, pointing to relevant text in the standard or other references that might be more understandable.

7. Ranjana and Lantsa
Document: L2/16-015 Towards an encoding for the Ranjana and Lantsa scripts - Pandey

Discussion: We reviewed this document, which recommends that any proposal for Ranjana must also take into account Lantsa, the name for the script outside of Nepal and India. As noted in the document, Lantsa has a way of representing semi-vowels in conjuncts that is not found in Ranjana, as well as script-specific invocation, punctuation, and other marks that will need to be accommodated in a proposal. The document also raises questions as to which glyph shapes should be chosen as the representative glyphs, as well as the relationship between Ranjana and Wartu, a similar script but one with a curved headstroke.

This document helps to move forward the discussion on the relationship between Ranjana, Lantsa, and Wartu. We look forward to further documents on Ranjana (/Lantsa), which explain the fixed-form consonants, and provide additional details on Wartu (with examples).

Recommendations: We recommend the UTC members take note of this document, and send any feedback to the author.

8. Soyombo
a. Soyombo JIHVAMULIYA and UPADHMANIYA
Document: L2/15-331 Proposal to encode JIHVAMULIYA and UPADHMANIYA for Soyombo – Pandey

Discussion: We reviewed this proposal, which requests two characters for Soyombo. The two characters were mentioned during the ad hoc meeting on Soyombo and Zanabazar Square in Tokyo (see L2/15-249, slide 23, and the ad hoc report L2/15-249). The analysis provided is similar to that given for Sharada.

The following comments were raised during the discussion:

- The proposal only provides one example, can additional examples be provided?
- Are the characters more like letters or cluster initials (with characters after them)?
- Do the symbols ever appear in isolation, such as in a text that discusses them?

Recommendations: We recommend the UTC review this proposal, and forward the comments above and any from UTC members to the author.

b. Soyombo Pluta
Document: L2/16-016 Proposal to encode the Soyombo mark PLUTA – Pandey

Discussion: We reviewed this proposal to add SOYOMBO MARK PLUTA, a vowel-lengthening mark that corresponds to U+0F85 TIBETAN MARK PALUTA, U+1139D GRANTHA SIGN PLUTA, etc. The *pluta* mark was mentioned during the Tokyo ad hoc meeting on Soyombo and Zanabazar Square as a character that was missing from the repertoire (see L2/15-248, slide 23, and the ad hoc report L2/15-249). The proposed character seems to be well-justified.

Recommendations: We recommend the encoding of Soyombo MARK PLUTA at U+11A9D, and suggest cross-references to other *pluta* characters be added in the names list.

c. Punctuation marks in Soyombo and Zanabazar Square
Document: L2/15-336 Comments on L2/15-249 regarding Soyombo and Zanabazar Square punctuation – Pandey

Discussion: We reviewed this document, which questioned a request in the Soyombo and Zanabazar Square Ad Hoc Meeting report from Tokyo (L2/15-249R) to add text to the script proposals indicating that Soyombo and Zanabazar Square use ꣎ U+ 0F1C TIBETAN SIGN RDEL DKAR GSUM and ꣏ U+ 0F1A TIBETAN SIGN RDEL DKAR GCIG. The ad hoc also recommended the characters be added to a ScriptExtensions.txt section in the revised proposals.

This document calls out the distinction between the use of these symbols as editorial marks in Soyombo and Zanabazar Square but as astrological symbols in Tibetan. We agree with the author that unification of these symbols is not prudent, and the characters should not be added to ScriptExtensions.txt. If additional information is provided later on these symbols as discrete editorial marks, they can be added as additional characters.

Recommendations: We recommend the UTC review this document.

9. Zanabazar Square
a. Zanabazar Square Subjoiner
Document: L2/15-342 Subjoiner for Zanabazar Square - Pandey

Discussion: We reviewed this proposal, which changes the encoding model for Zanabazar Square. As proposed, the SUBJOINER would be used exclusively for forming conjuncts, freeing the *virama* to act solely as a vowel killer. In Zanabazar Square, the *virama* appears exclusively for transliterating Lantsa or Tibetan.

In our opinion, encoding a separate SUBJOINER is a marked improvement in the encoding model, since it will enable representation of Zanabazar Square text without the use of ZWJ and ZWNJ.

Recommendations: We recommend the UTC accept ZANABAZAR SQUARE SUBJOINER.

b. Zanabazar Square Head Marks
Document: L2/15-341 Proposal to encode additional head marks for Zanabazar Square – Pandey

Discussion: We reviewed this proposal for two additional head marks for Zanabazar Square. The two atomic head marks match similar characters in Tibetan (cf. U+0F04 U+0F05). The two proposed head marks can take additional signs, such as *candrabindu* or *anusvara*. The proposal is sound in our view.

Recommendations: We recommend the UTC accept the two characters: ZANABAZAR SQUARE INITIAL DOUBLE-LINED HEAD MARK and ZANABAZAR SQUARE CLOSING DOUBLE-LINED HEAD MARK.

10. Old Sogdian
Document: L2/15-089 Preliminary Proposal to Encode the Old Sogdian Script

Discussion: We reviewed this proposal at an earlier script ad hoc meeting and made the following comments:

- §3 Roadmap for Sogdian scripts. This section is very useful. Sims-Williams had noted that 'Sogdian Uyghur' was not a good name, and preferred 'Late Sogdian cursive', so that name should be noted.
- §4.3 Encoding model
  - Remove this section, since the pattern for Old Sogdian should follow the model of other Aramaic scripts (which does not leave holes for characters that have not yet been found).
  - Remove the holes in the chart. (If the characters are found, they can be added, but they must be unambiguous.)
- §5.1 Letters
  - Show ALTERNATE HE in text
  - Provide more information on final NUN. Sims-Williams had written "Since ZAYIN and NUN are always distinct in final position…", it may be that final NUN should be separately encoded, as in Hebrew, Nabataean, and Palmyrene.
  - Should DALETH and RESH be distinguished? Show contrastive evidence of their use.
  - Consider making ALTERNATE AYIN the AYIN character, and, if DALETH and RESH are unified, consider changing the name of RESH to "DALETH-AYIN-RESH".

§9 Acknowledgements: Correct the minor typo on Nicholas Sims-Williams' name.

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

AFRICA
11. N'Ko
Document L2/15-338 Proposal to encode four N'Ko characters in the BMP – Everson

Discussion: We reviewed this document.

The first character that is proposed, TE-KERENDE, can be represented using U+2010 HYPHEN or U+2011 NON-BREAKING HYPHEN, but would need to be designed in a font on the baseline. Note that U+2010

HYPHEN is used in such a way in Arabic text. The other three characters are well-documented and straight-forward.

Recommendations: We recommend the UTC approve the following three characters:
07FD NKO DANTAYALAN
07FE NKO DOROME SIGN
07FF NKO TAMAN SIGN

12. Medefaidrin
Document: L2/16-020 Proposal for encoding the Medefaidrin (Oberi Okaime) script in the SMP – Rovenchak et al.

Discussion: We reviewed this proposal, which improves upon the earlier version (L2/15-298). Only a few issues remain:

- How are the letters pronounced, when read off a chart, and how would one represent the pronunciation in IPA? Since most scripts spell out the characters' names – except for the Latin script – this information is needed for the character names.
- Is there a correlation between the phonological analysis and the writing system, or is it arbitrary?
- Provide more information on the new character EXCLAMATION OH, explaining its use and showing it in context.

Recommendations: We recommend the UTC review this proposal and send comments to the authors.

13. Mandombe
Document: L2/16-019 Proposal for encoding the Mandombe script in the SMP – Rovenchak et al.

Discussion: We reviewed this document, which shows progress from the last version (L2/15-118) and demonstrates that the encoding is possible and implementable.

Comments raised during the discussion:

- We recommend more details on calendar symbols be provided.
- Names need to be spelled out, but they can be listed in an ancillary text file that is kept in synch with the main table in the proposal.
- The proposal should remove the PILUKAs.
- Although ELLIPSIS is identified as a separate character by the inventor of the script, it is representable by three MANDOMBE DOT characters, and hence should be removed.

Recommendations: We recommend the UTC members review this proposal and send the authors feedback.

14. Egyptian Hieroglyphs
Document L2/16-018 Proposal to encode three control characters for Egyptian Hieroglyphs – Glass and Richmond

Discussion: We reviewed this document, which builds off earlier documents (L2/15-069 and L2/15-123). The proposal requests three characters, which match the Manuel de Codage symbols "&" as ligature joiner, "*" as a horizontal joiner, and ":" as a vertical joiner.

With the encoding of these three characters, quadrats can be created in plain text, and Egyptologists will no longer need to rely on proprietary software for rendering, which has prevented active use of the Unicode-encoded characters.

Recommendations: We recommend the UTC review this document and accept the three characters:
13430 EGYPTIAN HIEROGLYPHIC SIGN LIGATURE JOINER
13431 EGYPTIAN HIEROGLYPHIC SIGN HORIZONTAL JOINER
13432 EGYPTIAN HIEROGLYPHIC SIGN VERTICAL JOINER

We also recommend an explanation of the examples from §5 (Attested quadrat structures) be added to the document, including a discussion of how to handle #25 and #26.

15. Arabic
Document: L2/15-329 Proposal to encode Quranic marks used in Quran published in Libya - Mussa A. A. Abudena

Discussion: We reviewed this proposal, which proposes 37 characters used in a writing style for the Koran. A comparison table on page 3 compares Khrraz (used in Saudi Arabia) versus Aldani.

A number of characters appear to be good candidates for encoding (such as 4-6, probably 8-11), a few are already encoded, and others are ligatures (30-36) and hence are not candidates for encoding. A careful study is needed to analyze the proposed characters, identify which new characters are encodable, and which may be stylistic variants. In a few cases, new pieces may need to be encoded.

Recommendations: We recommend the UTC review this document, and solicit feedback from other experts.

EAST ASIA
16. Hentaigana
Documents: L2/15-343 Revised Proposal of HENTAIGANA - Japan
L2/15-334 Hentaigana proposal by Nicholas Tranter (Note: The comments in L2/15-334 were written in response to L2/15-239, which was the earlier proposal from Japan.)

Discussion: We reviewed the revised proposal (L2/15-343), which has taken into account the comments provided by Nicholas Tranter (L2/15-334). Unlike an earlier version of the Hentaigana proposal (L2/15-239), this version no longer separates characters that are identical in shape but are used to represent different sounds; this seems to be an appropriate approach for encoding. The naming pattern adopted here, with the names HENTAIGANA LETTER A-1, HENTAIGANA LETTER A-2, etc., follows current syntax.

One outstanding issue involves those characters which have the same source Kanji and are not differentiated by the government source, but are differentiated in the academic source. Such an approach might suggest that two glyphs of the same character are being proposed for encoding because they vary calligraphically. Providing clear examples of the characters in contrast will help to evaluate such cases.

Examples include:

revised#	Old#	UnifiedOId#1	UnifiedOId#2	Proposed Glyph Reference	RevisedCharacterName	Phonetic Value
221	231				HENTAIGANA LETTER YA-1 ・Derived From 4E5F 也	や (U+3084)
222	232				HENTAIGANA LETTER YA-2 ・Derived From 4E5F 也	や (U+3084)
258	270				HENTAIGANA LETTER RO-1 ・Derived From 5442 呂	ろ (U+308D)
259	271				HENTAIGANA LETTER RO-2 ・Derived From 5442 呂	ろ (U+308D)

Recommendations: We recommend the UTC review this proposal, and send the proposal author comments. The UTC should also discuss HENTAIGANA LETTER E-1, which is encoded at U+1B001, and decide how to proceed.

NUMBER SYSTEMS

17. Tally marks

Document: L2/15-328 Proposal to encode tally marks – Lunde and Daisuke MIURA

Discussion: We reviewed this document, which proposed characters used for tallying.

For the TALLY DIGIT ONE through TALLY DIGIT FOUR, use of U+007C VERTICAL LINE would be the most likely method to represent tally marks one through four. However, TALLY DIGIT FIVE is required.

The ideographic characters are commonly used in East Asia. To represent IDEOGRAPHIC TALLY DIGIT TWO, we recommend use of U+1D36E COUNTING ROD TENS DIGIT SIX. The remaining four characters are good candidates for encoding.

The dot and dash tally digits seems to vary (cf. pp. 11-13) and the box tally digits are rather rare. Although the dot and dash and box tally digits are elements in tallying systems, we don't yet feel a clear case has been made that they are characters.

Recommendations: We recommend the UTC discuss this proposal, and accept the following five characters:
TALLY DIGIT FIVE in the Miscellaneous Symbols and Arrows block (U+2BD3 is the next spot available under a set of "Miscellaneous symbols", and before two-headed arrows)
IDEOGRAPHIC TALLY DIGIT ONE
IDEOGRAPHIC TALLY DIGIT THREE
IDEOGRAPHIC TALLY DIGIT FOUR
and
IDEOGRAPHIC TALLY DIGIT FIVE in the Counting Rod Numerals block. (The first open slot is at U+1D372.)

The following documents were postponed until the next script ad hoc meeting.
Siyaq
a. Diwani
Document L2/15-066 Proposal to Encode Diwani Siyaq Numbers (revised) - Pandey

Ottoman
Document L2/15-072 Proposal to Encode Ottoman Siyaq Numbers (revision 2) - Pandey

Unification of Diwani and Ottoman
Document: L2/15-340 Unification of 'Diwani' and 'Ottoman' Siyaq Numbers – Pandey

Arabic Siyaq
Document L2/16-017 Proposal to Encode Arabic Siyaq Numbers in Unicode - Pandey

b. Persian Siyaq
Document L2/15-122 Proposal to Encode Persian Siyaq Numbers in Unicode - Pandey

The following were not discussed:
L2/15-241 Proposal to encode Latin small capital letter Q - Severin Barmeier
L2/15-243 Representation of Fractional Signs in Kannada script - Srinidhi A, et al
L2/15-255 Request to change the representative glyph of Sharada Vowel Signs Vocalic L and Vocalic LL - Srinidhi A, et al