L2/20-250

TO: **UTC**
FROM: **Deborah Anderson, Ken Whistler, Roozbeh Pournader, Lisa Moore, Peter Constable, and Liang Hai**[1]
SUBJECT: **Recommendations to UTC #165 October 2020 on Script Proposals**
DATE: **October 1, 2020**

The Script Ad Hoc group met on August 28, September 11 and 25, 2020 in order to review proposals. The following represents feedback on proposals that were available when the group met. A table of contents is provided below.

[1] Also participating were Ben Yang, Craig Cornelius, Ned Holbrook, Jan Kučera, Andrew Glass, Marek Jeziorek, Norbert Lindenberg, Lawrence Wolf-Sonkin, Steven Loomis, Manish Goregaokar, and Lorna Evans.

I. EUROPE

1. Latin

Document: L2/20-266 Consolidated code chart including proposed phonetic and medieval characters – Everson

Comments: We reviewed an earlier version of this code chart. Because of the large number of revisions to the code points of characters approved at the July 2020 UTC meeting, the Script Ad Hoc group has since agreed that rescinding the previously approved Latin characters listed in the Recommendation below is warranted. Having a new, comprehensive proposal in hand will make approval of all of the characters cleaner -- both those reviewed earlier and the newer additions, and will be less likely to introduce errors. It will also result in less confusion for the Pipeline bookkeeping of approvals for future releases.

Collisions with earlier approved characters affect outstanding proposals for IPA modifier letters, modifier capital letters, Middle Scots S, Orumulum letters, and Sigmoid S. As a result, the Script Ad Hoc recommendations for those proposals are not included in this document.

Recommendation: We recommend the UTC make the following approval:

SAH-UTC165-R1: The UTC rescinds the following characters:

A7CB LATIN SMALL LETTER FENG DIGRAPH WITH TRILL
A7CC LATIN SMALL LETTER LEZH WITH RETROFLEX HOOK
A7CD LATIN SMALL LETTER TURNED Y WITH BELT
A7CE LATIN LETTER SMALL CAPITAL L WITH BELT
A7CF MODIFIER LETTER CAPITAL AA
A7D0 MODIFIER LETTER SMALL L WITH BELT
A7D1 MODIFIER LETTER SMALL LEZH
A7D2 MODIFIER LETTER SMALL L WITH RETROFLEX HOOK AND BELT
A7D3 MODIFIER LETTER SMALL LEZH WITH RETROFLEX HOOK
A7D4 MODIFIER LETTER SMALL TURNED Y WITH BELT
A7D5 MODIFIER LETTER SMALL CAPITAL L WITH BELT
A7D6 MODIFIER LETTER SMALL LS DIGRAPH
A7D7 MODIFIER LETTER SMALL LZ DIGRAPH
A7D8 MODIFIER LETTER SMALL FENG DIGRAPH
A7F0 LATIN SMALL LETTER ESH WITH DOUBLE BAR
A7F1 LATIN LETTER RETROFLEX CLICK WITH RETROFLEX HOOK
A7F2 LATIN SMALL LETTER T WITH TOP HOOK AND RETROFLEX HOOK
A7F3 LATIN SMALL LETTER TURNED R WITH LONG LEG AND RETROFLEX HOOK
A7F4 LATIN SMALL LETTER L WITH FISHHOOK
AB6C LATIN LETTER TURNED SMALL CAPITAL G
AB6D LATIN SMALL LETTER REVERSED K
AB6E LATIN SMALL LETTER REVERSED G
AB6F LATIN SMALL LETTER REVERSED ENG
10780 LATIN SMALL LETTER TURNED T WITH CURL
10781 LATIN LETTER INVERTED GLOTTAL STOP WITH CURL
10782 LATIN SMALL LETTER ESH WITH DOUBLE BAR AND CURL
10783 LATIN LETTER STRETCHED C WITH CURL
10784 LATIN LETTER SMALL CAPITAL TURNED K

10790 LATIN SMALL LETTER O WITH RETROFLEX HOOK
10791 LATIN SMALL LETTER I WITH STROKE AND RETROFLEX HOOK
10793 LATIN SMALL LETTER TESH WITH RETROFLEX HOOK
10794 LATIN SMALL LETTER L WITH BELT AND PALATAL HOOK
10795 LATIN SMALL LETTER ENG WITH PALATAL HOOK
10796 LATIN SMALL LETTER TURNED R WITH PALATAL HOOK
10797 LATIN SMALL LETTER R WITH FISHHOOK AND PALATAL HOOK
10798 LATIN SMALL LETTER EZH WITH PALATAL HOOK
10799 LATIN SMALL LETTER DEZH WITH PALATAL HOOK
1079A LATIN SMALL LETTER TESH WITH PALATAL HOOK
(Reference: Section 1 of L2/20-250 Script Ad Hoc Recommendations)

---

2. Todhri
Document: L2/20-188R Proposal for encoding the Todhri script in the SMP of the UCS -- Everson

Comments: We reviewed this revised proposal for the Todhri script, which has taken into account comments from the Script Ad Hoc. The author has added more text on how accents should be represented, with a list of what letter-accent combinations fonts should support on the bottom of page 3. It was noted that because a character was removed from an earlier version of the proposal, the number of characters proposed in 52, not 53 as on the proposal summary form.)

In our view, the repertoire is acceptable. The only outstanding question for the UTC involves whether the characters Todhri E <i> and U <o> should be decomposed or not. The options are listed in section 2.2 on page 4. (The opinion of the SAH was that the two characters should be atomically encoded without canonical decompositions.)

Recommendation: We recommend that the UTC approve the following, after discussing the decomposition of U+105C9 TODRI LETTER E and U+105EA TODRI LETTER U:
SAH-UTC165-R2: The UTC accepts 52 Todhri characters in a new Todhri block (U+105C0..U+105FF) for encoding in a future version of the standard. (Reference: L2/20-188R)

---

3. Vithkuqi
Document: L2/20-187R Proposal for encoding the Vithkuqi script in the SMP of the UCS -- Everson

Note: Below are comments on an earlier version of this document that the Script Ad Hoc reviewed.

Comments: We reviewed this proposal, which has been seen by the Script Ad Hoc group several times. The main issue centered around the two distinct B characters, since earlier images were not clear. The author was able to find and include high-quality scans, which also had Roman transliteration, thus confirming the existence of two 'B' characters.

The following are the comments from the discussion:

- We recommend the less common 'B' character (U+10571), which had been called BHE, be re-named BBE. The name BHE suggests a phonetic difference between it and BE, which is not clear. Note the pattern of doubling the consonant in the case of the H's (H, HHA).
- As a result of the name-change for BHE, section 5.1 should be revised.
- The Script Ad Hoc has no objections about encoding the four case-pairs of modern-use characters (GJA, RRE, XHE, and ZHE), although no evidence has been provided, other than a font by Edon Muhaxheri (shown on the top of page 4, with the characters circled in red). The author mentions the letters could appear in tattoos in the future. The UTC should discuss this.
- Modify the section of ordering, to indicate that case pairs should be distinguished at the tertiary weight level, as for other bicameral scripts.

Recommendation: We recommend that the UTC approve the following, after discussing modern use characters:
SAH-UTC165-R3: The UTC accepts 78 Vithkuqi characters in a new Vithkuqi block (U+10570..U+105BF) for encoding in a future version of the standard, with the edits as noted. (Reference: L2/20-187R)

II. AMERICAS

4. UCAS

Document: L2/20-255 Proposal to encode 16 additional characters to the Unified Canadian Aboriginal Syllabics -- King

Comments: We reviewed this proposal for additions to Unified Canadian Aboriginal Syllabics: twelve are in modern usage for Nattilik (Nattilingmiutut), a dialect of Western Canadian Inuktut, and four are historic characters used for Cree and Ojibwe. The proposal author has been working with the local communities in Canada, which have discovered gaps in the repertoire when trying to write their languages. UCAS covers some 38 languages, dialects, and subdialects. Some of the proposed characters have appeared in customized fonts with PUA encoding, but they have not yet been proposed for inclusion in Unicode.

The following are highlights of the discussion:

- All the proposed characters are attested in the figures. While the attestation for the historic characters is slim (i.e., two figures), the historic characters were part of the original Cree Syllabics orthography.
- The proposal also includes several letters of support for the modern-use characters. In addition, it provides keyboard layouts (figures 12-13) and a Nattilik Syllabics orthography chart (figure 14).
- The use of "Nattilik" in the names for the twelve modern characters seems acceptable; it is shorter than using "Nattilingmiutut."
- Because there are only ten spots available in the UCAS Extended block, a new block is needed. One column is available starting at U+11AB0, which we recommend be used for this set of characters, and we suggest the name "Unified Canadian Aboriginal Syllabics Extended-A." The ten available spots in the UCAS Extended block could be used in the future for possible encoding of other UCAS characters that are currently under investigation by Craig Cummings (such as for Kickapoo).
- There was discussion about names and annotations, since four proposed names end in O or OO, but the annotations have U or UU, representing local Nattilik spelling preferences. It was noted

that the character names in the UCAS blocks are already inconsistent.

Recommendation: We recommend that the UTC approve the following:
***SAH-UTC165-R4*:** The UTC accepts 16 Unified Canadian Aboriginal Syllabics characters in a new Unified Canadian Aboriginal Syllabics Extended-A block (U+11AB0..U+11ABF) for encoding in a future version of the standard.
11AB0 CANADIAN SYLLABICS NATTILIK HI
11AB1 CANADIAN SYLLABICS NATTILIK HII
11AB2 CANADIAN SYLLABICS NATTILIK HO
11AB3 CANADIAN SYLLABICS NATTILIK HOO
11AB4 CANADIAN SYLLABICS NATTILIK HA
11AB5 CANADIAN SYLLABICS NATTILIK HAA
11AB6 CANADIAN SYLLABICS NATTILIK SHRI
11AB7 CANADIAN SYLLABICS NATTILIK SHRII
11AB8 CANADIAN SYLLABICS NATTILIK SHRO
11AB9 CANADIAN SYLLABICS NATTILIK SHROO
11ABA CANADIAN SYLLABICS NATTILIK SHRA
11ABB CANADIAN SYLLABICS NATTILIK SHRAA
11ABC CANADIAN SYLLABICS SPE
11ABD CANADIAN SYLLABICS SPI
11ABE CANADIAN SYLLABICS SPO
11ABF CANADIAN SYLLABICS SPA
(Reference: L2/20-255)

We also recommend that the UTC make the following disposition:
Assigns an AI to the Roadmap Committee to request a new Unified Canadian Aboriginal Syllabics Extended-A block that extends from U+11AB0..U+11ABF. Reference: Section 4 of L2/20-250 (Script Ad Hoc Recommendations).

III. MIDDLE EAST
5.Arabic
5a. Three Symbols
Document: L2/20-245 Proposal to encode three Arabic symbols -- Khaled Hosny and Roozbeh Pournader

Comments: We reviewed this proposal for three Arabic characters that are known in the Arabic type community and were brought to the attention of Roozbeh Pournader by Khaled Hosny. While the characters are relatively rare, they clearly are in use, primarily in Arabic language materials. Most examples are from texts published in Egypt, but may be used in other countries. (For example, see the Persian text in figure 6, though the text was published in Egypt.)

The first character, ARABIC END OF TEXT MARK, ends a block of text, larger than a paragraph. The other two characters are currency signs, informally written as a dotless (or occasionally dotted) head of *jeem* above an amount, similar to the Arabic subtending and supertending marks at U+0600..U+0605. The symbols are not known to be used in official material, but do appear in advertising, on price tags, and handwritten texts. The glyphs show some variation in shape.

Note that the first character will close the last available spot in the Arabic block; the currency symbols are in the new Arabic block, Arabic Extended-B block.

Recommendation: We recommend that the UTC approve the following:
***SAH-UTC165-R5*:** The UTC accepts the following 3 characters for encoding in a future version of the standard:
061D ARABIC END OF TEXT MARK
0890 ARABIC POUND MARK ABOVE
0891 ARABIC PIASTRE MARK ABOVE
(Reference: L2/20-245)

5b. Kazakh, Kyrgyz, and Uyghur
Document: Request for annotations and possible glyph changes for Kazakh, Kyrgyz, and Uyghur (Draft) – Evans

Comments: The Script Ad Hoc reviewed a draft version of this document and provided feedback.

Recommendation: We recommend the UTC make the following disposition:
Assigns an AI to Ken Whistler and the Ed Committee to change the spelling "Uighur" to "Uyghur" in the names list annotations, to bring them in line with the current spelling conventions in the Core Specification. For 14.0. Reference: Section 5b of L2/20-250 (Script Ad Hoc Recommendations).

5c. Eastern Arabic Fraction
Document: L2/20-239 Public feedback (dated August 16 2020): Eastern Arabic Fractions -- Moemen Metwally

Comments: We reviewed this public feedback, in which the author requested vulgar fractions be encoded for Arabic Indic digits.

Vulgar fractions in Arabic should typically be handled as a sequence with U+2044 FRACTION SLASH, described in section 6.2 of the Core Spec. Note that the vulgar fractions encoded in the Number Forms block have been encoded for compatibility with legacy character sets, and are given compatibility decompositions with U+2044.

Recommendation: We recommend the UTC make the following disposition:
Assigns an AI to Roozbeh Pournader to respond to the author of Eastern Arabic Fractions feedback in L2/20-239 about the vulgar fractions and 'Egyptian' two. Reference: Section 5c of L2/20-250 (Script Ad Hoc Recommendations)

---

6. Book Pahlavi
Document: L2/20-246 Teeth and bellies: a proposed model for encoding Book Pahlavi -- Pournader

Comments: The Script Ad Hoc did not review this document, but the author requests those interested to send him any comments.

Recommendation: We recommend the UTC make the following disposition:
Notes this document (L2/20-246), but takes no further action. Reference: Section 6 of L2/20-250 (Script Ad Hoc Recommendations).

---

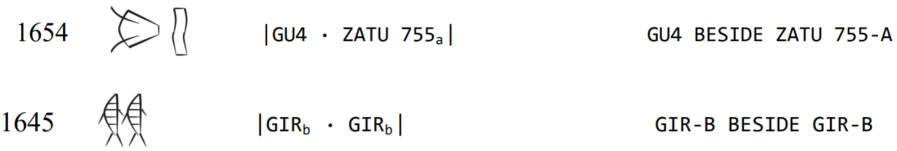
7. Proto-Cuneiform
Document: L2/20-193 Preliminary proposal to encode Proto-Cuneiform in Unicode -- Pandey

Note: Below are comments on an earlier version of this proposal that the Script Ad Hoc reviewed.

Comments: We reviewed this preliminary proposal for Proto-Cuneiform (PC), a script that had been earlier proposed by Michael Everson (L2/16-267 and L2/17-157), with additional information by Laura Hawkins (L2/19-284). The proposals from 2016 and 2017 were limited to the 200+ most frequent signs, but encoding the entire repertoire was deemed warranted in order to represent PC texts. There are 349 numeric signs, 1106 individual ideographic signs, and 640 complex signs. After removing 45 fragmentary signs, the total repertoire numbers 2050 characters. Anshuman Pandey reported that he will be preparing a separate document on the number systems of PC and Proto-Elamite.

The following comments were made:

- Include a section showing the difference between Proto-Cuneiform and Proto-Elamite signs.
- Include a section talking about those characters that appear beside one another, with actual text examples. Which would you handle atomically and which would be handled as sequences?



- Circulate the proposal to experts, pointing out that in character encoding, not all graphically compound characters need to be separately encoded as units: sequences can be used. For Sumero-Akkadian cuneiform, those sequences that had clear lexical significance ("complex sign") but were graphically compound were separately encoded.
- How do experts define a compound, and should a compound be encoded as an atomic character, or should it be represented by a sequence of two (or more) characters?
- In order to accommodate the full set of characters, we recommend PC be located in the range U+12580… U+12DFF, which will necessitate moving Cypro-Minoan to another location. We recommend U+12F90…U+12FFF for Cypro-Minoan.

Recommendation: We recommend the UTC make the following approval:
***SAH-UTC165-R6*:** The UTC moves the 98 approved characters for Cypro-Minoan at U+12700..U+12761 and its attendant Cypro-Minoan block (U+12700..U+1276F) to U+12F90..U+12FF1 in a Cypro-Minoan

block whose range extends from U+12F90…U+12FFF. Reference: Section 7 of L2/20-250 (Script Ad Hoc Recommendations).

SAH-UTC165-R7: The UTC extends the range of the Proto-Cuneiform block so the block extends from U+12580…U+12DFF. Reference: Section 7 of L2/20-250 (Script Ad Hoc Recommendations).

We also recommend the UTC make the following dispositions:
Assigns an AI to the SAH to forward the comments in section 7 of L2/20-250 (Script Ad Hoc Recommendations) to the author of L2/20-193.

Assigns an AI to the Roadmap Committee to request a new Proto-Cuneiform block that extends from U+12580… U+12DFF, and moves the Cypro-Minoan block to U+12F90…U+12FFF. Reference: Section 7 of L2/20-250 (Script Ad Hoc Recommendations).

---

8. Proto-Elamite
Document: L2/20-192 Preliminary proposal to encode Proto-Elamite in Unicode -- Pandey

Note: Below are comments on an earlier version of this proposal that the Script Ad Hoc reviewed.

Comments: We reviewed this preliminary proposal for Proto-Elamite. It and Proto-Cuneiform are believed to have developed independently, but ultimately from the same source system in Mesopotamia in the fourth millennium BCE.

In Proto-Elamite, there are two categories of characters: 58 numeric and 1578 general ideographic signs for a total of 1636 signs. The ideographic signs can be further sub-categorized into 1374 individual signs and 204 compound signs.

The following were noted during discussion:

- Provide more information on the writing system of Proto-Elamite, discussing the syntax of the signs and the method of ordering them.
- We recommend Proto-Elamite be located at U+1BD00..U+1C37F.
- The atomic approach to encoding characters is advocated in this proposal. Discuss the pros and cons of atomic encoding versus a model with format controls for stacking, etc. If new combinations are likely to be discovered, using the atomic model will force the community to request new characters. If the sign repertoire is well-known and relatively stable, the atomic model may be best. Using a blend of the two models, such as is the situation for Egyptian hieroglyphs, will require a judgement call in some edge cases. If the combinatorial model is used, fields would need to be defined (see example below).

- Explain in the proposal that the repertoire does not include all graphical elements found in characters if they are not individually attested (i.e.,  , found in M127, below, is not found as

a single, stand-alone character). If a name is needed to refer to the individual elements, a name could be created.

M127  M127 + M127

Recommendation: We recommend the UTC make the following approval:
***SAH-UTC165-R8*:** The UTC approves a new Proto-Elamite block that extends from U+1BD00..U+1C37F. Reference: Section 8 of L2/20-250 (Script Ad Hoc Recommendations).

We also recommend the UTC make the following disposition:
Assigns an AI to the SAH to forward the comments in section 8 of L2/20-250 (Script Ad Hoc Recommendations) to the author of L2/20-192.

Assigns an AI to the Roadmap Committee to request a new Proto-Elamite block that extends from U+1BD00..U+1C37F. Reference: Section 8 of L2/20-250 (Script Ad Hoc Recommendations).

IV. SOUTH AND CENTRAL ASIA
9. Ahom
Document: L2/20-258 Proposal to encode additional signs for the Tai Ahom script – Morey

Note: Below are comments on an earlier version of this document that the Script Ad Hoc reviewed.

Comments: We reviewed this proposal to add additional consonants to the Ahom script. The Ahom script is used to write the Tai Ahom language, which is being revived today. Seven characters appear in the Tai Ahom version of the publication *World Tipiṭaka in Tai Scripts: 15-Script Set*, a Pali language document. The *World Tipiṭaka*, which is lengthy and very important to the community, did not distinguish YA and JA, so an additional character, probably an AHOM LETTER YA, is needed, though no shape has been suggested. The user community has also requested an AHOM LETTER FA and there may be other letters that may be added later.

The following summarizes the discussion:

- Currently there is only room to add six additional characters in the Ahom block, so an additional column is needed. There are two options: either extend the current block by one column to U+1174F or add a new one-column block, “Ahom Extended” (or “Ahom Extended-A”) from U+11740..U+1174F. In either case, Zou will need to be moved over one column (i.e., so its range extends from U+11750..U+117AF) and the size of the Pyu allocation needs to be reduced to five columns (i.e., U+117B0..U+117FF).
- We recommend the seven proposed characters go in the first seven spots in the new column (or new block) (i.e., at U+11740..U+11746).

- Remove FA and YA from the proposal. As more evidence is provided, a proposal can be submitted later for FA and YA. We agree that good locations for FA and YA would be U+1171B..U+1171C, since these two available spots are at the end of the consonants.

Recommendations: We recommend that the UTC approve the following:
SAH-UTC165-R9: The UTC accepts 7 Ahom characters for encoding in a future version of the standard:
11740 AHOM LETTER CA
11741 AHOM LETTER TTA
11742 AHOM LETTER TTHA
11743 AHOM LETTER DDA
11744 AHOM LETTER DDHA
11745 AHOM LETTER NNA
11746 AHOM LETTER LLA
(Reference: L2/20-258)

SAH-UTC165-R10: The UTC either (a) extends the current Ahom block one column so the block extends from U+11700..U+1174F or (b) approves a new Ahom block with the name Ahom Extended (or Ahom Extended-A) at U+11740..U+1174F. Reference: Section 9 of L2/20-250 (Script Ad Hoc Recommendations).

We also recommend that the UTC make the following disposition:
Assigns an AI to the Roadmap Committee to request the Ahom block be extended one additional column (with range U+11700..U+1174F) or, depending upon the UTC decision, create a new Ahom Extended (or Ahom Extended-A) block from U+11740..U+1174F, move the Zou block over so it extends from U+11750..U+117AF, and reduce Pyu to five columns (i.e., U+117B0..U+117FF). Reference: Section 9 of L2/20-250 (Script Ad Hoc Recommendations).

---

10. Goyakanadi
Document: L2/20-264 Introducing the Goyakanadi Alphabet -- Shashank Shenoy Basty

Comments: We reviewed this introductory proposal for an historical script used to write the Konkani language, as well as Goan Marathi.

The following is a summary of the comments:
- Relying on handwriting is risky.
- Invite experts to assist on this proposal. For example, ask script experts to help identify the text elements for the Kannada and Goyakandi texts in figure 7 and compare them. For assistance from the Unicode side, ask those with experience on proposals (such as Srinidhi and Sridatta, Shriramana Sharma, or Vinodh Rajan).
- At this point, the typographic examples of Goyakanadi on pages 19 and 21 are irrelevant.
- Compare Goyakanadi characters with Kannada and Telugu in a side-by-side chart of all the graphemes, but add columns for the different Goyankanadi manuscripts. A good example is figure 13 on page 26, but the figure should be expanded to cover the whole script, with multiple manuscript sources. Such a comparison should be used to replace that in section 3.5.

Recommendation: We recommend the UTC make the following disposition:
Assigns an AI to the SAH to forward the comments in section 10 of L2/20-250 (Script Ad Hoc Recommendations) to the author of L2/20-264.

---

11. Kannada
Document: L2/20-260 Preliminary Proposal to encode KANNADA SIGN COMBINING ANUSVARA RIGHT TOP -- Shashank Shenoy Basty

Comments: We discussed this proposal for one Kannada character.

The following summarizes the comments:

- The name should be changed to KANNADA SIGN COMBINING ANUSVARA ABOVE RIGHT (cf. GRANTHA SIGN COMBINING ANUSVARA ABOVE, TELUGU SIGN COMBINING ANUSVARA ABOVE)
- On page 2 and 3, provide the source of the example in the caption, as well as for the other examples. Details of publication date/place, etc. can be provided in a separate bibliography section.
- Change the general category property on page 5 from Mn to Mc.
- Change the code point from U+0C84 to U+0CF3.
- Technically Indic positional categories and Indic syllabic categories are needed, but Unicode experts can do this.

Recommendation: We recommend the UTC make the following disposition:
Assigns an AI to the SAH to forward the comments in section 11 of L2/20-250 (Script Ad Hoc Recommendations) to the author of L2/20-260.

---

12. Old Uyghur
Document: L2/20-191 Final proposal to encode Old Uyghur in Unicode -- Pandey

Note: Below are comments on an earlier version of this document that the Script Ad Hoc reviewed.

Comments: We reviewed this proposal for Old Uyghur, which has been revised several times and undergone many rounds of review.

One of the sticking points has been how to handle cases where word-final letters with extended horizontal terminals touch the initial letter of the following word (section 4.2). Various options were discussed: would a dedicated word-breaker (or joiner) be an option? What do scholars want? Should the decision be in the hands of the author, who would decide whether a breaking space or a swash that joins words should be used?

The following summarizes the comments:

- Because the touching terminal does not change the following word-initial letter's joining form (i.e., the initial does not become a medial), the behavior is, in our opinion, stylistic. Circle examples, such as on page 31, and call out the examples in the discussion in 4.2.
- In our view, SPACE should be used for ordinary separation of words and ZWNJ may be an option when one wants to "connect" words as in manuscripts.
- Change the character names and any directional wording in the document, now that the default directionality is horizontal, no longer vertical.
- Check the properties and make sure they align with the proposed characters

Recommendation: We recommend the UTC make the following disposition:
Assigns an AI to the SAH to forward the comments in section 12 of L2/20-250 (Script Ad Hoc Recommendations) to the author of L2/20-191.

---

13. Tangsa
Document: L2/20-259 Proposal to add the Tangsa Script in the SMP of the UCS -- Morey

Note: Below are comments on an earlier version of this proposal that the Script Ad Hoc reviewed. The comments have been accommodated in the posted version of the proposal.

Comments: We reviewed this proposal for the Tangsa script. Earlier versions of the proposal have been seen by the Script Ad Hoc.

The following comments were noted:

- A description of how tones are named is provided. The tonal character names reflect what the users would recognize; they are based on the Romanized orthography of Rev. Gam Win, in which the tones are marked with letters -Q, -Z, -X, and -C. (Cf. Liangshan Yi naming scheme, where tones are also named with letters.) Annotations with tone numbers (i.e., Tone 1, Tone 2, etc.) could be added, if deemed useful.
- During discussion of the proposal, the author noted the use of "V" for short /a/ vowels in the characters U+16A78..U+16A7B. As a result, the author amended the discussion on pages 3 and 4 and in the character names in U+16A78..U+16A7B.
- The order of the vowels reflects that of the creator, Lakhum Mossang, that is, -Z, -C-, -Q, and -X. However, the vowel set TANGSA LETTER UEC U+16A94 to TANGSA LETTER UEX U+16A97, has a different order: -C, -Z, -Q, -X, which also is the order specified by the script's creator.
- Because of last-minute changes in the names of some characters and their order, some members of the Script Ad Hoc requested the proposal have one more round of review before approval is recommended.

It was noted that the new block is already on the SMP Roadmap.

Recommendation: We recommend that the UTC make the following disposition:
Assigns an AI to the SAH to add L2/20-259 to the Script Ad Hoc agenda for additional review.
Reference: Section 13 of L2/20-250 (Script Ad Hoc Recommendations).

V. SOUTHEAST ASIA, INDONESIA, AND OCEANIA

14. Hanunoo / Tagalog

Document: L2/20-257 Please reclassify the Philippine *pamudpod* -- Brennan

Comments: We reviewed this document, which basically would allow U+1734 HANUNOO SIGN PAMUDPOD, which is classified as a Pure_Killer in IndicSyllabicCategory.txt, to be formally extended for use in Tagalog. To achieve this, the author makes a number of specific requests: add a formal name alias PHILIPPINE PAMUDPOD to U+1734 HANUNOO SIGN PAMUDPOD, change the Script value for U+1734 from Hano to Zyyy (Common) in Scripts.txt, and change the character's gc value from Mn to Mc. He further suggests U+1734 HANUNOO SIGN PAMUDPOD be added to ScriptExtensions.txt, listing Hanunoo and Tagalog.

The following comments were made:

- Viramas are not commonly shared across scripts. For example, a separate Syloti Nagri virama was accepted in 2019, rather than "sharing" the Bengali virama (L2/19-024).
- If the character *were* to be shared between scripts, a formal name alias is probably not warranted; instead, an annotation could be used. A formal name alias is used in certain cases, such as "Corrections for serious problems in the character names" (see https://www.unicode.org/Public/UCD/latest/ucd/NameAliases.txt). Cf. U+AA6E MYANMAR LETTER KHAMTI HHA, for which a formal name alias seems appropriate (see Section 17 of this document).
- After discussion, the recommendation from the Script Ad Hoc was to recommend a new character be encoded, TAGALOG SIGN PAMUDPOD, unless evidence is provided that demonstrates a strong convention of using U+1734 HANUNOO SIGN PAMUDPOD. We recommend the location be U+1715 and gc=Mc. Hence the character would have the following properties:
  (UnicodeData.txt) 1715;TAGALOG SIGN PAMUDPOD;Mc;9;L;;;;;N;;;;;
  (IndicSyllabicCategory.txt) Pure_Killer
  (IndicPositionalCategory.txt) Bottom_And_Right
- The author has confirmed that "PAMUDPOD" is the native name for this character, stating the name was "borrowed from the Hanuno'o language into Tagalog, and is spelled identically in both languages."

Recommendation: We recommend that the UTC approve the following:
SAH-UTC165-R11: The UTC accepts U+1715 TAGALOG SIGN PAMUDPOD for encoding in a future version of the standard. (Reference: L2/20-257)

15. Kawi

Documents: L2/20-256 Preliminary Proposal to encode Kawi in the UCS -- Aditya Bayu Perdana and Ilham Nurwansah

Note: Below are comments on an earlier version of this proposal that the Script Ad Hoc reviewed. Some of the points raised have been addressed in the new version.

Comments: We reviewed this preliminary proposal for Kawi, a script that is used today by researchers, enthusiasts and others, including people using the script to write languages not found in "authentic" Kawi corpora. An earlier proposal was written by Anshuman Pandey in 2012 (L2/12-125).

The following summarizes points made during the discussion:

- Based on past experience with Indic scripts, we would recommend 2 viramas: a visible killer and a stacker. The authors have adopted this approach, naming one "KILLER" and the other "SUBJOINER."
- The ccc for KILLER and SUBJOINER should be 9.
- For RO, we recommend use of DIGIT TWO, with an explanation provided.
- We suggest that the visually atomic and composite forms of AA, II, UU shouldn't be unified, and the composite forms should be encoded as sequences, using terms such as "visually unique" and "visual composite" or "graphic composite". Do not add decomposition in Unicode data.
- Describe the pros and cons of different *repha* placements – before consonants, after first consonant, between consonant group and vowels, at end of syllable.
- For punctuation, provide at least two samples each from different sources, citing the relevant figures after each mark listed in section 5.6. Consider whether users need to distinguish them or treat some of them as equal.
- The Roadmap allocation of Kawi is currently 5 columns, but 6 columns are needed. As a result, we recommend the Kawi block be moved to U+11F00..U+11F5F, and Leke moved to U+11DB0..U+11DEF.

Recommendation: We recommend that the UTC approve the following:
SAH-UTC165-R12: The UTC moves the Kawi block to U+11F00..U+11F5F, and the Leke block be moved to U+11DB0..U+11DEF. Reference: Section 15 of L2/20-250 (Script Ad Hoc Recommendations).

We recommend the UTC make the following disposition:
Assigns an AI to SAH to forward the comments in section 15 of L2/20-250 (Script Ad Hoc Recommendations) to the authors of L2/20-256.

Assigns an AI to the Roadmap Committee to request the Kawi block be moved to U+11F00..U+11F5F, and Leke moved to U+11DB0..U+11DEF. Reference: Section 15 of L2/20-250 (Script Ad Hoc Recommendations).

---

16. Khmer
Document*:* L2/20-174 Public Feedback (dated July 2, 2020) from Kent Karlsson

Comments: We reviewed this Public Review feedback from Kent Karlsson. The issue concerns Table 16-8 Khmer Subscript Consonant Signs in *TUS* 13.0, where the glyph for Khmer Consonant Sign Coeng DA and Khmer Consonant Sign Coeng TA are identical, posing a problem for users. The author recommends the glyph for Coeng DA should match one based on KHMER LETTER DA.

The following comments were made:

- Provide the historical background for Khmer Consonant Sign Coeng DA, with evidence.

- Discuss whether Khmer Consonant Sign Coeng DA should be handled by fonts, a VS, or a separate character. What are the trade-offs of the different approaches?
- It was noted that the following link to one of the references in the feedback is no longer working: http://aefek.free.fr/iso_album/antelme_bis.pdf (pp25 and 26). Norbert Lindenberg, however, found the a copy in the Internet Archive: https://web.archive.org/web/20190814131541/http://aefek.free.fr/iso_album/antelme_bis.pdf

Recommendation: We recommend the UTC make the following disposition:
Assigns an AI to Norbert Lindenberg to work with Kent Karlsson on a proposal discussing Khmer Consonant Sign Coeng DA. Reference: Section 16 of L2/20-250 (Script Ad Hoc Recommendations).

---

17. Myanmar
Name Alias for MYANMAR LETTER KHAMTI HHA
Document: L2/20-263 Request for a formal name alias for uniAA6E Khamti HHA -- Mitchell

Comments: We reviewed this short document, which requested a formal name alias of type correction for U+AA6E MYANMAR LETTER KHAMTI HHA. The name is incorrect: it should be MYANMAR LETTER KHAMTI LLA. Evidence is provided confirming that the current name is wrong. (The topic was raised by the author in L2/20-162, item 2.)

Recommendation: We recommend that the UTC approve the following:
***SAH-UTC165-R13*:** The UTC accepts a formal name alias of type correction for U+AA6E MYANMAR LETTER KHAMTI HHA. The formal name alias will be: MYANMAR LETTER KHAMTI LLA. (Reference: L2/20-263)

VI. EAST ASIA
18. Kanbun
Documents: L2/20-232 Proposal to Encode 9 Additional Kanbun Marks -- Wáng Yifán
L2/20-235 Unihan Ad Hoc Recommendations for UTC #165 Meeting -- Lunde et al.

Comments: We briefly reviewed the proposal for nine Kanbun marks and the Unihan Ad Hoc Recommendations on the proposal (pp. 24-25). Evidence is provided, and the request is reasonable, in our view.

The proposed location for a new "Kanbun Extended" block that extends from U+16FB0..U+16FBF appears acceptable; this location and name were recommended in the Unihan Ad Hoc document. We agree that the UTC should discuss the font size of the glyphs in the Unicode charts (i.e., should the glyphs in the code charts appear in reduced size as is currently done in the Kanbun block, or should they be full size, following the practice of the Japanese type industry?). Since we agree with the Unihan recommendations, no formal character approval is repeated below. However, a request for the Roadmap Committee to allocate space for the new block is included.

Recommendation: We recommend that the UTC make the following disposition:

Assigns an AI to the Roadmap Committee to request a new Kanbun Extended block that extends from U+16FB0..U+16FBF. Reference: Section 18 of L2/20-250 (Script Ad Hoc Recommendations).

VII. SYMBOLS AND NUMERICAL NOTATION SYSTEMS

19. Currency Symbol

Document: L2/20-261 Proposal to add the currency sign for the KYRGYZ SOM to the UCS -- The National Bank of Kyrgyz Republic

Comments: We reviewed this proposal to add a Kyrgyz currency sign, requested by National Bank of Kyrgyz Republic. Examples are provided showing the sign in use. In order to have a name that is similar to other currency signs, we recommend SOM SIGN (cf. MANAT SIGN), and mention the country in an annotation.

Recommendation: We recommend that the UTC approve the following:
SAH-UTC165-R14: The UTC accepts the following character for encoding in a future version of the standard: U+20C0 SOM SIGN
Reference: L2/20-261

---

20. Kaktovik Numerals

Document: L2/20-262 Request to move the tentative location of Kaktovik numerals on the roadmap – Marín Silva

Comments: We reviewed this request to move Kaktovik numerals next to Mayan numerals, which we agree with the submitter (Eduardo Marín Silva) makes sense. To minimize the amount of re-shuffling, we recommend moving the Kaktovik numerals, which has two columns allocated to it in the SMP Roadmap, to U+1D2C0..U+1D2DF, so it is up against Mayan Numerals. With this change, the Flute allocation should be moved to U+1D250..U+1D26F, the current location of the Kaktovik numerals. The Pipa characters can remain where they are.

Recommendation: We recommend the UTC make the following disposition:
Assigns an AI to the Roadmap Committee to move the Kaktovik numerals block to U+1D2C0..U+1D2DF and put the Flute characters in the location U+1D250..U+1D26F. Reference: Section 20 of L2/20-250 (Script Ad Hoc Recommendations).

---

21. Persian Siyaq

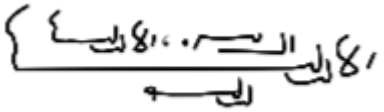
Document: L2/20-249 Proposal to Encode Persian Siyaq Numbers in Unicode (rev. Sept. 23) -- Pandey

Note: Below are comments on an earlier version of this proposal that the Script Ad Hoc reviewed.

Comments: We reviewed this proposal, which was last revised in late 2015 (L2/15-122R). The major changes are outlined on page 1.

The following summarizes the comments:

- Provide an explanation with rationale of why '3' and '6' have alternates encoded but not '80'.
- Emphasize that Persian Siyaq is used to represent currency and weights, but is not to be used to denote normal numbers.
- The glyphs in figure 36 are outliers, and shouldn't be relied on.
- How valid is Pihan as a source? What material did PIhan draw from?
- In section 5 "Considerations for Rendering" provide more examples of rendering. How would the following (from Figure 29, p. 49) be handled? Do such structures raise calligraphic issues, rather than being appropriately handled in plain text?



  How would the PLACEHOLDER MARK be used in '100' and numbers above? Are there examples?
- Should PERSIAN SIYAQ PLACEHOLDER MARK be a combining mark?
- Investigate further the qaran-specific shapes in figure 6 (page 26).
- Send a copy of the proposal to Shervin Afshar for his comments.

Recommendation: We recommend the UTC make the following disposition:
Assigns an AI to the SAH to forward the comments in section 21 of L2/20-250 (Script Ad Hoc Recommendations) to the author of L2/20-249.

VIII. OTHER TOPICS

22. *Variation Sequences for Combining Marks*

Documents:

L2/20-244 Variation sequences for combining marks-- Lindenberg
L2/20-247 Restrictions on base characters of variation sequences (L2/20-244) –Buff

Comments: We reviewed this request to make a technical change to the standard, changing the restriction against variation sequences with nonspacing combining marks to a restriction applying only those combining marks whose combining class is not 0. The document discusses the impact of VS on text segmentation, line breaking, and rendering, and provides draft text to the Core Spec. The initial impetus for this change arose during discussion of Khamti (L2/20-162).

When the Script Ad Hoc initially reviewed this document, there was agreement to support it. However, recent feedback was received from Charlotte Buff (L2/20-247), which requires careful consideration, so we recommend this document be returned to the Script Ad Hoc for further discussion.

Recommendation: We recommend the UTC make the following disposition:
Assigns an AI to the SAH to add L2/20-244 and L2/20-247 to the Script Ad Hoc agenda for further discussion. Reference: Section 22 of L2/20-250 (Script Ad Hoc Recommendations).

---

23. ***Other PRI Script Ad Hoc Topics***
Document: L2/20-239 Comments on Public Review Issues (July 20, 2020 - September 23, 2020)
Note: The Script Ad Hoc has not yet discussed the following two pieces of feedback as the document L2/20-209 has not yet been discussed. The PRI feedback will be discussed when L2/20-209 is taken up.

Date/Time: Thu Sep 3 16:18:05 CDT 2020
Name: Eduardo Marín Silva
Report Type: Feedback on an Encoding Proposal
Opt Subject: Suggestion to supplement L2/20-209 with named character sequences

> The proposal to add the characters needed for the kana Hokkien/Minnan orthography is in my opinion well formed and should not be considered "preliminary".
>
> That being said, I would suggest also proposing a set of named character sequences for the letters with combining marks, if and only if they have formal names. This is already the case for other kana letters with combining marks, and so it would make it all consistent.

Date/Time: Sat Sep 5 10:34:31 CDT 2020
Name: Ken Lunde
Report Type: Feedback on an Encoding Proposal
Opt Subject: Feedback on feedback on L2/20-209

> With regard to Eduardo Marín Silva's 2020-09-03 feedback on L2/20-209, I disagree with the proposal to add named character sequences. They are not necessary. The existing named character sequences for kana exist only because those combining sequences correspond to atomic characters in the JIS X 0213 standard, which is explicitly mentioned in NamedSequences.txt. Named character sequences for additional kana were once proposed in L2/16-133, but the UTC rejected them during UTC #147 for this reason:
>
> https://www.unicode.org/L2/L2016/16133-japanese-voiced-vowels.pdf
>
> The first paragraph of Section 1.1 of UAX #34, Unicode Named Character Sequences, captures this nicely:
>
> In some limited circumstances it is necessary to also provide a name for such sequences. The primary example is the need to have an identifier for a sequence to correlate with an identifier in another standard, for which a cross-mapping to Unicode is desired. To address this need, the Unicode Standard defines a mechanism for naming sequences and provides a short list of sequences that have been formally named. This list is deliberately selective: it is neither possible nor desirable to attempt to provide names for all possible sequences of Unicode characters that could be of interest.
>
> Regards...
> -- Ken