L2/20-169

TO: **UTC**
FROM: **Deborah Anderson, Ken Whistler, Roozbeh Pournader, Lisa Moore, Peter Constable and Liang Hai[1]**
SUBJECT: **Recommendations to UTC #164 July 2020 on Script Proposals**
DATE: **July 21, 2020**

The Script Ad Hoc group met on June 5 and July 10, 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, Andrew Glass, Marek Jeziorek, Norbert Lindenberg, Patrick Chew, Jan Kučera, Lawrence Wolf-Sonkin, Manish Goregaokar, and Lorna Evans.

I. EUROPE

1. Armenian

Documents:
L2/20-174 Comments on Public Review Issues (see full text of feedback in section 28)
L2/20-143 Uppercase of U+0587 ARMENIAN SMALL LIGATURE ECH YIWN -- Anderson

Comments: We reviewed the document L2/20-143, which responded to Public Review Feedback from Markus Scherer. The question involved the uppercase form of U+0587 ARMENIAN SMALL LIGATURE ECH-YIWN և. Currently, SpecialCasing.txt says it should uppercase to: U+0535 ARMENIAN CAPITAL LETTER ECH + U+0552 ARMENIAN CAPITAL LETTER YIWN, but the PRI Feedback says it should be: U+0535 ARMENIAN CAPITAL LETTER ECH + U+054E ARMENIAN CAPITAL LETTER VEW. The feedback further notes that Wikipedia says "The ligature և has no majuscule form; when capitalized it is written as two letters ԵՒ (classical) or ԵՎ (reformed)."

In our opinion, L2/20-143 provides the best answer (summarized on page one of that document):
U+0587 uppercases to U+0535 + U+0552 in Western Armenia (and Eastern Armenian speakers in Iran)
U+0587 uppercases to U+0535 + U+054E in Eastern Armenia ("Armenia" and those Armenians from former Soviet Union). L2/20-143 provides examples and more detailed information.

Recommendation: We recommend the UTC make the following disposition:

Assigns an AI to the SAH to forward L2/20-143 to the Properties and Algorithms Ad Hoc.

2. Cypro-Minoan

2a. Cypro-Minoan Script

Document: L2/20-154 Final proposal to encode the Cypro-Minoan script in the SMP --- Everson

Comments: We reviewed this proposal for Cypro-Minoan (CM), which was a revision of the 2016 proposal (L2/16-089) with changes based a meeting with experts (L2/19-166) and later email correspondence with experts.

The following summarizes the comments made:

- Reduce the block by one column, so the Cypro-Minoan block ends at U+1276F. Since it is unknown whether the two punctuation marks CM301 or CM302 appear in other Aegean scripts, it is advisable to keep them in the CM block.
- If experts want to propose the 21 CM0 signs found in ENKO Atab 001 (shown in figure 1) at a later point, a separate block can be allocated for them.
- An issue that has not been agreed upon by experts is the glyph change for U+1270C CM13, which varies from the main reference work, Olivier 2007. However, since the character's identity

is not in doubt, the revised glyph is retained in the proposal. This glyph change was agreed at a meeting in Paris in 2019. Michael Everson will provide the argument from Valério 2014 in a separate document on the font.

- How to handle numbers remains under study, since more research is needed. The proposal does not explicitly recommend unifying them with the Aegean Numbers.

Recommendation: We recommend that the UTC approve the following:

SAH-UTC164-R1: The UTC accepts 98 Cypro-Minoan characters for encoding in a future version of the standard, as amended during discussion at the SAH. Reference: L2/20-154.

2b. Cypro-Minoan Numbers

Document: L2/20-155 Considerations regarding Cypro-Minoan and Aegean numbers – Everson

Comments: We briefly reviewed this document that discussed Cypro-Minoan (CM) and Aegean numbers. Michael Everson references suggestions for CM numbers from an earlier draft of the proposal and discusses the configuration of the marks, referring to characters in the Aegean Numbers block and Linear B configurations.

Since the Cypro-Minoan numbers have been set aside for further study, no further action was deemed necessary.

Recommendation: We recommend the UTC make the following disposition:

Notes this document, but takes no further action, pending further work on the numbers by the author and the Cypro-Minoan experts.

2c. Cypro-Minoan Font

Document: L2/20-156 Considerations regarding a normalized Cypro-Minoan reference font – Everson

Comments: We reviewed this document, which showed different fonts for Cypro-Minoan. The font used in the main Cypro-Minoan reference work by Olivier is shown in figure 1, with a few changes by Michel Everson. This is the version strongly preferred by one expert (Valério). Figure 2 shows the CM characters whose strokes are monoline, based on the font style used for Linear B. However, this version was not preferred by any experts. Figure 3 contains hand-written glyphs by two experts. Michael Everson would like the default font for the code chart to be closer to the shapes in figure 3.

Because the glyphs are not normative, the discussion on the preferred font is not a blocker that will prevent Cypro-Minoan from being approved.

Recommendation: We recommend the UTC make the following disposition:

Notes this document, but takes no further action, pending further work on the fonts by the author and the Cypro-Minoan experts.

3. Latin

3a. Capital H with a line below in Unicode
Document: L2/20-174 Comments on Public Review Issues – Lippert (see full text of feedback in section 31)

Comments: We briefly reviewed this Public Review Feedback, in which the author requests a capital H with line below, corresponding to U+1E96 LATIN SMALL LETTER H WITH LINE BELOW. The character is needed for transliterating Egyptian hieroglyphs. The author also asks for upper- and lowercase *h* with U+032D COMBINING CIRCUMFLEX ACCENT BELOW, used for transliterating Demotic. The author admits the characters could be represented by the letter *h* and a combining diacritic, but fonts do not do a good job rendering the combinations.

Unicode has made a decision not to add any new precomposed characters, partly because the combinations of *h* and combining diacritic can already be represented by the sequences <0048 LATIN CAPITAL LETTER H, 0331 COMBINING MACRON BELOW>, <0048 LATIN CAPITAL LETTER H, 032D COMBINING CIRCUMFLEX ACCENT BELOW>, respectively. Those that are already in the Unicode (such as U+1E96 LATIN SMALL LETTER H WITH LINE BELOW) are present because they were in a pre-existing standard.

The main issue for the author appears to be getting fonts to render the sequences as desired. The author may need to test out various fonts to find one with an acceptable appearance and check with other Egyptologists. If the fonts do not provide the preferred appearance, the author can contact the font vendor and ask for changes. The request might get more attention if other Egyptologists also sign on to the request. (See also the FAQ https://www.unicode.org/faq/char_combmark.html#12b.)

Recommendation: We recommend the UTC make the following disposition:

> Assigns an AI to the SAH to forward the comments in section 3a of L2/20-169 (Script Ad Hoc Recommendations) to the author of the Public Review Feedback on "Latin capital H with line below" (reference: L2/20-174).

***3b.* Expansion of the extIPA and VoQS**
Document: L2/20-116R Expansion of the extIPA and VoQS -- Miller and Ball

Comments: We reviewed this final proposal to add 22 additional characters used to represent extended IPA and Voice Quality Symbols (VoQS).The Script Ad Hoc has regularly been reviewing revisions of the this proposal, earlier versions of which were seen at previous UTC meetings (L2/20-038, L2/20-039, L2/20-116). Characters include 4 characters in the Combining Diacritical Marks Extended block, 4 in the Latin Extended-E block (which will now be filled), and 4 Latin letters and 10 modifier letters in the Latin Extended-D block.

This final version addresses earlier questions and comments made by the Script Ad Hoc and is, in our view, mature.

Recommendation: We recommend that the UTC approve the following:

SAH-UTC164-R2: The UTC accepts the following 22 characters for encoding in a future version of the standard (reference: L2/20-116R):

1AC1 COMBINING LEFT PARENTHESIS ABOVE LEFT
1AC2 COMBINING RIGHT PARENTHESIS ABOVE RIGHT
1AC3 COMBINING LEFT PARENTHESIS BELOW LEFT
1AC4 COMBINING RIGHT PARENTHESIS BELOW RIGHT
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
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

We also recommend the UTC make the following disposition:

Assigns an AI to the names list editor to review the proposed annotations on page 4 of L2/20-116R.

Assigns an AI to Debbie Anderson to get a font from Kirk Miller for the extIPA and VoQS characters.

3c. Phonetic click letters

Document: L2/20-115R Unicode request for additional phonetic click letters -- Miller and Sands

Comments: We reviewed this proposal to add 7 phonetic symbols used to represent click consonants. Earlier versions of the proposal have been seen at previous Script Ad Hoc meetings and at the April/May UTC meeting (L2/20-115). Two of the proposed characters are in Latin Extended-D block and 5 in a new Phonetic Extensions Supplement-A block, which extends from U+10780…U+107BF. The new block is already on the SMP Roadmap.

Recommendation: We recommend that the UTC approve the following:

SAH-UTC164-R3: The UTC accepts the following 2 characters in the Latin Extended-D block for encoding in a future version of the standard:

A7F0 LATIN SMALL LETTER ESH WITH DOUBLE BAR
A7F1 LATIN LETTER RETROFLEX CLICK WITH RETROFLEX HOOK

and the following 5 characters in a new Phonetic Extensions Supplement-A block (U+10780…U+107BF)

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

(Reference: L2/20-115R)

We also recommend the UTC make the following disposition:

Assigns an AI to Debbie Anderson to get a font from Kirk Miller for the phonetic click characters.

3d. IPA Retroflex Letters, etc.

Document: L2/20-125R Unicode request for expected IPA retroflex letters and similar letters with hooks (revised) – Miller

Note: The following comments were made on an earlier version of the proposal; the comments were accommodated in the revised proposal, L2/20-125R.

Comments: We reviewed this proposal to add 13 characters used in various linguistic publications. The Script Ad Hoc already reviewed earlier versions of this proposal, which was also available at the last UTC meeting (L2/20-125).

The following comments were made:

- Remove “ligature” from the title of the document, replacing it with something like “fused”
- Replace “ligature” in the header with wording such as “additional character”
- List the proposed characters by code point order
- Remove U10792 LATIN SMALL LETTER DEZH WITH RETROFLEX HOOK from page one and take out the glyph in the chart on page 2, but retain the text “a gap in the Supplemental Plane is left for…”
- Doublecheck all figures (and figure 23).

We recommend this set of 13 characters be approved, after the author has modified his proposal, based on the comments above.

Recommendation: We recommend that the UTC approve the following:

SAH-UTC164-R4: The UTC accepts the following 3 characters in the Latin Extended-D block for encoding in a future version of the standard:

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

and the following 10 characters in the new Phonetic Extensions Supplement-A block:

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: L2/20-125R]

We also recommend the UTC make the following disposition:

Assigns an AI to Debbie Anderson to get a font from Kirk Miller for the IPA retroflex letters and other characters.

3e. Harrington diacritic

Document: L2/20-182 Unicode request for Harrington diacritic -- Miller and Henry

Comments: We reviewed this proposal for one character that was used in material written by J.P. Harrington.

According to the Harrington Database Project, Harrington penned over one million pages of linguistic and ethnographic notes on Native Americans and their languages. His materials are part of multiple projects today, including the Harrington DB Project. It also is used in revival language projects, such as the Purismeno Chumash dictionary. The proposed character, COMBINING NUMBER SIGN, is a common symbol in the works of Harrington. Accurately representing the Harrington material digitally is deemed to be very important. Ken Whistler voiced strong support for the character.

The proposal provides ample justification for the character, and includes a letter of support from the Western Institute for Endangered Language Documentation. We recommend approval of this character.

Recommendation: We recommend that the UTC approve the following:

SAH-UTC164-R5: The UTC accepts U+1AC6 COMBINING NUMBER SIGN
for encoding in a future version of the standard. (Reference: L2/20-182)

---

4. Todhri

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

Note: The Script Ad Hoc reviewed an earlier version of this proposal before submission to the UTC document register, and the comments below refer to that version.

Comments: We reviewed this preliminary proposal for the Todhri script, which is an historical, unicameral script used for the Albanian language. The proposal draws heavily on a book published by the late Robert Elsie on the script, published in 2016. Todhri and Vithkuqi (section 5, below) are two of the important historical scripts of Albania that are not yet encoded.

The following comments were made during discussion:

- To Michael Everson's knowledge, there has been no in-depth study of the accents in this script. Elsie reports (p. 12 of his 2016 book) that the author of the Todhri texts "often added a vertical or horizontal stroke over vowels to denote word stress." However, Elsie's transliteration only uses an acute to indicate stress.

  How should the accents be encoded: should they be script-specific or should the generic set of characters be used? (Note that the Elbasan script contained marks that looked like Greek breathing and/or accent marks. Because a complete analysis has not yet been done by scholars for Elbasan, the generic combining characters were used.)

- The letter E is distinguished from I by a dot; similarly, U is distinguished from O by a dot. Discuss the different encoding options for E and U:

  (a) Do not encode Todhri E and U separately; just represent them with Todhri I and O plus a combining dot above.

  (b) Encode Todhri E and U separately, but without canonical decompositions (to I and O + combining dot above).

  (c) Encode Todhri E and U separately, with canonical decompositions.

  The Script Ad Hoc did not come to any consensus on which option to take.

- Rename the proposal "Preliminary Proposal…"
- Change "4" on the top of page 2 (beside AS) to "3."
- On page 2, replace "A" with "2" for pre-nasalized character discussion and replace "B" with "3," to match the numbers in the chart on the top of the page.
- Is the ligature SHT rendered differently from the rendering of the two letters SH + T that are not ligated? Would encoding the atomic character SHT cause problems for those searching SH + T.
- The omega has a horizontal line above it in figures 1, 2, and 3. A line should be added above the glyph based on the figures.
- A dot appears in figure 1 and 2:
   Should the dot be added to the representative glyph?
- Mention in the caption for Figure 2 that Todhri is represented by the "Elbasan" column in the top chart.
- In the caption of figure 5, describe the meaning of colors in the glyphs.

Recommendation: We recommend the UTC make the following disposition:

Assigns an AI to the SAH to forward the comments in section 4 of L2/20-169 (Script Ad Hoc Recommendations) to the author of L2/20-188.

---

5. Vithkuqi

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

Note: Note: The Script Ad Hoc reviewed an earlier version of this proposal before submission to the UTC document register, and the comments below refer to that version

Comments: We reviewed this proposal, which revises the 2017 version of the proposal (L2/17-316), which was written before the death of a key expert, Robert Elsie. The 2017 proposal noted that the Kosovan artist Edon Muhaxheri had created an art exhibit with the script's letters and designed a new font. However, Muhaxheri had made some errors in his work by relying on Faulmann's faulty analysis in *Das Buch der Schrift*. Michael Everson has been in contact recently with Mr. Muhaxheri, and some corrections were made to the font. Michael Everson also worked with Muhaxheri to create additional letters for modern Albanian.

The following comments were made during discussion:

- The revised proposal includes details on the characters' ascenders and descenders (see page 4), which better reflect the original shapes of the letters.
- The brand-new characters are included in the image on page 4 and are annotated in the names list, but examples of actual use are needed. Without evidence, holes should be left in the chart until the characters can be shown to be in use by a community. Edon Muhaxheri reports that people have requested tattoos in the script, so images of the tattoos would be useful to have.
- Annotate the three paragraphs on the top of page 5 describing what they each indicate.

Recommendation: We recommend the UTC make the following disposition:

Assigns an AI to the SAH to forward the comments in section 5 of L2/20-169 (Script Ad Hoc Recommendations) to the author of L2/20-187.

II. AFRICA

6. Adrinkra

Documents:

L2/20-178 Adinkra request – Yaw

L2/20-179 Information on Current Usage of Adinkra (and Jan 2020 doc) -- Korankye (via Anderson)

Comments: We reviewed the two documents: a brief request from Yaw for encoding the Adinkra script, and a document with additional information from Charles Korankye. A full proposal with examples from in dictionaries, primers, etc. is needed. Information on submitting new proposals is described in: https://www.unicode.org/pending/proposals.html.

Recommendation: We recommend the UTC make the following disposition:

Assigns an AI to the SAH to forward the comments in section 6 of L2/20-169 (Script Ad Hoc Recommendations) to the authors of L2/20-178 and L2/20-179.

---

7. Egyptian Hieroglyphs

Document*:* L2/20-176 Cluster model for Egyptian Hieroglyphic Quadrats -- Glass

Comments: We reviewed this document, which outlines changes that need to be made by platforms or applications in order to provide full rendering support for shaping an Egyptian Hieroglyph quadrat. The cluster model should also be able to apply to Mayan Hieroglyphs. The document also discusses mirroring of the Egyptian hieroglyph script. The author, Andrew Glass, reported he will publish the cluster model described in this document on github.

Andrew Glass noted that Egyptian Hieroglyphs, like a number of other complex scripts, has no values in the IndicSyllabicCategory.txt data file in the UCD. This creates problems for rendering engines that rely on that property value for some aspects of rendering. A description of the cluster model in the Core Spec will provide some guidance for implementers.

Recommendation: We recommend the UTC make the following disposition:

Assigns an AI to Andrew Glass to provide text on the cluster model in the Core Spec for a future version of the standard, as described in L2/20-176 and section 7 of L2/20-169 (Script Ad Hoc Recommendations) .

---

8. Kore Sebeli

Document:

L2/20-180 Proposal for the encoding of Kore Sebeli --Guigon, Bangoura, and Sylla

Note*: L2/20-180 is a revised version of the proposal based on the comments below. The Script Ad Hoc has not yet reviewed the revised proposal. Many of the comments below have been accommodated in the posted version.*

Comments: We reviewed this proposal for a recently invented script for the Soso language, spoken in Guinea and surrounding countries.

The following comments were made during discussion (but see “Note” above):

- For a number of the proposed characters (symbols, punctuation or combining marks), we recommend using already encoded characters:
    - For plus, use U+22A4 DOWN TACK and for minus, U+22A5 UP TACK. It was noted that the two math symbols are identical to those use for Garay (L2/19-162), a script used in Senegal for Wolof.

- For equals, we recommend use of U+003A COLON. It was noted that COLON is also used as a mark of punctuation. If a distinction needs to be made between COLON and the sign for equals, provide examples showing the differentiation.
- For FINAL POINT , we recommend the sequence <002E, 002F, 002E>.

- For INVERSED EXCLAMATIVE MARK, a new character may be warranted. It was noted that the descender of U+00A1 ¡ INVERTED EXCLAMATION MARK extends below the baseline, whereas the samples shown of the Kore Sebeli mark sit on the baseline. Does the difference in positioning matter? If it doesn't make a difference, can INVERSED EXCLAMATIVE MARK be unified with 00A1 INVERTED EXCLAMATION MARK and, if not, why not?

- For the multiplication sign, we recommend either a new character or a sequence <00D7 MULTIPLICATION SIGN, 0308 COMBINING DIARESIS>. (The proposal can note the graphically similar character for the multiplication sign, U+1E8D LATIN SMALL LETTER X WITH DIAERESIS.)

- DIVISION SIGN:
  - The example in figure 1 (below) is similar to the Western long-division sign "⟌" ; it is used in a higher-level layout that wraps numbers, and is not in-line. Is the Kore Sebeli division sign used inline (such as "4 ÷ 4 = 1," for "÷")? If so, how would it be written? Can an image showing inline division expressions be provided?

  - The symbol proposed for division is graphically similar to 2A3D ⌙ RIGHTHAND INTERIOR PRODUCT and U+221F ∟ RIGHT ANGLE. Can the Kore Sebeli be unified with one of these?

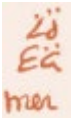
- Figure 8 contains characters that look like Latin comma and exclamation mark:

Are these Latin marks of punctuation? If so, add them to the list on p. 14.

- Vertical text. In Unicode, a script is assigned one default direction. Based on the proposal, the horizontal direction would be the default. Are there certain environments when vertical writing is used?

  The figures show two distinct kinds of "vertical" layout figs 3, 11 on the one hand, and fig 17 on the other. Which of these is still used? What are the contexts? What are the details regarding layout in each case? A section providing details is needed.

  This word appears to have two syllables:



  This word appears to have three syllables—first syllable on the top, and two syllables below:

This word appears to be five syllables—two on top and three below:

Is the reading direction
12
345

or

13
245?

Figure 17 appears to show actual vertical line direction, with each syllable written horizontally

- Create a chart with lowercase and uppercase characters appearing side-by-side. This will make review easier.

- For the diacritics (p. 14), use the three already encoded characters U+0307 COMBINING DOT ABOVE, U+0308 COMBINING DIARESIS, and U+1AB4 COMBINING TRIPLE DOT, since punctuation characters generally are used across many scripts. For the "combining double dot vertical," we recommend a new character be proposed, COMBINING TWO DOTS VERTICAL ABOVE. It will go into the Combining Diacritical Marks Extended block.

- Provide an analysis of those characters with diacritics and those without. For example, lowercase BA Ɛ has no dots, but lowercase GBE Ɛ̈ does. Uppercase DA takes a dot above (Ḍ), but the shape does not appear without the dot. Uppercase KA Ꞇ contrasts with uppercase KHO Ꞇ̇. Similarly, upper- and lowercase YE are differentiated from the casing pair for WE by two dots above. Such information is important to be able to know whether certain characters might better be represented as sequences with diacritics, so as to avoid duplicate encoding (i.e., using a base character with combining diacritics instead of using a single atomic character with dot diacritic).

- The placement of the dots in the proposed glyph  are at an angle, but in figure 7, they can appear horizontally when handwritten, i.e.,  . Is the angle important, or it is just a feature of the font?
- Figure 5 mentions that the script has developed over time, and the letter "o" has changed shape. Circle the glyphs in the figure.

Recommendation: We recommend the UTC make the following disposition:

Assigns an AI to the SAH to forward the comments in section 8 of L2/20-169 (Script Ad Hoc Recommendations) to the author of L2/20-180.

III. MIDDLE EAST

9. Arabic
9a. Kalasha
Document: L2/20-160 Proposal to include Kalasha Language alphabets (revised) - Rehmat Aziz Khan Chitrali
Background document: L2/20-091 Proposal to include Kalasha Language alphabets

Comments: The revised proposal includes examples from the 1999 dictionary, published in part by SIL, besides the images of a keyboard created by the proposal author.

The *Ethnologue* reports Latin as the primary script for this Indo-Iranian language of Pakistan, a decision that reportedly was made after the 1999 dictionary was published. Additional printed evidence of the Arabic script for this language is needed.

Recommendation: We recommend the UTC make the following disposition:
Assigns an AI to the SAH to forward the comments in section 9a of L2/20-169 (Script Ad Hoc Recommendations) to the author of L2/20-160.

9b. Indus Kohistani
Document: L2/20-157 Proposal to include Indus Kohistani Language alphabets -- Rehmat Aziz Khan Chitrali

Comments: We reviewed this proposal, which was written by the same author as the Kalasha language proposal (section 9a, above). The author has created a keyboard for "Indus Kohistani." The author states on page 1 that the proposed characters were used in a Shina Qaida (Primer Book). However, it was noted that Indus Kohistani (ISO 639-3 mvy) is a separate language from Shina Kohistani (ISO 639-3 pkl). It was noted that the captions to the examples read "excerpt from Indus Kohistani Qaida", but the links go the Shina Qaida.

Additional examples of "Indus Kohistani" are required from publications other than those of the author.

Recommendation: We recommend the UTC make the following disposition:
Assigns an AI to the SAH to forward the comments in section 9b of L2/20-169 (Script Ad Hoc Recommendations) to the author of L2/20-157.

IV. SOUTH AND CENTRAL ASIA

10. Brahmi

Documents:

L2/20-069 Proposal to Annotate Brahmi Sign Anusvara - Rajan and Sharma

L2/20-129 Comments on L2/20-069: Encoding of Tamil Brahmi Virama (U+11070) --- Ganesan

L2/19-402 Proposal to Encode 6 Characters in the Brahmi Block – Vinodh Rajan and Shriramana Sharma

Comments: We briefly reviewed L2/19-402, which proposed a separate character for BRAHMI SIGN OLD TAMIL VIRAMA, and L2/20-069, a later document which instead proposed annotating BRAHMI SIGN ANUSVARA.

Andrew Glass, the proposer of the Brahmi script, explained his reasons for supporting the separate encoding of the virama:

- While the Brahmi *anusvara* and Old Tamil virama have similar shape and one may have influenced the other historically, usage in Old Tamil is distinct: the virama is a pure killer. He noted that Old Tamil does not use *anusvara*, but that doesn't make the two equivalent.
- The position of the *anusvara* for the IndicSyllabicCategory=TOP, but the dot can move around. The Old Tamil virama generally has a right-side position and kerns to the right-side element, though the virama also can occur inside the base glyph. The virama is considered a spacing character by Andrew Glass.

We also briefly reviewed the document L2/20-129, which also supported a separate BRAHMI SIGN OLD TAMIL VIRAMA. The author of this document suggested orally that the Old Tamil virama could be represented with a ring instead of a solid dot, but evidence for the ring was not provided.

The outcome of the discussion was to recommend the separate encoding of BRAHMI SIGN OLD TAMIL VIRAMA.

Recommendation: We recommend that the UTC approve the following:

SAH-UTC164-R6: The UTC accepts U+11070 BRAHMI SIGN OLD TAMIL VIRAMA for encoding in a future version of the standard. Reference: L2/19-402.

---

11. Gurmukhi

11a. Bindi before Bihari

Documents:

L2/20-060 Feedback on Gurmukhi Bindi Before Bihari (L2/18-319, L2/19-167, L2/19-283) – Sarabveer Singh

L2/20-076 Summary of email discussion on Gurmukhi BINDI – Anderson

L2/20-104 Public Review Feedback from Singh

Comments: We reviewed the various documents on Gurmukhi and discussed with users the issues surrounding how to represent Gurmukhi *bindi* (U+0A02 ◌ਂ GURMUKHI SIGN BINDI) before *bihari* (U+0A40 ◌ੀ GURMUKHI VOWEL SIGN II). As described in L2/20-076, the *bindi* can occur both before and after *bihari,* but the current default placement is for the *bindi* after *bihari*, seen in left example below.

ਈਂਟ, ਧਿਆਈਂ

Note: The comments below on *bindi* apply also to *tippi* (U+0A70 ◌ੰ GURMUKHI TIPPI), i.e., the *tippi* can occur on the left or right side of the vowel sign.

The following summarizes the points raised during discussion:

One option is to handle *bindi* before *bihari* as a stylistic variation in the font, an approach that will be faster than waiting for an encoding solution. However, this approach will be dependent on fonts' support and text formatting, and will not be able to contrast the two forms (i.e., *bindi* before or after *bihari*) in plain text exchange, and thus will not actually fulfill the proposers' needs.

The second option, the distinct character sequence approach, which is the long-term, plain text solution, should ideally be based on orthographic contrast, because this special case may further complicate text shaping and users will need to wait for text shaping engines to be updated.

With the font-based approach, both forms are encoded in the sequence (a) below, and will be displayed as either "*bindi* after *bihari*" (the norm in ordinary texts) or "*bindi* before *bihari*" (the special form requested to be supported) according to font formatting (either through different fonts or OpenType Layout options).

(a) <consonant, 0A40 ◌ੀ GURMUKHI VOWEL SIGN II, 0A02 ◌ਂ GURMUKHI SIGN BINDI>
For an independent vowel *ii*:
<0A08 ਈ GURMUKHI LETTER II, 0A02 ◌ਂ GURMUKHI SIGN BINDI>)

(b) <consonant, 0A02 ◌ਂ GURMUKHI SIGN BINDI, 0A40 ◌ੀ GURMUKHI VOWEL SIGN II>
For an independent vowel *ii*:
<0A72 ੲ GURMUKHI IRI, 0A02 ◌ਂ GURMUKHI SIGN BINDI, 0A40 ◌ੀ GURMUKHI VOWEL SIGN II>)

With the distinct character sequence approach, sequence (a) will continue to represent the normal "*bindi* after *bihari*" form, while sequence (b) will be recommended to represent the special "*bindi* before *bihari*" form. This approach is useful because the two visually distinct representations are encoded with different character sequences. Currently sequence (b) is not generally supported by text shaping engines (especially the OpenType Layout).

Option (b)'s sequence for the independent vowel *ii* (<0A72 *iri*, 0A02 *bindi*, 0A40 *vowel sign ii>)* may appear to be disallowed in Table 12-16 of *TUS* (see below).

Table 12-16. Gurmukhi Vowel Letters

For	Use	Do Not Use
ਇ	0A07	<0A72, 0A3F>
ਈ	0A08	<0A72, 0A40>

However, the Script Ad Hoc participants did not consider the "Do Not Use" prescription for <0A72 *iri*, 0A40 *vowel sign ii*> as applying to <0A72 *iri*, 0A02 *bindi*, 0A40 *vowel sign ii*>. The prescription is believed to specifically address the confusability issue between <0A72 *iri*, 0A40 *vowel sign ii*> and U+0A08 *ii* (and the precomposed single character is preferred) when they may look identical.

With the distinct character sequence approach, the special appearance of having an extra *bindi* between U+0A72 *iri* and U+0A40 *vowel sign ii* ensures the precomposed U+0A08 *ii* cannot be used to create a confusable alternative encoding. Additional discussion may be added to the Core Spec to clarify this topic.

It was noted that once shaping engines are updated to allow the sequences with *bindi* preceding *bihari*, the *bindi* may overlap with *bihari* in most existing fonts, so the *bindi* may not be visible and hence might be confusable. This is, however, not an example of confusability caused by the architecture, but a common case where fonts are just not prepared for newly introduced usage of combining marks.

- CLDR (rather than UCA) is the proper place to specify tailoring of sequences that weight as single elements.

Recommendation: We recommend the UTC make the following disposition:

Assigns an AI to the Liang Hai to propose text for the Gurmukhi block introduction allowing for the sequence < consonant letter or U+0A72 ਲ਼ GURMUKHI IRI, U+0A02 ◌ਂ GURMUKHI SIGN BINDI, 0A40 GURMUKHI VOWEL SIGN II> and how it should be represented, noting that for many existing fonts the *bindi* may overlap with *bihari*. Reference: Section 11a of L2/20-169 Script Ad Hoc Recommendations.

11b. Addha Ya after Bihari

Document: L2/20-170 Rendering of 'Addha Ya' after 'Bihari' in Gurmukhi -- Sarabveer Singh

Comments: We reviewed this document, which requested the rendering of *addha ya* (i.e., half *ya*) <U+0A4D ◌੍ GURMUKHI SIGN VIRAMA, U+0A2F ਯ GURMUKHI LETTER YA> after a *bihari* (U+0A40 ◌ੀ GURMUKHI VOWEL SIGN II) be supported. The author proposed a ZWJ between *bihari* and *addha ya*

The example provided is:

Current rendering: ਕ੍ਰੀਜੇ Correct rendering: ਕ੍ਰੀਜੋ

<U+0A15 GURMUKHI LETTER KA, U+0A40 GURMUKHI VOWEL SIGN II, U+0A4D GURMUKHI SIGN VIRAMA, U+0A2F GURMUKHI LETTER YA,U+0A4B GURMUKHI VOWEL SIGN OO>

The author proposes inserting ZWJ (U+200D) after the vowel sign *ii* (*bihari*) but before the virama.

The following comments were noted:

- The current text in *TUS* (p. 482) reads:

  When U+0A2F GURMUKHI LETTER YA follows a dead consonant, it assumes a different form called *addha* in Punjabi, without the leftmost part, and the dead consonant returns to the nominal form, as shown in Table 12-17.

Table 12-17. Gurmukhi Conjuncts

ਦ + ◌੍ + ਯ	→	ਦ੍ਯ	(*dya*)	addha ya

[Additional *addha* forms are documented in Table 12-18, below]

ਸ + ◌੍ + ਥ	→	ਸ੍ਥ	(*stha*)	addha tha
ਸ + ◌੍ + ਮ	→	ਸ੍ਮ	(*sma*)	addha ma

- Are there additional *addha* forms besides those documented above (*ya*, *tha*, *ma*)?

- Although using a ZWJ to request a consonant letter's conjoining form (*addha ya*, in this case) aligns with the existing text presentation and shaping recommendations in *TUS*, existing text shaping engines (in particular, OpenType Layout ones) tend to insert a dotted circle (U+25CC) as a placeholder base in the absence of a valid base character for the conjoining sign. Therefore the proposed <ZWJ, virama, *ya*> encoding will not work well.

- Explain the poetic metrical considerations that may support an intentionally contrasted use of the two: वीजे and वीजे.

- There was discussion on whether this is a font or implementation issue.

Recommendation: We recommend the UTC make the following disposition:

Assigns an AI to the SAH to forward the comments in section 11b of L2/20-169 (Script Ad Hoc Recommendations) to the author of L2/20-170.

11c. Other changes to Gurmukhi

Document: L2/20-183 Proposed Changes to Gurmukhi – Irvanjit Singh and Manvir Singh

Comments: We reviewed this document, which identified three issues in Gurmukhi.

The following captures the comments:

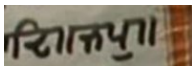
- The first issue reported by the authors is that there is no way to show subscript Gurmukhi numbers.

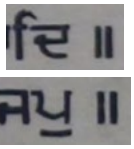
  [Comments by Script Ad Hoc]

  - If the authors want to propose the subscript numbers, find clear examples (circling the numbers and identifying them in the caption) and provide full information on the characters, including details on the placement and usage of the numbers. Show and discuss any difference between what appears in the manuscript vs. published version of texts.

- If some numbers, such as 7 and 8, are not attested, holes could be left for them in the code chart. Nine open spots are available at the end of the Gurmukhi block.
- The authors should consult L2/07-343, the proposal for Devanagari that included superscript numbers (now in the Devanagari Extended block).
- A general control character mechanism to create a subscript or superscript of any character is extremely unlikely to be approved.
- If music notation is required for plain text, create a proposal with provide evidence and rationale. Note that layout of musical notation is outside the scope of the Unicode Standard, and should be handled by higher-level protocols that make use of the graphical elements. (See discussion under Western Musical Notation in section 21.2 of the Unicode Standard).

- The second issue involves rendering U+0965 DEVANAGARI DOUBLE DANDA when it appears in historical *larivaar* Gurmukhi manuscripts (that is, manuscripts where words are written continuously with no breaks between words). In the manuscripts, the double dandas' two vertical strokes connect to the Gurmukhi script's headstroke when there are letters immediately preceding or following them. (See examples below.)

historical manuscript 

printed text with a modern font 

[Comments by Script Ad Hoc]
- The issue is not an encoding issue; it can be handled by a font.
- The Script Ad Hoc did not come to consensus whether text should be included in the Gurmukhi block intro in *TUS* about rendering of double dandas in manuscripts.

- The final issue involves YAYYA and HALF YAYYA
  The Script Ad Hoc was not able to get to this topic in any depth, but a few comments are captured below.

[Comments by Script Ad Hoc]
- A separate document should be written, with specific requests on what is being asked. Discuss what current implementations do (or don't do), and the rendering alternatives, and if the request is to encode a new character, justify the request based on discussion of current state and the options. Provide clearly identified examples, including, for example, the code points for the text in red on page 7.
- Fix the PDF so the correct characters appear.

Recommendation: We recommend the UTC make the following disposition:
Assigns an AI to the SAH to forward the comments in section 11c of L2/20-169 (Script Ad Hoc Recommendations) to the authors of L2/20-183.

---

12. Kaithi
Document: L2/20-151 Proposal to encode the KAITHI VOWEL SIGN VOCALIC R -- Srinidhi and Sridatta

Comments: We reviewed this proposal, which requested one Kaithi character. As noted in this document, the original Kaithi proposal by Anshuman Pandey did not propose the VOWEL SIGN VOCALIC R, though it is attested, because no independent vowel letter had been found. (Instead, KAITHI LETTER RA and KAITHI VOWEL SIGN II is used to represent the independent vowel letter.)

The authors of L2/20-151 have found several examples of use of VOWEL SIGN VOCALIC R, which, in our view, justifies encoding the character.

Recommendation: We recommend that the UTC approve the following:

SAH-UTC164-R7: The UTC accepts U+110C2 KAITHI VOWEL SIGN VOCALIC R for encoding in a future version of the standard. Reference: L2/20-151.

---

13. Limbu

Document: L2/20-184 Changing Indic Syllabic Category of Limbu Kehmphreng [*sic*] – Hosken

Comments: We reviewed this request to change the Indic syllabic category (InSC) for U+193A SIGN KEMPHRENG from Vowel_Dependent to an InSC property that will result in a Universal Shaping Engine (USE) category of Vowel Modifier. The author states that the current categorization causes KEMPHRENG to be stored before spacing vowels, whereas it needs to occur *after* spacing vowels. (He refers to storage order, though the InSC property doesn't impose a conformance requirement on storage order. However, in USE it would affect expected order of characters within a well-formed cluster, and that appears to be his concern.)

The author's goal is recategorization that would result in a USE categorization of VOWEL_MOD. He recommends Tone_Mark, but is open to other possibilities, as long as the result is a USE category of VOWEL_MOD.

The following comments were noted:

- The Limbu proposal L2/02-055 (WG2 N2410) describes the function of the SIGN KEMPHRENG as indicating a long vowel, but is itself neither a vowel or a tone mark. As a vowel modifier, it should follow a vowel in encoded representation.
- There is no Indic syllabic category for vowel modifier. Indic length marks are assigned InSC = Vowel_Dependent.
- Is the issue specifically a Universal Shaping Engine issue? Is there text in the Core Specification that needs to be changed?

Follow-up to the above discussion:

- Andrew Glass has confirmed that from the USE perspective, BINDU is a vowel modifier, and he added an override in 2019 for U+193A SIGN KEMPHRENG. The USE mapping to the VOWEL_MOD class via (Bindu) is now available in the latest Windows update (2004). (The USE overrides are publicly available, but the USE spec will need to be updated by Andrew Glass.) Andrew also commented "The naming semantics of InSC [IndicSyllabicCategory} may make some uncomfortable with formally putting this into the Bindu category. We could consider having a new bucket for Vowel_Modifier in InSC which USE would internally map to its VOWEL_MOD class."

Recommendation: We recommend the UTC make the following disposition:

Assigns an AI to the SAH to forward the comments in section 13 of L2/20-169 (Script Ad Hoc Recommendations) to the author of L2/20-184.

---

14. Old Uyghur

Document: L2/20-191 Final proposal to encode Old Uyghur in Unicode (July 2020) -- Pandey

Comments*:* We reviewed this revised version of the Old Uyghur proposal, which has been reviewed several times by members of the Script Ad Hoc and had review by various Old Uyghur experts. The Script Ad Hoc lauds the careful work of the author, particularly as the research involved materials that spanned nine centuries and a wide geographical area and reaching out to scholars.

The following summarizes comments that arose during discussion:

- The bulleted list of major changes on page 1 addresses several of the controversial points raised in earlier reviews of the Old Uyghur proposal. The revised proposal has tabled several supplemental characters, diacritics, and punctuation, for example, and recommends handling ambiguities present in cursive writing by markup. The revisions listed on page 1 remove the major stumbling blocks to encoding Old Uyghur, in the view of the Script Ad Hoc.
- Other suggested modifications:
    - Fill out section §8.2 Handling Ambiguity by providing explanation and examples of individual letters or by referring to earlier versions of the proposal which discussed this topic.
    - Add language to §4.2 Baseline extension noting that because the script is primarily vertical, the BASELINE EXTENDER is not unified with U+0640 ARABIC TATWEEL.
    - Until more information can be provided to justify encoding the space-filling terminal (§4.3), remove it from the list of proposed characters. Propose the character when a more thorough analysis can be provided.
    - In §9.4 Shaping properties for ArabicShaping.txt (page 40) list type C (Join_Causing) for U+10F82 OLD UYGHUR BASELINE EXTENDER and remove this character from the dual-joining table on page 29.

Recommendation: We recommend the UTC make the following disposition:

Assigns an AI to the SAH to forward the comments in section 14 of L2/20-169 (Script Ad Hoc Recommendations) to the author of L2/20-191.

---

15. Sinhala

Document: L2/20-137 Encoding of a character "Sinhala Chandrabindu"-- Gihan Dias

Comments: We reviewed this response from the ICT Agency of Sri Lanka liaison to Unicode, which asked that U+0D81 SINHALA SIGN CANDRABINDU be removed. The liaison representative states that though some Sinhala scholars were consulted on the character, the ICT Agency of Sri Lanka did not concur on encoding the character and its encoding makes the Sinhala block out of synch with Sri Lanka Standard 1134. The liaison further asked Unicode not encode characters in the Sinhala “code page” of BMP,

without agreement from ICT Agency of Sri Lanka (which then will coordinate with Sri Lanka Standards Institute).

The Script Ad Hoc noted that the U+0D81 SINHALA SIGN CANDRABINDU was published March 20, 2020, in Unicode 13.0. Because of Unicode's character encoding stability policy, once a character is encoded, it cannot be removed. Deprecating the character -- that is, identifying that a character is strongly discouraged in new documents -- is an approach that is limited to characters that pose significant architectural problems or ones causing implementation problems (see D13 in section 3.4, p. 90 of The Unicode Standard). One approach that is available, however, is to add an annotation to U+0D81 SINHALA SIGN CANDRABINDU that reads "used historically and not part of modern Sinhalese" or "not in modern usage."

Unicode is a standard for the writing system of various languages and its characters may or may not be part of a country's national standard. As such, the next version of SLS 1134 is not required to include all the characters in Unicode – a subset is certainly possible, as long as the code points in the national standard are aligned with those in Unicode.

To help resolve the issue at hand, we encourage the ICT Agency of Sri Lanka liaison to provide input on the proposed annotation. In addition, we recommend Lisa Moore be appointed as the Unicode liaison to ICT Agency of Sri Lanka in order to prevent any further miscommunication in the future.

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

Assigns an AI to Lisa Moore to respond to the author of L2/20-137 regarding the encoding of U+0D81 SINHALA SIGN CANDRABINDU.

---

16. Takri

Document*:* L2/20-149 Proposal to encode the TAKRI SIGN CANDRABINDU -- Srinidhi and Sridatta

Comments: We reviewed this proposal to add one Takri character. The document includes many examples.

The following captures comments raised during discussion:

- As noted in the Takri script proposal (L2/09-424), there is no standard form of the script, and "Takri" includes various regional varieties, with variation in glyph shapes and orthographic styles.
- Some of the modern glyphs (shown in the far right column of figure 4) appear to be significantly different from earlier periods. Given the range of variations shown in figure 4 and the fact that there is no standard "Takri," it is possible that some of the columns could be different scripts.
- In figure 1, identify the characters enclosed in the red box and provide the character sequence.
- The examples mainly show *candrabindu* over *om*. Does it occur over other letters? Should *om* be encoded atomically?
- Explain the other *candrabindu*-like structures in figure 2, identifying the base under the *candrabindu*, such "Aa" below. How should the following be encoded?

- Based on figure 4, does *candrabindu* only appear in old Takri?

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

Assigns an AI to the SAH to forward the comments in section 16 of L2/20-169 (Script Ad Hoc Recommendations) to the author of L2/20-149.

---

17. Telugu

17a. Two letters for Telugu

Documents:

L2/20-147 Unicode Expert committee – TVA [Tamil Virtual Academy] minutes -- Ponnavaiko et al.

L2/20-161 Response to the proposal to encode two letters in Telugu (L2/20-119) -- Muthu Nedumaran

Reference document:

L2/20-119 Proposal to Encode Two Letters in Telugu by Rajan

Comments: We reviewed the letter from Tamil Nadu, which was a response to L2/20-119. The document includes four recommendations, followed by detailed discussion in the minutes.

The following comments were raised during discussion:

- The goal of Unicode is to digitally represent text, so text is accurately represented in plain text. Pronunciation (mentioned in points ii and iii on pp. 1-2) is not taken into consideration in character encoding decisions.
- Recommendation iv (directly below) raised a valid point, in the group's view, but a number of questions and comments were raised (also below).

  Recommendation iv. Duplicating native characters of one language into another in an Indian context will have long term negative impact on all the languages of India that has huge number of languages having close relations to one another. Transcribed texts with mixed language characters are the simplest cases that exist in every language of India with several versions. If Unicode Consortium start encoding every such mix, it will impact the Indian language structure adversely over the period of time.

  - Comments on Recommendation iv:
    - When considering changes (such as to add Telugu to the set of scripts in the ScriptExtensions property), the scale of the problem needs to be taken into account: does the particular request affect a variant orthography, such as historical variants in an existing script which do not have widespread usage? Is it just allowing character to be quoted within another script? Or is it allowing the wholesale use of one script inside another? Because the individual Indic scripts have distinct shaping features, the latter could cause significant problems.
    - Script itemization, that is, the process in which text is broken down into script runs, could be a problem if the ScriptExtensions approach is adopted, because different scripts require different rendering, and the two Tamil characters in question are base characters that already have a strong script value ("Tamil," instead of "Common" or "Inherited"). (Vinodh Rajan reported the two characters do participate in Telugu shaping behavior.) How should implementations handle the Telugu-specific rendering of Tamil characters? From the view of one member of the Script Ad Hoc, specific Indic fonts could choose to support such forms if the itemization work were done at the platform level.

- We briefly reviewed the comments from Muthu Nedumaran, which proposed the two transcriptional letters may be unified to the existing Telugu characters that are graphically distinct but phonetically equivalent. He further proposed three potential solutions for requesting the transcriptional forms (i.e., locale-dependent font variation, manually requested stylistic set of the font, and standardized variation sequences).

- The Script Ad Hoc agreed to recommend the UTC remove the two characters at U+ 0C5B and U+0C5C from the Pipeline.

  It was not clear to the Script Ad Hoc what the best technical solution would be; a more thorough understanding of the technical implications of adding the ScriptExtensions property to the two characters is needed. Introducing the ScriptExtensions property for base characters with a strong script value may have major architectural impact and may not be properly implemented.

Recommendation: We recommend that the UTC approve the following:

SAH-UTC164-R8: The UTC retracts its approval of U+0C5B TELUGU LETTER TAMIL TRANSCRIPTIONAL LLLA and U+0C5C TELUGU LETTER TAMIL TRANSCRIPTIONAL RRA. Reference: L2/20-169 (Script Ad Hoc Recommendations).

We also recommend the UTC make the following disposition:

Assigns an AI to Lisa Moore to respond to the authors of L2/20-147.

17b. Telugu Underline Diacritic

Document: L2/20-120 Representing the Underline Diacritic in Telugu – Rajan

Comments: We reviewed this document which asks how to represent an underline diacritic in Telugu. This diacritic is used by linguists to denote the two phonemes (/æ/) and / æ:/). It is also used to extend the aspirated consonants and to denote Perso-Arabic phonemes.

Two options are presented for representing Telugu underline: either encode a new character, Telugu sign underscore, or unify the character with U+0952 DEVANAGARI STRESS SIGN ANUDATTA.

The following summarizes the comments made during discussion:

- *R10* (on p. 460 of *TUS* and repeated on page 4 of this document) states that modifying marks, including bindus [*anusvara*] and *svaras* apply to the syllable, and the order should be vowel sign, bindu [*anusvara* in this case], the *svaras* [*anudatta* here].
- We recommend *anudatta* be used for the underline diacritic with the existing encoding order.
- We recommend an AI be assigned to the Editorial Committee to review the wording on *anudatta* in the Core Spec and provide additional text (since it is only briefly mentioned on page 466). Also we recommend it is made clear that *anudatta* is covered by R10. Lastly, we recommend the specific behavior of *anudatta* be documented in the Telugu chapter.

Recommendation: We recommend the UTC make the following disposition:

Assigns an AI to the SAH to forward the comments in section 17b of L2/20-169 (Script Ad Hoc Recommendations) to the author of L2/20-120.

Assigns an AI to the Editorial Committee to review wording on *anudatta* in the Core Spec and provide additional text, as described in Section 17b of L2/20-169 (Script Ad Hoc Recommendations).

---

18. Tulu (Tigalari)
Document*:* L2/20-177 Tulu Unicode Minutes and Charts -- Dr. U.B. Pavanaja et al.
Related document:
L2/17-378 Preliminary proposal to encode Tigalari script in Unicode - Murthy and Rajan

Comments: We briefly reviewed this document, which includes the November 2019 minutes of a meeting of the Karnataka Tulu Sahitya Academy, in which the character set of Tulu was approved. A number of further actions were outlined in the minutes on page 1, including creating a Tulu font, sending the font to Unicode, and following up with Unicode on any issues. The contact to Unicode on Tulu will be Dr. U.B. Pavanaja.

The document includes a list of proposed character and code points, however it doesn't include enough analysis for it to be considered as a valid proposal for character encoding. The document is followed by the minutes and code charts that are signed by various members of the Karnataka Tulu Sahitya Academy on page 7-12.

The following comments were made during discussion:

- Another proposal for Tigalari (L2/17-378) has been seen by the UTC and Script Ad Hoc. The other proposal has received extensive comments from the Script Ad Hoc (see pp. 17-12 of L2/18-039). These comments should be taken incorporated in a proposal. Note that the authors of L2/17-378 are working on a revision of the document. So as not to duplicate work, the authors of L2/20-177 should correspond with the authors of L2/17-378, and try to work together.

Recommendation: We recommend the UTC make the following disposition:
Assigns an AI to the SAH to forward the comments in section 18 of L2/20-169 (Script Ad Hoc Recommendations) to the authors of L2/20-177.

SOUTHEAST ASIA, INDONESIA, AND OCEANIA

19. Javanese, Balinese, Sundanese
Document*:* L2/20-150 Syllabic category of Balinese Surang, Javanese Layar, and Sundanese Panglayar -- Bayu, Nurwansah, and Lindenberg

Comments: We reviewed this document, which was a response to an error report from Richard Wordingham and Action Item 143-A56a (investigate the feedback from R.S. Wihananto re inconsistency in InSC for Khmer ROBAT and Javanese, Balinese, and Sundanese *rephas*, in feedback on PRI #297, and suggest solution).

The document goes through the various points raised by Wihananto and Wordingham, providing rationale (and evidence, if needed) to support a number of requests:

- Change in the Indic Syllabic Category for three characters (U+1B03 BALINESE SIGN SURANG, U+A982 JAVANESE SIGN LAYAR, and U+1B81 SUNDANESE SIGN PANGLAYAR) to Consonant_Final;
- Modify annotations for the three characters (above) in the names list, and add an annotation for BALINESE SIGN SURANG and JAVANESE SIGN LAYAR;
- Update text in the "Behavior of *ra*" section of the Balinese block intro (§17.3 of *TUS*) and Figure 17-2 to remove the current recommendation of font-dependent *repha* for Kawi texts (<*ra*, virama, base> instead of the normal <base, *surang*>), due to the unnecessary complexity and confusability it brings to Balinese text processing. Details are spelled out on page 1 of L2/20-150).

The Script Ad Hoc agreed to the steps recommended in this proposal. Andrew Glass likewise agreed with the proposed changes.

Recommendation: We recommend that the UTC approve the following:

SAH-UTC164-R9: The UTC makes a change in the Indic Syllabic Category for U+1B03 BALINESE SIGN SURANG, U+A982 JAVANESE SIGN LAYAR, and U+1B81 SUNDANESE SIGN PANGLAYAR from Consonant_Succeeding_Repha to Consonant_Final in IndicSyllabicCategory.txt. Reference: L2/20-150.

We also recommend the UTC make the following dispositions:

Assigns an AI to the Editorial Committee to make the changes to section 17.3 of *TUS* as outlined in L2/20-150 and discussed in section 19 of L2/20-169 (Script Ad Hoc Recommendations).

Assigns an AI to the names list editor to review the proposed annotations for U+1B03 BALINESE SIGN SURANG, U+A982 JAVANESE SIGN LAYAR, and U+1B81 SUNDANESE SIGN PANGLAYAR on page 1 of L2/20-150 and discussed in section 19 of L2/20-169 (Script Ad Hoc Recommendations).

---

20. Myanmar

20a. Khamti

Document: L2/20-162 Notes on Khamti -- Ben Mitchell et al.

Comments: We reviewed this document, which makes several requests to be able to represent the writing conventions of the Khamti, Aiton, Phake, and Khamyang languages of India (and, for Khamti, Myanmar). These languages use the Myanmar script. The document also seeks input on questions (posed on page 2), and provides examples supporting the requests.

Questions and comments are noted below.

1. Dotted forms

Khamti users prefer dotted style on certain letters. A 2015 proposal recommended either disunifying the

dotted forms from non-dotted Myanmar characters (L2/15-257) or variation sequences; the latter was approved (see the document L2/15-320). Variation sequences are needed for 8-9 additional characters (listed in 1a). In addition, ten other characters need dots, but they are non-spacing combining marks (gc=Mn) or spacing marks (gc=Mc) (section 1b). Of those listed in 1b, only 103A MYANMAR SIGN ASAT has ccc=9, all others are ccc=0.

While the dot that appears in Khamti Shan glyphs is mentioned in §16.3 *TUS* (page 650), there is no mention of the standardized variation sequences, unlike text on variation sequences for Phags-Pa, Manichaean, Mongolian, and math.

[Comments by Script Ad Hoc]

- Support for dotted forms in fonts (with VSes) is inconsistent in fonts today, as reported by Craig Cornelius. This is causing problems for users. Most fonts in use today are not Unicode-based.
- The dotted forms appear to be presentation variants. Users would like to be able to access them. However, most common fonts support the major languages by default, which leaves out many minority language users.
- Is it possible to predict a nonspacing mark's dot from the base's VS? (It was noted that when a base is not dotted, the nonspacing mark may still be dotted, cf. 'fire' on page 5.)
- Currently, a VS is only allowed after a base or spacing combining mark. The UTC would need to agree to allow VS after any non-combining mark. The main concern involves canonical re-ordering.
- Various options could be considered:
    - disunify all dotted characters (though it was noted this approach would go against existing practice).
    - approve the proposed set of variation sequences listed in 1a and spacing marks in 1b, but encode the non-spacing combining marks as new characters.
    - delegate the variation to the font level
    - ask the UTC to relax its rule about a VS after NSM, and approve the entire set of new variation sequences proposed in 1a and 1b.
- We recommend the authors prepare a proposal for the UTC on the possible options, considering the edge cases and implications for normalization and segmentation.

2. Incorrect name for U+AA6E MYANMAR LETTER KHAMTI HHA

The character U+AA6E MYANMAR LETTER KHAMTI HHA is reported to be incorrect; instead, it should be MYANMAR LETTER KHAMTI LLA.

[Comments by Script Ad Hoc]

- Ask Martin Hosken about the origin of the name.
- We recommend the proposers suggest an annotation first, then if the name is documented as a true name, propose a formal name alias.

3. Khamti Numerals

Additional research still needs to be done to verify with users that the characters are correct.

4. Recommendations sought: consonant *ra* and *-ai* vowel

The vowel sign *-ai* and the consonant *ra* can be represented by three different characters in Unicode, as shown below:

vowel sign -ai:

U+1032 ◌ဲ MYANMAR VOWEL SIGN AI (standard Burmese usage, recommended in UTN 11),
U+1086 ◌ႆ MYANMAR VOWEL SIGN SHAN FINAL Y (for Shan),
U+109D ◌ႝ MYANMAR VOWEL SIGN AITON AI (for Aiton)

The preferred dotted form is ◌ႆ, closest visually to U+109D ◌ႝ

consonant ra:

U+101B ရ MYANMAR LETTER RA (recommended in UTN 11, dotted form preferred for Khamti digital texts),
U+AA73 ꩳ MYANMAR LETTER KHAMTI RA (historic form or revised form?; not familiar to any collaborators),
U+AA7A ꩺ MYANMAR LETTER AITON RA (historic form cited in L2/08-276, and recommended by authors of this document for manuscripts in Aiton).
(In addition, U+101C လ MYANMAR LETTER LA may be used in handwritten texts for *ra*.)

How can standardized encoding be achieved in such a situation?

[Comments by Script Ad Hoc]

- An annotation could be added to the three characters, noting that they should be treated the same in searching and sorting.

- In cases where different code points represent small variants of letters (which may, perhaps, stand for the same sound), the Unicode Collation Algorithm (UCA) typically collates them next to each other, but does not equate them completely. Collation implementations would rely on CLDR for further tailoring, if, for example, complete equivalence is needed in some context.

  Currently in the UCA, U+1032 MYANMAR VOWEL SIGN AI is next to U+109D MYANMAR VOWEL SIGN AITON AI, but not beside U+1086 MYANMAR VOWEL SIGN SHAN FINAL Y. U+101B MYANMAR LETTER RA is next to U+AA73 MYANMAR LETTER KHAMTI RA and U+AA7A MYANMAR LETTER AITON RA.

  We recommend that U+1086 be collated next to U+109D, as the relationship between those two characters was not originally recognized when the current DUCET was drafted.

  It was noted that since the UCA default algorithm by itself cannot do a complete job of string ordering for languages using the Myanmar script, and must be augmented by syllabic based ordering, there may not be much benefit it trying to further adjust the default ordering specifically for Khamti.

5. Recommendations sought: Syllable reduplication
This topic was not yet discussed by the Script Ad Hoc. Ben Mitchell reported offline that additional information is needed on this item.

Recommendation: We recommend the UTC make the following dispositions:

Assigns an AI to the SAH to forward the comments in section 20a of L2/20-169 (Script Ad Hoc Recommendations) to the author of L2/20-162.

Assigns an AI to Ken Whistler to prepare a proposed update of UTS #10 for Unicode 14.0, to include all new repertoire for 14.0, and to include the change in collation for U+1086, based on the discussion in section 20a of L2/20-169 (Script Ad Hoc Recommendations).

20b. Arakanese and Mon

Document: L2/20-163 Notes on alternate letterforms in Arakanese and Mon languages-- Ben Mitchell et al.

Comments: We briefly reviewed this document, which requests assistance on how to handle historic Arakanese and Thai Mon letterforms that vary from standard Burmese: should separate characters be encoded, variation selectors be proposed, OpenType features (locl and language tags) used or should no action be taken (i.e., handle at the font level)?

[Comments by Script Ad Hoc]

- The Arakanese language today uses the Myanmar script, but documents written before mid-twentieth century use different forms of the letters. The Mon language in Thailand is written in the Myanmar script but with differences in several letterforms. In our view, encoding separate characters would be one option, or font variation (i.e., locl feature in OpenType) would be a different approach.

Recommendation: We recommend the UTC make the following disposition:

Assigns an AI to the SAH to forward the comments in section 20b of L2/20-169 (Script Ad Hoc Recommendations) to the author of L2/20-163.

---

21. Western Cham

Documents:

L2/20-061R2 Final Proposal to encode Western Cham (rev. May 25 2020) -- Martin Hosken

Comments:

L2/20-185 Comments on Western Cham proposal -- Patrick Chew

L2/20-186 Outstanding Western Cham Issues – Anderson

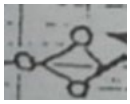
L2/20-104 Public Review Feedback from Doug Ewell on Arabic characters (see full text of feedback in section 29, below)

Comments: We reviewed the latest Western Cham final proposal alongside the "Outstanding Western Cham Issues" document (L2/20-186).

The author of the proposal relayed the desire of the Imam and his community to have the proposal move forward quickly, although Script Ad Hoc members mentioned that the first possible opportunity for publication would be Unicode 15 in September 2022. It was also noted that the text on the Pipeline of New Characters cautions users "Use of proposed or accepted characters is at implementers' own risk; the repertoire, names and code points of some characters may change before they are finally published in the Unicode Standard."

The following is a summary of the discussion:

- The Script Ad Hoc recommends moving the 8 lunar symbols from Arabic Mathematical Alphabetic Symbols to the first 8 spots in a new Arabic Supplemental Symbols block that extends from U+1EF00..U+1EF3F. This recommendation addresses the Public Review feedback comment from Doug Ewell in L2/20-104, which questioned putting the lunar symbols in the Arabic Mathematical Alphabetic Symbols block. (The entire PRI comment is contained in section 29 of this report.)
- We recommend postponing U+1E263 SIGN TANA TAMAT AYET, since its only attestation is figure 6. Could it be unified with U+2055 FLOWER PUNCTUATION MARK? Martin Hosken will ask members of the other user community (Pérez Pereiro, López Cortina, and Leb Ke) if they can provide any attestations. (Note: Debbie Anderson reported the group headed by Pérez Pereiro, López Cortina, and Leb Ke is working on a separate Unicode proposal.)
- U+1E264 SIGN TANA TAMAT TAKUE is also only attested in one figure. It should also, in our view, be removed from the proposal, until more evidence is provided.
- U+061D ARABIC END OF TEXT was considered by the Script Ad Hoc to possibly be a sequence of U+1E263 SIGN TANA TAMAT AYET during an earlier review (L2/20-122). At the July 10, 2020 meeting the Script Ad Hoc noted that the character appears to be a spacing filling mark in figure 23, where the shape is lengthened or shortened depending upon the available space. After discussion, the Script Ad Hoc agreed a single punctuation mark was appropriate, and implementations can adjust it to fill the available space in a line if needed.
- U+1E25F TRIPLE DANDA should be renamed END OF TEXT and be moved to another location. It was noted that an example of WESTERN CHAM TRIPLE DANDA is contained on page 3 of Patrick Chew's document (L2/20-185). Hence, a WESTERN CHAM TRIPLE DANDA that corresponds closely to that found in Eastern Cham (U+AA5F) should be considered for encoding at U+1E25F.
- The symbol in figure 23 (below) does not need to be encoded, according to Martin Hosken. Add a note in the caption mentioning this.



- Add a caption to the handwritten lists in figures 24-25 giving the context of the lists, noting when they were written. Is this the full set of Imam-approved characters? (See PPA note below.)
- Figure 19 includes an error mark, which is not proposed. Add a section in the proposal listing such characters, since they can be added to ScriptExtensions.txt.
- Provide an example of PPA. (Note: This character does not appear in the Imam's list.)
- The question by Martin Hosken (p. 9 of the proposal) on the Indic Properties of U+1E233 and U+1E236 and USE is postponed for later discussion.
- In our opinion, the overall approach in the proposal aligns with the Unicode model. Additional characters can be added to the repertoire when evidence is provided.

- We highly encourage the two user communities to continue to engage with one another and resolve their issues. A proposal that meets with the approval of both groups is needed for moving the approval process along more quickly. A letter confirming the group headed by Pérez Pereiro, López Cortina, and Leb Ke agrees with the core set of characters and/or agrees to support the proposal is desired.

Recommendation: We recommend the UTC make the following disposition:

Assigns an AI to the SAH to forward the comments in section 21 of L2/20-169 (Script Ad Hoc Recommendations) to the author of L2/20-061R2.

We also recommend the UTC make the following disposition:

Assigns an AI to the Roadmap Committee to request a new Arabic Supplemental Symbols block that extends from U+1EF00..U+1EF3F. Reference: Section 21 of L2/20-169 (Script Ad Hoc Recommendations).

VI. EAST ASIA

22. Kana

Document: L2/20-152 Base Character of HIRAGANA LETTER ARCHAIC WU -- Gross
Reference doc:
L2/19-381 Proposal to Encode Missing Japanese Kana -- Gross

Comments: We reviewed this short document, which requested the following annotation can be added to U+1B11F: “derived from 6C59 汙.”

As mentioned in the January Script Ad Hoc recommendations (L2/20-046), Japan had requested evidence showing that the base character of U+1B11F HIRAGANA LETTER ARCHAIC WU is U+7D06. The Script Ad Hoc recommendations noted that “the request from Japan for evidence showing the base character of HIRAGANA LETTER ARCHAIC WU … need not hold up encoding the character; that information can be added as an annotation later." (The character U+1B11F HIRAGANA LETTER ARCHAIC WU was approved by the UTC in January 2020.)

The proposer of U+1B11F HIRAGANA LETTER ARCHAIC WU was in touch with a classical Japanese researcher and they concluded the base character is U+6C59. Discussion and evidence is provided in https://kobunworld.blog.fc2.com/blog-entry-5.html.

Recommendation: We recommend the UTC make the following disposition:

Assigns an AI to the names list editor to include the proposed annotation, as described in L2/20-152.

---

23. Tangut

Document: L2/20-166 Tangut Glyph Modifications and Corrections (DRAFT) -- West and Zaytsev

Comments: We briefly reviewed this lengthy document, which proposed modifications for 72 Tangut components and 1,493 Tangut ideographs (in tables 1 and 7, respectively). The proposed changes reflect a systematic distinction between components with joined strokes versus those with unjoined strokes. The document also proposes glyph corrections for 2 Tangut components (table 4) and 32 Tangut ideographs (table 5). This document is a follow-up to discussion at WG2 meeting #68 in 2019, where it was noted that glyph shape issues affect a large number of Tangut ideographs.

The document requests careful review by Tangut experts, with the goal for the glyph changes to be incorporated in Unicode 14.0 in September 2021. We encourage further review of the proposed changes by experts.

Recommendation: We recommend the UTC make the following disposition:

Notes this document (L2/20-166), and encourages any Tangut experts to review it, submitting comments back to the authors of L2/20-166 .

SYMBOLS

24. Blissymbolics

Document: L2/20-140 Toward encoding Blissymbolics in Plane 1 (supersedes L2/98-364 = WG2 N1866) -- Everson

Comments: We briefly reviewed this "FYI" document, which corrected errors in the original 1998 Blissymbolics proposal (L2/98-364). A new, greatly expanded proposal for Blissymbolics will be forthcoming.

Recommendation: We recommend the UTC make the following disposition:

Notes this document (L2/20-140), but takes no further action.

---

25. Klingon

Document: L2/20-181 Proposal to encode Klingon in Unicode -- Shoulson and Litaer

Background documents:
L2/97-273 Proposal to encode Klingon Plane 1 (=WG2 N1643) – Everson
L2/16-329 pIqaD (Klingon) and its Usage -- Shoulson

Comments: We reviewed this proposal for Klingon. The early Klingon proposal from Michael Everson dates to 1997 (L2/97-273=N1643). In 2001, the 1997 proposal generated a response document L2/01-212, which led to a nonapproval that was recorded in May 2001. Mark Shoulson penned a later Klingon proposal in 2016, with many examples (L2/16-329). The minutes from UTC #149 in November 2016 recorded an action (149-A103) "Respond to submitter that it looks like there is sufficient usage to justify encoding Klingon as a script. UTC would need clear proof that Paramount would not pursue legal action against the Unicode Consortium, or anyone who implements the script."

The following comments were made during discussion:

- It was noted that there was a lawsuit involving Paramount that raised the issue of Klingon and an amicus brief (on the Klingon script, see page 12). The lawsuit was settled in 2017.
- We request the proposal author provide information on why trademark and copyright are no longer an issue, pointing to the notice of non-approval. A disclaimer from Paramount stating they have no interest in IP rights to the encoding of the proposal is needed.
- Provide some background on the Klingon Language Institute, which has a font for Klingon.
- Add the date to the proposal.

Recommendation: We recommend the UTC make the following disposition:

Assigns an AI to the SAH to forward the comments in section 25 of L2/20-169 (Script Ad Hoc Recommendations) to the author of L2/20-181.

---

26. Legacy Computing Symbols

Document: L2/20-158 Request for additional annotations in Symbols for Legacy Computing – Bettencourt

Comments: We reviewed this request for three suggested annotations to characters in the Symbols for Legacy Computing block. The proposed annotations appear to be too detailed for the names list. A more appropriate location might be in the Symbols for Legacy Computing block intro (§22.7 of *TUS*) or a detailed Technical Note.

Recommendation We recommend the UTC make the following dispositions:

Assigns an AI to the Editorial Committee to review the information contained in L2/20-158 and see if text can be added to §22.7 of *TUS*, or whether a Technical Note would be more appropriate. Reference: Section 26 of L2/20-169 (Script Ad Hoc Recommendations).

---

27. Music Symbols

Document: L2/20-159 Proposal to encode two accidentals for Iranian classical music – Pournader

Comments: We reviewed this proposal for two accidental characters invented in the early twentieth century for Iranian classical music.

The code points appear to be acceptable. This proposal provides ample attestations of the characters, along with the required character property information. The document proposed the ON bidirectional class, based on the fact that three other common symbols used in similar contexts are ON, and because the proposed characters frequently occur in right-to-left contexts.

Recommendation: We recommend that the UTC approve the following:

SAH-UTC164-R10: The UTC accepts the following two characters for encoding in a future version of the standard:
1D1E9 MUSICAL SYMBOL SORI
1D1EA MUSICAL SYMBOL KORON
Reference: L2/20-159.

VIII. PUBLIC REVIEW FEEDBACK (L2/20-174)

28. Public Review Feedback on Armenian
NOTE: This Public Review Feedback is handled above under 1. Armenian.

Date/Time: Thu Apr 23 13:30:44 CDT 2020 Contact:markus.icu@gmail.com
Name: Markus W Scherer
Report Type: Error Report
Opt Subject: uppercase of U+0587 ARMENIAN SMALL LIGATURE ECH YIWN

> We have received a bug report claiming that the uppercase form of U+0587 is wrong [The entry in] SpecialCasing.txt means that the ligature small *ech-yiwn* uppercases to capital *ech+yiwn*=0535+0552. The report says that it should uppercase to capital *ech+vew*=0535+054E.
>
> Wikipedia says "The ligature has no majuscule form; when capitalized it is written as two letters Եւ (classical) or ԵՎ (reformed)."
>
> Can someone confirm this? If true, should we change Special Casing.txt. to use the "reformed" uppercasing? Should implementers (e.g., ICU) offer both versions? Under what conditions? Please advise.

---

29. Public Review Feedback on Western Cham
NOTE: This Public Review Feedback is handled above under 21. Western Cham.

L2/20-104 Public Review Feedback from Doug Ewell on Arabic characters

> L2/20-061 proposes, among other characters, a group of eight characters for Western Cham lunar month names (ARABIC SYMBOL ONE DOT LUNAR MONTH through ARABIC SYMBOL SEVEN DOTS LUNAR MONTH), to be placed in the Arabic Mathematical Alphabetic Symbols block at code points U+1EEF8 through U+1EEFF.
>
> The Arabic Mathematical Alphabetic Symbols block was intended for stylistic variations of existing Arabic letters, to be used in special mathematical contexts. It is analogous to the Mathematical Alphanumeric Symbols block for existing Latin and Greek letters and digits. It is not intended for encoding of new “normal” characters. The proposed characters are “special” in that they are used only in Western Cham and only for lunar month names, but they are not “mathematical”; they are not used to represent variables, constants, sets, etc. in mathematical expressions.
>
> Both the text and the proposed Unicode properties show that the proposed

characters are not stylistic variations of existing Arabic letters, and do not follow the pattern of other characters in this block:

1EEF8;ARABIC SYMBOL ONE DOT LUNAR MONTH;So;0;ON;;;;;N;;;;;
cf.
1EE00;ARABIC MATHEMATICAL ALEF;Lo;0;AL;<font> 0627;;;;N;;;;;

They are “symbols” (So), not “letters” (AL), and are not <font> varieties of existing letters.

In the revision history, it was noted that these characters were moved in Revision 3 (November 2019) from the proposed Western Cham block to this block. Item 6 in the section “Repertoire” includes an inadvertent lingering reference to ARABIC SYMBOL SEVEN DOTS LUNAR MONTH being encoded at U+1E26F.

I recommend moving these eight symbols back into the proposed Western Cham block, as they were before Revision 3. I have no objection at all to encoding these symbols, only to this particular proposed location.

---

30. Public Review Feedback on Khmer
NOTE: The following came in too late to be discussed by the Script Ad Hoc. It will be on the agenda for a future Script Ad Hoc.

Date/Time: Thu Jul 2 15:46:27 CDT 2020
Name: Kent Karlsson
Report Type: Error Report
Opt Subject: KHMER CONSONANT SIGN COENG DA should look like KHMER LETTER DA, not like KHMER LETTER TA

Regarding:
http://www.unicode.org/versions/Unicode13.0.0/ch16.pdf
Table 16-8. Khmer Subscript Consonant Signs

This table gives for
17D2 178A khmer consonant sign coeng da
a glyph that is identical to that of
17D2 178F khmer consonant sign coeng ta

Actually, COENG DA did have, and should still have, a (range of) glyph derived from the (range of) glyph for KHMER CONSONANT DA.

The current "recommendation" (if that is what that table is) leads to that neither the author nor the reader of a text knows which of the two (COENG DA or COENG TA) is used in a text, as both looks like COENG TA. Further, one cannot represent (with that "recommendation") texts that really do have a

COENG DA that looks similar to a DA. COENG DA really did have its own glyph based on the glyph for DA. Having a separate (preferably DA-shape based) glyph for COENG DA will both make it possible for authors and readers to see (without checking the character code) whether a COENG DA or a COENG TA is used, and also makes historical as well as modern spelling using COENG DA possible.

(Introducing a "KHMER ARCHAIC COENG DA" or similar, which has been floated as a possibility, is not a good idea. It does not solve the first problem, and would be a strange and unnecessary "solution" to the second problem.)

I got two references from Richard Wordingham, both showing a "DA-shaped" COENG DA:

* http://aefek.free.fr/iso_album/antelme_bis.pdf (pp25 and 26)
* http://www.khmerfonts.info/fontinfo.php?font=1507

So the use of a "COENG TA"-glyph where one used to use "COENG DA" should be seen as a spell change, not a "glyph merger" or whatever.

Changing (correcting) fonts to use a "DA"-like glyph for "COENG DA" may reveal some (in modern view) spell errors, but that is as it should be.

Conclusion: in table 16-8, change the glyph in the line for
17D2 178A khmer consonant sign coeng da
to a subscript glyph based on the glyph for KHMER LETTER DA.

---

31. Public Review Feedback on Latin
NOTE: This Public Review Feedback is handled above under 3. Latin

Date/Time: Tue Jun 16 04:38:17 CDT 2020
Name: Sandra Lippert
Report Type: Feedback on an Encoding Proposal
Opt Subject: capital H with line below etc.

Dear Sirs and Madams,

I hope I chose the correct category for this - I did not find "proposing an encoding."

I am an Egyptologist, and while I am very glad that in the last years, almost all of the special glyphs we need for translitterating ancient Egyptian have been added to Unicode, I am very much puzzled why there is still no capital H with a line below in Unicode, even though the corresponding lowercase letter (U+1E96) exists. This was clearly an oversight, but why did it not get fixed since? It cannot be that no-one ever pointed it out: in my search for answers, I came upon a discussion thread

from 18 years ago
(https://unicode.unicode.narkive.com/8rfiWRgg/capital-letter-h-with-line-below)
where this problem was already mentioned, but nothing seems to have been done about it since. There, it was suggested that one combine U+0048 and U+0331, but this works only in a very limited number of fonts because the combining macron below is sometimes too large or too narrow for capital H and is often shifted to one side instead of being centered coreectly.

And while we are at it: the glyphs for capital and lower case h with ^ underneath (necessary for translitterating demotic texts) are also absent from Unicode, and again, adding a combining circumflex below (U+032D) does not work in a lot of fonts because it is not centered correctly. Sometimes, it works in the regular font but "slips off" to one side as soon as one switches to italics, which is standard for egyptological translitteration. This is not a very fancy letter either, and its "cousin," Ṯ/ṯ (U+1E70 / U+1E71), also used in translitterating demotic, is already present, so it would be very helpful if it was finally encoded as well.

Thank you in advance for considering my request. I am looking forward to hearing from you,

kind regards,

Sandra Lippert
Directrice de recherche
CNRS, Paris (UMR 8546-AOrOc)
