L2/23-012

TO: **UTC**
FROM: **Deborah Anderson, Ken Whistler, Roozbeh Pournader, and Peter Constable**[1]
SUBJECT: **Recommendations to UTC #174 January 2023 on Script Proposals**
DATE: **January 17, 2023**

The Script Ad Hoc group met on November 18 and December 16, 2022, and January 6, 2023, in order to review proposals. The following represents feedback on proposals that were available when the group met.

Table of Contents

[1] Also participating were Craig Cummings, Lorna Evans, Manish Goregaokar, Liang Hai, Ned Holbrook, Kushim Jiang, Frank van de Kasteelen, James Kirby, Jan Kučera, Robin Leroy, Norbert Lindenberg, Steven Loomis, Ksenya Samarskaya, Cheon Hyeong Sim, Harald Tveiten, Cheon Hyeong Sim, Lawrence Wolf-Sonkin, Michel Suignard, and Ben Yang. The text for the comments and recommendations was based on notes taken by Debbie Anderson, Lawrence Wolf-Sonkin, Liang Hai, and Norbert Lindenberg.

A. PROPOSALS REQUIRING UTC ACTION

I. EUROPE

1 Sidetic

Action: For UTC discussion and decision

Document: L2/23-019 Revised proposal to encode Sidetic in Unicode – Anshuman Pandey

Recommendations: We recommend the UTC make the follow disposition:
SAH-UTC174-R1: Reserve the code points U+10940..U+1095F for a new Sidetic block, with 29 code points reserved for Sidetic characters as described in L2/23-019 and section 1 of L2/23-012.

Comments: We reviewed this revised proposal, which has been reviewed by experts. The current naming system (SIDETIC LETTER N1, etc., with no hyphen between N and 1) was confirmed to be the one used by experts. N indicates the numbering system of Nollé 2001; cf. figures 17 and 18.

The SAH recommended only two columns for this script. If an additional column is needed later for additional characters, it can be added. (Anshuman Pandey reported that additional letters appear on

coins that may need to be added. Experts also reported new inscriptions with possible new letters.) The current Roadmap shows Sidetic beside Numidian, for which no proposal has been received.

The SAH considers Sidetic to be mature, based on this proposal and we recommend the UTC reserve code points for this script. The proposal is ready for review by PAG.

II. MIDDLE EASTERN SCRIPTS

2 Arabic

2a Two Quranic Characters

Action: For UTC discussion and decision

Document: L2/22-281R Proposal to encode Two Quranic Arabic Characters – Rikza F. Sh.

Recommendation: We recommend the UTC make the following disposition:
SAH-UTC174-R2: Reserve the code points for U+10EFB ARABIC SMALL LOW NOON and U+10EC5 ARABIC SMALL YEH BARREE WITH TWO DOTS BELOW, based on L2/22-281R and section 2 of L2/23-012.

The following actions are recommended:
Action Item for Lorna Evans and Roozbeh Pournader: Mention the ccc value of U+10EFB ARABIC SMALL LOW NOON and U+08D9 ARABIC SMALL LOW NOON WITH KASRA in UTR #53. (Reference: section 2a of L2/23-012)
Action Item for Liang Hai and the EdComm: Investigate whether to include text about the error of ccc value of U+08D9 ARABIC SMALL LOW NOON WITH KASRA in the Core Spec. (Reference: section 2a of L2/23-012).

Comments: We reviewed this proposal, which replaces L2/22-153 Proposal to Encode Three Quranic Arabic Characters by the same author. (The July 2022 Script Ad Hoc recommendations L2/22-128 had recommended the proposal author revise his earlier proposal and break it into two: one requesting two characters (with different code points) and a second document with the request to allocate space for 300 unencoded Arabic characters. The author followed the SAH recommendations.)

The two proposed characters appear reasonable to us, and are in the locations the SAH had recommended, U+10EFB and U+10EC5.

After our initial review, we received feedback from PAG alerting the SAH about an error in the proposal L2/22-281: U+10EFB ARABIC SMALL LOW NOON had ccc=230 (above) and it should have been ccc=220 (below). The error was corrected and the revised proposal is posted as L2/22-281R.

The PAG also discovered the same issue applies to an already encoded character, U+08D9 ARABIC SMALL LOW NOON WITH KASRA (published in Unicode 9.0, 2016). The error was originally caught by

Lorna Evans (http://www.unicode.org/review/pri350/) in May 2017, but it was too late in the encoding process and could not be corrected before publication. Note that David Corbett also noted the error in his comments on PRI #359 (on Proposed Draft UTR #53 from 2018).

For U+08D9, an action item is needed to explore whether text should be added to the Core Spec on the ccc error for U+08D9.

2b Unencoded Arabic Characters

Action: For UTC discussion and decision

Document: L2/22-284 List of Unencoded Arabic Characters – Rikza F.Sh.

Recommendations: We recommend the UTC make the following disposition:
SAH-UTC174-R3: Accept a new block allocation for Arabic Extended-D that extends from U+10D90..U+10E5F (Reference: L2/22-284 and section 2b of L2/23-012).

The following actions are recommended:
Action Item for Ken Whistler and the Roadmap Committee: Update the Roadmap to accommodate a new Arabic Extended-D block from U+10D90..U+10E5F, moving Byblos to U+1EB90..U+1EBFF (Reference: L2/22-284 and section 2b of L2/23-012)

Comments: We reviewed this document, which lists a large number of new characters (with examples and description), as had been mentioned in the author's earlier document (page 11 of L2/22-153). Based on the number of new characters, the author requests a new block be allocated that accommodates ca. 304 characters.

Roozbeh Pournader has made an initial review of the new characters and the document has also been reviewed by Marijn van Putten, a Quranic expert. In their view, many characters are eligible, but not all. According to Pournader, space for 222 new Arabic additions is needed. Hence, we agree that a new block is warranted, but only 13 columns are needed. To accommodate the new block, we recommend Byblos be moved down to U+1EB90..U+1EBFF.

The document includes honorifics, which can be accommodated currently in the Arabic Presentation Forms-A block. That block has 25 spots available. The document also includes a few Latin characters for transliteration of Quranic text, Latin "ligatures", and two modifier letters.

We encourage the proposal author or others to prepare full proposals, with attestation, descriptions, and properties of each character.

III. SOUTH AND CENTRAL ASIAN SCRIPTS

3 Bengali

Action: For UTC discussion and decision

Document: L2/22-268R Revised Proposal to Encode Alternate BA for the Bengali Language – Rajan and Chakraborty
Related document: L2/22-278 Usages of Bengali Letter Alternate BA for Pali Language – Srinidhi and Sridatta

Recommendation: We recommend the UTC make the following disposition:
SAH-UTC174-R4: Reserve the code point U+09FF for BENGALI LETTER ALTERNATE BA, based on L2/22-268R and section 3 of L2/23-012.

Comments: We revisited the Bengali alternate BA character topic.

Background: The SAH recommendations for the October 2022 UTC (L2/22-248), suggested the sequence <09AC BENGALI LETTER BA, 09BC BENGALI SIGN NUKTA> be made available to users in a special font. John Hudson instead recommended adding a new character. The issue concerns how to represent the alternate Bengali BA character, a letter that appears in Sanskrit printed works. The representative glyph of that character is identical to the existing U+09F1 BENGALI LETTER RA WITH LOWER DIAGONAL, which is used in modern Assamese orthography. However, the conjoining behavior is very different, making unification not at option.

Four options were considered during SAH discussion:

- Handle the Bengali alternate BA in a font, making no change in texts
- Use U+09BC BENGALI SIGN NUKTA (as recommended by the SAH earlier)
- Use a Variation Selector for U+09F1 BENGALI LETTER RA WITH LOWER DIAGONAL
- Encode a new character.

The implementers in the SAH advocated for a new character: relying on special fonts is not reliable; use of nukta will cause problems as the fallback form would look wrong; VS is not supported by the current Indic shaping engines.

While the different options all pose potential problems, the group arrived at consensus to recommend encoding of a new character.

The SAH has done an assessment and considers U+09FF BENGALI LETTER ALTERNATE BA, as proposed in L2/22-268R, to be a mature proposal. The proposal is ready for review by PAG.

---

4 Chisoi

Action: For UTC discussion and decision

Document: L2/22-218R3 Proposal to Encode Chisoi in the Universal Character Set -- Biswajit Mandal

Recommendation: We recommend the UTC make the following disposition:
SAH-UTC174-R5: Reserve the code points U+16D80..U+16DAF for a new Chisoi block, with 40 code points reserved for Chisoi characters as described in L2/22-218R3 and section 4 of L2/23-012.

Comments: We reviewed this revised document dated November 1, 2022. The author has made the change requested in the SAH recommendations (L2/22-248), with an example in section 3.4 (page 5) showing that when a consonant is modified by both a SISO and JARAHA, the sequence should be in visual order, not logical order: order should be <C, SISO, JARAHA>

In our view, the script is ready for encoding.

The proposal was sent to PAG, which caught an error. The linebreak property (page 9) for the digits was "AL" but they should have been NU. The error has been corrected and the revised proposal has been posted. A typo was also corrected in the same line (DGIT was corrected to DIGIT).

---

5 Sharada

Action: FYI with action to record

Recommendations: The following actions are recommended:
Action Item for Liang Hai: Add Do Not Use table for U+1118E SHARADA LETTER AI and U+111C4 SHARADA OM to Core Spec and noting that AI should not be represented by a sequence, but OM should use a sequence.
Action Item for Roozbeh Pournader: Add U+111C4 SHARADA OM and the discouraged sequence for SHARADA LETTER AI in his draft Do Not Use data file.

Comments: In discussion about a forthcoming proposal for Sharada additions to represent the Kashmiri language, the SAH noted some issues with SHARADA OM and SHARADA LETTER AI that need to be addressed in the Core Spec. In addition, Roozbeh Pournader should add SHARADA OM to his Do Not Use data file.

---

6 Tolong Siki

Action: For UTC discussion and decision

Document: L2/23-024 Proposal to encode the Tolong Siki script in Unicode -- Pandey
Related document:
L2/10-106 Preliminary Proposal to Encode the Tolong Siki Script in the UCS – Pandey

Recommendation: We recommend the UTC make the following disposition:
SAH-UTC174-R6: Reserve the code points U+11DB0..U+11DEF for a new Tolong Siki block, with 54 code points reserved for Tolong Siki as described in L2/23-024 and section 6 of L2/23-012.

Comments: We reviewed this updated proposal for Tolong Siki, a script used to write Kurukh, a Dravidian language spoken in India.

The author unified TOLONG SIKI SIGN MITALA (combining dot above) and TOLONG SIKI SIGN SULA (combining dot below) with the existing Combining Diacritical Marks (U+0307 and U+0323), as recommended by the Script Ad Hoc. In the version the SAH saw, Script Extensions for combining diacritics were added. A sentence or two that these combining diacritics are used in the script should be included in the Tolong Siki block intro. However, combining diacritics are not included in Script Extensions. (The Script Extensions data have been removed in the posted version of the proposal.)

Not unifying TOLONG SIKI SIGN SELA (the vowel length sign) with COLON (U+003A) is reasonable, because it has a glyph variant with round circles and should have a distinct general category (Lm vs. Po). Similarly, not unifying TOLONG SIKI SIGN HECAKA (glottal stop) with VERTICAL LINE (U+007C) was also reasonable.

Anshuman Pandey reported he received evidence showing the use of Tolong Siki in Bengali. He will add these examples to the version of the proposal that will be posted.

We deem Tolong Siki proposal to be mature, and recommend the UTC reserve code points for the script. The proposal is ready for review by PAG.

---

7 Tulu

Action: FYI with action to record

Document: L2/23-021 Additional Document for Tulu Unicode Proposal – Dr. Akash Raj Jain

Recommendations: The following actions are recommended:
Action Item for Jan Kučera: Investigate which Tulu characters can be unified with those in Tulu-Tigalari (and which cannot) and look into the architectural issues of unifying modern Tulu with Tulu-Tigalari.
Action Item for Debbie Anderson: Ask the Tulu-Tigalari proposal authors whether “Tulu” would be acceptable as the script’s name.

The UTC should also discuss whether the possibility of unifying Tulu with Tulu-Tigalari warrants postponing the inclusion of Tulu-Tigalari in ISO/IEC 10646 beyond Amendment 2, or else adopting a different font for the code chart. If the UTC decides Tulu-Tigalari should be removed from the CDAM2, ballot comments should be made requesting its removal.

Comments: We reviewed this document on Tulu from the Karnataka Tulu Sahitya Academy (KTSA).

Background: In January 2022, the UTC approved Tulu-Tigalari, which was identified as being used to represent the archaic script. Tulu-Tigalari is contained in CDAM2. The proposal for Tulu from Karnataka Tulu Sahitya Academy (KTSA) is intended to represent modern-day use of the script.

The authors have addressed many of the questions and concerns (on pages 36ff.), which the Script Ad Hoc report cited (section 12 of L2/22-023).

Discussion points that were noted:

- Based on an analysis by Jan Kučera, many of the glyph shapes are the same or similar, but three characters differ between KTSA and Tulu-Tigalari:
  TULU LETTER E is the same as the alternative form of TULU-TIGALARI LETTER EE
  TULU VOWEL SIGN E is the same as TULU-TIGALARI VOWEL SIGN EE
  TULU VOWEL SIGN O is the same as TULU-TIGALARI VOWEL SIGN OO
- The KTSA document proposes script-specific digits, which agree with script-specific digits reported in Tulu-Tigalari proposal (L2/22-031, p. 59). However, the Tulu-Tigalari proposal recommended use of Kannada digits for historical use (L2/22-031, p. 36), and noted that further study was needed for Tulu digits (referring to figures 21 and 22 of L2/22-031).
- The KTSA has been working on the orthography since 2007, with script behaviour changing even between this and their previous proposal (see note on p. 20 and cf. chart on Fig. 22 p. 122 in L2/21-188), and photographical evidence of the script usage does not always reflect the newest proposal. This raises the question whether the script is complete and stable enough to proceed with encoding.
- The document states that historic conjunct forms are not used much today (page 38).

Possible approaches that were mentioned during SAH discussion:

- Encode the Tulu-Tigalari characters, but use glyphs that are closer to those identified with modern Tulu. With this approach, historical script users may need specialized fonts.
- Encode modern characters, with unique historic characters in a separate block (cf. Meetei Mayek with modern characters in one block and historic characters in a separate Extensions block).
- Separately encode the scripts (cf. Bengali and Maithili)

The SAH group generally favored the first option, though research on specific issues is needed before a decision can be made, such as:

- identification of which characters would be unifiable (and which characters are not)
- resolving the architectural issues (i.e., conjunct formation) of modern vs. historic script usage.
- check whether it is acceptable to use the script name "Tulu" for both the modern and historical usage.

The SAH discussed holding off publication of Tulu-Tigalari in Unicode until the feasibility of can be assessed. If the UTC prefers this approach, ballot comments on the next CDAM ballot will be needed.

IV. SCRIPTS FROM SOUTHEAST ASIA, INDONESIA, AND OCEANIA

8 Tai Yo

Action: For UTC discussion and decision

Document: L2/22-289R Final Proposal to encode the Tai Yo Script– Viet Khoi Nguyen et al.

Recommendation: We recommend the UTC make the following disposition:
SAH-UTC174-R7: Reserve the code points U+1E6C0..U+1E6FF for a new Tai Yo block, with 55 code points reserved for Tai Yo characters as described in L2/22-289R and section 8 of L2/23-012.

Comments: We reviewed this proposal for one of the two scripts used to write the Tai Yo language of Vietnam. In the preliminary script proposal L2/22-152 and the later version L2/22-208, the script's name was "Yo Lai Tay." The name of the script was one outstanding issue for members of the SAH and PAG, since this name is inconsistent with other Tai scripts (Tai Viet, Tai Le, Tai Tham) and also because the inclusion of "Lai" 'script' in the name is not generally permitted.

- One of the co-authors communicated with the other authors and they have come to an agreement on the script name as Tai Yo, as long as the code chart includes the community's preferred name (such as Yo Lai Tay, Lai Tay, or Lay Tai).
- The PAG had asked why the fullwidth ASCII variants (U+FF10 … U+FF19) are recommended. Only these characters provide the expected behavior.
- The proposed sample character (page 13, script metadata) has been changed to a more unique shape.

The SAH has done an assessment and considers the Tai Yo script to be ready to encode.

The proposal was posted and forwarded to PAG, which caught a small error in the code point for the representative glyph. The code point has been corrected and is now posted as L2/22-289R.

V. SYMBOLS

9 Sanban Sign

Action: FYI with action to record

Document: L2/22-207R Updated proposal to encode the Sanban Sign for Chinese folk music and local operas – Eiso Chan

Recommendations: The following action is recommended:
Action Item for Debbie Anderson: Relay the feedback from section 9 of L2/23-012 to the proposal author of L2/22-207R.

Comments: We reviewed this revised proposal for SANBAN SIGN. On pages 5-6, the author addressed comments from the October 2022 Script Ad Hoc report .

The following summarizes the discussion on this proposal:

- After lengthy discussion, the group was not convinced on the need for a new character. It would be more expedient for users to make use of other characters that are already available to indicate a distinct signature in musical scores.

- The proposal author, who is a regular contributor to IRG work on Han unification, responded to earlier SAH suggestions that SAH did not properly consider the etymology of the sign. Given the specialized usage of this character, the group questioned whether criteria that are appropriate for Han were appropriate for this specialized character.

- On page 2 of the document, it says: "Earlier, the researchers only used the Han character [U+6563] (săn) at the same position of the scores to record this kind rhythm form." It appears that the concept of a non-metric time signature was originally indicated in the score with U+6563, a Han character, but it was later abbreviated to either U+5344, U+5EFE or U+30B5. For the glyphs shown on Table 0.1 on page 5, users can employ:
  (for glyph 1) U+5344 (or alternatively, U+3039 HANGZHOU NUMERAL 20)
  (for glyph 2) U+5EFE
  (for glyph 3) U+30B5 KATAKANA LETTER SA.

  Such usage would reflect likely past practice, which would have found appropriate glyphs in existing CJK fonts for such usage in scoring.

- Since the character appears in scoring, specialized software will be needed, so specialized font and software will be needed in any event.

- Having a new CJK unified ideograph that unifies the shapes of U+5344, U+5EFE, and U+30B5 with source glyph shapes overlapping already existing encoded characters would not seem appropriate. Is this the reason John Jenkins did not consider this as a CJK Unified Ideograph?

VI. PUBLIC REVIEW FEEDBACK

10 Legacy Computing Symbols

Action: For UTC discussion and decision

Document: L2/23-007 Comments on Public Review Issues (October 27, 2022 - January 5, 2023)

Date/Time: Tue Dec 6 07:20:49 CST 2022
Name: Charlotte Buff
Report Type: Other Document Submission
Opt Subject: L2/21-235R: Issues concerning two legacy computing characters

Regarding the “Symbols for Legacy Computing Supplement” repertoire that is slated for release in a future version of the standard:

• Is the proposed U+1CDB7 BLOCK OCTANT-3478 unifiable with existing U+1FB97 HEAVY HORIZONTAL FILL? U+1FB97 is annotated with the alias “upper middle and lower one quarter block”, which is the same block arrangement that results from octants 3, 4, 7, and 8 being set.

• Is the name LARGE TYPE PIECE RAISED UPPER RIGHT ARC for U+1CE35 correct? Looking at the other “upper arc” pieces (U+1CE1A and U+1CE24), shouldn’t its name be LARGE TYPE PIECE RAISED UPPER*LEFT* ARC instead?

Recommendations: We recommend the UTC make the following disposition:
SAH-UTC174-R8 Accept the name change for U+1CE35 from LARGE TYPE PIECE RAISED UPPER RIGHT ARC to LARGE TYPE PIECE RAISED UPPER LEFT ARC. (Reference: section 10 of L2/23-012)

Action Item for Markus Scherer and PAG: change the name of U+1CE35 from LARGE TYPE PIECE RAISED UPPER RIGHT ARC to LARGE TYPE PIECE RAISED UPPER LEFT ARC in UCD.txt. (Reference: section 10 of L2/23-012)
Action Item for Ken Whistler and the UTC: Update the Pipeline with the name change for U+1CE35 (Reference: section 10 of L2/23-012)
Action Item for Ken Whistler and the EdComm: Add a cross-reference between U+1CDB7 BLOCK OCTANT-3478 and U+1FB97 HEAVY HORIZONTAL FILL in the names list. (Reference: section 10 of L2/23-012)

Comments: We reviewed this feedback on the Legacy Computing additions, whose repertoire is slated for Unicode 16.0.

In our opinion, unifying U+1CDB7 BLOCK OCTANT-3478 with U+1FB97 HEAVY HORIZONTAL FILL is not warranted. It isn’t possible to tell, for example, if the aspect ratio is the same in fonts. A cross-reference, however, can be added in the names list.

We agree the name change for U+1CE35 LARGE TYPE PIECE RAISED UPPER RIGHT ARC is warranted. Doug Ewell, a key author in the original proposal, was agreeable to this name change.

VII. RECS ON 16.0 AND BEYOND

The SAH is agreeable to the Release Management plan that Unicode version 16.0 contain only scripts and characters that are in in the ISO/IEC 10646 CDAM2, which are indicated in green on the Pipeline as of January 2023.

The UTC should discuss the processing of characters and scripts. In this set of recommendations, the Script Ad Hoc only recommends the code points for characters and scripts be reserved, but does not yet recommend the UTC approve them.

The Script Ad Hoc also discussed criteria for posting documents.

B. DOCUMENTS NOT REQUIRING UTC ACTION (by script, in alphabetical order)

11 Arabic

Arabic Additions for Quranic orthographies

Document: L2/22-251 Review of some Arabic additions for Quranic orthographies – Salim Al Mandhari

Comments: This document provides comments on five symbols that were not proposed in L2/19-306, but which needed more research.

The information provided will be useful for future proposals, and we thank the author for his document.

It was noted that the names of the characters (with ETHHAR, EDGAHM, EKHFA, EQLAB, which derive from L2/15-329) would probably need to be re-named. Many Arabic names in Unicode are today based on a graphical description of the character.

The author has been thanked for his contribution.

---

12 Cyrillic

CYRILLIC CHE WITH HOOK

Documents: L2/22-280 Proposal to encode Cyrillic Che With Hook -- Manulov
L2/23-015 Comments on CYRILLIC CHE WITH HOOK's use in Khanty and Tofa – Anderson

Comments: We reviewed this request to add a new case pair for CYRILLIC LETTER CHE WITH HOOK. Feedback provided by Tapani Salminen and info from the Tofa (Tofalar) community (via Rustam Yusupov of Tofasatan.ru) in L2/23-015 suggests the character is not currently needed for representing text in Khanty and Tofa (Tofalar). Stronger justification is needed before adding this case pair, specifically clear evidence of contrast between CHE WITH DESCENDER vs. CHE WITH HOOK. Is the character being used widely and is it standardized?

A question to consider in such cases: For a small community, would adding a variant make the representation of the language better or worse? (If the form with hook is a glyph variant and could be handled by fonts, then adding the new case pair with a hook -- when a descender is also used for the same letter-- would add to confusion.)

Debbie Anderson has already forwarded the comments to the proposal author.

13 Khmer

Khmer Encoding Structure

Documents: L2/22-290 Khmer Encoding Structure (Nov. 2022) – Hosken et al.
L2/23-025 Khmer orthographic syllables – Lindenberg
Related document:
L2/23-038 Orthographic Syllable Structure in Core Spec – Pournader

Comments: Norbert Lindenberg introduced the two documents:

- L2/22-290, which is a revision of Martin Hosken's earlier 2021 document on the encoding structure of Khmer (L2/21-241).
- a draft of L2/23-025 Khmer orthographic syllables by Lindenberg, which includes proposed changes to the Core Spec, based on actionable portions of "Khmer Encoding Structure."

The Khmer orthographic syllables document seen by the SAH included BNF, which sparked a long discussion about including BNF in the Core Spec, a location that is likely to be read by implementers. The SAH generally recommended publishing the BNF information in a Unicode Technical Note first, as an easy and readily accessible first step. However, a UTN doesn't have the weight of the Core Spec or other specifications. The topic about how to handle orthographic syllabic structure of complex scripts in a future spec should be revisited in the January 2023 UTC meeting (see L2/23-038 which can act as the basis of discussion).

There was also lengthy discussion on backwards compatibility. Most experts agreed that we want to make sure existing Khmer text should render as it has been, instead of dotted circles starting to appear suddenly.

SAH members and other interested parties are encouraged to read L2/22-290.

Two Khmer Characters

Document: L2/23-018 Preliminary Proposal to Update Properties for Two Khmer Characters – Steven Loomis

Comments: We discussed an early version of this preliminary proposal, which was written in response to a comment from Makara SOK, who reported that U+17D4 KHMER SIGN KHAN did not cause a sentence break (though it should). Steven Loomis found the character had no Unicode property identifying that it as causing a sentence break. The annotation for U+17D4 does,however, state "functions as a full stop, period." Similarly, the annotation for U+17D5 KHMER SIGN BARIYOOSAN is "indicates the end of a section or text."

The following points were raised in discussion:

- We recommend the author ask experts if U+17D4 KHMER SIGN KHAN is also used in abbreviations.
- In PropList.txt, assign both characters U+17D4 KHMER SIGN KHAN and U+17D5 KHMER SIGN BARIYOOSAN as Sentence_Terminal (which in turn will cause an update to SentenceBreakProperty.txt).

The proposal has been sent to PAG for their review.

14 Latin

Initial Teaching Alphabet

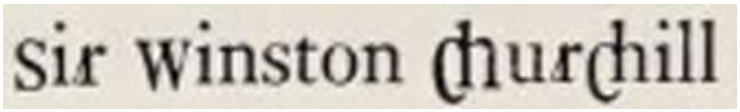
Document: L2/22-286 Proposal to Encode Latin characters for Initial Teaching Alphabet -- Manulov

Comments: We reviewed this preliminary proposal for the Initial Teaching Alphabet. The characters appeared in the exploratory proposal by Michael Everson in L2/08-428, but with two additional characters. The alphabet was designed to assist children in learning to read English. According to Wikipedia, it has fallen out of use, although there is a ITA foundation.

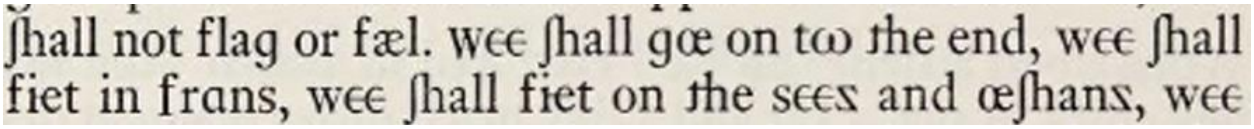
The following comments arose during discussion:

- ITA characters were proposed in L2/08-428, but in that exploratory proposal, two additional characters were included but are not proposed here: an alpha looking letter ("at" on page 8) and an omega without a loop ("oot" on page 8). How should those be represented?
- In our view, casing should not be introduced for these characters. Figure 13 shows drop caps being invented for capitalization as well as use of larger sized letters.

  Cf. "Sir Winston Churchill":

  Cf. "We" in the first line versus last word on the second line:

- If the proposed characters are encoded, only lowercase letters should be encoded.
- It was noted that the ITA represents one's group's pronunciation of English, though English pronunciation can vary considerably.

Note: Kirk Miller worked on an ITA proposal in 2020, and has offered to collaborate with Nikita Manulov on the ITA proposal.

Debbie Anderson has already forwarded the comments to the proposal author.

15 Maya Hieroglyphs (Classic)

Document: L2/23-020 Preliminary Classic Maya Syllabary and List of Classic Maya Signs - Vail

Comments: We reviewed this preliminary list of Classic Maya signs, based on work by Dr. Gabrielle Vail and team. The list comprises 450 signs (both Syllabograms and Logograms) from Classic Maya materials from the Central, Southern and Western regions. The work complements the list of Codical Maya signs in L2/20-248 by Carlos Pallán.

The characters are not being formally proposed, but this document reports on current progress. The current goal for encoding Maya Hieroglyphs is to propose the Codical signs first, since they are from a small, well-defined corpora (i.e., three codices), and then add Classic signs, whose written materials number in the thousands. For those signs that occur in both periods but differ in glyph shape, a font can be used to display the selected period (i.e., a Classic font vs. Codical font), much as is done for CJK.

According to Andrew Glass, there do not appear to be structural differences between the Codical and Classic signs and glyph blocks. The reading order is well-established, and the encoding order will follow the Top-to-Bottom, Left-to-Right order.

Comments:

- The names should be rationalized, especially if the naming conventions vary between Codical and Classic (and different catalog conventions). Note that Unicode character names cannot have apostrophes. (Apostrophes are used to denote glottal stops in Mayan.)
- A structural analysis is needed. For example, 261 appears to be the left-hand piece of 260. How will these be represented in the encoding?

260	K'UH		AMCv2	god, holy
261	K'UH		AMCv3	god; holy

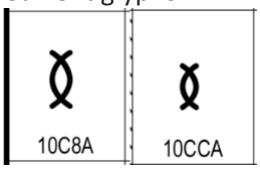
---

16 Old Hungarian

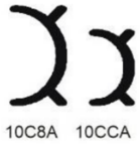
Document: L2/22-285 Discussion and proposal of a compromise solution (discussion of L2/21-115) – Viktor Kovács and „Természetesen" Association, Solt

Comments: We briefly reviewed this lengthy document that discusses topics raised in L2/21-115 and L2/21-246, and gives background on the encoding of Old Hungarian. The author requested the glyphs be changed for U+10C8A OLD HUNGARIAN CAPITAL LETTER CLOSE E and U+10CCA OLD HUNGARIAN SMALL LETTER CLOSE E, a request shared by the author of L2/21-115.

Current glyphs:



Proposed new glyphs:



After an off-line discussion between Michael Everson and the author of this document (L2/22-285), the author withdrew his request. Michael Everson's rationale for not changing the glyphs was:

> We encoded CLOSE E (Ë) according to the evidence of the primary sources. See Figure 3 in [L2/]12-168. You will see H drawn curvy (though in other sources it is angular), Ë drawn curvy, and E, which looks like a) with two diagonals piercing the).
> What [L2/21-115 has] done is try to invent a new glyph for Ë, looking like) with two diagonals NOT piercing the), but that's not what we find in the historical sources, and that is what we encoded.
>
> [The authors of L2/21-115] have moved glyph ambiguity from Ë/H to Ë/E (and É!) and there's no justification for that. Yes, Ë/H look very similar in the Rudimenta, but that's the reality.
> In principle they could make their own fonts with their new glyph but Ë, H, and E are distinctive in the Rudimenta.

In sum, no change is requested of the UTC.

---

17 Ranjana

Documents: L2/23-028 Preliminary proposal to encode Ranjana in Unicode – Pandey
Related documents:
L2/16-015 Towards an encoding for the Ranjana and Lantsa scripts - Pandey
L2/13-243 "Proposal to Encode Ranjana Script in ISO/IEC 10646" (Dev Dass Manandhar, Samir Karmacharya and Bishnu Chitrakar)
L2/09-192 "Preliminary proposal for encoding the Rañjana script in the SMP of the UCS" (Michael Everson)

Comments: We reviewed this preliminary proposal. The proposal unifies Ranjana with Lantsa, a variant used in Tibet, which also includes a variety of regional styles. One might compare Ranjana and Lantsa, both calligraphic scripts, with the variety of calligraphic forms that can be seen in varieties of blackletter styles in the Latin script. (A decision on whether Vartu should also be unified will need to be made.)

The following comments were made during discussion:

- At the beginning of the document, add content on the benefits of unifying Ranjana and Lantsa, showing sources and making the case that they are more similar than different.
- Section 4: identify the rows with names (Ranjana / Lantsa).
- Section 4.3 (pages 12-13): the long u glyph is incorrect.
- Use Myanmar as the model for naming characters, instead of Tibetan or Zanabazar Square. Rename FINAL RA and FINAL YA to be MEDIAL RA and MEDIAL YA
- Add examples from manuscripts.
- Show most frequent conjuncts, comparing Ranjana vs. Lantsa.

These notes above have been sent to the proposal author.

---

18 Symbols

18a Universal Feed Symbol

Document: L2/22-283 Proposal to encode Universal Feed Symbol – Pierre Marshall

Comments: We reviewed this proposal to encode a “universal feed symbol.”

The following comments were made during discussion:

- The proposal does not provide examples in running text, only a description of the icon.
- In our view, this is an icon, not a text element. The sign doesn’t change size depending upon the size surrounding the text. When examples of the sign are provided demonstrating the use of the sign as a text element, the proposal could be taken up again.
- The reasoning for not accepting this character is comparable to that recorded for the External Link Sign (from the Archive of Notices of Non-Approval):

> Proposals to encode a character for the "external link sign", which is often seen as a graphic element indicating a link to a document located external to the website where the page using the external link sign resides. (See L2/06-268, L2/12-143, L2/12-169.)
>
> **Disposition**: The UTC rejected the proposals to add "external link sign", most recently in L2/12-169. It is unclear that the entity in question is actually an element of plain text, given the inevitable connection to its function in linking to other documents, and thus its coexistence with markup for links. Furthermore, the existing widespread practice of representing this sign on web pages using images (often specified via CSS styles) would be unlikely to benefit from attempting to encode a character for this image. (This notice of non-approval should not be construed as precluding alternate proposals which might propose encoding a simple shape-based symbol or symbols similar in appearance to the images used for external link signs, should an appropriate plain-text argument for the need to encode such a simple graphic symbol be forthcoming.) References to UTC Minutes: [131-C26], May 10, 2012.

Debbie Anderson will forward the comments to the proposal author.

18b Transnistrian Ruble

Document: L2/23-022 Proposal to encode Transnistrian Ruble Sign - Manulov

Comments: We reviewed this request for a currency symbol**.**

The following comments were made:

- There is no evidence of the proposed character being used as a text element, the samples all show a logo. Please provide plain text samples with the symbol inline. Examples could come from handwritten signs.
- Provide references to [1], [2], and [9]. (They appear to come from English and Russian Wikipedia pages.)
- The sign was earlier mentioned in L2/19-291, so this should be cited in the proposal.
- Contact should be made with the user community, in particular communication with the bank.

Debbie Anderson will forward the comments to the author of the proposal

---

19 Tai Viet additions for Jinping Dai

Document: L2/23-023 Proposal to encode additional Tai Viet characters for the Jinping Dai – Kushim Jiang
Related documents:
L2/22-210 Final proposal to add 22 characters for Tai Don writing systems – Kushim Jiang
L2/22-273 Response to Eiso Chan's Comments on the Tai Don or Tai Khao encoding – Jim Brase
L2/22-272 Comments on the Tai Don or Tai Khao encoding – Eiso Chan
L2/22-099 Comments on Updated proposal to encode the Tai Don script in UCS - Brase (Nov. 2021)
(See SAH comments in L2/22-068, page 16f.)
L2/22-098 Updated proposal to encode the Tai Don script in UCS -- Kushim Jiang (Sept. 2021 version)
L2/20-208 A response to Kushim Jiang, "Preliminary proposal to encode the Tai Khao script in UCS" -- Jim Brase
L2/20-207 Preliminary proposal to encode the Tai Khao script in UCS -- Kushim Jiang
(See SAH comments in L2/21-016R, page 30)
L2/08-217 Writing Tai Don: Additional characters needed for the Tai Viet script – Brase
L2/06-041.Unified Tai Script for Unicode. -- Ngô Trung Việt, Jim Brase.

Comments: We reviewed this proposal for Tai Don, which incorporated some changes made by Kushim JIANG based on comments from Liang Hai and Lawrence Wolf-Sonkin.

- Feedback for the author was to approach the script based on the shape of the graphic elements, rather than by the sound.
- Check whether specific letters of Tai Don's regional variants should be unified with Tai Viet in the way the Tai Viet proposers originally recommended. Verify that point before examining how this specific region variant of Tai Don (Jinping Dai) should be unified with other Tai Don variants and Tai Viet.

20 Other script and character topics in process

The following script and character topics are in process:

- Bima
- Incung
- Feedback on Ra + vocalic vowels in Kannada and Bengali scripts
- Seal
- Nagari