TO: UTC
FROM: Deborah Anderson, Ken Whistler, Roozbeh Pournader, and Peter Constable[1]
SUBJECT: Recommendations to UTC #173 November 2022 on Script Proposals
DATE: October 28, 2022

The Script Ad Hoc group met on August 19, September 16, and October 14, 2022, 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 Simon Cozens, Craig Cummings, Lorna Evans, Andrew Glass, Manish Goregaokar, Liang Hai, Ned Holbrook, Frank van de Kasteelen, Jan Kučera, Norbert Lindenberg, Kamal Mansour, Harald Tveiten, 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, Norbert Lindenberg, Jan Kučera, and Ben Yang.

A. PROPOSALS REQUIRING UTC ACTION

I. EUROPEAN SCRIPTS

1 Latin
1a CAPITAL LETTER F WITH STROKE

Action: For UTC discussion and decision

Document: L2/22-196 On LATIN CAPITAL LETTER F WITH STROKE --- Moyogo Jacquerye

Recommendations: We recommend the UTC make the following disposition:
SAH-UTC173-R1: Accept the glyph change for U+A798 LATIN CAPITAL LETTER F WITH STROKE, as documented in L2/22-196, for Unicode 15.1.

The following actions are recommended:
Action Item for Debbie Anderson and the Script Ad Hoc: Create a glyph erratum for U+A798 LATIN CAPITAL LETTER F WITH STROKE, based on L2/22-196.
Action Item for Ken Whistler and the EdComm: Modify the annotations for U+2043, U+A798, and U+1799 as suggested in L2/22-196.

Comments: We reviewed this proposal which recommended the representative glyph for U+A798 LATIN CAPITAL LETTER F WITH STROKE be changed, based on research provided by the proposal author. The current glyph was based on one sample from a 1930 document by Diedrich Westermann on the Ewe language for a voiceless bilabial fricative. However, the current glyph differs from the shape used in 1930 and the proposed glyph was used in a Bible publication. The new glyph also better matches the character's name. Michael Everson was agreeable to the change.

In our view, the proposed change is reasonable.

(Michel Suignard reports he does not need a font.)

1b SMALL LETTER Y WITH SHORT RIGHT LEG

Action: For UTC discussion and decision

Document: L2/22-199 On LATIN SMALL LETTER Y WITH SHORT RIGHT LEG -- Moyogo Jacquerye

Recommendations: We recommend the UTC make the following disposition:
SAH-UTC173-R2: Accept the glyph change for U+AB5A LATIN SMALL LETTER Y WITH SHORT RIGHT LEG, as documented in L2/22-199, for Unicode 15.1

The following action is recommended:
Action Item for Debbie Anderson and the Script Ad Hoc: Create a glyph erratum for U+AB5A LATIN SMALL LETTER Y WITH SHORT RIGHT LEG, as described in L2/22-199.

Comments: We reviewed this request to change the glyph for U+AB5A LATIN SMALL LETTER Y WITH SHORT RIGHT LEG, which was originally proposed as part of the Teuthonista set of characters (L2/11-202). The 2011 Teuthonista proposal showed the glyph in its italic form. Research by the proposal author Denis Moyogo Jacquerye identified the shape of the glyph in its roman form and proposed changing the glyph to the roman shape. We agreed with the request. Michael Everson also agreed to the glyph change.

(Michel Suignard reports he does not need a font.)

1c SMALL LETTER BLACKLETTER O WITH STROKE

Action: FYI with action to record

Document: L2/22-198 On LATIN SMALL LETTER BLACKLETTER O WITH STROKE --- Moyogo Jacquerye

Recommendations: The following action is recommended:
Action Item for Ken Whistler and the EdComm: Add an annotation to U+AB3E LATIN SMALL LETTER BLACKLETTER O, clarifying the source of the character and its preferred usage for Unicode 15.1. (Reference: L2/22-198).

Comments: We discussed this proposal to change the glyph for U+AB3E LATIN SMALL LETTER BLACKLETTER O WITH STROKE. The request is based on additional research by the proposal author, who provided samples of the glyphs at higher resolution than those used in the original proposal for Teuthonista (L2/11-202).

The current Unicode representative glyph is in blackletter style (as also reflected in its name), but the samples with better resolution are clearly not blackletter. The author also noted that the symbol was not limited to German dialectology, but was used as a general phonetic symbol, so he proposed an annotation, "not used in German dialectology." The symbol appears to be a regular slashed o (i.e., U+00F8 LATIN SMALL LETTER O WITH STROKE).

The author spoke with Michael Everson, who co-wrote the Teuthonista proposal, and Everson noted that none of the co-authors, who are German dialectologists, have objected to the character, its glyph, or its name to date, so no glyph change should be made.

II. AFRICA

2 Egyptian Hieroglyphs

Action: FYI with action to record

Document: L2/22-265 Considerations concerning Egyptian Hieroglyphs extension – Suignard

Recommendations: The following action is recommended:
Action Item for Debbie Anderson, Andrew Glass, and EdComm: Update section 11.4 of the Core Spec (p. 449) for Unicode 16.0, clarifying that use of EGYPTIAN HIEROGLYPH INSERT AT MIDDLE is the approach for representing single groups, and the enclosure model should be used for representing more than one group. (Reference: L2/22-265 and section 2 or L2/22-248)

Comments: We discussed this document by Michel Suignard which raised a number of issues relating to his work on the Egyptian Hieroglyph extensions proposal.

Background on this document
A group of Egyptologists from Europe have been working with Michel Suignard, Andrew Glass, and Debbie Anderson since November 2020. A meeting with Egyptologists in January 2023 in Berlin will offer an opportunity for experts to provide their input on the topics below.

Productive font rendering
Andrew Glass has been working on a proto-type font, which has demonstrated possibilities (and limits) of present-day font rendering. Egyptologists understand that absolute positioning is not a hard requirement, and the document L2/21-248 acknowledged the role of JSesh for texts that require fine palaeographic representation. Andrew Glass mentioned that he can update the font as needed to suit the community's needs.

In general, the Egyptologists are in favor of a productive mechanism over atomic encoding, when applicable. (Some guidelines on when atomic encoding is recommended are contained in this section and in the section on "Token versus abstract signs and encoding strategy" on page 5.)

Compound signs and their definition in Unicode code points
The SAH agreed that sequences need to be documented for typical compound signs found in collections like those represented by Hieroglyphica and JSesh, since there could be multiple sequences to define a compound. Michel Suignard also recommends documentation of sequences in cases where overly/insertion/arrangement is identified as a compound sign by Egyptologists.

Compound signs that can use either enclosure or insertion
The SAH agreed that text in the Core Spec needs to be clarified that the middle insertion is the preferred model for inserting a single group of signs. (See Action Item captured above.)

Definition of a container, or can a container be defined as a sequence of code points
Andrew Glass noted that there is sparse documentation on mirroring and other controls, but he has updated the cluster definition in the Universal Shaping Engine.

Additional consideration concerning container sizes
The SAH agreed that bases that are not already in the repertoire should be proposed.

For container sizes, the font will take care of necessary sizing, so there is no need to redraw glyphs to be wider or narrower (or to have multiple versions encoded based on size).

Token versus abstract signs and encoding strategy
In the view of the SAH, the following examples should be handled at the font level, and not at the encoding level, unless they are contrasted in the same doc with different semantics. In short: a different font should be used to get the different variants (unless contrasted).

Some examples (X4, O33):
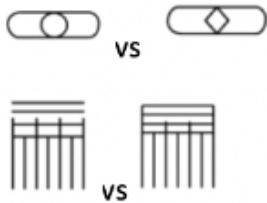
Transliteration
The Gardiner transliteration system is acceptable to the SAH. It was noted that with the Core Spec moving to HTML, any transliteration system will work, though there are restrictions on the names list. Consistency would be preferred, so as to avoid "ransom note" effect as much as possible.

Character names
The options for character names are listed below:

Naming convention	Syntax	Examples EGYPTIAN HIEROGLYPH [EXTENDED]
Unikemet extension	[A-IKZ, AA, NU, NL]ddd	A0071
IFAO based	[A-IKZ,AA]-dd-ddd	A-01-003
Code point	hhhhh	13460
PUA	hhhhh	F0002

The goal is to get the characters encoded with stable names.

The SAH considered use of PUA code points in character names to be an undesirable option. We understand that they may be the stable identifiers in current use in the database used for preparation of this proposal, but PUA code points are by definition private use and the UTC has never used them in assigning character names. If newly encoded hieroglyphs are given names based on code points rather than well-established catalog identifiers, then it is far better to use the actual assigned code point for that, rather than a PUA code point. The convention of using the assigned code point in a character name, as for CJK UNIFIED IDEOGRAPH-4E00 or TANGUT IDEOGRAPH-17000 is well established practice for the Unicode Standard.

The Unikemet extension is already inconsistent, so it is probably not a good candidate.

An issue with the IFAO scheme is if people will want to insert characters and re-catalog code points in the future.

In the view of the SAH, the use of the code point as the name would be the best option. Of course, the meeting in Berlin in January will provide an opportunity for the experts there to weigh in with their preference.

Ancillary information and what to put in the code chart
To reduce extraneous information in the names list – which should not be considered a place for entering detailed information -- a separate database is recommended, though this will involve more tooling and likely take additional time. A tab-delimited text file with limited basic information will likely accompany the first Unicode proposal for the Egyptian Hieroglyph extensions.

It was mentioned that the signs in the code charts may not be the most common one. To represent text from a certain location, such as Dendara, a specialized font should be used. Variants can be documented.

III. MIDDLE EASTERN SCRIPTS

3 Arabic

Action: FYI with action to record

Document: L2/22-221 On ARABIC LETTER KAF WITH DOT BELOW – Febri Muhammad Nasrullah

Recommendations: The following action is recommended:
Action Item for Ken Whistler and the EdComm: Add an annotation to U+08B4 ARABIC LETTER KAF WITH DOT BELOW based on L2/22-221, with the appropriate cross-references for Unicode 15.1.

Comments: We reviewed this request to add an annotation to U+08B4 ARABIC LETTER KAF WITH DOT BELOW, noting the character is used in Pegon. The evidence presented confirms the character is used in Pegon. There was some discussion about the wording "also used in," but the sub-head already specifies Arwi and hence provides context.

IV. SOUTH AND CENTRAL ASIAN SCRIPTS
4 Malayalam and Sinhala

Action: For UTC discussion and decision

Documents: L2/22-214 Use of ARDHAVISARGA in Malayalam and Sinhala – Srinidhi and Sridatta
Note: The following document was received after the SAH had met and was not yet discussed:
L2/22-269 Response to L2/22-214, Use of ARDHAVISARGA in Malayalam and Sinhala
--Gihan Dias

Recommendations: The following action is recommended:
Action Item for Roozbeh Pournader and the UTC: Add the values {Sinh, Mlym} to U+1CF2 in ScriptExtensions.txt for Unicode 15.1. (Reference: L2/22-214)

Comments: The proposal authors provide examples of U+1CF2 VEDIC SIGN ARDHAVISARGA in Malayalam and Sinhala, and request the values be added to ScriptExtensions.txt.

No new characters were being requested. The addition of scripts to ScriptExtensions.txt only documents that the characters have been used in the scripts listed. The evidence provided in L2/22-214 confirms the use of U+1CF2.

5 Ol Onal

Action: For UTC discussion and decision

Document: L2/22-151R Proposal to encode the Ol Onal script – Mandal and Kučera
Note: A few additional comments from Properties and Algorithms group (PAG) were received after the SAH had met. The comments are contained in PAG recommendations, L2/22-244 and have been addressed in L2/22-151R.

Recommendations: We recommend the UTC make the following disposition:
SAH-UTC173-R3: Accept 44 Ol Onal characters at U+1E5D0..U+1E5FF, with names, properties, and glyphs as given in L2/22-197, in a new Ol Onal block that extends from U+1E5D0..U+1E5FF for a future version of the standard. (Reference: section 5 of L2/22-248)

The following actions are recommended:
Action Item for Ken Whistler and UTC: Update the Pipeline to include 44 Ol Onal characters in a new Ol Onal block that extends from U+1E5D0..U+1E5FF. (Reference: section 5 of L2/22-248 and L2/22-197)
Action Item for Jan Kučera (and EdComm) and Biswajit Mandal: Send the Ol Onal font to Michel Suignard. (Reference: section 5 of L2/22-248)

Background on the script:
Ol Onal is used to write Bhumij, a Munda language of the Austroasiatic family that is spoken primarily in the Indian states of Jharkhand, West Bengal, and Odisha. The language is commonly written in the Odia, Bengali, Devanagari, and Latin scripts. The script was invented between 1981 and 1992. Although originally a casing script, only uppercase is used today. There is no evidence of lowercase being used. This proposal has been reviewed by the SAH a number of times and was revised based on comments from the SAH. Several changes have been incorporated as recommended by the SAH in L2/22-128.

In our view, the script is mature.

Note: After the SAH met, additional comments from PAG were made (see L2/22-244); changes were subsequently made to the proposal and been incorporated in L2/22-151R.

6 Sharada

Action: FYI with action to record

Document: L2/22-213 Use of Fractions and Rupee Mark in Sharada script – Srinidhi and Sridatta

Recommendations: The following action is recommended:
Action Item for Roozbeh Pournader and the UTC: Add the value {Shrd} to U+A830..U+A835, U+A838 in ScriptExtensions.txt for Unicode 15.1. (Reference: L2/22-213)

Comments: We reviewed this request to add Sharada to ScriptExtensions.txt for 7 different North Indic characters (U+A830..U+A835, U+A838), based on evidence provided in the documents.

7 Tulu-Tigalari

Action: For UTC discussion and decision

Document: L2/22-260 Proposal to encode three characters in Tulu-Tigalari -- Srinidhi and Sridatta
Related documents:
L2/17-182 Comments on encoding the Tigalari script (Section 5: New characters which require encoding)
L2/22-031 Updated proposal to encode the Tulu-Tigalari script in Unicode

Recommendations: We recommend the UTC make the following disposition:
SAH-UTC173-R4: Accept U+113D2 TULU-TIGALARI GEMINATION MARK and U+113D3 TULU-TIGALARI SIGN PLUTA, with names, glyphs, and properties as given in L2/22-260, for a future version of the standard. (Reference: section 7 of L2/22-248)

The following actions are recommended:
Action Item for Ken Whistler and UTC: Update the Pipeline to include U+113D2 TULU-TIGALARI GEMINATION MARK and U+113D3 TULU-TIGALARI SIGN PLUTA (Reference: section 7 of L2/22-248 and L2/22-260)
Action Item for Debbie Anderson and EdComm: Send the Tulu-Tigalari font to Michel Suignard. (Reference: section 7 of L2/22-248)

Comments: We reviewed this proposal to add three Tulu-Tigalari characters:
TULU-TIGALARI GEMINATION MARK
TULU-TIGALARI SIGN PLUTA
TULU-TIGALARI LETTER ARCHAIC II.

Tulu-Tigalari was approved for encoding at the Jan 2022 UTC meeting (#170) and is already included in CDAM2 (see repertoire in L2/22-134).

Evidence of all three proposed characters is provided. It was noted that the document's authors had already identified the SIGN PLUTA and the GEMINATION MARK as potential characters in their comments on "Tigalari" in L2/17-182. A space had been reserved for GEMINATION MARK in the Tulu-Tigalari script proposal (L2/22-031) at the proposed code point (U+113D2).

Notes from discussion:

- The Tulu-Tigalari script proposal authors, Vaishnavi Murthy and Vinodh Rajan, support the proposed characters.
- A question was posed whether the SIGN PLUTA was part of the orthographic syllable, but Vinodh Rajan, one of the proposal authors for the Tulu-Tigalari script proposal (L2/22-031), replied it was not.
- Regarding ARCHAIC II: The handling of independent vowel letters in Tulu-Tigalari was discussed at length in SAH meetings. The outcome of the SAH discussion on TULU-TIGALARI LETTER II is reflected in the Tulu-Tigalari script proposal (L2/22-031, page 34) in which TULU-TIGALARI LETTER II was proposed with a canonical decomposition:

11383   ꯍ   TULU-TIGALARI LETTER II

≡ 11382 ꯍ   113C9 ꯍ

and text stating that rarer variants, shown below, should be handled as sequences.

ꯍ , ꯍ

In footnote 23 of the Tulu-Tigalari proposal, the authors state, "The rare form of Tulu-Tigalari II: °O° is being examined by the authors and not included in this proposal as it can be updated at a later date."

Because "ARCHAIC II" raised architectural issues, we recommend the UTC hold off encoding it at this time. However, both the GEMINATION MARK and the SIGN PLUTA were considered reasonable by the SAH.

V. SCRIPTS FROM SOUTHEAST ASIA, INDONESIA, AND OCEANIA
8 Kawi

Action: For UTC discussion and decision

Document: L2/22-236 Proposal to encode KAWI SIGN NUKTA -- Febri Muhammad Nasrullah
Related documents:
L2/22-093 Review of Kawi code chart of Unicode 15 Alpha (PRI 442) – Kinaṇṭi, et al. (Note: The proposed character was mentioned on page 3 of this document in the section, "POSSIBILITY OF OTHER UNPROPOSED CHARACTERS.")
Note: L2/22-250 Canonical combining class for nukta characters – Lindenberg was submitted after the SAH meeting and was not reviewed by the SAH.

Recommendations: We recommend the UTC make the following disposition:
SAH-UTC173-R5: Accept U+11F5A KAWI SIGN NUKTA, with name, glyph, and properties as given in L2/22-236 for a future version of the standard. (Reference: section 8 of L2/22-248)

The following actions are recommended:

Action Item for Ken Whistler and UTC: Update the Pipeline to include U+11F5A KAWI SIGN NUKTA. (Reference: section 8 of L2/22-248 and L2/22-236)
Action Item for Norbert Lindenberg and EdComm: Send the Kawi font for nukta to Michel Suignard. (Reference: section 8 of L2/22-248)

Comments: We reviewed this proposal to add KAWI SIGN NUKTA, which was not in the original Kawi proposal (L2/20-284R).

Note: The following two comments were based on an earlier version of L2/22-236 that was seen by the Script Ad Hoc. The version now posted in the document register has already made changes based on the comments in the following two bullets.

The following remarks were made during discussion:
- The proposed glyph has three dots above it, but the example on page 3 has four dots above it, and the example on the bottom of page 2 has two dots and a slanted line. The proposal should explain the variation.
- The proposed ccc value is 7, which is traditional for nuktas. However, Norbert Lindenberg questioned this assumption and recommended ccc=0, because the assignment of ccc=7 has not been consistently applied and doesn't appear to serve any clear purpose. During discussion there was no objection to assigning ccc=0 for this character.

  Norbert Lindenberg agreed to write a one-page document on the canonical combining class of nuktas. This has been done and is posted L2/22-250. It has been sent to PAG for review.

9 Yo Lai Tay

Action: For UTC discussion and decision

Document: L2/22-208 Final Proposal to encode the Yo Lai Tay Script – Viet Khoi Nguyen et al.
Note: A few additional comments from the Properties and Algorithms Group (PAG) were received after the SAH had met. The comments are contained in PAG recommendations L2/22-244.
.

Recommendations: We recommend the UTC make the following disposition:
SAH-UTC173-R6: Accept 55 Yo Lai Tay characters at U+1E6C0..U+1E6FF, with names, glyphs, and properties as given in L2/22-208, in a new Yo Lai Tay block that extends from U+1E6C0..U+1E6FF for a future version of the standard. (Reference: section 9 of L2/22-248)

The following actions are recommended:
Action Item for Ken Whistler and UTC: Update the Pipeline to include 55 Yo Lai Tay characters in a new Yo Lai Tay block that extends from U+1E6C0..U+1E6FF. (Reference: section 9 of L2/22-248 and L2/22-208)
Action Item for Debbie Anderson and EdComm: Send the Yo Lai Tay font to Michel Suignard. (Reference: section 9 of L2/22-248)

Comments: We reviewed this proposal. Changes were made based on comments from last Script Ad Hoc meeting. In our view, the proposal is deemed mature, so we recommend the script for encoding.

VI. PUBLIC REVIEW FEEDBACK
10 Syloti Nagri

Action: FYI with action to record

Document: L2/22-243 Comments on Public Review Issues
      Date/Time: Wed Sep 14 09:07:32 CDT 2022
      Name: David Corbett
      Report Type: Other Document Submission
      Opt Subject: Relative order of U+A802 and U+A823
      L2/02-388 says "The correct encoded representation for this diphthong follows the phonological ordering: < Syloti Nagri dependent a, Syloti Nagri dvisvara sign >". U+A802 SYLOTI NAGRI SIGN DVISVARA has Indic_Positional_Category=Top. U+A823 SYLOTI NAGRI VOWEL SIGN A has Indic_Positional_Category=Right. The usual order of Indic vowel signs in Unicode is left, top, bottom, right. Therefore, it seems like U+A802 should actually precede U+A823, but on the other hand Unicode often orders marks phonetically, so maybe U+A823 should precede U+A802.
      Which order should Syloti Nagri text use? The standard should explicitly explain which order to use.

Recommendation: The following action is recommended:
Action Item for Debbie Anderson: Forward the comments in section 10 of L2/22-248 to David Corbett in response to his feedback [Wed Sep 14 09:07:32 CDT 2022], contained in L2/22-243.

Comments: Norbert Lindenberg noted that the block intro on Syloti Nagri in the Core Spec doesn't include many of the details that appeared in the 2002 proposal (L2/02-388).

Syloti Nagri was added in Unicode version 4.1, which had no update to the Core Spec. The Core Spec section on Syloti Nagri was added in Unicode version 5.0, a version that included many new script additions, and which made it difficult for editors to provide content on the new additions. As a result, the Core Spec is under-specified. Note, too, that Indic Syllabic Categories were not added until Unicode 6.0.

Further research is needed on this topic, and we invite the feedback author to do the research and write a document recommending specific changes to the Core Spec.

VII. Summary of Script Ad Hoc and Properties and Algorithms Group Joint Meeting

A summary to the joint meeting between members of the Script Ad Hoc and the Properties and Algorithms Group is posted as L2/22-266. Discussion covered (a) the process of handling of properties in proposals, (b) the addition of modifier letters for phonetic transcriptional systems, and (c) composites (in particular, the encoding models for composites when equivalent combining sequences exist).

VIII. RECS ON 15.1 and 16.0

The SAH is agreeable to a plan for 15.1 that no new scripts be added, and no changes are made to the Core Spec. In our view, none of the approved characters or scripts are urgent. We understand that version 16.0 however would include new scripts and character additions.

The SAH also agreed to the addition of 5 IDS characters in 15.1. For 15.1, the SAH recommends (in this document) a few annotation additions and glyph changes. The SAH is also supportive of changes related to line breaking at orthographic syllables as discussed in document L2/22-080R2 by Norbert Lindenberg (with changes affecting UAX #14).

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

11 Arabic

Document: L2/22-261 Proposal for extensions to the Arabic block -- Mohamed Naji

Comments: We reviewed this proposal for Arabic extensions which contains three parts:
Part 1 asks for characters to be able to write scientific text, specifically 48 "closing forms" that connect to a preceding character, have the same properties as the final form of a character, and have glyphs that are distinguished from the isolate forms by a small tic-mark.
Part 2: proposes 10 new digits based on the idea of al-Khwarizmi, with Arabic script property and bidi class Arabic Number. The rationale cited is to help resolve problems in scientific text and to handle phone numbers, where U+0030..U+0039 are used.
Part 3: proposes a new sign drawn by the prophet Mohamed (see glyph on p. 19).

The following remarks were made during discussion:
General:
- The proposal should provide documented use of existing characters that need to be encoded, rather than creating new inventions.
- The Arabic glyphs should be in the style of the current code charts.

Part 1: In essence, this request is for a brand-new form of markup indicating where adjacent words start and end for scientific units and annotation. However, the problem is already solvable by inserting a ZWNJ between the two final letters of a word, an approach that is used universally for the rendering of Persian compound words. An alternative approach is to use a particular font or stylistic layout option (e.g., collapsing regular word spaces and triggering a variant letter shape glyph substitution).

Part 2: It was noted that Arabic numerals, Indian, and "European" numerals all run left to right. The problem regarding telephone numbers may be because it involves a sequence of numbers, rather than a sequence of digits (see https://r12a.github.io/scripts/arabic/arb.html#expressions). The different dividers may play a role. It may be that the proposer's computing environment could also play a part. One member of the Script Ad Hoc is collecting additional information, which is needed before a response can be provided. In our view, however, creating new novel digits is not the best approach.

Part 3: Provide examples of the proposed character from texts, citing the sources.

Debbie Anderson will forward the comments to the author of L2/22-261.

---

12 Badugu

Document: L2/22-220 INTRODUCING THE BADUGU SCRIPT – Biswajit Mandal

Comments: We briefly reviewed this "introducing" document, which is very preliminary. We thank the author for his work and encourage him to continue working on the script proposal.

Debbie Anderson has already forwarded comments to the proposal author.

---

13 Bengali

Document: L2/22-268 Proposal to Encode Alternate BA for the Bengali Language – Rajan and Chakraborty

Comments: We reviewed this proposal to encode a new character BENGALI LETTER ALTERNATE BA. The letter is proposed in order to represent the Brahmic letter VA which appears in Sanskrit-origin words. In current mainstream Bengali orthography, both VA and BA are spelled with U+09AC BENGALI LETTER BA (since Bengali is missing /v/). However, representing Sanskrit texts is problematic, since VA and BA require a clear orthographic distinction. The proposed BA letter follows normal BENGALI LETTER BA behavior.

Assamese and Oriya innovated new letters to differentiate the two phonemes, by modifying the glyph for BA: Assamese has U+09F1 (BA with lower diagonal) and Oriya has U+0B35 (BA with nukta). Bengali publications innovated a BA letter by inserting an internal diagonal into the glyph for BA, a glyph that is

identical to U+09F0 BENGALI LETTER RA WITH MIDDLE DIAGONAL, which was encoded specifically for Assamese.

Various options were discussed. Rather than encode a new character, the SAH recommended use of the sequence <09AC BENGALI LETTER BA, 09BC BENGALI SIGN NUKTA>, which would be available to users in a special font. This was deemed preferable to use of U+09F0 BENGALI LETTER RA WITH MIDDLE DIAGONAL with special shaping for Sanskrit texts.

Vinodh Rajan will revise his document based on the comments above.

---

14 Brahmi

Document: L2/22-219 Proposal to encode MINIMIZING VIRAMA in Brahmi script – Srinidhi and Sridatta

Comments: We reviewed this document proposing a "minimizing virama" for Brahmi.
Brahmi has two viramas: BRAHMI VIRAMA (U+11046), which is used as a vowel-killer and to form conjuncts, and OLD TAMIL VIRAMA (U+11070), which is used only to kill vowels (and does not form conjuncts). The proposed character has no distinct glyph, but acts to create a small, unligated vowelless consonant appearing to the lower right-hand side of the preceding consonant.

- We discussed different encoding model options. The group agreed that four atomic characters should be proposed for vowel-less final M, N, K and T. The set of characters is apparently very limited, so separately encoding the characters was deemed most acceptable. Other options considered included a new virama character (or special modifier) that creates a reduced form of a consonant or a model with ZWJ and Virama, as in Oriya (requiring no new characters).
- The characters' names should reflect their relationship to the normal MA, NA, KA, and TA characters.
- Can the authors confirm that these are the only such characters?

Debbie Anderson will forward comments to proposal authors, cc'ing Jan Kučera.

---

15 Chisoi

Documents:
L2/22-218R Proposal to Encode Chisoi in the Universal Character Set – Mandal
Related document:
L2/21-183 Proposal to Encode Chisoi in the Universal Character Set – Mandal
Comments on above contained on pp. 12-13 of L2/21-174

Comments: We reviewed this revised proposal for Chisoi, dated Sept. 11, 2022, which accommodated earlier comments from the SAH. One change has not yet been made in the version seen by the Script Ad Hoc: In Section 3.4: When a consonant is modified by both a SISO and JARAHA, the sequence should be in visual order, not logical order. (i.e., the order should be <C, SISO, JARAHA>).

There was an outstanding question about collation, since it does not match the dictionary order in figure 19(e). However, it appears the dictionary should not be taken as a reliable source, so the proposed collation order is probably acceptable.

Jan Kučera has forwarded comments to the proposal author.

---

16 Cyrillic
16a Old Caucasian Orthographies

Document: L2/22-262 Proposal to encode 23 Cyrillic characters for old Uslar's Caucasian Alphabets Alphabets – Nikita Manulov

Comments: We reviewed this proposal for 23 Cyrillic characters used for pre-revolutionary Caucasian alphabets, referencing alphabets for Abkhazian, Lak, Tabasaran, Avar, Chechen, Dargin, Lezgin, Tsakhur).

The following comments were made based on the discussion:
- Provide justification for encoding these additional letters. Are there identifiable digital projects that need the additional letters? If so, identify the projects and mention members of the user community who have been consulted.
- Provide further background on the letters, identifying the scope of usage. Were they only used from 1870-1917? Are the examples from the only books in existence? What communities of users were communicating with these letters?
- Map the proposed letters against what is intended for Georgian and Armenian alphabets.
- Do research and identify the range of glyphs as found in various texts and provide at least one example for each character (see following comment).   Provide a table showing the range of glyphs and different styles (i.e., handwritten vs. hooked vs. curled glyphs) and justify the glyph design selected as the representative glyph.
- For each proposed letter, cite the relevant figure (i.e., CYRILLIC SMALL LETTER CURLED ROUNDTOP EL **[figure 1]**). Circle the particular glyphs in the images and note the cited characters in the captions.
- It was noted that the proposed shape of one character did not match the figure: the glyph for CYRILLIC SMALL LETTER CURLED ROUNDTOP EL in figure 1, is as follows:
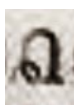
   but the proposed glyph is
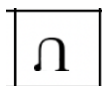
- The reference to "Uslar" in the proposal title presumably refers to Baron Peter von Uslar, who did research into Caucasian languages (source: Wikipedia) The proposal does not mention the connection or explain what "old" Uslar refers to, but such information would be helpful.

Debbie Anderson will send the above comments to the proposal author.

_____

16b Old Udmurt

Document: L2/22-217 Proposal to encode missing Cyrillic case pair for old Udmurt – Nikita Manulov

Comments: We reviewed this proposal to add a Cyrillic case pair for old Udmurt.

- The representative glyph is not justified by the examples; the cross bar should be on the baseline, based on the example from https://vivaldi.nlr.ru/bx000000815/view/?#page=12
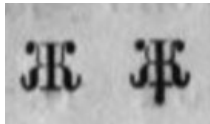
- We invite more actual research and discussion about the relationship between this case pair and other ZHE modifications in Cyrillic (i.e., ZHE with descender, ZHE with breve, ZHE with diaeresis). Why is this one different? What were other orthographies using for this sound?
- As noted in the Wikipedia article on Udmurt alphabets, there were several Cyrillic alphabets devised for Udmurt and other languages; this particular alphabet did not appear to survive into the twentieth century (http://udmurt.info/texts/periodist.htm). Why does this case pair need to be interchanged? Provide strong rational for carrying the case pair in an international standard.
- The ISO form states no contact with the user community has been made. Getting support from users would be important in making a stronger case.
- Provide more sources and images with better resolution. Harald Tveiten located a source with better resolution: https://rusneb.ru/catalog/000200_000018_RU_NLR_DIGIT_30059/

Debbie Anderson has already forwarded comments to the proposal author.

_____

17 Kore Sebeli

Documents: L2/22-222 Proposal for the encoding of « KORE SEBELI » – Mohamed Bentoura Bangoura et al.
Related documents:
L2/21-209 Proposal for the encoding of « KORE SEBELI » - Mohamed Bentoura Bangoura et al
L2/21-174 SAH Comments on L2/21-209
L2/21-016R SAH Comments on L2/20-180

[L2/20-180](#) Proposal for the encoding of « KORE SEBELI » - Mohamed Bentoura Bangoura et al
[L2/20-169](#) SAH Comments on an earlier version of L2/20-180

Comments: We reviewed this proposal for a neoscript, created to write the Soso language, which is spoken in Guinea and surrounding countries. The creator of the script, Mohamed Bentoura Bangoura, is one of the three co-authors, along with his student Mohamed Lamine Sylla, and a font designer Lucille Guignon, who wrote the proposal.

Tbe following points were raised during SAH meetings
- We discussed the pros and cons of the encoding model for Kore Sebeli, as presented by Ben Yang. The group agreed that the Arabic model (i.e., encoding the characters atomically) was the preferable approach over a Latin model (use common combining marks) or the N'Ko model (use script-specific combining marks). The Arabic model was the simplest.
- p. 8 Provide justification for two forms of lowercase DA: Are they used by different communities? Are they fully equivalent?
- p. 10 The following link does not work:
  http://sumale.vjf.cnrs.fr/phono/AfcheTableauOrtho2N.php?choixLangue=soso
  A working archived version of the document is available at:
  https://web.archive.org/web/20220119094829/https:/sumale.vjf.cnrs.fr/phono/AfficheTableauOrtho2N.php?choixLangue=soso
- p. 14 The right-hand glyph is the same as GNÉ. Is this an error?
- p. 15 Provide examples of RIGHT SINGLE QUOTATION MARK in the Kore Sebeli script.
- p. 15 (and page 5) Separately encoding the KORE SEBELI INVERTED EXCLAMATION MARK is acceptable, since it varies from the currently encoded character U+00A1; the Kore Sebeli character sits on the baseline.
- p. 15 (and page 5) Math symbols for multiplication, division, plus and minus should be separately encoded. There should be no decomposition for multiplication sign (cf. U+2A30 MULTIPLICATION SIGN WITH DOT ABOVE, which has not decomposition).
- p. 15 The EQUALS sign should be unified with U+003A COLON.
- Add a section that identifies confusable characters. In it, note that FUIN is potentially confusable with the digit six.
- Are there consistent rules for casing?
- Provide selected examples from the various publications to show the range of documents using the script (i.e., from printed works by Ousmane Camara, Ibrahima Sory Khalide Camara, and Mohamed Sylla).

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

------

18 Latin
18a SMALL LETTER TURNED O OPEN-O and SMALL LETTER TURNED O OPEN-O WITH STROKE

Documents: [L2/22-271R](#) On LATIN SMALL LETTER TURNED O OPEN-O and LATIN SMALL LETTER TURNED O OPEN-O WITH STROKE --- Moyogo Jacquerye
Related documents:

L2/22-128 Recommendations to UTC #172 July 2022 on Script Proposals (pages 6-7)
L2/22-097 On LATIN SMALL LETTER TURNED O OPEN-O and LATIN SMALL LETTER TURNED O OPEN-O
WITH STROKE – Moyogo Jacquerye
L2/22-109 On TURNED O OPEN-O and TURNED O OPEN-O WITH STROKE – Everson
(See SAH comments in L2/22-128)

Comments: We reviewed this document that comments on Everson's response document L2/22-109
regarding LATIN SMALL LETTER TURNED O OPEN-O and LATIN SMALL LETTER TURNED O OPEN-O WITH
STROKE. The SAH's last set of recommendations L2/22-128 did not recommend any change to the
current glyphs, as had been proposed in L2/22-097.

In summary, the characters are very rare, and the background, outlined in L2/22-271, describes the
complexity of the situation.

Denis Moyogo Jacquerye has revised the version seen by the SAH, and it is now posted as L2/22-271R.

18b Ben Franklin's Phonetic Alphabet

Document: L2/22-215 Proposal to encode missing Latin letters from Benjamin Franklin's phonetic
alphabet – Nikita Manulov
Related document:
L2/08-428 Exploratory proposal to encode Germanicist, Nordicist, and other phonetic characters in the
UCS

Comments: We reviewed this request for 4 Latin letters used in Benjamin Franklin's phonetic alphabet.
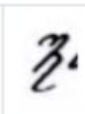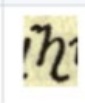The characters were proposed earlier by Michael Everson in L2/08-428.

The following comments were made during discussion:
- Add a reference to the earlier proposal, L2/08-428. The glyphs in this proposal appear to be
  very close to those in L2/08-428 and two characters have the exact same names. The newly
  proposed name with SERIF is not appropriate in this case, in our opinion.
- The proposed glyphs appear to come from low-resolution images. Provide better images and
  make changes to the glyphs reflecting the improved images. (See detailed comments below.)
- Identify the user community and explain why these letters need to be encoded.
- To identify the characters from the examples, the characters need to be de-italicized. Can they
  already be represented with current characters?
- If deemed eligible, the letters should go into the SMP, not the BMP. The BMP is being reserved
  for characters actively being used today, particularly for current languages needing
  representation.

Comments from a member of SAH:

1. In this proposal, Franklin's letter *dh* called LATIN SMALL LETTER H WITH RETROFLEX HOOK AND
   LONG SERIF AT TOP (= L2/08-428's LATIN SMALL LETTER H WITH RETROFLEX HOOK AND
   ANGLED ASCENDER) has a "long serif at top" on the top left (L2/08-428's "angled ascender" on
   the top left) but this is not at all what is visible in Benjamin Franklin handwritten

correspondence or in printed works. Franklin's has a notched terminal instead of the long serif or angled ascender. See for example:

Handwritten sample of Franklin's dh from facsimile of letter from Benjamin Franklin to Polly Stevenson, 20 July 1768, in Nicola Twilley, *Six new letters for a reformed alphabet*, 2006 (link)	
Printed sample of Franklin's dh from "A reformed mode of spelling", in *Political, miscellaneous, and philosophical pieces*, 1779, p. 474 (link)	
Printed sample of Franklin's dh from "A scheme for a new alphabet", in *The writings of Benjamin Franklin*, vol. 5, 1906, p. 176 (link)	

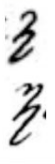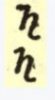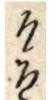Now compare them with the glyph of the character proposed:

Glyph proposed in L2/08-428	
Glyph proposed in this proposal (1.)	

2006 link: http://www.benfranklin300.org/_etc_pdf/Six_New_Letters_Nicola_Twilly.pdf ;
1779 link: https://archive.org/details/politicalmiscell00franrich/page/474/mode/1up ;
      1906   k: https://archive.org/details/writingsofbenjam05franuoft/page/176/mode/1up

2. In this proposal, Franklin's letter *th* called LATIN SMALL LETTER H WITH LONG RIGHT LEG (= L2/08-428's LATIN SMALL LETTER H WITH LONG RIGHT LEG) has a long right leg. However, this does not correspond with Franklin's shape for the letter in handwritten form in correspondence or the printed form in Franklin's "A scheme for a new alphabet, "in *The writings of Benjamin Franklin*, 1906 where the only difference between Franklin's letter *dh* and Franklin's letter *th* is the top left notched terminal on dh and an italic curved terminal on th. Both have the same bottom right hook in those works except in Franklin's "A reformed mode of spelling," in *Political, miscellaneous, and philosophical pieces*, 1779 where *dh* has an exaggerated hook and th has a simple hook similar to an italic bottom terminal. It's not clear whether Franklin intended the hook to be different.

Handwritten Franklin's th and of Franklin's dh	
Printed Franklin's th and Franklin's dh (swapped) from "A scheme for a new alphabet", in *The writings of Benjamin Franklin*, vol. 5, 1906, p. 176 (link)	
Printed Franklin's th and of Franklin's dh from "A reformed mode of spelling", in *Political, miscellaneous, and philosophical pieces*, 1779, p. 488 (link)	

1906 link: https://archive.org/details/writingsofbenjam05franuoft/page/176/mode/1up
1779 link: https://archive.org/details/politicalmiscell00franrich/page/n488/mode/1up

Debbie Anderson has already forwarded comments to the proposal author.

18c Latin letters for Old Adyghian

Document: L2/22-216 Proposal to Encode Latin characters for old Adyghian -- Nikita Manulov

Comments: We reviewed this proposal to add 13 Latin characters to support old Adyghian orthography from 1927-1938. The document is still preliminary.
The following were raised during discussion:

- The author needs to make a case for encoding new characters. For example, are there users who are digitizing texts and have a problem representing the texts?
- What is the encoding model? Are small capitals stylistic or not?
- Provide a full analysis of old Adyghian, showing all the letters of the orthography and their casing relations. It was noted that some images are unicameral, but in others, the letters are case paired.
- The ISO forms says contact with the user community has been made. List the individuals contacted.
- Which of the proposed characters look like Cyrillic?
- Provide a bibliography and sources for the figures. What is the source of figure 2?

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

---

19 Nüshu

Documents:
L2/22-138 Proposal to update the reference glyphs in the Nüshu block and the Ideographic Symbols and Punctuation block -- Kushim Jiang
L2/22-263 Feedback to Nushu glyph update proposed by L2/22-138 – Suzuki Toshiya
L2/22-264 Nüshu feedback on L2/22-138 (Proposal to update the reference glyphs in the Nüshu block and the Ideographic Symbols and Punctuation block) -- Endo Orie

Comments: We reviewed in L2/22-138, a request to update the font for 397 Nüshu characters, using the glyphs from the Noto Traditional Nüshu font. The document lists new forms for 22 characters.

We also reviewed the documents from Suzuki and Prof. Endo Orie.
- Suzuki requested China be asked about the font change. Debbie Anderson forwarded the question to Chen Zhuang to ask Prof. Zhao, who originally proposed the script. Chen Zhuang responded on Sept. 19, 2022, that Prof. Zhao was ok with the font updates in L2/22-138.
- Prof. Endo Orie, however, is against the change of font, who asked two Nüshu "transmitters," who prefer the current code chart font. She also mentions that the script was originally handed down from person to person, so there is no fixed standard. In addition, changing the shape of 22 characters will create a "ransom" effect, essentially changing a sub-collection of glyphs to a different style.

The SAH agrees with Prof. Endo Orie, that preserving the glyphs in the code chart as they stand is important and should be retained.

It was further noted that there are two Noto fonts: Noto Traditional Nüshu (by Pr Liu Zhao and Kushim Jiang) and Noto Sans Nüshu (by Lisa Huang), but the two projects didn't actively communicate and the difference between them is not clear.

Debbie Anderson will forward comments to the proposal author.

---

20 Proto-Cuneiform

Document: L2/22-239 Revised proposal for Proto-Cuneiform – Pandey
Related documents:
L2/21-184 Proto-Cuneiform: Comparison of Sign Images and Glyphs -- Pandey
L2/20-193 Preliminary proposal to encode Proto-Cuneiform in Unicode -- Pandey
L2/19-284 Proposal to Encode Proto-Cuneiform in the SMP of the UCS -- Everson
L2/17-157 Proposal to encode Proto-Cuneiform in the SMP of the UCS -- Everson
L2/16-267 Preliminary proposal to encode Proto-Cuneiform in the SMP of the UCS -- Everson

Comments: We reviewed this revised proposal for Proto-Cuneiform. In the proposed repertoire, numeric signs appear first, followed by ideographic signs, and then compounds (which are further subdivided into atomic and composable). The composable compounds are analyzed into their

component elements, using format control characters similar to those being used for Egyptian Hieroglyphs (i.e., begin and end group, joiners, overlay, reverse [=EGYPTIAN HIEROGLYPH MIRROR HORIZONTALLY], and container [= EGYPTIAN HIEROGLYPH INSERT AT MIDDLE]). The repertoire includes variants

The following summarizes the discussion:

- The analysis of compounds into their constituents will be very useful for font technology to identify the component glyphs and the rules for building the glyphs. However, for encoding model, the SAH recommends atomic encoding of the characters, so the model fits with that already being used for cuneiform (both for Cuneiform and Early Dynastic Cuneiform). Hence the SAH recommends the encoding of 2095 characters. (In contrast to Proto-Cuneiform, the Egyptologists have been lobbying for an approach that can handle the fairly productive way characters were being created by scribes, hence the requirement for format controls.)
- On page 136, change the line breaking property for U+126E7..U+12CA0 to Lo (not Nl).
- The analysis in this document could be published separately (though it was noted this has not been done elsewhere for cuneiform).
- The numeric signs could be separated into a different block or be part of a single block. Cuneiform has a separate "Cuneiform Numbers and Punctuation" block, but it is not a requirement to have them in separate blocks.

Anshuman Pandey will seek feedback from experts and revise the proposal based on the comments above.

---

21 Sidetic

Document: L2/22-235 Revised proposal to encode the Sidetic script in Unicode – Pandey
Related documents:
L2/21-111 Preliminary proposal to encode the Sidetic script in Unicode – Pandey
L2/19-106 Introducing the Sidetic Script – Pandey

Comments: We reviewed this updated document, which proposes 29 characters (three more than in L2/21-111) and changes the names from phonetically-based names to numbers used by scholars (based on Nollé, hence N before the number). The present proposal has incorporated feedback from the experts Michaela and Christian Zinko, received Sept. 2022.

The version viewed by the SAH had names with hyphens (i.e., N-1, etc.), which is technically valid. If scholars use the hyphen, it can be retained. However, Anshuman Pandey reports the current citation practice is without hyphens, so the names without hyphens are incorporated in the version in the document register, L2/22-235.

This proposal L2/22-235 has been sent out to the experts Michaela and Christian Zinko and Alfred Rizza for review. (Michaela Zinko reports that a new inscription will provide additional information, and a

meeting in November on the inscription will afford additional insights from other experts.) The question on how to handle inscriptions on coins (as discussed in section 4.1) has also been forwarded to experts.

---

22 Syloti Nagri

Documents: L2/22-212 Proposal to Encode Syloti Nagri Numerals – Mandal
Related documents:
L2/22-225 Supportive letter for Syloti Numerals – Md Islam
L2/21-187 Proposal to Encode Syloti Nagri Numerals -- Mandal and Mowbray
L2/21-174, page 14 SAH comments on L2/21-187.
L2/21-140 Proposal to Encode the Sylheti Nagri Numerals in the UCS - Mowbray
L2/21-130, p. 13 SAH comments on L2/21-140
L2/02-388 Documentation in support of proposal for encoding Syloti Nagri in the BMP – Lloyd Williams et al.

Comments: We reviewed this revised document, dated July 12, 2022.
Background: When Syloti Nagri was proposed, the proposers (including Constable, co-author) did not believe evidence supported the numerals at the time (page 19 of L2/02-388), though a set of digits appear in the primers of Prof. Erhasuzzaman, comprising Bengali digits, Latin, and Arabic-Indic digits.

More evidence is provided (specifically, in figures 11-17), but all the sources for the figures are not clearly identified or listed in the bibliography.

The SAH did not have consensus that the evidence was compelling enough at this stage to recommend approval for the digits.

Questions and comments for the proposal author:
- Describe the groups supporting the effort (listed on page 2) and identify their audience(s). An official letter of support from an institution or language group would be helpful.
- More attested use would be more compelling: Can the proposal author provide newspapers showing their use? Are the digits being taught in schools today?
- Provide sources for all the figures and references in the bibliography.
- Note that Surma Faror Khobor site, cited on page 2 of the proposal, has European and Arabic digits, which doesn't support the case for the Syloti Nagri digits.
- Compare shapes to document any changes through time.
- Identify sources for figures and include printed works in the bibliography.
- Provide prose description for the figures. (Note that the Surma Faror Khobor site shows Arabic and European digits, so it is not useful as an example for Syloti Nagri numerals.)

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

Note: Unicode Consortium has received many letters of support for the proposal from the community. Rather than post all the letters individually in the document register, the proposal author will cite the supporters in the revised proposal.

---

23 Symbol: Sanban sign

Document:L2/22-207 Proposal for Sanban Sign for Chinese folk music and local operas -- Eiso Chan (See also comment below from Eiso Chan, citing IRG docs L2/22-253 and L2/22-254)

Comments: We reviewed this proposal for one character to be added to the Musical Symbols block. The character is used as a beat sign, to record a rhythm form.

The following points were raised during discussion:

- The examples in the proposal are from scores, which are not plain text.

- We recommend use of the KATAKANA LETTER SA U+30B5, which is supported in Chinese fonts. If the character appears within text and needs to be distinguished from Japanese text, then the character would be a candidate for disunification.
- On the etymology of the character, see https://en.wikipedia.org/wiki/Sa_(kana), which cites the Katakana letter SA.

Note: After the SAH met, the following information was forwarded by Eiso Chan:

The IRG editorial report (L2/22-254) relays the following: "The editors had no objection to encode the Sanban sign as a musical symbol separately from CJK characters." The "IRG Meeting #59 Recommendations and Action Items" (L2/22-253) reports: "IRG has no objection to encode Sanban signs as a complete set of musical symbols under a symbol notation system, separate from CJK characters."

---

24 Symbol: Russian Fire Hydrant symbol

Document: L2/22-211 Proposal to encode Russian Fire Hydrant Symbol – Nikita Manulov

Comments: We reviewed this proposal for the Russian fire hydrant symbol. The symbol is a ligated acronym of Cyrillic "P" and "K." The acronym comes from two words shown in figure 5 (page 3).

The following comments were made:

- Provide plain text examples, which may be able to demonstrate the need for standardization as a symbol that needs to be interchanged, as opposed to a signage convention. Note that signage per se is not a guarantee that a symbol should be encoded. For example, the lit match in figure 6 is not encoded. Also, note the set of ISO 7000 symbols, which appear on equipment, https://www.iso.org/obp/ui/#iso:pub:PUB400001:en and are largely not encoded in Unicode.
- Has contact been made with the user community? The ISO form response is "no," but contact with the user community would be helpful.
- Number the pages.

In sum, additional information is needed before a decision can be made regarding the encoding of this character.

Debbie Anderson has already forwarded comments to the proposal author.

----

25 Tai Don

Document: L2/22-210 Final proposal to add 22 characters for Tai Don writing systems – Kushim Jiang
Related documents:
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 L2/22-210, a revised proposal, which builds off of earlier proposals by the same author (i.e., L2/20-207 and L2/22-098).

The following comments were made:

- To assist reviewers, it would be helpful to include an "Executive Summary" at the top of the document, clarifying that it proposes adding 22 characters to the Tai Viet script in order to represent the Tai Don and Tai Dam writing systems.
- Section 2.1: Script blocks and characters cannot be renamed after they have been published (cf. https://www.unicode.org/faq/casemap_charprop.html#23 and name stability policy: https://www.unicode.org/policies/stability_policy.html) However, additional names can be included in the note at the top of the names list, cf. the note above the names list for Tai Tham: https://www.unicode.org/charts/PDF/U1A20.pdf

- Section 2.3: Use character names with TAI VIET. A subhead or annotation can identify those characters used in the Tai Don or Tai Dam orthography.
- The figures in section 2 need additional explanation (see detailed questions below) and clearer formatting.
- Section 2.7: Give stronger rationale on why the different fish symbols in section should be encoded. Note that every proposed character should have an example in running text. Add a note beside each proposed symbol, pointing to the example with the symbol.
- Confirm the proposal author is in touch with the user community and they are being actively consulted on the proposal.
- Include Indic property data (cf. the Indic property data for Tai Viet. Cf. https://www.unicode.org/Public/UCD/latest/ucd/IndicSyllabicCategory.txt https://www.unicode.org/Public/UCD/latest/ucd/IndicPositionalCategory.txt
- In the charts, it would be more helpful to the Script Ad Hoc to include character names alongside code points.

Answer other questions posed in comments below (by Lawrence Wolf-Sonkin and Liang Hai):

2.1:
- "Specifically, the current character subset includes a complete set of alphabets for the Tai Dam writing system, with other three pairs of consonant letters used in the Tai Don writing system."
    - "the current subset" refers to the contents of this proposal?
    - which "other three pairs of consonant letters" are we referring to?
- All the sentences in the second paragraph of 2.1 needs clarification
    - Is he suggesting changing the name of Tai Viet (which we cannot do)? [See comment above]

2.2:
- What does "soon-to-be-completed" mean here? Is the script undergoing a lot of recent changes?

2.4
- We should have some proper headers explaining what the columns represent in the parlance of the proposal.
    - Explain column 3 (why is it semi-sliced off). We're assuming it's Tai Viet
    - Column 4 should say Tai Dam
    - Column 5-9 should say Tai Don.
    - Add two new columns for codepoints and notes for what the situation is for the character (is this a new proposed character, an existing character, etc.)
- the above comments apply also to figure 7

2.4 unification process section:
    - We need a further justification of the unifications and disunifications, especially those that differ from Jim Brase's document. Namely, does the user community consider these graphemes, even ones that are the shapes from neighboring regions, to be shapes they would recognize and/or produce

2.8 Collation
- Some of the ordering seems to be..... fishy
    - Namely, the fish takes a secondary collation weight (is this intentional? If so, why?)
- A.1 Normalization of Initials
- Is the 3rd and 4th entries U+AA80 and U+AAC3 meant to be described as phonetically identical?
- A.3 Matching Final
- Why do two entries [uŋ] and [um] use different placeholder bases other than U+AA80?

The above comments have been sent to the proposal author.

---

26 Wuyue Zhenxing Tu(Fu)

Document: L2/22-209 Proposal to encode the pictograph Wuyue Zhenxing Tu(Fu) Character --- WEI Xitong

Comments: We briefly reviewed this request for characters for five famous mountains in Chinese traditional culture. The characters originally appeared on a historic map of the mountains drawn by Taoist priests The characters are found on stele, books, amulets, coins, furniture, and decorations.

There was no consensus from the group regarding the encoding of the 5 characters. One concern voiced was that there doesn't appear to be any conventional representation for them.
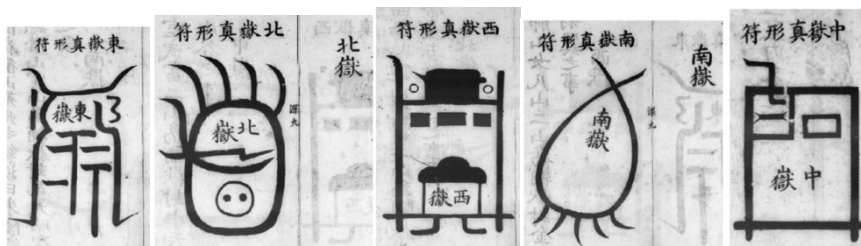
Comments for proposal author:

- Provide text about the variation of the symbols; this could help font designers
- Provide examples of the characters in running text.
- Explain why these should be encoded as characters? Do they have special status?
- We suggest the author keep China involved in the discussion.

We received the following comments from Chen Zhuang for the China NB, in which he states China is not supportive of this proposal as it stands:

> After consulting some of my experts, I think that the discussions are needed. It seems that the proposed 5 shapes are the most common ancient logos of the 5 moutains. Firstly, I want the author to give more evidences showing these 5 shapes are widely, overwelmingly accepted by people. Secondly, I concern that how we process the endless Taoist or Buddhist pictogrpahs similar like these. Currently, don't support this proposal.
>
> Attached [below] another series of Wuyue Zhenxing Tu pictures

It was noted during SAH discussion that the images from Chen Zhuang differ from the proposed glyphs:
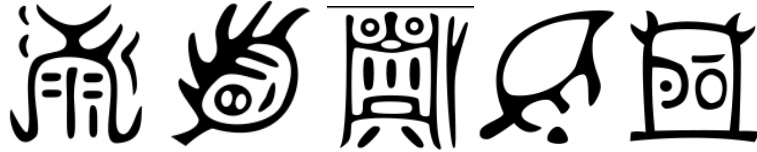
Debbie Anderson will forward the comments above to the proposal author.

---

27 Yi Ideographs

Document:
L2/22-183 Introduction to encode Yi Ideographs - Kushim Jiang
Related document:
L2/08-193 Preliminary proposal to encode Classical Yi characters (abridged. A link to the full proposal including ~2048 pages of charts is 137MB, is available on the Unicode doc registry) -- China NB

Comments: The introductory document proposes work on Yi ideographs. In general, the group was in favor of making progress on Yi ideographs.

Some years ago, Debbie Anderson, Andrew West and Michael Everson had done some initial work on Sani Yi and Chuxiong Yi.

Debbie Anderson has already forwarded comments to the proposal author.

---

28 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
- L2/22-251 Review of some Arabic additions for Quranic orthographies

---

C. OTHER TOPICS: "Do Not Use" Datafile

FYI: Roozbeh Pournader is starting a datafile with Do Not Use data.