L2/22-023

TO: **UTC**

FROM: **Deborah Anderson, Ken Whistler, Roozbeh Pournader, and Peter Constable**[1]

SUBJECT: **Recommendations to UTC #170 January 2022 on Script Proposals**

DATE: **January 22, 2022**

The Script Ad Hoc group met on October 22, November 12 and 19, December 10, 2021 and January 7, 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 Fred Brennan, Lorna Evans, Andrew Glass, Liang Hai, Ned Holbrook, John Hudson, Richard Ishida, Marek Jeziorek, Jan Kučera, Norbert Lindenberg, Kamal Mansour, Lisa Moore, Lawrence Wolf-Sonkin, and Ben Yang. The text for the comments and recommendations was based on notes taken by Debbie Anderson, Fred Brennan, Norbert Lindenberg, and Jan Kučera.

I. EUROPE

1 Cyrillic

1a Cyrillic Letter Multiocular O

Document: L2/22-002 Proposal to revise the glyph of CYRILLIC LETTER MULTIOCULAR O -- Everson

Comments: We reviewed this document, which recommended a change to the glyph for U+A66E CYRILLIC LETTER MULTIOCULAR O. The example in the original proposal, L2/07-003r, was a fuzzy image with 10 eyes. However, the code chart glyph had 7 eyes, which was an error. According to Étienne FD, the sign only occurred in one text.

Ralph Cleminson, a co-author on the original proposal, commented that "the seven- and ten-eyed forms can be considered glyph variants" and the manuscripts vary. He also mentioned that more often than not, the character is not employed at all, instead texts use ʘ CYRILLIC LETTER BINOCULAR O. Clarification from Cleminson seems warranted, since he suggests that both the 7 and 10-eyed shapes occur. His comments could also be interpreted to mean that MULTIOCULAR O is a variant of the BINOCULAR O.

In our view, the glyph should be changed now, but a one-line annotation should be proposed to explain that the earlier version of the chart had a 7-eyed version.

Recommendations: We recommend the UTC make the following disposition:
SAH-UTC170-R1: Approve a glyph change for U+A66E CYRILLIC LETTER MULTIOCULAR O from a 7-eyed glyph to a 10-eyed glyph for a change in Unicode 15.0. (Reference: L2/22-002)

Action Item for Michael Everson: Provide Michel Suignard with a glyph and propose an annotation for U+A66E CYRILLIC LETTER MULTIOCULAR O, describing the change from 7-eyed glyph to a 10-eyed glyph. (Reference: L2/22-002 and Section 1a of L2/22-023)
Action Item for Michael Everson and Debbie Anderson: Contact Ralph Cleminson and get clarification on his comments. (Reference: L2/22-002 and Section 1a of L2/22-023)
Action Item for Debbie Anderson and Editorial Committee: Create a glyph erratum for U+A66E CYRILLIC LETTER MULTIOCULAR O. (Reference: Section 1a of L2/22-023 and L2/22-002)

1b Cyrillic Modifier Letters

Document: L2/22-010 Addendum II to L2/21-107, Cyrillic modifier letters

Comments: We reviewed this request for two Cyrillic letters, which are an addition to the set of Cyrillic characters approved at the July UTC meeting, based on proposal L2/21-107, in the new Cyrillic Extended-D block.

The two characters are: MODIFIER LETTER CYRILLIC SMALL LETTER STRAIGHT U WITH STROKE, used in Kazakh, and COMBINING CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I. Examples of the two proposed characters are provided.

Recommendation: We recommend that the UTC make the following disposition:
***SAH-UTC170-R2*:** Accepts the following two characters, as documented in L2/22-010, for a future version of the standard:
1E06D MODIFIER LETTER CYRILLIC SMALL STRAIGHT U WITH STROKE
1E08F COMBINING CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I

Action Item for Ken Whistler: Update the Pipeline to include U+1E06D MODIFIER LETTER CYRILLIC SMALL STRAIGHT U WITH STROKE and U+1E08F COMBINING CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I, as documented in L2/22-010.
Action Item for Debbie Anderson: Confirm that Michel Suignard has the font for the Cyrillic additions. (Reference: L2/22-010 and Section 1b of L2/22-023)

2 Latin

2a African Reference Alphabet

Documents:
L2/21-231 On the 1978 version of the African Reference Alphabet -- Marín Silva
L2/21-247 Feedback on African Reference Alphabet (L2/21-231) -- Denis Jacquerye

Comments: We reviewed L2/21-231, a document on the 1978 version of the African Reference Alphabet, which was proposed at a UNESCO-sponsored conference in Niamey, Niger. The alphabet was revised in 1982. The author gives his opinion on characters in the 1978 version of the alphabet.

The following comments were made:

- What evidence is there that this alphabet is being used today? Are there works being digitized in which the characters need to be supported in an international standard? What is the use-case for a plain text encoding based on this 1978 document?
- The document should clearly state what is being requested, i.e., “The following four characters are being proposed…”, with full information (properties, references, etc.).
- Any proposed annotations should be clearly stated, with accompanying justification. (Note that the Summary on page 4 states that four annotations are proposed, but the list doesn’t include U+01B7 on the bottom of page 3.) Note that annotations are intended to make clear the identity and use of characters, but they are not meant to be an encyclopedia-type reference of typefaces.

We also reviewed L2/21-247, comments on L2/21-231 by Denis Jacquerye, which we recommend be forwarded to the author of L2/21-231.

Recommendation: We recommend the UTC make the following disposition:
Action Item for Rick McGowan: Relay the comments in Section 2a of L2/22-023 as well as a link to L2/21-247 to the author of L2/21-231.

2b Casing Pair used by Some African Orthographies

Document: L2/21-229 Exploratory document on a problematic casing pair used by some African orthographies – Marín-Silva

Comments: We reviewed this document, which proposes different approaches to handle the casing pair Бб "that according to Wikipedia is used by some African orthographies [though]… [the author] couldn't confirm the veracity of those claims." The document discusses five possible encoding models: one that would involve changing casing relations (A1), a second option involving use of a pair of characters for the old Zhuang orthography (A2), a third option which uses a current case pair but involves a different glyph form for the uppercase (A3), a fourth option involving use of Cyrillic letters for languages that use Latin letters (B), and lastly encoding a new Latin case pair (C).

The following comments were made:

- Relying solely on Wikipedia as evidence is not advisable.
- "In this document I discuss the following casing pair (Бб)": Clarify exactly which characters are being referred to at the beginning of the document, rather than the glyphs.
- "Here we discuss different encoding models to support this orthography." Which orthography is being referred to here (i.e., Clement Doke for Shona or one of the many others mentioned)?
- No actual usage in current digital data or orthographic practice is provided. In our view, speculating on what might be helpful to an unidentified audience is not constructive. We recommend the author submit proposals that address problems real-world users are encountering, citing sources outside of just Wikipedia.
- Note that the document L2/08-034R had suggested an annotation for U+0181 LATIN CAPITAL B WITH HOOK be added, stating that a variant glyph of U+0181 is used in some Liberian orthographies with appearance of U+0182. This was based on evidence found in Toma and Dan/Gio. The variant of U+0181 is now included in SIL fonts. However, there is no information on whether the variant is actually being used today.

Recommendation: We recommend the UTC make the following disposition:
Action Item for Rick McGowan: Relay comments in Section 2b of L2/22-023 to the author of L2/21-229.

2c Closed Insular G

Documents:
L2/21-242 Glyph Corrections for U+AD70 LATIN CAPITAL LETTER CLOSED INSULAR G and U+AD71 LATIN SMALL LETTER CLOSED INSULAR G – Baker
L2/22-004 On the glyph of LATIN LETTER CLOSED INSULAR G -- Everson and West
L2/21-243 Comment on "On the glyph of LATIN LETTER CLOSED INSULAR G" by M. Everson and A. West -- Baker

Comments: We reviewed this set of documents discussing the shapes of LATIN LETTER CLOSED INSULAR G (U+AD70 and U+AD71).

The identity of the case pair for LATIN LETTER CLOSED INSULAR G is not in doubt. Because the code charts serve as a model for font providers, there is good reason to have the representative glyph in the charts be accurate. However, these documents reflect a disagreement on the glyph shape. If other outside experts are consulted and agree on changing the glyph, then the subject can be returned to the Script Ad Hoc for discussion. However, at this time no change is needed, in our opinion.

Michael Everson is invited to follow up with experts (such as Dr. Colleen Curran at Oxford).

Recommendation: We recommend that the UTC make the following disposition:
Notes these documents but takes no further action.

II. AFRICA

3 Adinkra

Document: L2/21-237 Response to Unicode Technical Committee -- Korankye, Adinkra Alphabet Encoding Committee

Comments: We reviewed this response to the Script Ad Hoc comments in L2/21-016 (page 18), which had addressed the Adinkra draft proposal L2/21-020.

The following were comments from the Script Ad Hoc:

- More actual usage of the script in the community is required, particularly showing *continued* usage through time.
- There is no evidence of any publications outside of the creator. Printed publications are needed (besides those of the creator).

In sum, we invite a proposal once widespread usage in communities can be shown and a strong case is made that the script needs to be encoded for digital interchange.

Recommendation: We recommend the UTC make the following disposition:
Action Item for Debbie Anderson: Relay comments in Section 3 of L2/22-023 to the author of L2/21-237.

---

4 Egyptian Hieroglyphs

4a Format Control Characters

Document: L2/21-248 Additional control characters for Ancient Egyptian hieroglyphic texts – Glass et al.

Comments: We reviewed this revised proposal for additional characters – including control characters – for Egyptian Hieroglyphs. Earlier versions of the proposal have been reviewed by the Script Ad Hoc.

The following are comments on the proposed characters shown on Table 2 (page 3):

V011D (one addition to Egyptian Hieroglyphs main chart)
The addition of U+1342F EGYPTIAN HIEROGLYPH V011D is justified, in our opinion. It fills the last available code point in the current Egyptian Hieroglyph block.

Insertion controls
The Script Ad Hoc agreed with the proposed three controls for insertions at the top, bottom, and middle. Any ambiguity should be handled by orthographic checking, and Egyptologists will need to figure out the orthographic rules, but such checking shouldn't be enforced as part of the Unicode model.

Enclosure controls
The set of four proposed format controls for enclosures was deemed reasonable. They will better represent text for Egyptologists, including those working with hieratic.

Mirror control
We agreed with the proposed single control for mirroring, which will be done in the font at the sign level. If the control follows a symmetrical sign, it is not mirrored, and the control appears visibly to let the user know the control is present.

Blank signs
Two characters, HALF BLANK and FULL BLANK, seemed reasonable (with glyphs including the abbreviations "FB" and "HB"). (On the glyphs, see below "Font.")

Lost signs
Four signs for lost signs ("atomic shades") were agreeable to the SAH. Validating encoding sequences will rest with the users of the system.

We agree with the proposed model of using a Variation Selector to extend the "lost" sign, so the shading fills the entirety of the available surface, as opposed to the default behavior, whereby white space appears between lost signs. (See Section 4b, below)

Damage (/Uncertainty) modifiers
Based on SAH recommendation, 15 code points are proposed for all combinations of sign shades indicating damage, with a limit of one after each sign (see page 14 of the proposal for rationale). (On the glyphs, see below "Font.")

Rotation
Rotation will be handled by variation selectors, as was recommended by the SAH. (See Section 4b, L2/22-012, which includes a list of variation sequences for those signs that will be rotated at 90, 180, and 270 degrees. The list would be updated as more rotatable characters are identified.) Note: In sequences with mirroring and rotation, mirroring comes after rotation.

Brackets
The Script Ad Hoc agreed that common Western punctuation characters should be used as brackets, thereby allowing them to participate in Egyptian shaping. The common characters are part of general editorial practice, even outside Egyptology. Andrew Glass has filed a bug with the Microsoft Office team since currently common punctuation is not working with Egyptian hieroglyphs in Word, though such signs do work in the DWrite and Harfbuzz shaping engines.

Font

Michel Suignard will create the font, which may vary from what the glyphs in the code chart on page 3.

- Note that the FULL BLANK and HALF BLANK should have a dashed box around them to be consistent with the other Egyptian Hieroglyph format controls.
- The dotted box conventions around the 15 damaged signs (U+13447..U+13455) should be discussed by the UTC and, if adopted, be documented in Chapter 24.1 of TUS.

  Compare the following:
  - proposed dotted box glyph for Egyptian Hieroglyph damaged sign (gc=Mn):

  - dashed box for combining marks (but with no visible display) for variation selectors (U+FE00.. U+FE0F):

  - dashed box surrounding a dotted circle to indicate combining marks that move the position of the vowel around the nucleus in Miao (U+16F8F..U+16F92):

  - U+2B1A DOTTED SQUARE (gc=So):

  - dotted circle that is iconic of the face shape for Sutton SignWriting faces (U+1DA00 ff.):

  - dotted circle in Combining Diacritical Marks for Symbols; in the example below, a large graphic shape is conceptually laid over another base character:

Roadmap Allocation

Since the current allocation for Egyptian Hieroglyph Format Controls has only 7 open slots, we recommend the block be extended two more columns from its current allocation U+13430..U+1343F to U+13430..U+1345F, and leave two columns empty (U+13460..U+1347F). As a result, the Egyptian Hieroglyph extensions should start at U+13480.

Recommendations: We recommend the UTC make the following dispositions:
SAH-UTC170-R3: Accepts 30 Egyptian Hieroglyph characters, as documented in Tables 29-31 of L2/21-248, for a future version of the standard. (Reference: Section 4a of L2/22-023)

SAH-UTC170-R4: Accepts the extension of the Egyptian Hieroglyph Format Controls block from the current allocation U+13430..U+1343F to U+13430..U+1345F. (Reference: Section 4a of L2/22-023)

Action Item for Michel Suignard to create a font for the Egyptian Hieroglyph characters, based on Section 4a of L2/22-023, UTC discussion, and page 3 of L2/21-248.
Action Item for Ken Whistler: Update the Pipeline to include 30 Egyptian Hieroglyph characters, as documented in Tables 29-31 of L2/21-248.
Action Item for Ken Whistler and Debbie Anderson: Confirm the Roadmap changes described in Section 4a of L2/22-023 are incorporated in the Roadmap (i.e., extend Egyptian Hieroglyph Format Controls from U+13430..U+1343F to U+13430..U+1345F, and leave U+13460..U+1347F empty).

4b Variation Sequences for Egyptian Hieroglyphs

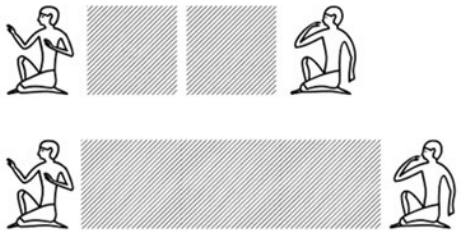
Document: L2/22-012 Rotations of Egyptian Hieroglyphs to be Registered in Unicode – Werning (with a .txt file by Andrew Glass attached to the PDF listing Rotations for Standardized Variants)

Comments: We reviewed this document from Daniel Werning, which identifies those characters in Unicode that have been found in rotated positions. The data identifying the characters and their rotations are drawn from the Thesaurus Linguae Aegyptiae, Pyramid Text project, Karnak Project, Athribis project and Kom Ombo project.

The list includes rotations of 90°, 180°, and 270°, using variation selectors FE00, FE01, and FE02, as described on page 14 of L2/21-248. The list also includes sequences of other angles (30°, etc.) using FE03. The FE03 set of sequences is not yet being proposed.

Attached to the PDF is a plaintext file with 98 additions to StandardizedVariants.txt. This list contains:

- 94 sequences to be used for rotations of 90°, 180° and 270°. The rotations are clockwise for text normally rendered left-to-right but counterclockwise when text is mirrored right-to-left.
- 4 sequences (at the bottom of the list) are used to denote expanded forms for “lost” symbols, that is, when any of the 4 “lost signs” (U+13443..U+13446) appear in a sequence with U+FE00, they expand to fill the whitespace between the signs. In the example below, the top sequence would be <13000, 13443, 13443, 13000> but the bottom sequence would be <13000, 13443 FE00, 13443, 13000>.



The names list will reference the rotations, but the glyphs will be suppressed in the charts.

Recommendation: We recommend the UTC make the following disposition:
SAH-UTC170-R5: The UTC accepts 98 standardized variants (94 rotations and 4 expanded lost signs), as documented in the attachment to L2/22-012, for a future version of the standard.

Action Item for Ken Whistler: Update the Pipeline to include the 98 Standardized Variation Sequences, as documented in the attachment to L2/22-012.

---

5 Garay

Document: L2/22-030 Consideration of the encoding of Garay with updated user feedback -- Rovenchak et al.

Comments: We reviewed this document which provides additional information and user feedback on the earlier Garay script proposal (L2/16-069), as well as further clarifications to L2/19-163 and the Script Ad Hoc comments in L2/18-168.

The following were noted during discussion:

- We believe the bidi class AN is correct for numbers and R for letters.
- Section 5 Collation order: In this section, details are needed on how strings would be ordered, such as in a dictionary. For example, should OLD NA and OLD KA appear at the end of the alphabet or should they be considered variants of KA / NA (essentially equivalents to KA / NA and interfiled in an ordered listing)? Are vowels treated as distinct letters at the beginning of an alphabet? Are VOWEL SIGN E, GEMINATION MARK, etc. considered secondary? (The fact that the letters are ordered based on their numerical values is not important for collation.)
- In section 5, clarify that “Vowel diacritics… are placed in the third columns” refers to columns in the code chart (and not Table 1).
- Provide sources on the figures, if possible.
- Is Garay most likely to occur with Arabic or Latin?

Recommendations: We recommend the UTC make the following disposition:
Notes this document but takes no further action; the SAH comments have already been conveyed to the author.

III. MIDDLE EAST

6 Arabic

6a Arabic Alef with Right Hamza

Document: L2/22-035 Proposal to encode Arabic Alef with Right Hamza used in Quran published in Iran - Lateef Sagar Shaikh

Comments: We reviewed this proposal to encode an Arabic *alef* with right *hamza*, which is found in some Qurans from Iran.

In our opinion, a new character is not needed, as *alef* with right *hamza* can already be represented by the sequence <0621, 0627> for the isolated shape and <0640, 0654, 0627> for the final shape. The use of U+0640 ARABIC TATWEEL and U+0654 ARABIC HAMZA ABOVE for the final shape is discussed in the subsection "Quranic Texts" on page 394 in section 9.2 Arabic of *TUS*

The proposal mentions U+0676 ARABIC LETTER HIGH HAMZA WAW can be used "without any issue," but this character -- as well as U+0675, U+0677 and U+0678, all characters with a right high hamza -- are not recommended for use. As a result, U+0676 ARABIC LETTER HIGH HAMZA WAW should not be used as a model (see page 394 in section 9.2 Arabic of *TUS* and the Arabic block names list),

Recommendation: We recommend the UTC make the following disposition:
Action Item for Debbie Anderson: Forward the comments in Section 6a of L2/22-023 to the author of L2/22-035.

6b Balochi

Document: L2/21-238 Response to SAH re: Proposal to add four new Arabic characters for Balochi language (L2/19-320) – Qazi Rehan

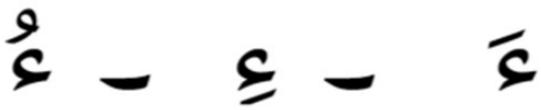
Comments: We reviewed this feedback from Qazi Rehan, which is a response to comments from the Script Ad Hoc recommendations (L2/19-343), which in turn contained comments on a proposal to encode four new Arabic characters for Balochi by Qazi Rehan (L2/19-320).

The author has two main points.
1. In section 1, he reports that in a font used on Facebook he cannot get U+064F ARABIC DAMMA, U+0650 ARABIC KASRA, or U+064E ARABIC KATHA to appear with U+0621 ARABIC LETTER HAMZA. Instead, he only sees the stand-alone *hamza* (U+0621).

[SAH comment]

> Although they seem to "disappear," they are actually mispositioned, overstriking the *hamza* and affecting its shape.

The author requests new characters for *hamza with damma above*, *hamza with fatha above*, and *hamza with kasra above* (below).



[SAH comment]

> This request is not justified. The problem is the font, so the author should contact the vendor of the font he is using and ask for the appropriate character sequences to be properly positioned/displayed.

2. In section 2, the author repeated his request for a “yeh hamza above with Arabic kasra” shown below.

The Script Ad Hoc had earlier suggested that the letter could be an alternate glyph for U+06D3 ARABIC YEH BARREE WITH HAMZA ABOVE. The author suggests instead his new character can be written with the sequence <U+0626 ئ ARABIC LETTER YEH WITH HAMZA ABOVE, U+08F6 ◌ۦ ARABIC KASRA WITH DOT BELOW, U+06D2 ے ARABIC LETTER YEH BARREE>.

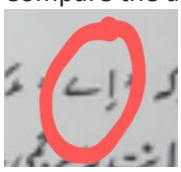
[SAH comments]

U+08F6 in the sequence appears to be an error, perhaps for U+0650 ARABIC KASRA?

In order to make a decision on how to proceed, more information is needed, with answers to the following questions:

- What is the pronunciation?
- The manuscript image on page 3 shows an extra tooth:

Compare the above with:



Provide a higher-resolution image of the manuscript on page 3 and translation of the text.

Recommendation: We recommend the UTC make the following disposition:
Action Item for Roozbeh Pournader and Debbie Anderson: Relay the comments in Section 6b of L2/22-023 to the author of L2/21-238

6c Chinese National Body Comments

Document: (see ballot comments below)
Background documents:
L2/20-289 Request for glyph changes and annotations for Kazakh, Kyrgyz, and Uyghur - Evans
L2/21-050 Chinese comments on WG2 N5155 (= L2/20-289) - China NB
L2/21-098 Response to China NB comments on WG2 N5155 (UTC document L2/21-050) (WG2 N5162)

Comments: We reviewed the ballot comments from the China National Body on Amendment 1 of ISO/IEC 10646:2020. The comments refer to the following new text under the new sub-heading “Digraphic letters for Kazakh” in the names list (which first appeared in AMD 1 and Unicode 14.0):

Digraphic letters for Kazakh

Use of these characters is discouraged. They were encoded for Kazakh digraphs, but their decompositions do not reflect the preferred order of representation.

0675 ٵ ARABIC LETTER HIGH HAMZA ALEF
≈ 0627 ا 0674 ◌ٴ

0676 ۄ ARABIC LETTER HIGH HAMZA WAW
≈ 0648 و 0674 ◌ٴ

0677 ۅٟ ARABIC LETTER U WITH HAMZA ABOVE
≈ 06C7 ۇ 0674 ◌ٴ

0678 ۍ ARABIC LETTER HIGH HAMZA YEH
≈ 064A ي 0674 ◌ٴ

Below are the ballot comments from the China NB:

> The note "Use of these characters is discouraged. ……" for U+ U+0675 through U+0678 should be removed.
>
> These four letters are not equal to the corresponding target sequences, and the values and orderings are different. This paragraph will mislead CLDR and the input method, and the strings which the end users input and output will be different from the previous. The names of Kazakh compatriots in China always refer to the common Kazakh words written with the Arabic script, these four letters are commonly used. The note "Use of these characters is discouraged. ……" introduces hidden dangers to the Chinese administration concerning residents and residency.

We agree the decomposition mappings for these four characters are problematic, but they cannot be changed (see stability policy). The new text in the names list has apparently been interpreted as meaning the letters should not be used when writing Kazakh. However, the text was intended to convey that to represent these letters, use of U+0675..U+0678 is discouraged, but the *correct order of decomposed sequences* should be used. To clarify the text, we suggest annotations identifying the following recommended sequences to be used:

0675 ARABIC LETTER HIGH HAMZA ALEF
* should be represented using the sequence 0674 0627
0676 ARABIC LETTER HIGH HAMZA WAW
*should be represented using the sequence 0674 0648
0677 ARABIC LETTER U WITH HAMZA ABOVE
*should be represented using the sequence 0674 06C7
0678 ARABIC LETTER HIGH HAMZA YEH
*should be represented using the sequence 0674 0649

Recommendation: We recommend the UTC make the following disposition:
Since the comments have been forwarded to the project editor for ISO/IEC, who has incorporated them into the draft disposition of ballot comments to CDAM1 in WG2 N5173, we recommend the UTC notes the comments in Section 6c of L2/22-023 but takes no further action.

6d Lam-Alef Ligature for al-Dani

Document: L2/22-025 Clarification on spelling of lam-alef ligatures for al-Dani -- Lorna Priest Evans

Comments: We reviewed this document that requests clarification on the spelling of *lam-alef* ligatures for the al-Dani orthography. The issue arose as a result of the question posed in the 2021 Public Review feedback submitted by Patrik Sjöwall (see SAH recommendations L2/21-130), who asked how an attached *fatha* or dot would behave in a *lam-alef* ligature.

Evans researched the topic and noticed that the *lam-alef* in al-Dani varies from its presentation in Hafs (the default Koranic tradition): in Hafs, the *lam* in a *lam-alef* ligature is on the right, but in al-Dani, *lam* is on the left. Putting diacritics on the *lam* in a *lam-alef* ligature would hence vary between the two orthographies.

We agreed the difference in the two orthographies needs to be documented in the Core Spec with examples. Also, we recommend the text specify that logical order should be used to represent al-Dani style texts. In addition, we recommend the current wording about the *lam-alef* ligature in section 9.2 "Arabic Ligatures: Ligature Classes" of *TUS* should be loosened, since not all styles of Arabic ligate *lam* and *alef* and the *lam-alef* ligatures are not always considered obligatory.

Recommendation: We recommend the UTC make the following disposition:
Action Item for Lorna Evans: Provide text with examples of *lam-alef* ligature with diacritics for the Hafs and al-Dani orthographies and propose wording for the Ligature Classes subhead of chapter 9.2. (Reference: L2/22-025)

6e Quranic Superscript Alef Motahafar

Documents:
L2/21-204 Proposal to encode Quranic Superscript Alef Motahafar used in Quran published in Libya -- Lateef Shaikh
L2/21-239 Comments on L2/21-204 Quranic Superscript Alef Motahafar used in Quran published in Libya -- Putten

Comments: We reviewed L2/21-204, which proposed an ARABIC SUPERSCRIPT ALEF MOTAHAFAR character used in the North African tradition of writing the Quran. The proposed character was discussed earlier (see October 2021 Script Ad Hoc Recommendations L2/21-174) and appeared in earlier documents (as X7 in L2/19-306 and characters #28 and #29 in L2/15-329).

On the question of encoding, various options were discussed.

- Marijn van Putten (L2/21-239) was against encoding the character, instead recommending unifying it with U+0670 ARABIC SUPERSCRIPT ALEF and having the font handle the rendering of superscript *alef* next to *lam*. However, the approach to Arabic encoding in the past has been that if two characters look different enough, they should be separately encoded, even if they are semantically equivalent.
- Another suggestion was to unify it with U+076A ARABIC LETTER LAM WITH BAR or a new LAM-plus-mark character.

- Roozbeh Pournader favored a new combining character, which concurred with Ben Yang's analysis of the various options.

We recommend the author of L2/21-204 provide justification for the character as combining as opposed to a precomposed, atomic character, based on the analysis by Ben Yang.

On the name, we find the proposed name "Alef Motahafar" as not immediately transparent, so an English name was recommended.

Suggested names include:
ARABIC COMBINING ALEF OVERLAY
COMBINING ARABIC ALEF OVERLAY
ARABIC ALEF OVERLAY
ARABIC SMALL ALEF
ARABIC STRIKING SUPERSCRIPT ALEF FOR LAM

(For a name that does not include *lam*, an annotation should be included in the names list that the character is only used to annotate *lam*.)

If the character is encoded, we recommend the block introduction discuss how the character interacts with combining marks.

Ben Yang also requested evidence of typographic forms of *dagger alif* when rendering "Allah" in various Quranic traditions.

We also recommend the author change the codepoint to be in the SMP, at U+10EFC.

Recommendations: We recommend the UTC make the following disposition:
Action Item for Ben Yang: Relay the comments above (including his analysis) to the proposal author of L2/21-204 and ask him to update his proposal.

---

7 Linear Elamite

Document: L2/21-233 Preliminary proposal to encode Linear Elamite in Unicode -- Pandey

Comments: We reviewed this preliminary proposal for Linear Elamite. This document is intended to introduce the script to the Script Ad Hoc and UTC and make a change to the Roadmap to better reflect the size of the repertoire. Currently Linear Elamite extends from U+1C380..U+1C3CF. Based on this document, we recommend the allocation be changed to U+1C380..1C4FF.

Recommendations: We recommend the UTC make the following disposition:
Action Item for Ken Whistler and the Roadmap Committee: Update the Roadmap to reflect the allocation for Linear Elamite from U+1C380..U+1C3CF to U+1C380..1C4FF. (Reference: Section 7 of L2/22-023)

IV. SOUTH AND CENTRAL ASIA

8 Devanagari

Document: L2/21-240 Proposal to encode the Devanagari letter Rajasthani SA in Unicode -- Biswajit Mandal

Comments: We reviewed this request to encode DEVANAGARI LETTER RAJASTHANI SA.

The following reflects comments raised during discussion:

- Include a section on conjunct formation and give an example of what a half-form would look like. For implementers it would be useful to also list attested conjuncts or syllables.
- "Rajasthani" in the character name suggests the character is only used for the Rajasthani language (though note that "Rajasthani" is a macrolanguage in Ethnologue). According to the proposal, the character is used for other languages in the Rajasthani region. No good alternative name was suggested by the Script Ad Hoc, however.
- Can the author provide contrastive evidence of the proposed character and श्ा (other than in a chart)?
- The glyph should be made compatible with other Devanagari characters in the code charts.

Recommendation: We recommend the UTC make the following disposition:
Action Item for Debbie Anderson: Relay comments in Section 8 of L2/22-023 to the author of L2/21-240.

9 Kannada and Telugu

Document: L2/22-006 Proposal to encode ARCHAIC SHRII in Kannada and Telugu -- Srinidhi and Sridatta

Comments: We reviewed this proposal to add an ARCHAIC SHRII character for Kannada and Telugu. The sign is reported to have evolved from a form in the 11th-12th centuries.

The proposal includes many examples of the ARCHAIC SHRII character vs. the typical way to write SHRII (i.e., figures for Kannada, 4, 10, 11, 15-18). ARCHAIC SHRII has features of a symbol: it can begin or end texts and is used as a space filler.

The following summarizes the comments:

- In our opinion, compatibility decompositions are not needed
- Cibu found a similarly shaped SHRII character in Malayalam. He is invited to write a proposal for it.
- Rather than encoding a single character for the different scripts -- which raises the question where it would be located -- we agree that two separate characters be encoded, following the precedent set by the "siddham" sign (Telugu U+0C77; Devanagari U+A8FC, Sharada U+111DB,

Kannada U+0C84) and OM. The naming follows that of OM (i.e., the name does not contain "SIGN").
- The glyph for Telugu may need to be adjusted to fit with the current Telugu code chart glyphs.

Recommendation: We recommend that the UTC make the following disposition:
***SAH-UTC170-R6*:** Accepts the following two characters, as documented in L2/22-006, for a future version of the standard:
0CDC KANNADA ARCHAIC SHRII
0C5C TELUGU ARCHAIC SHRII

Action Item for Ken Whistler: Update the Pipeline to include U+0CDC KANNADA ARCHAIC SHRII and U+0C5C TELUGU ARCHAIC SHRII, as documented in L2/22-006.
Action Item for Debbie Anderson and Srinidhi/Sridatta: Provide a font to Michel Suignard (Reference: L2/22-006 and Section 9 of L2/22-023.)

---

10 Mongolian

Document: L2/21-244 Proposal to encode 4 Mongolian characters in UCS – Kushim Jiang

Comments: We reviewed this proposal for four Mongolian characters, which the author noticed when reviewing a manuscript. The four characters are identified as "Ali Gali" letters, a set of letters which were used to transcribe Tibetan and Sanskrit texts.

The following captures the comments made during discussion:

- The proposal author has relayed to Liang Hai that the proposed TODO ALI GALI PHA character is potentially a variant of the existing U+184C TODO PA, but there are manuscripts that distinguish the two forms. In addition, the China national standard for Todo, GB/T 36649-2018, has included the forms as second and third forms of U+1892 ALI GALI PA. This information should be discussed in the proposal.
- Provide clarification on the relation between TODO ALI GALI PHA and the existing U+184C TODO PA.
- Provide an analysis for ALI GALI O and ALI GALI OO and discuss whether these could be handled as sequences.
- In charts 2 and 3, explain more fully what the "Comments" column indicates. What is "[P]"?

We welcome continued discussion on how to represent Ali Gali texts.

Recommendation: We recommend the UTC make the following disposition:
Action Item for Liang Hai: Relay comments in Section 10 of L2/22-023 to the author of L2/21-244.

---

11 Sunuwar

Document: L2/21-157R Proposal to encode the Sunuwar script in Unicode -- Pandey

Comments: We reviewed this revised version of the Sunuwar script proposal, which has been seen several times by the Script Ad Hoc.

The following summarizes the recent changes, based on discussion with the Script Ad Hoc:

- U+003A COLON has been recommended for the vowel length mark (*laissi*).

- For the *taslathenk*, U+0310 COMBINING CANDRABINDU is suggested, though the glyph for COMBINING CANDRABINDU varies from the shape of the *taslathenk*, so a font would need to change the glyph to better reflect the original shape. The *taslathenk* -- like other diacritics listed in section 7.2 -- were part of an orthographic experiment by Rapaca and were not widely adopted.

  Since the other members of the set of legacy diacritics (*sangmilu*, *sangkirs*, and *sangrums*) can readily be represented with common diacritics, we agree that U+0310 COMBINING CANDRABINDU could be used. If an encoding issue arises, a script-specific character could be proposed, if needed.

Recommendations: We recommend the UTC make the following disposition:
***SAH-UTC170-R7*:** Accepts 44 Sunuwar characters in a new Sunuwar block (U+11BC0..U+11BFF), as documented in L2/21-157R, for encoding in a future version of the standard.

Action Item for Ken Whistler: Update the Pipeline to include 44 Sunuwar characters, as documented in L2/21-157R.
Action Item for Debbie Anderson and Anshuman Pandey: Send Michel Suignard a font. (Reference: L2/21-157R)

---

12 Tulu / Tulu-Tigalari

Documents:
L2/22-034 Proposal to Encode Tulu Script in Unicode -- Dr. Akash Raj Jain and Karnataka Tulu Saahithya Academy (KTSA)
L2/22-033 Approving TULU UNICODE SCRIPTS by UNICODE CONSORTIUM -- Beluru Sudarshana
L2/22-031 Updated proposal to encode the Tulu-Tigalari script in Unicode -- Murthy/Rajan
L2/22-032 Further Response to Tulu Academy Documents – Murthy/ Rajan
(See also Section 21f Feedback section below for feedback on Tulu-Tigalari)
Related documents:
L2/20-279 Comments on differences between Tulu and Tigalari proposals -- Kučera
L2/21-019 Proposal to encode Tulu -- Pavanaja
L2/21-188 Tulu documents -- Akashraj Jain
L2/21-189 Tulu Lipi Parchaya (translation) -- Radhakrishna Bellur, Nischith Ramakunja

Comments: We reviewed the various Tulu and Tulu-Tigalari documents. Two separate proposals were submitted: one for Tulu from the Karnataka Tulu Saahithya Academy (KTSA) for modern Tulu and one

for Tulu-Tigalari from Vaishnavi Murthy and Vinodh Rajan to aid in the digitization of manuscripts. Two supplemental documents were also submitted, also discussed below.

Tulu proposal from KTSA (L2/22-034) and letter of support (L2/22-033)
The new document from Karnataka Tulu Saahithya Academy (KTSA) proposes "Tulu" be encoded. The discussion of the name appears on page 6 and pp. 24-27. Some examples of inscriptions are presented on pp. 8-10 (with small images). To see examples from modern signage one needs to refer to L2/21-188 (pp. 101ff.) and the discussion about the script (with images of inscriptions) in L2/21-189. A letter from the advisor to the Chief Minister of Karnataka State in support of the KTSA proposal was also received (L2/22-033).

The following comments were noted:

- The current KTSA Tulu proposal does not provide a summary code chart for easy reference, but refers to glyphs printed one-per-page in L2/21-188 (on pp. 1-99). This approach makes it very difficult to follow the overall structure of the proposal. For example, a reviewer cannot easily see all the vowel signs that are proposed, nor can they quickly grasp the independent vowels versus dependent and "part-vowel formations" being proposed, or any other signs that are proposed. Adding to the confusion, the list of characters in that document contains glyphs beyond what is included in Unicode code charts (e.g., it includes some conjuncts [pp. 74-88]) and does not include code points.

- The current KTSA proposal appears to contain similar script information to that submitted earlier, L2/21-019, but because the font isn't embedded in the latest proposal, the glyphs for Tulu don't appear, which makes it difficult to do a comprehensive review. In addition, having evidence spread across different documents (such as L2/21-188 and L2/21-189) makes it challenging to review. Feedback based on the review by Jan Kučera in L2/20-279 has not been addressed, most notably the missing evidence for some of the characters at the top of page 7. This is most important, as evidence for all proposed characters needs to be clearly documented.

- The proposed code points (U+11B50..U+11BAF) are not on the Roadmap for Tulu. Any subsequent proposal should use code points XXXX0, XXXX1, etc. Such a proposal should include the list of characters, their names and properties, and all the evidence in one document, as well as specific details on the difference in the behavior between modern Tulu and Tulu-Tigalari (as proposed in L2/22-031).

- In our analysis, the "Tulu" script for modern Tulu appears to be based on the historic script proposed by Murthy/ Rajan, which is intended specifically for digitization of manuscripts. If specific evidence is provided identifying differences in the behavior of modern Tulu, separate encoding for modern Tulu may be considered. To state it in another way, if the modern script deviates in behavior from the historic script, that might provide evidence suggesting that encoding a separate, modern Tulu script may be needed.

Response from Murthy/Rajan to SAH comments (L2/22-032)
We reviewed this document from authors Murthy/Rajan, which responded to questions posed in the Script Ad Hoc recommendations L2/21-174 (which in turn responded to documents submitted by the Karnataka Tulu Academy).

The authors confirmed their proposal covers all the manuscript and epigraphical evidence, and included an example showing how the Academy proposal would not represent an inscription well. They also did not agree that a joiner just for historic use would be a simple solution, as other changes would also be needed.

Tulu-Tigalari proposal (L2/22-031)

The proposal from Murthy and Rajan is "for the archaic Tulu-Tigalari script as seen used predominantly in hand-written manuscripts" (page 5). It is a revised version of L2/21-210, and it takes into account comments from the October 2021 Script Ad Hoc recommendations (page 16 of L2/21-174). Changes in the document included removal of PUSHPA and TIDDU from the proposal.

(Note: The first two comments below have already been incorporated in the posted version of the proposal, L2/22-031.)

- After a lengthy discussion in the Script Ad Hoc, we recommend canonical decompositions for the vowels U+11383 LETTER II, U+11385 LETTER UU, U+1138E LETTER AI, U+11391 LETTER AU, U+113C5 VOWEL SIGN AI. (VOWEL SIGN OO and VOWEL SIGN AU already have canonical decompositions.) With these decompositions, Do Not Use tables are not necessary.
- A minor correction on page 13 is needed to clarify that AU length mark is not used on its own.
- On the name: Calling the script just "Tigalari" could cause confusion, because in some parts of the Karnataka state "Tigalari" refers to Tamil Grantha (see also figure 5 in the proposal). An explanation in the block introduction could be added to explain the script's name and describe its usage, which is historic. Editorial comments such as these are often used when documenting complex scripts encoded in the Unicode Standard.

In our view, the Tulu-Tigalari proposal is technically sound and provides the details necessary for encoding this archaic script.

Recommendations: We recommend the UTC make the following disposition:
SAH-UTC170-R8: Accepts 78 Tulu-Tigalari characters in a new Tulu-Tigalari block U+11380..U+113FF. as documented in L2/22-031, for a future version of the standard.

Action Item for Ken Whistler: Update the Pipeline to include 78 Tulu-Tigalari characters, as documented in L2/22-031.
Action Item for Vaishnavi Murthy and Debbie Anderson: Send a font to Michel Suignard. (Reference: L2/22-031)
Action Item for Debbie Anderson: Reply to KTSA with feedback regarding creating a more accessible proposal with a clear list of characters and glyphs and identification of any differences in behavior or appearance with the historic Tulu-Tigalari writing system and other comments from Section 12 of L2/22-023.
Action Item for Norbert Lindenberg: Propose text in section 2.11 of the Core Spec on how to handle sequences with multiple left-reordering dependent vowels. (Reference: Section 12 of L2/22-023)

V. SOUTHEAST ASIA, INDONESIA, AND OCEANIA

13 Khmer

Document: L2/21-241 Khmer Encoding Structure – Hosken et al.

Comments: We reviewed this document on Khmer's encoding structure that presented questions and background on Khmer issues.

The following highlights the discussion:

- In terms of background, the current Khmer text in the Unicode Standard dates to 4.0 in 2003. The Unicode 4.0 acknowledgements note that the official Cambodia representatives and members of the Japanese National Body were very involved.
- To more effectively get a response from the Script Ad Hoc and UTC, we recommend re-framing the questions to ask specific questions. For example, instead of asking "How complex do we want to make the consonant shifter rules?" it would be preferable to offer different options (with pros and cons), and ask the Script Ad Hoc for their input on which option they recommend. If specific changes are requested for the Core Spec, provide specific wording. Note that Core Spec changes could be done relatively soon, unless there is a technical change (which would need the UTC's okay).
- One key issue raised was how to handle Middle Khmer, which is more complicated than Modern Khmer or Old Khmer. Because the number of Middle Khmer users is very limited (compared to Modern Khmer), a number of participants recommended focusing on Modern Khmer, postponing Middle Khmer and Old Khmer until later.
- It was mentioned that Middle Khmer might align with the Tham support being discussed for the Universal Shaping Engine.
- Options mentioned for Middle Khmer include:
    - Use of language tags for Middle Khmer. However, language tags were cautioned against. Language information is not available or easily lost in many environments. Supporting them in OpenType in this way would require major changes.
    - Use of separate script tags, which would let fonts identify themselves as supporting Modern or Middle Khmer (similar to Indic 1 vs Indic 2). As a result, opting into different validation might work better.
    - It might be better to focus support on Modern Khmer only, and have fonts hack in support for Middle Khmer e.g., by removing dotted circles, as some SIL Myanmar fonts do already.
- Currently, Unicode, OpenType and Cambodian keyboard documentation disagree on cluster structures, and implementations disagree on cluster structure validation. A well-defined standard cluster structure is needed, but where is the most appropriate location for such information (CLDR or Core Spec)?

  There was no agreement on whether Unicode needs a general framework for cluster validation before Khmer-specific regular expressions can go into the spec.

Recommendations: We recommend the UTC make the following disposition:
Action Item for Norbert Lindenberg: Relay the comments in Section 13 of L2/22-023 to the proposal authors.

VI. EAST ASIA

14 Ideographic Complex Scripts

Document: L2/21-165 Preliminary proposal on encoding method for ideographic complex scripts(s) - Eiso Chan

Comments: We reviewed section 3 of this document, which proposed an encoding method for ideographic complex script(s), including early Chinese organic chemical characters, characters used in transcriptions of Sanskrit, Tibetan and Tangut, and Jianzi Musical Notation.

In our view, the various sections are different and should be handled separately. These are discussed below.

3.1.1. Early Chinese chemical characters
Rather than provide a new approach to constructing the chemical characters with joiners, we recommend the author provide a list of the atomic signs, and the IRG can decide whether to encode them or not.

3.1.2. Sanskrit and Tibetan transcription
In our view, a strong case to interchange such transcriptions as text has not been made. Rather, such text can be handled by sequences of characters (similar to *fanqie*, which uses two characters to represent a sound in Chinese). For the transcriptions, an image could be employed, alongside the sequence of characters.

3.1.3. Tangut
The Tangut transcriptions may require atomic encoding. However, the document contains only a small sample of the Tangut material. More information is needed, with analysis, and a full proposal.

3.1.4. Jianzi Musical Notation
Jianzi Musical Notation should be handled within the overall discussion of musical notation. Unlike *fanqie*, Jianzi is a closed system. It was noted that musical systems are not well-supported in implementations.

Also, future documents should refer to the earlier proposal for Jianzi L2/19-107 [=WG2 N5041] (discussed in SAH recommendations L2/19-286), and the comments from China National Database of Characters Program (WG2 N5074)

Recommendation: We recommend the UTC make the following disposition:
Action Item for Debbie Anderson or Liang Hai: Relay comments in Section 14 of L2/22-023 to the author of L2/21-165.

VII. SYMBOLS, PUNCTUATION, AND NOTATIONAL SYSTEMS

15 Blissymbols

Document: L2/22-003 On radicals for lexicography in Blissymbols – Everson

Comments: We reviewed this document, which provided background on radicals that are used for ordering and searching Blissymbolic characters. Approximately one-third of Bliss radicals are used only as radicals. This document asks whether 15 radicals listed on pages 13-14 should be encoded, since they were used historically in texts, but have since been "retired."

We agree they should be encoded: the radicals appeared in published materials and would allow users to cite them when discussing the history of sorting Bliss symbols.

Recommendations: We recommend the UTC make the following disposition:
Notes this document but take no further action.

---

16 Dwarf Planet Symbols

Document: L2/21-224 Unicode request for dwarf-planet symbols -- Miller

Comments: We reviewed this request for five dwarf planet symbols.

The following points were raised during discussion:

- Criteria for encoding new symbols includes their usage in text, particularly as evidenced in use by people (such as people discussing the symbol), or to fill out a set of currently encoded characters. In this case, the characters do fill out a set. The evidence provided is "text-ish" (such as figure 7). Inclusion in a font is not itself a strong rationale.
- The characters are not intended to serve as emoji.
- We recommend the characters be located at the end of the Alchemical Symbols block (i.e., from U+1F77B..U+1F77F).

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

SAH-UTC170-R9: Accepts the following five characters, as documented in L2/21-224, for a future version of the standard:
1F77B HAUMEA
1F77C MAKEMAKE
1F77D GONGGONG
1F77E QUAOAR
1F77F ORCUS

Action Item for Ken Whistler: Update the Pipeline to include five dwarf-planet symbols, as documented in L2/21-224.

---

17 Legacy Computing Symbols

Document: L2/21-235 Proposal to add further characters from legacy computers and teletext (Dec 2021 proposal)

Comments: We reviewed this proposal for 731 characters used on home computers manufactured from ca. mid-1970s to the mid-1980s and symbols used in the teletext broadcasting standard developed in the early 1970s. Mapping tables between the legacy character sets and the proposed symbols are provided.

The Script Ad Hoc has seen earlier versions of this proposal.

The proposed repertoire contains additions to the following blocks: Control Pictures, Supplemental Arrows-C, and Symbols for Legacy Computing, and a new block, Symbols for Legacy Computing Supplement (U+1CC00..U+1CEAF).

The following points were raised during discussion:

- The document adequately addresses the question whether more such requests will appear in the future in section 8 (page 5), at least to most members of the Script Ad Hoc.
- The focus of this proposal is on devices designed to interchange data with other devices (hence it does not cover arcade games, gaming consoles, etc.).
- The proposal does not include the Powerline symbols (proposed in L2/19-068R and discussed in SAH recommendations L2/19-343), as they are not related to legacy computer symbols.
- Note that the outlined Latin capital letters (U+1CCD6..U+1CCEF) have been given general category property Lu. These stylized Latin letters should probably ***not*** be treated as uppercase Latin letters, but instead should follow the pattern of the Enclosed Alphanumerics, which have gc=So. The outlined digits (U+1DDF0..U+1CCF9) have numeric properties as digits and a compatibility decomposition, which is acceptable.

  A revision to the proposal should be made addressing the general category property of these symbols and the properties should be changed appropriately.

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

***SAH-UTC170-R10*:** Accepts 731 legacy computing symbols, as documented in L2/21-235, for a future version of the standard, but changing the gc for the outlined Latin capital letters U+1CCD6..U+1CCEF from Lu to So.

***SAH-UTC170-R11*:** Accepts a new block allocation, Symbols for Legacy Computing Supplement (U+1CC00..U+1CEAF) (Reference: L2/21-235)

Action Item for Ken Whistler: Update the Pipeline to include 731 legacy computing symbols, as documented in L2/21-235.
Action Item for Ken Whistler and Debbie Anderson: Confirm the Roadmap is updated with Symbols for Legacy Computing *Supplement* (U+1CC00..U+1CEAF) (Reference L2/21-235)
Action Item for Debbie Anderson and Rebecca Bettencourt: Provide Michel Suignard with a font. (Reference L2/21-235)

Action Item for Doug Ewell and Rebecca Bettencourt: Update the proposal with discussion on the gc properties for the outlined Latin capital letters (U+1CCD6..U+1CCEF) and to adjust the properties accordingly (from gc=Lu to gc=So). (Reference: Section 17 of L2/22-023)

---

18 Lot of Fortune and Eclipse Symbols

Document: L2/22-005 Unicode request for Lot of Fortune and eclipse symbols -- Miller

Comments: We reviewed this document, which requested three astrological characters. Examples are provided. The document also discusses why Lot of Fortune should not be unified with U+2297 CIRCLED TIMES.

Recommendation: We recommend the UTC make the following disposition:
SAH-UTC170-R12: Accepts the following three characters, as documented in L2/22-005, for a future version of the standard:
1F774 LOT OF FORTUNE
1F775 OCCULTATION
1F776 LUNAR ECLIPSE

Action Item for Ken Whistler: Update the Pipeline with three astrological symbols, as documented in L2/22-005.

---

19 Punctuation delete mark

Document: L2/21-245 Proposal for the inclusion of the DELETE SIGN for proofreading & discussion of the intended use and behavior of already encoded signs -- Marín Silva

Comments: We reviewed this document that argues for a DELETE SIGN.

The following comments were raised during discussion:

- While we agree the concept of “deletion mark” exists -- as does a sign (“signifier”) for the concept -- the existence of a concept and sign does not necessarily mean that the sign should be encoded as a character in plain text.
- The deletion mark is one of the copyediting marks making up one *layer* of text (cf. use of color in figures 4-9). Encoding such a mark would be useful as part of a list of copyediting marks, for example, or as part of a system similar to the set of graphic elements that have been encoded for music, which require higher-level protocols to represent musical data and musical scores.
- The open slot at U+2065 between INVISIBLE PLUS and LEFT-TO-RIGHT ISOLATE is a code point in a range reserved for characters with the Default_Ignorable_Code_Point property, and hence would not be a suitable candidate for the character in any event.

(Note: L2/21-245 is a duplicate of L2/21-230)

Recommendation: We recommend the UTC make the following disposition:
Action Item for Rick McGowan: Relay comments in Section 19 of L2/22-023 to the author of L2/21-245.

---

20 Smalltalk

Document: L2/21-234 Proposal to add characters from Smalltalk (with attachments)

Comments: We reviewed this proposal to add 5 characters for compatibility with versions of the Smalltalk programming language. Smalltalk was originally developed in the 1970s. Mapping tables between Smalltalk and the proposed characters are attached to the proposal.

The Script Ad Hoc has seen earlier versions of this proposal.

Three characters that were proposed in earlier versions of the proposal are now recommended to be handled as sequences with ZWJ. The three characters are: APOSTROPHE S OPERATOR, LEFT AND RIGHT PARENTHESIS, and RIGHTWARDS ARROW TO LEFT PARENTHESIS. The authors of the proposal report they don't foresee a problem with sequences and ZWJ. The Smalltalk community is invited to document the recommended representation with these sequences.

Recommendation: We recommend that the UTC make the following disposition:
***SAH-UTC170-R13*:** Accepts the following 5 Smalltalk symbols, as documented in L2/21-234, for a future version of the standard:
1CEB0 HORIZONTAL ZIGZAG LINE
1CEB1 KEYHOLE
1CEB2 OLD PERSONAL COMPUTER WITH MONITOR IN PORTRAIT ORIENTATION
1CEB3 BLACK RIGHT TRIANGLE CARET
1F8B2 RIGHTWARDS ARROW WITH LOWER HOOK

Action Item for Ken Whistler: Update the Pipeline with five Smalltalk symbols as documented in L2/21-234.
Action Item for Debbie Anderson and Rebecca Bettencourt: Provide Michel Suignard with a font. (Reference: L2/21-234)

VIII. PUBLIC REVIEW FEEDBACK

21a Arabic Presentation Forms-A

Document: L2/21-169 Comments on Public Review Issues (July 20 - Sept 25, 2021)

Comments: We reviewed the two pieces of feedback from Brian Sullendar (timestamps: Wed Sep 8 04:10:47 CDT 2021, Thu Sep 9 02:50:37 CDT 2021), who found characters in Arabic Presentation Forms-A (U+FC03 and U+FBF9 / U+FBFA and U+FC68 [see below]) that all share the same compatibility decompositions, but have different glyphs. He wondered if this was an error.

FC03 ﰃ ARABIC LIGATURE YEH WITH HAMZA ABOVE WITH ALEF MAKSURA ISOLATED FORM
≈ <isolated> 0626 ئ 0649 ى

FBF9 ﯹ ARABIC LIGATURE UIGHUR KIRGHIZ YEH WITH HAMZA ABOVE WITH ALEF MAKSURA ISOLATED FORM
≈ <isolated> 0626 ئ 0649 ى

FBFA ﯺ ARABIC LIGATURE UIGHUR KIRGHIZ YEH WITH HAMZA ABOVE WITH ALEF MAKSURA FINAL FORM
≈ <final> 0626 ئ 0649 ى

FC68 ﱨ ARABIC LIGATURE YEH WITH HAMZA ABOVE WITH ALEF MAKSURA FINAL FORM
≈ <final> 0626 ئ 0649 ى

The characters cited derive from an unanalyzed collection of ligatures that came from IBM Egypt and date to Unicode 1.1. The decompositions for U+FBF9 and U+FBFA date to Unicode 3.0 and those for U+FC03 and U+FC68 date to earlier versions.

Note that characters with two identical compatibility decompositions may be displayed with different glyphs: cf.

1D41A 𝐚 MATHEMATICAL BOLD SMALL A
≈ <font> 0061 a latin small letter a

1D44E 𝑎 MATHEMATICAL ITALIC SMALL A
≈ <font> 0061 a latin small letter a

The characters cited in the feedback are different forms of ligatures that happen to share the same sequence, but may have distinct presentation forms. Users should avoid using the compatibility Arabic ligatures in the Arabic Presentation Forms-A block, which were encoded primarily for compatibility.

To address the question posed in the feedback, annotations should be added to the relevant Arabic Presentations Forms-A characters, mentioning that the same sequence may have distinct presentation variants, and different forms of ligatures may be needed.

Recommendations: We recommend the UTC make the following disposition:
Action Item for Roozbeh Pournader: Provide annotation to the names list in Arabic Presentation Forms-A, mentioning that the same sequence may have distinct presentation variants, and different forms of ligatures may be needed. (Reference Section 21a of L2/22-023)

21b Latin Additional Letters

Document: L2/22-018 Comments on Public Review Issues (Sept 25, 2021 - Jan 17, 2022): Feedback October 13, 2021 from Eduardo Marín Silva “Suggestion on the encoding of Latin Theta”

Feedback received:

In the document L2/21-206, it is suggested to encode a Latin casing pair for Theta. The difference with the Greek pair Θθ (0398 and 03B8) is that the capital form, always has a horizontal stroke that touches both sides of the letter, while the Greek letter can have a shorter stroke with its own serifs. Another similar pair is the Latin Ɵɵ (019F and 0275), the difference with this pair, is that the lowercase is at x height, while the orthography requires a tall glyph like the proper Greek small theta. Encoding a new pair is problematic, since phonetic notations already use the Greek codepoint (03B8) for the same sound. I propose some possible solutions.

1. Use the Greek pair: In order to force the preferred glyph for Latin based orthographies, a SVS can be added, called "latin form" or "long stroke form". This would mean that the default glyph is still what the Greek users expect, and the small Theta remains untouched.
2. Use the Latin barred o pair: Similarly, in order to force the preferred glyph on the lowercase, a SVS can be added, called "theta form", "tall form" or "elongated form". This has the benefit of keeping the text completely Latin. Characters that are confusable with others, but only in certain contexts, is not new.
(Deciding between 1 or 2 depends of what the users prefer in case the default glyph has to be displayed; either an uppercase with a shorter stroke or a shorter lowercase)
3. Just encode a small Latin Theta and make it an alternate lowercase to 019F: Such a solution is my least preferred one, since it has the same downsides as just encoding a new Latin pair.
4. Bite the bullet and encode the new pair: It wouldn't be the first time confusable characters are disunified due to problematic casing relations.

All other letters in the document are acceptable, but I would rename the first pair as LATIN CAPITAL/SMALL LETTER REVERSED GLOTTAL STOP. They should be disunified on the same basis of the regular glottal stop pair.

Comments: We briefly reviewed the feedback.

We recommend these alternative suggestions in the feedback be forwarded to Denis Moyogo Jacquerye for consideration without comment.

Also, it was noted that use of Standardized Variation Sequences for bases that are part of a case-pair is an architectural issue that has not yet been addressed. (Note: The Script Ad Hoc comments on Latin theta are contained in L2/21-174.)

Recommendation: We recommend the UTC make the following disposition:
Action Item for Rick McGowan: Relay comments in Section 21b of L2/22-023 to the author of October 13, 2021 feedback on Latin theta contained in L2/22-018.
Action Item for Debbie Anderson: Relay the feedback from Eduardo Marín Silva in Section 21b of L2/22-023, with no comment to Denis Moyogo Jacquerye.

21c Legacy Malayalam Characters

Document: L2/22-018 Comments on Public Review Issues (Sept 25, 2021 - Jan 17, 2022): Feedback Nov. 16 2021 from Jack Varanelli “Unicode request for legacy Malayalam”

Comments: We reviewed feedback from Jack Varanelli, who noted that the character U+1DF27 LATIN SMALL LETTER N WITH LEFT HOOK has the same name as U+0272, a voiced palatal nasal. However, the Latin characters for Malayalam represent retroflex consonants, not palatals. (The Latin character for Malayalam was one of a set of six that were proposed in L2/21-156 and were approved at the Oct. 2021 UTC meeting.)

The character for the palatal, U+0272 ɲ, has a baseline left hook, presumably based on U+0321  COMBINING PALATALIZED HOOK BELOW. Most other “WITH LEFT HOOK” characters in Unicode also have a baseline left hook (U+019D Ɲ, U+0272 ɲ, U+0528 Ԩ, U+0529 ԩ, and U+1DAE ᶮ), except U+AB52 ꭒ SMALL LETTER U WITH LEFT HOOK, which more closely resembles the Latin character for Malayalam.

In order to make clear the Latin Malayalam characters have no relation to the palatalized hook, we recommend that all six new characters be renamed from “…WITH LEFT HOOK” to “..WITH MID-HEIGHT LEFT HOOK.”

Recommendations: We recommend the UTC make the following disposition:
SAH-UTC170-R14: Modify the names of the following six characters (which had been earlier accepted as consensus 169-C6) from:
1DF25 LATIN SMALL LETTER D WITH LEFT HOOK
1DF26 LATIN SMALL LETTER L WITH LEFT HOOK
1DF27 LATIN SMALL LETTER N WITH LEFT HOOK
1DF28 LATIN SMALL LETTER R WITH LEFT HOOK
1DF29 LATIN SMALL LETTER S WITH LEFT HOOK
1DF2A LATIN SMALL LETTER T WITH LEFT HOOK

to:
1DF25 LATIN SMALL LETTER D WITH MID-HEIGHT LEFT HOOK
1DF26 LATIN SMALL LETTER L WITH MID-HEIGHT LEFT HOOK
1DF27 LATIN SMALL LETTER N WITH MID-HEIGHT LEFT HOOK
1DF28 LATIN SMALL LETTER R WITH MID-HEIGHT LEFT HOOK
1DF29 LATIN SMALL LETTER S WITH MID-HEIGHT LEFT HOOK
1DF2A LATIN SMALL LETTER T WITH MID-HEIGHT LEFT HOOK
Reference: Section 21c of L2/22-023.

Action Item for Ken Whistler: Update the Pipeline with name changes to six Latin characters for legacy Malayalam, as documented in Section 21c of L2/22-023.

21d Old Hungarian

Document: L2/21-246 Feedback: Proposal for a compromise of the recent Old Hungarian proposal -- Eduardo Marin-Silva

Comments: This feedback offered suggestions based on the Old Hungarian proposal (L2/21-115). (Note the SAH comments on L2/21-115 are contained in L2/21-130.)

We reviewed this document which had suggested a note at the top of the names list, glyph changes for CLOSED EH, annotations and name aliases for various characters, and support for additional characters proposed in L2/21-115.

The following comments were made during discussion of the feedback:

- Some editorial suggestions in sections 1, 3, and 7 will be taken under advisement by the names list editor. Note that formal name aliases for certain characters in section 3 would not be appropriate. Formal name aliases are reserved for actual errors.
- On the glyph changes in 2 and the recommendations to encode characters in sections 4-6, we seek more input from other Old Hungarian experts on the advisability of adding more characters and making the glyph changes.

Recommendations: We recommend the UTC make the following disposition:
Action Item for Rick McGowan: Relay comments in Section 21d of L2/22-023 to the author of L2/21-246.
Action Item for Ken Whistler and the Editorial Committee: Take into account the editorial suggestions in sections 1, 3, and 7 of L2/21-246.

21e Symbol for PLAY

Document: L2/22-018 Comments on Public Review Issues (Sept 25, 2021 - Jan 17, 2022): Feedback Sept. 29, 2021 from Marín Silva “On the apparent arbitrary exclusion of the play button as a distinct character on the face of duplicates”

Feedback received:

> The relevant symbols discussed are old (since at least the period where cassette players where popular); they were later adopted in so many contexts, that they could be said to be universal representations of their respective functions.
> Naturally, since they were (and still are) so important, most of them were assigned a Unicode codepoint on the "Miscellaneous Technical" block, with some being apparently duplicated. Here I proceed to discuss those:
>
> 
>
> Both of them seem to serve the same purpose, with the second set having the term "ISOSCELES RIGHT TRIANGLE" being applied instead of simply "TRIANGLE" to distinguish them. Both sets are isosceles and have right angles so the differences in name are not helpful. In practice, it seems like the first set tends to have consistent advance width with padding at all sides, while the other set tends to have a tight advance width with respect to the glyph, which means that the up and down arrows end up slightly wider than the left and right ones. If this is the "true" difference between them, then the name chosen does not reflect that, and it is unclear why they couldn't be unified anyway. i.e. why was it important to have both sets?

23F9 ⌑ BLACK SQUARE FOR STOP, 25A0 ■ BLACK SQUARE, 25FC ■ BLACK MEDIUM SQUARE, 2B1B ■ BLACK LARGE SQUARE, 2BC0 ■ BLACK SQUARE CENTRED and 1F532 ⌑ BLACK SQUARE BUTTON:
Out of all of them, the most generic is 25A0, perhaps it was disunified into 23F9 because on user interfaces it is important that all buttons have the same width, while 25A0 was free to lack padding at both sides. 2BC0, the "centered" one forms part of a set, where "centered" just means the figures have consistent padding at both sides. The last character is disunified on account of the different function in UI's where it has a dual and represent a selected or unseleceted button. Similarly, while disunifying on account of size makes sense, either 25FC or 2B1B could have been used for the "stop" function if only one of them was declared to be it.
23FA ⌑ BLACK CIRCLE FOR RECORD, 25CF ● BLACK CIRCLE, 26AB ⬤ MEDIUM BLACK CIRCLE and 1F534 🔴 LARGE RED CIRCLE:
A similar situation to the "stop" symbol applies to the "record" one, with one caveat; the symbol is often shown with a red color. With this in mind, not only does it make sense to disunify it from 25CF, it also makes sense to disunify it from 26AB and 1F53A, on account of the stability of their colors. So there are no problematic disunifications here. Except maybe 25CF ● and 2022 • BULLET, but that is independent of the issue at hand.
The only symbol to NOT be disunified was the "play" symbol, the closest matches being 25B6 ⌑ BLACK RIGHT-POINTING TRIANGLE and 2BC8 ► BLACK MEDIUM RIGHT-POINTING TRIANGLE CENTRED. It makes little sense to disunify the symbols already discussed, but not this one. Whatever rationale applied to the other characters, should also apply to this one.
I therefore highly recommend to encode a new symbol, The glyph would harmonize great with the other symbols, since it can have a smaller glyph and the padding necessary at the same time. Disunification also has the benefit of allowing fonts to depict the symbols inside an enclosure by default, since that is what users often expect.
I suggest the name BLACK RIGHT-POINTING TRIANGLE FOR PLAY or BLACK RIGHT-POINTING EQUILATERAL TRIANGLE FOR PLAY. If a separate document needs to be written for it I would gladly do so.

Comments: We reviewed this feedback on the need for a symbol for PLAY and have the following comments:

- We do not see that the current actual digital representation of PLAY is causing a problem in the representation of text or emoji.
- This feedback is not a proposal.

Recommendation: We recommend the UTC make the following disposition:
Action Item for Rick McGowan: Relay comments in Section 21e of L2/22-023 to the author of Sept. 29, 2021 feedback on PLAY contained in L2/22-018.

21f Tulu-Tigalari

Document: L2/22-018 Comments on Public Review Issues (Sept 25, 2021 - Jan 17, 2022): Feedback Sept. 27, 2021 from Marín Silva “On the Tiddu mark and Virama+Repha of Tulu-Tigalari”

Feedback:

> I would also like to suggest to encode one more character, to reproduce the behavior on page 34, where the Virama and the Repha can fuse, despite them not being adjacent in the sequence. Instead, I propose encoding another character called: TULU-TIGALARI VIRAMA WITH REPHA. This would reduce the complexity necessary to input this character. It can have the same properties as the Virama and be placed at 113DE, so no characters need to be shifted from their current positions

Comments: We briefly discussed this feedback to add a VIRAMA WITH REPHA character.

In our opinion, such a character is not needed. It could create problems for the encoding of the script, because it combines characters that could be far from each other in the representation. We would not be able to properly use canonical decompositions or simple "Do Not Use" tables to take care of such alternative encodings. Modern font technology is fully capable of ligating or properly positioning the components of the proposed combined character. (The Tulu-Tigalari proposal discusses *repha* and virama on page 25 of their proposal L2/22-031, and says the combination should be handled at the font level.)

(Note that the feedback above was one of two pieces received; the other feedback was on TIDDU, which has since been removed from the proposal, so it is not discussed in the Recommendations.)

Recommendation: We recommend the UTC make the following disposition**:**
Action Item for Rick McGowan: Relay comments in Section21f of L2/22-023 to the author of Sept. 27 2022 feedback on Tulu-Tigalari contained in L2/22-018.

IX. OTHER FEEDBACK

22 Bopomofo: Change of Vertical_Orientation property for Bopomofo Tone Marks

Document: L2/22-037 Change of Vertical_Orientation property for Bopomofo Tone Marks
(Note: This particular document was not reviewed by the Script Ad Hoc, but it requests an action item be recorded that the SAH had earlier recommended.)

Comments: This document contains an analysis by Ken Whistler of 2020 feedback originally sent to the CJK & Unihan group on the possible change of Vertical_Orientation property for Bopomofo, and then tracks what happened afterwards. The document was created in response to a query from Ken Lunde and others about the status of AI 164-A57.

In sum: Discussion of AI 164-A57 at the October 2020 Script Ad Hoc led to a separate recommendation for an action item that is cited in Section 21 of L2/21-016 of the Script Ad Hoc Recommendations. However, this later action item was never recorded, because the UTC did not take up Section 21 during UTC #166. To resolve the situation, we recommend the UTC record the action item, below.

Recommendation: We recommend the UTC make the following disposition**:**
Action Item for Liang Hai: Write up a document on the background of the situation re: Bopomofo tone marks. (Reference: Section 21 of L2/21-016 Script Ad Hoc Recommendations and Section 22s of L2/22-023).

X. RECOMMENDATIONS FOR UNICODE 15.0 (ETC.)

23a Recommendations for 15.0

The Script Ad Hoc recommends the following characters for inclusion in Unicode 15.0:
Egyptian H format controls L2/21-248 (30 characters)
Egyptian Hieroglyph Variation Sequences L2/22-012 (98 Sequences)
Cyrillic modifier letters L2/22-010 (2 characters)
Dwarf planets L2/21-224 (5 characters)
Lot of Fortune and Eclipse symbols L2/22-005 (3 characters)

23b Recommendations for a future version

We recommend the following scripts and characters for a future version of the standard:
Sunuwar L2/21-157R (44 characters)
Tulu-Tigalari L2/22-031 (78 characters)
Kannada Archaic SHRII and Telugu Archaic SHRII L2/22-006 (2 characters)
Smalltalk L2/21-234 (5 characters)
Legacy computing symbols L2/21-235 (731 characters)

23c Draft Candidates for 15.0 on Pipeline and in CDAM 1

The following are scripts and characters already identified on the Pipeline as Draft Candidates for 15.0 and are contained in CDAM1:
Nag Mundari
Kawi
Kaktovik numerals
Devanagari bhale mindu
Cyrillic modifiers
Latin letters with mid-height left hook (note name change from CDAM1)
Khojki additions
Arabic additions (note 3 characters get moved to new block, a change not yet shown in CDAM1)
Lao Yamakkan
Kannada Sign Combining Anusvara Above Right
HIRAGANA LETTER SMALL KO
KATAKANA LETTER SMALL KO
CJK Extension H