

**TO:** UTC

**FROM:** Deborah Anderson, Ken Whistler, Roozbeh Pournader, Lisa Moore, and Liang Hai<sup>1</sup>

**SUBJECT:** Recommendations to UTC #166 January 2021 on Script Proposals

**DATE:** January 14, 2021

The Script Ad Hoc group met on October 19 and 30, November 9 and 20, December 7 and 18, 2020 and January 5 and 11, 2021, in order to review proposals. The following represents feedback on proposals that were available when the group met.

## Table of Contents

|   |    |
|---|----|
| I. EUROPE .....   | 3  |
| 1 Cistercian.....   | 3  |
| 2 Cypro-Minoan .....  | 3  |
| 3 Latin.....  | 4  |
| 3a. Phonetic characters (extIPA, VoQS, phonetic click letters, IPA retroflex letters and similar letters with hooks)..... | 4  |
| 3b. IPA Modifier Letters - Pulmonic.....  | 6  |
| 3c. IPA Modifier Letters – Non-Pulmonic.....  | 8  |
| 3d. Addendum to Unicode request L2/20-253 .....   | 8  |
| 3e. Modifier Latin Capital Letters .....  | 9  |
| 3f. Subscript Modifier Letters .....  | 9  |
| 3g. Phonetic Punctuation and Diacritics .....   | 10 |
| 3h. Additional Para-IPA Letters.....  | 11 |
| 3i. Dezh with Retroflex Hook .....  | 11 |
| 3j. Old Polish Nasal Vowel Letter .....   | 12 |
| 3k. Medieval Punctuation .....  | 12 |
| 3l. Ten Characters for Middle English (Ormulum) .....   | 13 |
| 3m. SIGMOID S.....  | 14 |
| 3n. Two Characters for Middle Scots.....  | 14 |

---

<sup>1</sup> Also participating were Fred Brennan, Patrick Chew, Craig Cornelius, Lorna Evans, Andrew Glass, Manish Goregaokar, Ned Holbrook, John Hudson, Marek Jeziorek, Jan Kučera, Norbert Lindenberg, Kamal Mansour, Lawrence Wolf-Sonkin, and Ben Yang.

|  |    |
|--|----|
| 3o. Modifier Letter Heavy Prime.....   | 15 |
| 3p. OBLIQUE HYPHEN .....   | 16 |
| 4 Todhri .....   | 16 |
| 5 Vithkuqi.....  | 17 |
| II. AMERICAS .....   | 18 |
| 6 Mayan Hieroglyphs .....  | 18 |
| III. AFRICA.....   | 18 |
| 7 Adinkra .....  | 18 |
| 8 Egyptian Hieroglyphs .....   | 18 |
| 8a. Glyph changes to Egyptian Hieroglyphs block .....                          | 18 |
| 8b. Proposal to add one column to the Egyptian Hieroglyph Format Controls..... | 19 |
| 8c. Summary from Zoom calls on Egyptian Hieroglyphs.....                       | 20 |
| 9 Ethiopic .....   | 20 |
| 10 Kore Sebeli .....   | 21 |
| IV. MIDDLE EAST .....  | 21 |
| 11 Arabic .....  | 21 |
| 11a. Glyph changes and annotations for Kazakh, Kyrgyz, and Uyghur.....         | 21 |
| 11b. Sindhi and Behdini Kurdish .....  | 23 |
| 11c. Damma over Damma in Tunisian Quran .....                                  | 23 |
| V. SOUTH AND CENTRAL ASIA.....   | 24 |
| 12 Kannada .....   | 24 |
| 13 Mundari Bani.....   | 24 |
| 14 Old Uyghur .....  | 25 |
| 15 Tangsa .....  | 26 |
| 16 Tulu .....  | 27 |
| 17 Vedic.....  | 27 |
| VI. SOUTHEAST ASIA, INDONESIA, AND OCEANIA .....                               | 28 |
| 18 Kawi.....   | 28 |
| 18a. Kawi script proposal.....   | 28 |
| 18b. KAWI VOWEL SIGN AA .....  | 29 |
| 19 Tai Khao.....   | 30 |
| 20 Western Cham .....  | 31 |
| VII. EAST ASIA.....  | 31 |

|  |    |
|--|----|
| 21 Bopomofo .....  | 31 |
| 22 Kana.....   | 32 |
| 23 Kanbun .....  | 33 |
| 24 Tangut .....  | 33 |
| VIII. SYMBOLS, PUNCTUATION, AND NOTATIONAL SYSTEMS ..... | 34 |
| 25 Math Calligraphic Alphabets .....                     | 34 |
| 26 Persian Siyaq .....                                   | 34 |
| IX. OTHER TOPICS.....                                    | 35 |
| 27 Variation Sequences .....                             | 35 |
| 28 CCC of Znamenny Notation Characters .....             | 35 |
| X. PUBLIC REVIEW FEEDBACK .....                          | 36 |
| XI. RECOMMENDATIONS FOR UNICODE 14.0.....                | 41 |

## I. EUROPE

### 1 Cistercian

**Document:** [L2/20-290](#) Background for Unicode consideration of Cistercian numerals – Miller

**Comments:** We reviewed this background document on Cistercian numerals. The numerals were developed at the end of the 13<sup>th</sup> century and used primarily in Europe from the 13<sup>th</sup> to the 15<sup>th</sup> centuries. A project to digitize Cistercian manuscripts at Western Michigan University is not requesting the characters be in Unicode, so this is just an informational document.

It was noted that images or a non-Unicode based font can be used today for Cistercian numerals. A future proposal would need to make the case for the numerals to be interchanged as plain text. Also, complex rendering rules would also be needed, since the numbers are built on grids, much like [Mandombe](#).

**Recommendation:** We recommend the UTC make the following disposition:  
Notes this document (L2/20-290) but takes no further action.

---

### 2 Cypro-Minoan

**Documents:** [L2/20-156R](#) Considerations regarding a normalized Cypro-Minoan reference font (revised WG2 N5137R) – Everson

*Related document:* [L2/20-154](#) Final proposal to encode the Cypro-Minoan script in the SMP (WG2 N5135) – Everson

**Comments:** We reviewed this document, which describes the author's work to develop a normalized monoline Cypro-Minoan font that could be used in the code charts. The result is contained in figure 4 (page 6), which has been okayed by Cypro-Minoan experts.

The document also proposes (p. 2) adding one additional character into the repertoire, CM075B, and changing the glyph for CM075 from a box with a cross to just a box glyph. As explained in the document, the glyph without the cross occurs many times in the CM2 and CM3 groups, but the box with a cross occurs only in CM1, and not as often. The exact relationship between the two shapes has not been confirmed. The Cypro-Minoan experts were okay with the changes.

Cypro-Minoan was approved at the July 2020 UTC meeting (consensus [164-C8](#)). At the October 2020 UTC meeting, the Cypro-Minoan block was moved to U+12F90..U+12FFF (consensus [165-C16](#)). Because the new character is inserted into the repertoire, code points have shifted. As a result, a new recommendation should be made.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R1:** The UTC accepts one new Cypro-Minoan character, U+12FCE CYPRO-MINOAN SIGN CM075B, and moves U+12FCE..U+12FF1 to U+12FCF..U+12FF2. With the additional character, the total number of new characters in the Cypro-Minoan block will be 99. Properties for the Cypro-Minoan characters are as documented in [L2/20-154](#) (though code points need adjusting) and glyphs are as shown in Figure 4 of [L2/20-156R](#). (Reference: [L2/20-156R](#) and [L2/20-154](#))

We also recommend that the UTC make the following disposition:

Assigns an AI to Ken Whistler to update the Pipeline. (Reference: Section 2 of [L2/21-016 Script Ad Hoc Recommendations](#)).

---

## 3 Latin

### 3a. Phonetic characters (extIPA, VoQS, phonetic click letters, IPA retroflex letters and similar letters with hooks)

**Documents:** [L2/21-021](#) Reference doc numbers for [L2/20-266R](#) "Consolidated code chart of proposed phonetic characters" and IPA etc. code point and name changes – Anderson

[L2/20-266](#) Consolidated code chart of proposed phonetic characters (WG2 N5148) – Everson and Miller

**Comments:** We reviewed [L2/21-021](#) which contains code charts by Michael Everson and Kirk Miller and an accompanying document by Kirk Miller (pp. 7-8).

The charts (pp. 1-6) contain characters from 5 different proposals: [L2/20-115R](#), [L2/20-116R](#), [L2/20-125R](#), [L2/20-252R](#), [L2/20-253R](#). The characters are now located in two new blocks, Latin Extended-F and Latin Extended-G.

The characters from [L2/20-115R](#), [L2/20-116R](#), and [L2/20-125R](#) were approved at the July 27, 2020 UTC meeting, but were rescinded at the October 2020 UTC because the characters were re-arranged and, in a few cases, had their names changed.

The source proposals for the phonetic characters are listed in [L2/21-021](#). Note that the original documents [L2/20-115R](#), [L2/20-116R](#), and [L2/20-125R](#) do not have the correct code points (and in some cases, new names); the consolidated chart and list below reflect the new names and code points and should be used as a reference. (Those characters from [L2/20-252R](#) and [L2/20-253R](#) are listed separately under 3b and 3c, below. The character names and code points in those two proposals are correct, and can serve as references.)

For background, a comparison of the name and codepoint changes between what the UTC approved at the July 2020 meeting versus the new positions and names is included on pages 7-8 of [L2/21-021](#).

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R2:** The UTC accepts 38 phonetic characters for encoding in a future version of the standard, with properties as documented in [L2/20-115R](#), [L2/20-116R](#), and [L2/20-125R](#) (though code points may need adjusting) and glyphs and names in [L2/21-021](#):

10780 MODIFIER LETTER SMALL CAPITAL AA  
10790 MODIFIER LETTER SMALL FENG DIGRAPH  
10799 MODIFIER LETTER SMALL LS DIGRAPH  
1079A MODIFIER LETTER SMALL LZ DIGRAPH  
1079B MODIFIER LETTER SMALL L WITH BELT  
1079C MODIFIER LETTER SMALL CAPITAL L WITH BELT  
1079D MODIFIER LETTER SMALL L WITH RETROFLEX HOOK AND BELT  
1079E MODIFIER LETTER SMALL LEZH  
1079F MODIFIER LETTER SMALL LEZH WITH RETROFLEX HOOK  
107A1 MODIFIER LETTER SMALL TURNED Y WITH BELT  
1DF00 LATIN SMALL LETTER FENG DIGRAPH WITH TRILL  
1DF01 LATIN SMALL LETTER REVERSED SCRIPT G  
1DF02 LATIN LETTER SMALL CAPITAL TURNED G  
1DF03 LATIN SMALL LETTER REVERSED K  
1DF04 LATIN LETTER SMALL CAPITAL L WITH BELT  
1DF05 LATIN SMALL LETTER LEZH WITH RETROFLEX HOOK  
1DF06 LATIN SMALL LETTER TURNED Y WITH BELT  
1DF07 LATIN SMALL LETTER REVERSED ENG  
1DF08 LATIN SMALL LETTER TURNED R WITH LONG LEG AND RETROFLEX HOOK  
1DF09 LATIN SMALL LETTER T WITH HOOK AND RETROFLEX HOOK  
1DF0A LATIN LETTER RETROFLEX CLICK WITH RETROFLEX HOOK  
1DF0B LATIN SMALL LETTER ESH WITH DOUBLE BAR  
1DF0C LATIN SMALL LETTER ESH WITH DOUBLE BAR AND CURL  
1DF0D LATIN SMALL LETTER TURNED T WITH CURL  
1DF0E LATIN LETTER INVERTED GLOTTAL STOP WITH CURL  
1DF0F LATIN LETTER STRETCHED C WITH CURL  
1DF10 LATIN LETTER SMALL CAPITAL TURNED K

1DF11 LATIN SMALL LETTER L WITH FISHHOOK  
1DF12 LATIN SMALL LETTER DEZH WITH PALATAL HOOK  
1DF13 LATIN SMALL LETTER L WITH BELT AND PALATAL HOOK  
1DF14 LATIN SMALL LETTER ENG WITH PALATAL HOOK  
1DF15 LATIN SMALL LETTER TURNED R WITH PALATAL HOOK  
1DF16 LATIN SMALL LETTER R WITH FISHHOOK AND PALATAL HOOK  
1DF17 LATIN SMALL LETTER TESH WITH PALATAL HOOK  
1DF18 LATIN SMALL LETTER EZH WITH PALATAL HOOK  
1DF1A LATIN SMALL LETTER I WITH STROKE AND RETROFLEX HOOK  
1DF1B LATIN SMALL LETTER O WITH RETROFLEX HOOK  
1DF1C LATIN SMALL LETTER TESH WITH RETROFLEX HOOK  
(Reference: Section 3a of L2/21-016 Script Ad Hoc Recommendations)

We also recommend that the UTC make the following disposition:  
Assigns an AI to Ken Whistler to update the Pipeline. (Reference: Section 3a of L2/21-016 Script Ad Hoc Recommendations).

---

### 3b. IPA Modifier Letters - Pulmonic

**Document:** L2/20-252R Unicode request for IPA modifier-letters (a), pulmonic – Miller and Ashby

**Comments:** We reviewed this document which requested 35 pulmonic modifier letters for IPA. (A pulmonic consonant is one that is produced with airstream from the lungs.) The set of proposed characters has been seen before by the Script Ad Hoc.

This proposal includes a letter of support from the International Phonetic Association, which requested complete support for modifiers for the current IPA letters. The proposed repertoire also includes eleven characters with no examples, but are covered by the IPA “complete coverage” request. Of those eleven characters with no examples, a few are considered “accidental gaps” in the inventory--the fact that they are missing in examples of attested usage is purely coincidental. Several are implicit characters, not shown in IPA charts, but using systematic modifications for possible combinations of articulatory features. Several are retired IPA letters that are no longer recommended by IPA, although still used by some authors.

The document also includes evidence for separately encoding MODIFIER LETTER SMALL CAPITAL H (figure 62, page 31). Earlier the Script Ad Hoc had recommended unification of this character with U+1D78 MODIFIER LETTER CYRILLIC EN, but the italic version of Cyrillic character is distinct. We agree that separate encoding of MODIFIER LETTER SMALL CAPITAL H is warranted.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R3:** The UTC accepts 35 modifier letter characters in a new Latin Extended-F block (U+10780..U+107BF) for encoding in a future version of the standard, with glyphs and properties as documented in L2/20-252R:

10781 MODIFIER LETTER SUPERScript TRIANGULAR COLON  
10782 MODIFIER LETTER SUPERScript HALF TRIANGULAR COLON  
10783 MODIFIER LETTER SMALL AE  
10784 MODIFIER LETTER SMALL CAPITAL B  
10787 MODIFIER LETTER SMALL DZ DIGRAPH  
10788 MODIFIER LETTER SMALL DZ DIGRAPH WITH RETROFLEX HOOK  
10789 MODIFIER LETTER SMALL DZ DIGRAPH WITH CURL  
1078A MODIFIER LETTER SMALL DEZH DIGRAPH  
1078B MODIFIER LETTER SMALL D WITH TAIL  
1078E MODIFIER LETTER SMALL REVERSED E  
1078F MODIFIER LETTER SMALL CLOSED REVERSED OPEN E  
10791 MODIFIER LETTER SMALL RAMS HORN  
10792 MODIFIER LETTER SMALL CAPITAL G  
10795 MODIFIER LETTER SMALL H WITH STROKE  
10796 MODIFIER LETTER SMALL CAPITAL H  
10797 MODIFIER LETTER SMALL HENG WITH HOOK  
107A0 MODIFIER LETTER SMALL TURNED Y  
107A2 MODIFIER LETTER SMALL O WITH STROKE  
107A3 MODIFIER LETTER SMALL CAPITAL OE  
107A4 MODIFIER LETTER SMALL CLOSED OMEGA  
107A5 MODIFIER LETTER SMALL Q  
107A6 MODIFIER LETTER SMALL TURNED R WITH LONG LEG  
107A7 MODIFIER LETTER SMALL TURNED R WITH LONG LEG AND RETROFLEX HOOK  
107A8 MODIFIER LETTER SMALL R WITH TAIL  
107A9 MODIFIER LETTER SMALL R WITH FISHHOOK  
107AA MODIFIER LETTER SMALL CAPITAL R  
107AB MODIFIER LETTER SMALL TC DIGRAPH WITH CURL  
107AC MODIFIER LETTER SMALL TS DIGRAPH  
107AD MODIFIER LETTER SMALL TS DIGRAPH WITH RETROFLEX HOOK  
107AE MODIFIER LETTER SMALL TESH DIGRAPH  
107AF MODIFIER LETTER SMALL T WITH RETROFLEX HOOK  
107B0 MODIFIER LETTER SMALL V WITH RIGHT HOOK  
107B2 MODIFIER LETTER SMALL CAPITAL Y  
107B3 MODIFIER LETTER GLOTTAL STOP WITH STROKE  
107B4 MODIFIER LETTER REVERSED GLOTTAL STOP WITH STROKE  
(Reference: L2/20-252R)

We also recommend that the UTC make the following disposition:  
Assigns an AI to Ken Whistler to update the Pipeline. (Reference: L2/20-252R)

---

### 3c. IPA Modifier Letters – Non-Pulmonic

**Document:** [L2/20-253R](#) Unicode request for IPA modifier letters (b), non-pulmonic – Miller and Ashby

**Comments:** We reviewed this document for non-pulmonic modifiers (i.e., ejectives, implosive, and click consonants). It has been seen by the Script Ad Hoc before.

The proposed non-pulmonic characters are rare, probably due to the limited number of transcription conventions used for non-European languages. The proposal includes a letter from the International Phonetic Association, which requests support for modifier letters for all current and implicit IPA letters – including support for implosives and click releases. (Support for ejectives is not required, since modifier ejectives can already be represented by existing Unicode characters.)

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R4:** The UTC accepts 11 modifier letter characters in a new Latin Extended-F block (U+10780..U+107BF) for encoding in a future version of the standard, with glyphs and properties as documented in L2/20-253R:

10785 MODIFIER LETTER SMALL B WITH HOOK  
1078C MODIFIER LETTER SMALL D WITH HOOK  
1078D MODIFIER LETTER SMALL D WITH HOOK AND TAIL  
10793 MODIFIER LETTER SMALL G WITH HOOK  
10794 MODIFIER LETTER SMALL CAPITAL G WITH HOOK  
10798 MODIFIER LETTER SMALL DOTLESS J WITH STROKE AND HOOK  
107B5 MODIFIER LETTER BILABIAL CLICK  
107B6 MODIFIER LETTER DENTAL CLICK  
107B7 MODIFIER LETTER LATERAL CLICK  
107B8 MODIFIER LETTER ALVEOLAR CLICK  
107B9 MODIFIER LETTER RETROFLEX CLICK WITH RETROFLEX HOOK  
(Reference: L2/20-253R)

We also recommend that the UTC make the following disposition:  
Assigns an AI to Ken Whistler to update the Pipeline. (Reference: L2/20-253R)

---

### 3d. Addendum to Unicode request L2/20-253

**Document:** [L2/21-040](#) Addendum to Unicode requests for IPA modifier letters, L2/20-252 pulmonic and L2/20-253 non-pulmonic – Miller

**Comments:** This document provides evidence for U+107B5 MODIFIER LETTER BILABIAL CLICK, a non-pulmonic consonant that was proposed in [L2/20-253R](#). The MODIFIER LETTER BILABIAL CLICK did not have evidence of use when it was originally proposed. (The rationale for encoding the set of non-pulmonic consonants was based on the need for completeness of IPA coverage and potential need.) An example of a variant of a pre-Kiel click letter is also provided, but it has not yet been proposed, pending



feedback from the IPA Council. The proposal also includes examples of other pulmonic characters proposed in [L2/20-252R](#).

**Recommendation:** We recommend the UTC make the following disposition:  
Notes this document (L2/21-040) but takes no further action.

---

### 3e. Modifier Latin Capital Letters

**Document:** [L2/20-251](#) Unicode request for modifier Latin capital letters – Miller and Cornelius

**Comments:** We reviewed this proposal for three modifier capital letters, two of which (C, F) are used in the Chatino orthography (Oaxaca, Mexico) and for para-IPA transcription (i.e., <sup>ɸ</sup>V is used to indicate falling toneme on a vowel and N<sup>c</sup> for a generic post-stopped nasal). The third modifier capital letter (Q) is used in the phonemic transcription of Japanese. Examples are provided for the proposed characters.

We agree that both MODIFIER LETTER CAPITAL C and MODIFIER LETTER CAPITAL F should be located in the BMP, since they are used actively in orthographies today. The proposed code points appear acceptable. If possible, we recommend these be included in characters for Unicode 14.0.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R5:** The UTC accepts the following 3 characters for encoding in a future version of the standard, with glyphs and properties as documented in L2/20-251:

A7F2 MODIFIER LETTER CAPITAL C

A7F3 MODIFIER LETTER CAPITAL F

A7F4 MODIFIER LETTER CAPITAL Q

(Reference: L2/20-251)

We also recommend that the UTC make the following disposition:  
Assigns an AI to Ken Whistler to update the Pipeline. (Reference: L2/20-251)

---

### 3f. Subscript Modifier Letters

**Document:** [L2/21-043](#) Unicode request for subscript modifier-letter support – Miller

**Comments:** This proposal covers subscript modifiers in a variety of traditions. The proposed subscript modifier letters were never in official IPA usage.

During discussion there was concern that some of the figures do not make a clear case for the proposed subscripts to be carried in plain text. Rather, the evidence shows mixed orthographic conventions. Figures 7 and 8, for example, demonstrate a subscripting convention for segmental or morphological

analysis that is already representable as rich text. A more convincing scenario would be evidence of two authors communicating a transcription via print using the same subscript convention.

**Recommendation:** We recommend the UTC make the following disposition:  
Notes this document (L2/21-043) but takes no further action.

---

### 3g. Phonetic Punctuation and Diacritics

**Document:** L2/21-042 Unicode request for phonetic punctuation & diacritics – Miller

**Comments:** We reviewed this proposal for 15 additional phonetic punctuation and diacritic characters. The examples provided evidence of usage of the characters and seemed reasonable.

The author was requested to include a paragraph adding the Bidi\_Mirroring\_Glyph property values (for BidiMirroring.txt of the UCD) of the eight bracket and parentheses characters, then send the updated proposal for posting in the document register. (*Note: This action has been done.*)

Two characters, 11AB0 MODIFIER NUMBER SIGN and 11AB1 MODIFIER DOLLAR SIGN, have code points that conflict with approved characters (UCAS additions for Nattilik and historic Cree). At this point, we recommend holding off on these two characters.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R6:** The UTC accepts 13 phonetic punctuation and diacritic characters for encoding in a future version of the standard, with glyphs and properties as documented in L2/21-042:

1AC5 COMBINING SQUARE BRACKETS ABOVE  
1AC7 COMBINING INVERTED DOUBLE ARCH ABOVE  
1AC8 COMBINING PLUS SIGN ABOVE  
1ACD COMBINING DOUBLE PLUS SIGN ABOVE  
1ACE COMBINING DOUBLE PLUS SIGN BELOW  
2E55 LEFT SQUARE BRACKET WITH STROKE  
2E56 RIGHT SQUARE BRACKET WITH STROKE  
2E57 LEFT SQUARE BRACKET WITH DOUBLE STROKE  
2E58 RIGHT SQUARE BRACKET WITH DOUBLE STROKE  
2E59 TOP HALF LEFT PARENTHESIS  
2E5A TOP HALF RIGHT PARENTHESIS  
2E5B BOTTOM HALF LEFT PARENTHESIS  
2E5C BOTTOM HALF RIGHT PARENTHESIS  
(Reference: L2/21-042)

We also recommend that the UTC make the following dispositions:  
Assigns an AI to Ken Whistler to update the Pipeline. (Reference: L2/21-042)

---

### 3h. Additional Para-IPA Letters

**Document:** L2/21-041 Unicode request for additional para-IPA letters – Miller

**Comments:** This request for three characters is a supplement to a much larger request L2/20-125R, which requested IPA retroflex letters and similar letters with hooks. Attestation of uppercase versions of c with retroflex hook is needed before it can be accepted.

The proposed characters present evidence of their usage and seem reasonable. After discussion whether the three letters should be grouped together, the Script Ad Hoc recommended small s with curl be located in Latin Extended-F block at U+107BA, with the name MODIFIER LETTER SMALL S WITH CURL, and the other two be located in the Latin Extended-G block. (*Note: The proposal author has updated the proposal with the new code point and name, and the proposal is now posted as L2/21-041.*)

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R7:** The UTC accepts 3 phonetic characters for encoding in a future version of the standard, with glyphs and properties as documented in L2/21-041:

107BA MODIFIER LETTER SMALL S WITH CURL

1DF1D LATIN SMALL LETTER C WITH RETROFLEX HOOK

1DF1E LATIN SMALL LETTER S WITH CURL

(Reference: L2/21-041)

We also recommend that the UTC make the following dispositions:

Assigns an AI to Ken Whistler to update the Pipeline. (Reference: L2/21-041)

---

### 3i. Dezh with Retroflex Hook

**Document:** L2/21-004 Unicode request for dezh with retroflex hook – Miller and Everson

**Comments:** We reviewed this proposal for LATIN SMALL LETTER DEZH WITH RETROFLEX HOOK, which includes the opinions of John Esling and Michael Ashby (former and current presidents of the IPA, respectively). Both Esling and Ashby support the request, though no printed evidence is provided.

We recommend 1DF19 LATIN SMALL LETTER DEZH WITH RETROFLEX HOOK be encoded. but the UTC should discuss this, since the proposal has no examples.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R8:** The UTC accepts one phonetic character for encoding in a future version of the standard, glyphs and properties as documented in L2/21-004, after discussion:

1DF19 LATIN SMALL LETTER DEZH WITH RETROFLEX HOOK (Reference: L2/21-004)

We also recommend that the UTC make the following disposition:  
Assigns an AI to Ken Whistler to update the Pipeline. (Reference: [L2/21-004](#))

---

### 3j. Old Polish Nasal Vowel Letter

**Document:** [L2/21-039](#) Proposal to include the Old Polish nasal vowel letter in ISO/IEC 10646 and The Unicode Standard – Bunčić

**Comments:** We reviewed this proposal to add a case pair for a nasal vowel in Old Polish. The lowercase letter appears in manuscripts from the 12th to the 16th centuries and in modern scholarly texts. The proposal makes a clear case, in our view, providing evidence for the nasal vowel. The proposal also discusses allographs.

Both upper- and lowercase forms are proposed, though only the lowercase is attested. The requestor does note that capitalization is generally is used in Old Polish, and having a capital form will allow capitalization features shared with other letters.

The name was discussed. One suggestion was “horned o”, which is a translation of the native name of the letter. Because of potential confusion with characters whose names contain “with horn” (e.g., U+01A1 used for Vietnamese), an annotation was recommended.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R9:** The UTC accepts two Latin characters for encoding in a future version of the standard, with glyphs and properties as documented in [L2/21-039](#):

U+A7C0 LATIN CAPITAL LETTER OLD POLISH O

U+A7C1 LATIN SMALL LETTER OLD POLISH O

(Reference: [L2/21-039](#))

We also recommend that the UTC make the following disposition:  
Assigns an AI to Ken Whistler to update the Pipeline. (Reference: [L2/21-039](#))

---

### 3k. Medieval Punctuation

**Documents:** [L2/20-270R](#) Proposal to add two mediaeval punctuation characters – Everson  
[L2/20-278](#) Comment on [L2/20-270](#) regarding two puncti – Brennan

**Comments:** We reviewed this proposal for two medieval punctuation marks, originally proposed in [L2/16-125](#). (Cf. Script Ad Hoc Recommendations on the 2016 proposal in [L2/16-156](#).) We also reviewed the document [L2/20-278](#).

The characters appear to be justified and reflect a distinction used consistently. Modern medieval scholars would benefit from encoding the characters.

There was agreement that the names MEDIEVAL EXCLAMATION MARK and MEDIEVAL QUESTION MARK were appropriate. This naming convention aligns with MEDIEVAL COMMA (but not PUNCTUS ELEVATUS, which has no good English equivalent). There was some hesitation on the reference glyph provided for MEDIEVAL EXCLAMATION MARK.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R10:** The UTC accepts 2 medieval punctuation characters for encoding in a future version of the standard, with glyphs and properties as documented in L2/20-270R, after discussion:

2E53 MEDIEVAL EXCLAMATION MARK

2E54 MEDIEVAL QUESTION MARK

(Reference: L2/20-270R)

We also recommend that the UTC make the following disposition:

Assigns an AI to Ken Whistler to update the Pipeline. (Reference: L2/20-270R)

---

### 3l. Ten Characters for Middle English (Ormulum)

**Document:** L2/20-268 Revised proposal to add ten characters for Middle English to the UCS – Everson

**Comments:** This 2019 proposal had not been reviewed earlier. It requests 10 characters used in the lengthy 12c Middle English poem called the *Ormulum*, whose creator, Orm, created his own spelling system. The orthography provides valuable insights into English phonology of the time period.

Comments that arose during discussion:

- For triple acute, provide a cross-reference to U+1DFB COMBINING DELETION MARK, which was encoded in Unicode 9.0 for Newa.
- The SAH had earlier recommended eight of the proposed characters, except for LATIN CAPITAL LETTER DOUBLE THORN and LATIN CAPITAL LETTER DOUBLE WYNN, which had no evidence.
- Because the argument for case pairs regularly arises, it might be helpful to have a document discussing the formal criteria for the addition of case pairs, which can guide making decisions when no evidence is presented. It was also noted that newly created uppercase forms for letters only attested with lowercases in historical texts can be problematic regarding what exact glyph shape they should actually take.

Based on the discussion, the general consensus was to recommend the UTC approve LATIN CAPITAL LETTER DOUBLE THORN and LATIN CAPITAL LETTER DOUBLE WYNN. One member of the SAH did not agree.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R11:** The UTC accepts 10 characters used for Middle English for encoding in a future version of the standard, with glyphs and properties as documented in L2/20-268, after discussion:

1AC9 COMBINING TRIPLE ACUTE ACCENT

1ACA COMBINING LATIN SMALL LETTER INSULAR G

1ACB COMBINING LATIN SMALL LETTER INSULAR R  
1ACC COMBINING LATIN SMALL LETTER INSULAR T  
A7D0 LATIN CAPITAL LETTER CLOSED INSULAR G  
A7D1 LATIN SMALL LETTER CLOSED INSULAR G  
A7D2 LATIN CAPITAL LETTER DOUBLE THORN  
A7D3 LATIN SMALL LETTER DOUBLE THORN  
A7D4 LATIN CAPITAL LETTER DOUBLE WYNN  
A7D5 LATIN SMALL LETTER DOUBLE WYNN  
(Reference: L2/20-268)

We also recommend that the UTC make the following disposition:  
Assigns an AI to Ken Whistler to update the Pipeline. (Reference: L2/20-268)

---

### 3m. SIGMOID S

**Document:** L2/20-269 Proposal to add two SIGMOID S characters for mediaeval palaeography (WG2 N5146) – Everson

**Comments:** We reviewed this proposal to add a case pair for a “sigmoid s” used in the orthography of Middle Scots, which dates from 1450 to 1700. It is one of three shapes of the letter s (long s, regular s and sigmoid s), and is reportedly used to date texts. All three can be found in texts, such as in figure 4. We recommend the UTC approve the characters.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R12:** The UTC accepts 2 sigmoid S characters for encoding in a future version of the standard, with glyphs and properties as documented in L2/20-269:

A7D8 LATIN CAPITAL LETTER SIGMOID S  
A7D9 LATIN SMALL LETTER SIGMOID S  
(Reference: L2/20-269)

We also recommend that the UTC make the following disposition:  
Assigns an AI to Ken Whistler to update the Pipeline. (Reference: L2/20-269)

---

### 3n. Two Characters for Middle Scots

**Document:** L2/20-267 Proposal to add two characters for Middle Scots to the UCS – Everson

**Comments:** We reviewed this revised version of L2/20-267, which now includes text bolstering the argument against unification with LATIN CAPITAL LETTER SHARP S (see section 2.1 on page 2).

Discussion on this document mentioned the following points:

- The lowercase is clearly attested (see below). The examples of uppercase, in our view, are also evident: it appears in proper names and at the beginning of sentences (see below).

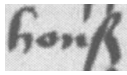


Figure 11 with lowercase Middle Scots S in ‘house’



Figure 15 with uppercase Middle Scots S “Sir” at beginning of sentence

- Unifying MIDDLE SCOTS S with LATIN LETTER SHARP S would result in text processes that replace the uppercase letter with SS, which is never appropriate for this letter. (Cf. “Unconditional mappings” in SpecialCasing.txt for LATIN LETTER SHARP S.)
- Also, the distribution of MIDDLE SCOTS S differs from LATIN LETTER SHARP S: SHARP S occurs normally medially and finally, but LATIN CAPITAL LETTER MIDDLE SCOTS S is commonly found in initial position.

Based on our review of the information in this document, we recommend the encoding of LATIN CAPITAL LETTER MIDDLE SCOTS S and LATIN SMALL LETTER MIDDLE SCOTS S.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R13:** The UTC accepts 2 Middle Scots S characters for encoding in a future version of the standard, with glyphs and properties as documented in L2/20-267:

A7D6 LATIN CAPITAL LETTER MIDDLE SCOTS S

A7D7 LATIN SMALL LETTER MIDDLE SCOTS S

(Reference: L2/20-267)

We also recommend that the UTC make the following disposition:

Assigns an AI to Ken Whistler to update the Pipeline. (Reference: L2/20-267)

---

### 3o. Modifier Letter Heavy Prime

**Document:** L2/20-286 Prime Proposal: Modifier Letter Heavy Prime – Erik Carvalhal Miller

**Comments:** We reviewed this request for a “heavy prime” character, which appears in some American dictionaries to mark primary stress.

The author notes that the requested character may appear to be a bolded presentation form of U+02B9 MODIFIER LETTER PRIME or U+2032 PRIME. Indeed, the examples of “heavy prime” appear where font distinctions are being made: some dictionaries use bold primary stress mark, and non-bold marks for secondary stress. Fine details of the shapes and angles of the glyphs can be dependent upon the fonts.

In our view, the examples are not evidence of a need for a new character. Instead, we recommend use of the existing characters U+2032 PRIME and U+02B9 MODIFIER LETTER PRIME. To provide the exact duplication of text, facsimiles or rich text should be employed.

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

Assigns an AI to Debbie Anderson to relay feedback to the proposal author of [L2/20-286](#).

(Reference: Section 3n of L2/21-016 Script Ad Hoc Recommendations)

---

### 3p. OBLIQUE HYPHEN

**Document:** [L2/21-036](#) Proposal to add the OBLIQUE HYPHEN to the UTC – Everson

**Comments:** We discussed this proposal for a single oblique hyphen. This character was used in the Middle Ages and Renaissance. In modern-day texts, such as the Oxford University Press, the oblique hyphen may be used as an end-of-line hard hyphen.

It was noted that the correct general category should be “Pd,” and not “Po.”

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R14:** The UTC accepts 1 punctuation character for encoding in a future version of the standard, with glyph and properties as documented in L2/21-036:

2E5D OBLIQUE HYPHEN

(Reference: L2/21-036)

We also recommend that the UTC make the following dispositions:

Assigns an AI to Michael Everson to update the proposal and change the property for OBLIQUE HYPHEN from Po to Pd, change figure “5” to figure 4, and send the revised proposal for posting in the document register. (*Note: This action has been done.*)

Assigns an AI to Ken Whistler to update the Pipeline. (Reference: L2/21-036)

---

## 4 Toghri

**Document:** [L2/20-188R2](#) Proposal for encoding the Toghri script in the SMP of the UCS – Everson

**Comments:** We reviewed this latest version of the proposal, which expanded the section on ordering, and moved the six Greek-based letters to the end of the repertoire. (The move should not impact collation.)

Based on public review feedback from David Corbett, two pairs of characters have had their names changed (E>EI and EH>E), so Elbasan, Vithkuqi, and Toghri are all consistent in their character naming conventions for the Albanian vowels, which are represented as <e> and <ë> when Albanian is written in its modern Latin orthography.



The SAH had recommended approval of the Todhri script in its [Oct. 2020 recommendations](#), but the UTC assigned an Action Item to Roozbeh Pournader and Debbie Anderson to discuss the decomposition of TODHRI LETTER EI and TODHRI LETTER U, before the script is considered ready for approval. Because the decomposition document is not yet ready, we recommend the UTC defer a decision on approving the script at this time.

**Recommendation:** We recommend the UTC make the following disposition:  
Notes this document (L2/20-188R2) but takes no further action.

---

## 5 Vithkuqi

**Document:** [L2/20-187R2](#) Proposal for encoding the Vithkuqi script in the SMP (WG2 N5138R2) – SEI/Everson

**Comments:** We reviewed this revised proposal, which has incorporated two name changes (based on the October 1, 2020 public review feedback from David Corbett):

10577 VITHKUQI CAPITAL LETTER E > 10577 VITHKUQI CAPITAL LETTER EI

10578 VITHKUQI CAPITAL LETTER EH > 10578 VITHKUQI CAPITAL LETTER E

Figure 15 is new; it compares modern Latin-based Albanian alphabet to the letters (or digraphs) for Elbasan, Todhri, and Vithkuqi.

It was noted that the SAH had recommended approval of the Vithkuqi script in its [Oct. 2020 recommendations](#), but the UTC #165 did not yet approve the script, and had instructed Deborah Anderson to respond to the author “UTC does not support encoding newly invented modern characters without evidence of usage in text,” which has been done.

The Script Ad Hoc recommends the newly created characters, GJA, RRE, XHE, and ZHE (highlighted yellow in the code chart) be removed, due lack of any evidence in print. We recommend holes for the characters can left open, until evidence is provided.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R15:** The UTC accepts 70 Vithkuqi characters in a new Vithkuqi block (U+10570..U+105BF) for encoding in a future version of the standard, with glyphs and properties as documented in L2/20-187R2, but leaving reserved code points at U+1057B, U+1058B, U+10593, U+10596, U+105A2, U+105B2, U+105BA, U+105BD.

(Reference: Section 5 of L2/21-016 Script Ad Hoc Recommendations; [L2/20-187R2](#))

We also recommend that the UTC make the following disposition:  
Assigns an AI to Ken Whistler to update the Pipeline. (L2/20-187R2)

## II. AMERICAS

### 6 Mayan Hieroglyphs

**Document:** L2/20-248 Updated List of Characters for Mayan Codices – Pallan

**Comments:** This is an FYI document, updating L2/19-171. The set of Mayan characters will eventually be expanded once data from Classic period material is incorporated, and a future Unicode proposal will be based on the more complete repertoire. One comment was mentioned during discussion: the author should explain the first two characters in any future update.

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

Assigns an AI to Debbie Anderson to relay feedback to the proposal author of L2/20-248.

(Reference: Section 6 of L2/21-016 Script Ad Hoc Recommendations).


---


## III. AFRICA

### 7 Adinkra

**Document:** L2/21-020 Adinkra [draft proposal] – Charles Korankye, Adinkra Alphabet Encoding committee

**Comments:** The proposed symbols are simplifications and alterations of the Adinkra symbols as usually found:

Adinkra symbol *bese saka*: 

Alphabet letter “v”: 

While the Adinkra symbols have been used in art, fabrics, pottery, and other media, their use today as an alphabet requires more evidence: how many people are using the alphabet today? How many books in the alphabet have been produced? (See similar questions posed for Kore Sebeli, below.)

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

Assigns an AI to Debbie Anderson to relay feedback to the proposal author of L2/21-020.

(Reference: Section 7 of L2/21-016 Script Ad Hoc Recommendations).

---

### 8 Egyptian Hieroglyphs

#### 8a. Glyph changes to Egyptian Hieroglyphs block

**Document:** L2/21-028 Glyph changes to Egyptian Hieroglyphs for Unicode 14.0 with proposed annotations – Anderson and Suignard

**Comments:** This document contains proposed glyph changes to the Egyptian Hieroglyphs block code chart (with glyphs highlighted in blue). The changes were reviewed by Daniel Werning of the Wortschatz der ägyptischen Sprache project in Berlin. The modified glyphs basically revert to those of Gardiner, and reflect the forms in Unicode version 5.2. In addition, this document also includes a comparison document (pp. 19-27) by Daniel Werning showing the glyphs from Unicode 5.2, version 9.0 glyphs (which are the same as 13.0), and glyphs from various fonts. The newly proposed glyphs are added in the right-hand margin.

The document also includes proposed annotations to the names list, which were provided by Werning and modified by Michel Suignard.

Discussion on use of atomic characters versus sequences for Egyptian Hieroglyphs will be handled in a separate document.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R16:** The UTC approves the proposed Egyptian Hieroglyph glyph changes for version 14.0 and notes that an erratum notice should be posted to document the changes.

(Reference: L2/21-028)

We also recommend the UTC make the following disposition:

Assigns an AI to the Editorial Committee to issue a glyph erratum notice on the Egyptian Hieroglyph characters in L2/21-028. (Reference: L2/21-028)

---

## 8b. Proposal to add one column to the Egyptian Hieroglyph Format Controls

**Document:** L2/21-005 Proposal to add one column to the Egyptian Hieroglyph Format Controls block – Glass

**Comments:** This document requests one additional column for Egyptian Hieroglyph Format Controls that will extend the range of the current block to U+1344F. The current block of format controls for Egyptian Hieroglyphs (U+13430..U+1343F) has only seven open spots, and Andrew Glass anticipates more than seven characters will be needed for rendering Egyptian text and transcription. (Discussion of new controls is detailed in the “Summary from Zoom calls” document, see below in Section 8c.)

A number of questions were raised about the control characters, particularly those for handling damaged signs. After discussion, the Script Ad Hoc recommended reserving a column for added format controls, but leaving it empty for now. Once characters have been approved for this column, the Egyptian Hieroglyph Format Controls block can formally be extended to end at U+1344F.

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

The UTC requests the Roadmap Committee adjust the Egyptian Hieroglyphs Extended-A block by one column, so it starts at U+13450, instead of U+13440, leaving one column unallocated at U+13440..U+1344F. (Reference: Section 8b of L2/21-016 Script Ad Hoc Recommendations).

## 8c. Summary from Zoom calls on Egyptian Hieroglyphs

**Document:** [L2/21-035](#) Summary from Zoom calls on Egyptian Hieroglyphs – Anderson and Glass

**Comments:** This document outlined approaches for handling the rendering of Egyptian hieroglyphs text and transcription, as discussed amongst a group of Egyptologists, Andrew Glass, Michel Suignard and Debbie Anderson.

There appears to be agreement on handling enclosures as outlined in this document.

For middle insertion, it appears one control character is needed, but discussion is ongoing. Teasing apart what should be handled by insertion versus kerning still needs to be discussed.

For mirroring and rotation: For those pairs of characters whose semantics are distinct when mirrored (i.e., legs walking forward have a different meaning from legs walking backward), separately encoding each character would be one approach. Andrew Glass prefers registering a set as variation sequences, over a format control character. For rotations, a question was raised how many rotations are needed? When a proposal for additional control characters is written, providing clear examples will be helpful.

**Recommendation:** We recommend the UTC make the following disposition:  
Notes this document ([L2/21-035](#)) but takes no further action.

---

## 9 Ethiopic

**Document:** [L2/21-037](#) Modern Gurage Orthography Additions to Ethiopic Script – Fekeda Menuta et al.

**Comments:** We reviewed this proposal to add 28 characters to represent the modern Gurage orthography. The orthography was endorsed by the Gurage Zone Administration Council in 2019 and was introduced in the public school system in September 2020.

This proposal has incorporated changes recommended by the Script Ad Hoc, including a comparison of the old versus new orthography, a revision of the collation section, and an explanation of the relation between Gurage and Sebatbeit (the latter being a term that appears currently in headers and in four character names). A few minor additional changes were recommended, affecting the collation section and further explanation on the use of Sebatbeit, which can be used to guide annotations and headers in the names list.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R17:** The UTC accepts 28 Ethiopic characters for the Gurage orthography in a new Ethiopic Extended-B block (U+1E7E0..U+1E7FF) for encoding in a future version of the standard, with glyphs and properties as documented in [L2/21-037](#). (Reference: [L2/21-037](#))

We also recommend that the UTC make the following dispositions:

Assigns an AI to the Roadmap Committee to include the new Ethiopic Extended-B block (U+1E7E0..U+1E7FF).

Assigns an AI to Daniel Jacob to update the proposal with the recommended changes and send in the proposal for posting in the document register. *(Note: This action has been done.)*

Assigns an AI to Daniel Jacob to provide a font to Michel Suignard.

Assigns an AI to Ken Whistler to update the Pipeline. (Reference: [L2/21-037](#))

---

## 10 Kore Sebeli

**Document:** [L2/20-180](#) Proposal for the encoding of Kore Sebeli –Guigon, Bangoura, and Sylla

**Comments:** Two main issues were discussed: the encoding model for the script and evidence of the script's use.

Two models are possible: encode the letter skeletons and use combining diacritics for the dots or encode the letters with dots atomically. We recommend the latter, unless it can be shown that the dots are used generatively to create additional new letter/dot combinations.

Several questions were raised around the script's use:

- How big is the user community?
- How many people are using the script daily to write the language?
- Are there publications not produced by the script's creator?
- Is the script being used for newspapers, pamphlets, and in books?
- Can the proposers provide evidence the script is being used and its use growing?
- How many people are learning the script in schools today? Are there any graduates of the school yet and, if so, how many?

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

Assigns an AI to Debbie Anderson to relay feedback to the proposal author, including comments from the July 2020 Script Ad Hoc Recommendations [L2/20-169](#) (pp. 10-12), which were not yet accommodated. (Reference: Section 10 of [L2/21-016](#) Script Ad Hoc Recommendations).

---

## IV. MIDDLE EAST

### 11 Arabic

#### 11a. Glyph changes and annotations for Kazakh, Kyrgyz, and Uyghur

**Documents:** [L2/20-289](#) Request for glyph changes and annotations for Kazakh, Kyrgyz, and Uyghur – Evans

[L2/20-293](#) U+0641 information for Uyghur, Kazakh and Kyrgyz– Eiso Chan et al.

**Comments:** We reviewed [L2/20-289](#), which clearly lays out proposed changes with rationale and examples. The document has already accommodated comments from the Script Ad Hoc.

---

We agree with the requested glyph changes, annotations, changes to ArabicShaping.txt and proposed wording for Chapter 9 of the Core Spec. The UTC should note that the characters U+0675..U+0678, which are recommended as being discouraged, may be candidates for deprecation.

Eiso Chan et al. (L2/20-293) agreed with nearly all of the recommended changes in L2/20-289, but had comments on ARABIC LETTER FEH (U+0641). The document L2/20-293 concerns how “f” is rendered in Chinese standards and in FounderType fonts used in Xinjiang, the region in China where the Arabic-script Kazakh, Kyrgyz, and Uyghur writing systems are used. He noted that while Chinese standards and some fonts have glyphs for U+0641 FEH that resemble U+06A7 QAF WITH DOT ABOVE, two fonts (in Table 2) contain a glyph for that resembles U+0641 FEH in the code chart. Eiso recommended an annotation for U+0641 FEH, mentioning that a glyph similar to U+06A7 QAF WITH DOT ABOVE is also used. However, Roozbeh Pournader remarked that though U+0641 is used for QAF WITH DOT ABOVE, it is an abuse of Unicode and should not be encouraged. Further discussion on the model is needed. A separate document will be forthcoming from Evans on U+0641.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R18:** The UTC accepts glyph changes for the following characters for Unicode 14.0:

U+06C5, U+FBE0, U+FBE1 (Reference: page 4 of [L2/20-289](#))

U+0677, U+06C7, U+FBD7, U+FBD8, U+FBDD (Reference: page 4 of [L2/20-289](#))

U+0674..U+0678 (Reference: page 5 of [L2/20-289](#))

We also recommend the UTC make the following dispositions:

Forwards the proposed annotations to the names list editor for the following characters:

U+0626 (Reference: page 3 of [L2/20-289](#))

U+06C7 (Reference page 4 of [L2/20-289](#))

U+0675..U+0678

for Unicode 14.0 (Reference page 5 of [L2/20-289](#))

Assigns an AI to the Editorial Committee to issue a glyph erratum notice for U+06C5, U+FBE0, U+FBE1, U+0677, U+06C7, U+FBD7, U+FBD8, U+FBDD, U+0674..U+0678 for Unicode 14.0. (Reference: [L2/20-289](#))

Assigns an AI to Roozbeh Pournader to change ArabicShaping.txt for:

U+06C5 from WAW WITH BAR to WAW WITH LOOP (Reference: page 4 of [L2/20-289](#))

U+0677 and U+06C7 from WITH DAMMA ABOVE to WITH COMMA ABOVE

for Unicode 14.0 (Reference: page 4 of [L2/20-289](#))

Assigns an AI to Lorna Evans and the Editorial Committee to prepare text for the Core Spec for Unicode 14.0, based on proposed wording on U+0626 on page 3 of [L2/20-289](#), and the wording regarding U+0674..U+0678 on page 5 of [L2/20-289](#).

Assigns an AI to Ken Whistler to adjust the weights in the DUCET for U+0675..U+0678 for Unicode 14.0 (Reference page 6 of [L2/20-289](#))

## 11b. Sindhi and Behdini Kurdish

**Document:** L2/20-288 Request for annotations for Sindhi and Behdini Kurdish – Evans

**Comments:** We reviewed this revised “Request for annotations...”, which incorporated comments from the Script Ad Hoc. The requests provide rationale for the proposed changes, with evidence. We agree with the proposed changes.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R19:** The UTC accepts glyph change for U+06FE ARABIC SIGN SINDHI POSTPOSITION MEN for Unicode 14.0 (Reference: page 2 of L2/20-288).

We also recommend the UTC make the following dispositions:

Assigns an AI to Lorna Evans and the Editorial Committee to prepare text for the Core Spec for Unicode 14.0 on REVERSED COMMA and REVERSED SEMICOLON in Sindhi (reference page 1 of L2/20-288), Sindhi *meem* (reference: page 2 of L2/20-288), and Behdini *heh* (reference: page 3 of L2/20-288).

Forwards the proposed annotation to the names list editor for U+0645 ARABIC LETTER MEEM for Unicode 14.0 (Reference: page 1 of L2/20-288).

---

## 11c. Damma over Damma in Tunisian Quran

**Document:** L2/20-292 Proposal to encode Damma over Damma used in Quran published in Tunisia – Lateef Sagar

**Comments:** DAMMA OVER DAMMA is a *dammatan*, used to represent the sound ‘-un’ in Arabic. L2/99-419R shows the “normal” *dammatan* (below left, U+064C) and the open *dammatan* (below, right, U+08F1), which are usually used contrastively:



Based on this proposal, the vertical *dammatan* is the Tunisian version of the normal *dammatan*, potentially a glyph variant of U+064C ARABIC DAMMATAN. Indeed, TUS mentions (section 9.2, page 369) that U+064C ARABIC DAMMATAN can be rendered “using two *dammatahs* vertically stacked.” The proposal shows its use contrastively with the open *dammatan*.

We recommend Lateef provide the information at the end of the Koran where it explains the use of the characters. Why can’t the normal *dammatan* (U+064C) be used, and could this form be a variant? If the proposed character should be separately encoded, show contrastive use of the proposed character and the normal *dammatan*. If it is proposed, the combining class should be the same as U+064C.

**Recommendation:** We recommend the UTC make the following disposition:  
Assigns an AI to Debbie Anderson to relay feedback to the proposal author of L2/20-292.  
(Reference: Section 11c of L2/21-016 Script Ad Hoc Recommendations).

---

---

## V. SOUTH AND CENTRAL ASIA

### 12 Kannada

**Document:** L2/20-228R Proposal to encode Kannada sign Nakaara Pollu – Srinidhi and Sridatta

**Comments:** This document proposes KANNADA LETTER NAKAARA POLLU. Nakaara Pollu is an historical vowelless form of NA. The UTC has previously approved a Nakaara Pollu for Telugu (proposed in L2/20-084). Shriramana Sharma had earlier recommended the character be handled at the font level with ZWNJ to prevent conjunct formation (L2/13-228). However, he has no objection to the current proposal, given the UTC’s decision to encode TELUGU LETTER NAKAARA POLLU.

Like TELUGU LETTER NAKAARA POLLU, Kannada Nakaara Pollu can be represented today as a font-level variation, but is not generally supported in fonts. It is not used in day-to-day texts and poses no clustering or re-ordering issues.

We recommend the UTC accept KANNADA LETTER NAKAARA POLLU at U+0CDD, so it will match the position in the code charts for TELUGU LETTER NAKAARA POLLU (U+0C5D).

It was noted that Liang Hai had an Action (163-A49) to update the Telugu block introduction to include U+0C5D TELUGU LETTER NAKAARA POLLU. A similar update should be made to the Kannada block intro for this historical character.

**Recommendation:** We recommend that the UTC approve the following:  
**SAH-UTC166-R20:** The UTC accepts 1 Kannada character for encoding in a future version of the standard, with glyph and properties as documented in L2/20-228R:  
0CDD KANNADA LETTER NAKAARA POLLU  
(Reference: L2/20-228R)

We also recommend that the UTC make the following disposition:  
Assigns an AI to Ken Whistler to update the Pipeline. (Reference: L2/20-228R)

Assigns an AI to Liang Hai to update the Kannada block intro in the Core Spec to include U+0CDD KANNADA LETTER NAKAARA POLLU. (Reference: Section 12 of L2/21-016 Script Ad Hoc Recommendations)

### 13 Mundari Bani

**Document:** L2/21-031 Proposal to Encode the Mundari Bani Script in the Universal Character Set – Wolf-Sonkin and Mandal



**Note:** *The comments below are based on the last version the Script Ad Hoc saw. The comments have been accommodated in the latest revision of L2/21-031.*

**Comments:** This proposal for the Mundari Bani script, an alphabet invented in the latter part of the 20th century to write Mundari, a Munda language in the Austroasiatic language family. It is used primarily in the Indian states of Jharkand, West Bengal, and Odisha. The language can also be written in the Devanagari, Bangla, Odia, and Latin scripts. The Mundari Bani script is reported to be taught in 65 schools.

The following are comments on this proposal:

- Add dotted circles for combining characters in the chart.
- Mention both U+002D HYPHEN-MINUS and U+2010 HYPHEN, and that no script-specific character for the dash punctuation is proposed.
- We agree with the proposal author that the historical shapes need not be separately proposed at this time. If a compelling case is made for them later, a proposal can be submitted.
- We recommend putting OJOD after the letters, at U+1E4EB.

The author of the proposal relayed that community groups in Jharkand and Odisha are lobbying to have Mundari included in the Eighth Schedule of the Indian Constitution, so it can become a scheduled language of India.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R21:** The UTC accepts 42 Mundari Bani characters in a new Mundari Bani block (U+1E4D0..U+1E4FF) for encoding in a future version of the standard, with glyphs and properties as documented in L2/21-031.

We also recommend that the UTC make the following dispositions:

Assigns an AI to Ken Whistler to update the Pipeline. (Reference: L2/21-031)

Assigns an AI to the Roadmap Committee to include the new Mundari Bani block (U+1E4D0..U+1E4FF).

Assigns an AI to Lawrence Wolf-Sonkin to provide Michel Suignard with a font.

Assigns an AI to Debbie Anderson to forward a link to the proposal to the Government of India. (*Note: This AI has been done.*)

---

## 14 Old Uyghur

**Document:** L2/20-191 Final proposal to encode Old Uyghur – Pandey

**Comments:** After many rounds of review by the Script Ad Hoc, the Old Uyghur proposal is, in our opinion, mature. The author has incorporated changes as recommended by the Script Ad Hoc.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R22:** The UTC accepts 26 Old Uyghur characters (U+10F70..U+10F89) in a new Old Uyghur block that extends from U+10F70..U+10FAF for encoding in a future version of the standard, with glyphs

and properties as documented in L2/20-191.  
(Reference L2/20-191)

We also recommend that the UTC make the following dispositions:

Assigns an AI to Ken Whistler to update the Pipeline. (Reference: L2/20-191)

Assigns an AI to Anshuman Pandey and Debbie Anderson to provide Michel Suignard with a font.

Assigns an AI to Debbie Anderson to forward a link to the proposal to China NB. *(Note: This AI has been done.)*

---

## 15 Tangsa

**Document:** L2/21-027 Proposal to add the Tangsa Script in the SMP – Morey

**Comments:** We reviewed this revised Tangsa script proposal, now with a fuller background on the history of the use of the script and a sample of the primer presented to the Director of Elementary Education, Government of Arunachal Pradesh in November 2020. After publication of the primer, it is expected the book will be approved for use in teaching.

The collation order should be confirmed by the user community. A recommended approach is to compile a wordlist of ca. 30 words with a mix of letters starting with consonants and vowels and ask members of the community if the order is ok. In addition, provide the wordlist in the Tangsa script alongside a Romanized version, and see if users require the two orders to be the same.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R23:** The UTC accepts 89 Tangsa characters (U+16A70...U+16ABE and U+16AC0..U+16AC9) in a new Tangsa block that extends from U+16A70..U+16ACF for encoding in a future version of the standard, with glyphs and properties as documented in L2/21-027.  
(Reference L2/21-027)

We also recommend the UTC make the following dispositions:

Assigns an AI to Stephen Morey and Debbie Anderson to provide Michel Suignard with a font.

Assigns an AI to Stephen Morey and Debbie Anderson to provide Ken Whistler with the user community's preferred collation order of Tangsa by July 20, 2021. (Reference: Section 15 of L2/21-016 Script Ad Hoc Recommendations).

Assigns an AI to Debbie Anderson to forward a link to the proposal to the Government of India. *(Note: This AI has been done.)*

Assigns an AI to Ken Whistler to update the Pipeline. (Reference L2/21-027)

## 16 Tulu

**Documents:** L2/20-279 Comments on differences between Tulu and Tigalari proposals – Kučera

L2/21-019 Proposal to encode Tulu – Karnataka Tulu Sahitya Academy

L2/17-378 Preliminary proposal to encode Tigalari script in Unicode – Murthy and Rajan

*Related document:*

L2/21-007 Tulu Support Letter from Dr. Guru Prasad

**Comments:** We reviewed the “Proposal to encode Tulu” from the Karnataka Tulu Sahitya Academy, alongside Jan Kučera’s helpful comparison of the Tulu proposal and that of the Tigalari proposal by Murthy and Rajan. A separate document by Vaishnavi Murthy was made available to the group, which contained a conjunct/ligature rendering chart.

The following comments on the Tulu proposal from the Karnataka Tulu Sahitya Academy were raised during discussion:

- Provide background on how non-traditional characters were invented and are currently used, and identify which were invented (vs. those that are attested)
- Provide attested variants of letters directly from materials (using Murthy’s [conjunct/ligature chart](#) as a reference)—not normalized redrawn forms.
- Review Jan Kučera’s comparison document.
- It was noted how the letters *ṛa* and *ḷa* seem to be swapped.

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

Notes these documents (L2/20-279, L2/21-019, L2/17-389 and L2/21-007) but takes no further action.

---

## 17 Vedic

**Documents:** L2/21-006 ISC assignment for *gomukha* characters – Glass and Pandey

*Background docs:*

L2/18-035 Encoding model issues with the Vedic *gomukha* characters – Sharma

L2/18-287 Further remarks on the encoding model of Vedic *gomukha* characters – Sharma

L2/17-098 Request for editorial updates to Indic Scripts – Srinidhi and Sridatta

Cf. Script Ad Hoc recommendations L2/18-300, page 7.

**Comments:** We discussed this request from Andrew Glass, who discovered no Indic Syllabic Categories were assigned for U+1CE9..U+1CEC. Andrew offered to investigate and propose assignments for U+1CE9..U+1CEC, as well as the ISC for other Vedic Extensions. (Note: The lack of ISC for these characters and 1CEE..1CF1 was earlier noted by Norbert Lindenberg, see F8 in L2/20-240.) (Note: Andrew Glass has already been added to AI 165-A49, concerning the Indic shaping properties for Vedic characters.)

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

Notes this document (L2/21-006) but takes no further action.

---

---

## VI. SOUTHEAST ASIA, INDONESIA, AND OCEANIA

### 18 Kawi

#### 18a. Kawi script proposal

**Documents:** [L2/20-284R](#) Proposal to encode Kawi – Aditya Bayu Perdana and Ilham Nurwansah

*Related documents:*

[L2/20-283](#) Repha representation for Kawi – Lindenberg

[L2/21-026](#) Comments on Kawi, L2/20-256 – Sinclair

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

The following highlight points of discussion:

- Based on a comment and evidence provided by Iain Sinclair ([L2/21-026](#)), a graphically distinct AU could be encoded, even though it is based on single attestation. A code point has been reserved for the character (U+11F11), until more examples are provided. The representative glyphs now reflect the early style, as was recommended by Sinclair.
- Kawi *repha* was discussed at length within the Script Ad Hoc. The document from Norbert Lindenberg ([L2/20-283](#)) was found to be extremely useful in documenting the different ways *repha* is handled by various scripts in Unicode. The group and Lindenberg agreed that representations that rely on the use of ZWJ for distinctions have fallen out of favor, as such control characters often cause problems in text input and processing.

For Kawi, three options were discussed for the representation of *repha* or final -r in Indic syllabic category:

1) Consonant\_Preceding\_Repha was the option that the Script Ad Hoc ended up supporting, largely because that value is already used by two other scripts supported in the Universal Shaping Engine (USE), and at least HarfBuzz implements it correctly. The comment in the USE spec on *repha* not being supported applies to Windows' implementation, not to HarfBuzz. Support in CoreText is unclear (Lindenberg reported later he tested both Windows and CoreText afterwards; they're both broken.) The encoding order is also consistent with the common phonetic order.

2) Consonant\_Succeeding\_Repha was considered risky because Kawi would be the only script supported in USE using that class, and would require updates to both USE documentation and implementations to become functional.

3) Consonant\_Final required nontrivial reordering to reproduce the phonetic order. Also, using this class would require the *repha* to interact with many other signs in final position in the syllable.

- One member of the Script Ad Hoc suggested that the visually distinct and the composite forms of KAWI LETTER AA should both be represented by <LETTER A, VOWEL SIGN AA>, because the contrastive use as presented was not strong enough. This did not raise a concern with other members of the group.
- The Script Ad Hoc recommended that the alternate forms of the vowels *-ai* and *-au* (section 5.3.4) be represented using sequences of two KAWI VOWEL SIGN E characters, rather than as font-dependent stylistic variants. Encoding the left part of the alternate form atomically was considered, but not recommended because of the high likelihood that users would use two KAWI VOWEL SIGN E characters anyway (cf. U+0DDB SINHALA VOWEL SIGN KOMBU DEKA). This recommendation has been incorporated in the proposal. As a result of this decision, the specification of the Universal Shaping Engine needs to be updated to support reordering of multiple pre-base vowels, and implementations may need to be updated as well. (Testing shows that all current USE implementations already allow sequences of two pre-base vowels and reorder them to before the base, but the resulting order when using different pre-base vowels varies between implementations.)

Lindenberg reported he is currently working on a test implementation of the script.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R24:** The UTC accepts 87 Kawi characters in a new Kawi block (U+11F00..U+11F5F) for encoding in a future version of the standard, with glyphs and properties as documented in L2/20-284R. (Reference: L2/20-284R)

We also recommend the UTC make the following disposition:

Assigns an AI to Liang Hai to investigate the use of multiple pre-base vowels in clusters of Brahmic scripts, especially rendering of sequences of different pre-base vowels. (Reference: Section 18a of L2/21-016 Script Ad Hoc Recommendations).

Assigns an AI to Ken Whistler to update the Pipeline. (Reference: L2/20-284R)

Assigns an AI to the proposal authors and Norbert Lindenberg to provide Michel Suignard with a Kawi font. (Reference: L2/20-284R)

---

## 18b. KAWI VOWEL SIGN AA

**Document:** [L2/20-287](#) Comments on L2/20-284 regarding disunification of KAWI VOWEL SIGN AA – M. Mahali Syarifuddin

**Comments:** This document recommends revisiting the disunification of KAWI VOWEL SIGN AA into two code points: U+11F34 KAWI VOWEL SIGN AA and U+11F35 KAWI VOWEL SIGN ALTERNATE AA. Bayu reported that the distinction is sometimes orthographic, but it is also sometimes just decorative.

Because the distinction is sometimes orthographic, the overall consensus was to retain the two characters, as proposed.

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

Assigns an AI to Norbert Lindenberg to relay feedback to the author of L2/20-287. (Reference: L2/20-287)

---

## 19 Tai Khao

**Documents:** L2/20-207 Preliminary proposal to encode the Tai Khao script – Kushim Jiang





Feedback:

L2/20-208 A response to Kushim Jiang, "Preliminary proposal to encode the Tai Khao script" – Jim Brase

**Comments:** We reviewed the preliminary proposal and accompanying response for the Tai Khao script, used to write the Tai Dón language, spoken in China, Vietnam, and Laos.

The following summarizes the comments:

- The encoding model for Tai Khao could follow the Indic model or that of its closely related script, Tai Viet, which has visual order. If Indic, a case would need to be made.
- Provide a comparison chart with Tai Viet.
- In section 2.4, xx42 and xx3F can be removed; these seem to be combinations of a mark on the preceding letter plus an ordinary consonant letter in order to indicate a coda. Instead, propose an inverted circumflex on its own.
- Discuss the background on the mark that appears in, for example, LOW CHO and HIGH CHO (with comparison with Tai Viet):

| Tai Khao |   |                          | Tai Viet  |      |
|----------|---|--------------------------|---|------|
| xx0C     |  | TAI KHAO LETTER LOW CHO  |  | AABC |
| xx0D     |  | TAI KHAO LETTER HIGH CHO |  | AABD |

Should the mark be separately encoded, and a decomposition be provided for characters such as xx0D? Comparison with Tai Viet and Thai would be useful.

- Section 2.7:
  - o Provide in-line examples for each proposed mark of punctuation (i.e., not just a chart, but use of the character in a line of text), including COMBINING DOT ABOVE which seems to be missing (or is the proposal suggesting the general character U+0307 should be used?). Cite the figures for each mark on page 6, and in the figures, circle the punctuation and note it in the caption. Example for page 6: "xx5D ... Same as period dot. See figure XX"
  - o COMBINING DOT ABOVE: list what characters it is used with.
  - o The ornamental characters in xxFE and xx5F need examples and additional analysis. In our view, the second and third shapes for xx5F appear to be quite different graphically, though they

may have the same semantics. It is possible the three shapes may need to be separately encoded, if adequate evidence is provided.

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

Assigns an AI to Debbie Anderson to relay feedback to the proposal author of L2/20-207.

(Reference: Section 19 of L2/21-016 Script Ad Hoc Recommendations).

---

## 20 Western Cham

**Documents:** L2/20-285 Comments on Final Proposal to encode Western Cham (L2/20-061) – Leb Ke  
L2/20-061r3 Final Proposal to encode Western Cham in the UCS – Hosken

**Comments:** We reviewed the comments from Leb Ke and the “final” Western Cham proposal, dated November 18, 2020.

The following were comments that arose during discussion:

- The code point and name of ARABIC END OF TEXT now need to be changed. The UTC recently approved ARABIC END OF TEXT MARK at the same code point, U+061D (L2/20-245).
- Given the latest unresolved evidence that the repertoire may not be complete, we consider the proposal not yet ready for approval.
- Patrick Chew will write up a document identifying characters that he considers to be missing from the proposal and include any other comments he may have.

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

Notes these documents and comments (L2/20-285 and L2/20-061r3) but takes no further action.

---

## VII. EAST ASIA

### 21 Bopomofo

**Documents:** L2/20-175 #164 Properties & Algorithms recommendations, F5 Bopomofo Tone Marks

*Related documents:*

Reference implementation: [https://cmex-](https://cmex-30.github.io/Bopomofo_on_Web/testpage/index.html?fbclid=IwAR38r52vf7IOAo0Y3KzwtTA6qGXFewaMSsNe1OGBRqWuyO9txyVI4hjHXKE)

[30.github.io/Bopomofo\\_on\\_Web/testpage/index.html?fbclid=IwAR38r52vf7IOAo0Y3KzwtTA6qGXFewaMSsNe1OGBRqWuyO9txyVI4hjHXKE](https://cmex-30.github.io/Bopomofo_on_Web/testpage/index.html?fbclid=IwAR38r52vf7IOAo0Y3KzwtTA6qGXFewaMSsNe1OGBRqWuyO9txyVI4hjHXKE)

N5088R Proposal to encode two Bopomofo combining tone marks – Selena Wei/TCA et al.

**Comments:** We reviewed item F5, which is feedback from Erika J. Etemad on the orientation of Bopomofo tone marks. The Bopomofo tone marks were discussed at the 2019 WG2 meeting, in relation to a proposal by Selena Wei, TCA, et al. (N5088R).

The following summarizes the discussion:

- Liang Hai to write up the background on the situation. A UTN on Bopomofo encoding and shaping may be needed. Is there an encoding issue or a property change issue that the UTC needs to decide on?
- Based on the discussion, there seemed to be support of existing characters, though perhaps in a different order. For new text, a couple of characters may need to be encoded, such as the COMBINING YANG DEPARTING TONE MARK and the COMBINING YIN DEPARTING TONE MARK, which were proposed in [N5088R](#).

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

Assigns an AI to Liang Hai to 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).

## 22 Kana

**Document:** [L2/20-209R](#) Final proposal to encode Taiwanese Kana in the UCS – Fred Brennan

**Comments:** This is a proposal for Taiwanese Kana, a katakana-based writing system originally created by Japanese linguists to write Taiwanese Hokkien. Hokkien is considered a dialect of Min Nan Chinese, which is spoken in China, Taiwan, and other locations.

This proposal has been reviewed by the Script Ad Hoc a few times. This version incorporates revisions to the character names and minor edits as recommended by the Script Ad Hoc. In our view, it is mature. An earlier version of the proposal was sent to experts in Japan by Ken Lunde; they responded in November that they have no position on the proposal.

The script is typically written vertically, but an example of horizontal typesetting is on page 15. For the dotted and overline characters, no widely accepted names are in use. The request from Marin Silva for named sequences ([L2/20-239](#)) could be evaluated on its own merits, if a proposal were made, but no proposal has been put forward. However, Ken Lunde noted ([L2/20-239](#)) that named sequences are used in certain circumstances when the sequence corresponds to an atomic character in another standard, which is not the case here.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R25:** The UTC accepts 13 Kana characters (U+1AFF0..U+1AFFE) in a new block Kana Extended-B in the range U+1AFF0..U+1AFFF for encoding in a future version of the standard, with glyphs and properties as documented in L2/20-209R. (Reference: L2/20-209R)

We also recommend the UTC make the following disposition:

Assigns an AI to Ken Lunde to add an entry in VerticalOrientation.txt for 1AFF0..U+1AFFE as “U”, after the script is published. (Reference: L2/20-209R)

Assigns an AI to Ken Whistler to update the Pipeline. (Reference: L2/20-209R)

Assigns an AI to the proposal author to provide a font to Michel Suignard. (Reference: L2/20-209R)



Assigns an AI to the Roadmap Committee to include a new Kana Extended-B block (U+1AFF0..U+1AFFF). (Reference: L2/20-209R) *(Note: This AI has been done in the draft Roadmap, version 13-0-5.)*

---

## 23 Kanbun

**Document:** L2/20-276 Suggestion concerning tentative roadmap placement of Kanbun Extended block – Moist

**Comments:** We reviewed this request for the Roadmap Committee to reserve U+2FE0..U+2FEF in the BMP for the Kanbun Extended block, instead of U+16FB0..U+16FBF in the SMP (as recommended in Script Ad Hoc Recommendations L2/20-250).

A strong case would need to be made to place characters on the BMP, and in our view, the single open column at U+2FE0..U+2FEF should be used for characters with a valid case for encoding on the BMP. The Kanbun Extended block does not, in our opinion, fit this criterium. Ken Lunde also agrees with this view.

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

Assigns an AI to Ken Lunde to relay feedback to the proposal authors. (Reference: L2/20-276)

Assigns an AI to the Roadmap Committee to reserve U+16FB0..U+16FBF for Kanbun Extended block.

---

## 24 Tangut

**Document:** L2/21-034 Feedback on L2/20-166 Tangut Glyph Modifications and Corrections – Sun Bojun

*Related document:* L2/20-166 Tangut Glyph Modifications and Corrections – West and Zaytsev

**Comments:** We reviewed the feedback from Dr. Sun Bojun and colleagues on L2/20-166, a document that proposed modifications to 72 Tangut components, 1,493 Tangut ideographs, as well as glyph corrections for 2 Tangut components and 32 Tangut ideographs. The proposed changes came about as a result of Unicode 13.0 disunifications (proposed in L2/19-207 and accepted at the July 2019 UTC meeting).

The July 2020 Script Ad Hoc report (L2/20-169, section 23, p. 31) had recommended experts review the proposed changes. The Script Ad Hoc wanted confirmation that experts understood the ramifications of the proposed glyph changes. The feedback document confirms that the experts consulted are in agreement with L2/20-166. Andrew West has also confirmed he has a font for the glyph changes and would like the changes incorporated in Unicode 14.0.

**Recommendation:** We recommend that the UTC approve the following:

**SAH-UTC166-R26:** The UTC accepts the Tangut glyph changes in L2/20-166.

(Reference: L2/20-166)

---

---

---

## VIII. SYMBOLS, PUNCTUATION, AND NOTATIONAL SYSTEMS

### 25 Math Calligraphic Alphabets

**Documents:** [L2/20-275R](#) Proposed variation sequences for math calligraphic letters – Sargent

*Related document:* [L2/20-281](#) Recent evolution of math alphabetic calligraphic script style – Hudson

**Comments:** We reviewed this document, which proposes 104 variation sequences for math calligraphic letters.

The representative glyphs for the regular and bold script characters in the Mathematical Alphanumeric Symbols block (and a small set of regular script characters in the Letterlike Symbols block) are in the roundhand style. The chancery style is more common today. The two script styles were unified when the math alphanumerics were added in Unicode 3.1, but the two styles are no longer always considered interchangeable, as shown by examples in the document. (The document from John Hudson, [L2/20-281](#), gives insight into the background, showing the evolution of styles over time)

Different options were discussed on how to handle math calligraphic alphabets. One option was to propose separate new characters for those characters not currently shown in the code charts. The consensus of the Script Ad Hoc, however, was in favor of Variation Sequences.

The differences between the two styles were not discernable in lowercase, so the lowercase forms are not being proposed. Lowercase forms could be handled by variation sequences later, if deemed necessary.

In our view, sequences for bold should not be proposed, since their use is not attested. As a result, we recommend only 52 variation sequences be approved.

**Recommendation:** We recommend that the UTC approve the following, after discussion:

**SAH-UTC166-R27:** The UTC accepts 52 variation sequences to distinguish roundhand and chancery style mathematical script alphabetic characters. (Note: This does not include distinguishing sequences for boldface.) (Reference: [L2/20-275R](#))

---

---

### 26 Persian Siyaq

**Document:** [L2/20-249](#) Proposal to Encode Persian Siyaq Numbers in Unicode – Pandey

**Comments:** Below are comments raised during discussion.

- In section 2.2, mention this set of characters are not used as stand-alone characters.
- Remove TWO TOMAN in 2.4
- Add a conversion chart.
- In 2.6, add “PREFIXED” to PERSIAN SIYAQ NUMBER NINETY.

- In 2.13, discuss how the decision was made for PLACEHOLDER as a spacing mark (and not a combining mark), with justification on why the choice was made.
- Provide an analysis of large numbers on how they work. Do they have structure? Depending upon the analysis, a recommendation can be made whether format controls are needed or not.
- Send the proposal to Shervin Afshar.
- Address comments from David Corbett from Sept. 30

**Recommendation:** We recommend the UTC make the following disposition:  
Assigns an AI to Debbie Anderson to relay feedback to the proposal author of L2/20-249.  
(Reference: Section 26 of L2/21-016 Script Ad Hoc Recommendations).

---

## IX. OTHER TOPICS

### 27 Variation Sequences

**Documents:** L2/20-244R Variation Sequences for combining marks – Lindenberg

*Feedback and Related Documents:*

L2/20-247 Restrictions on base characters of variation sequences (L2/20-244) – Charlotte Buff

L2/21-011 Comments on Public Review Issues (September 23, 2020 - January 8, 2021)

L2/21-012 UTC #166 properties feedback & recommendations

**Comments:** We reviewed L2/20-244R, and agree with the general principle expressed in this document. We also recommend the topic be included in the Core Spec.

In our view, the exact details on variation sequences for combining marks is more appropriate for other groups (see full details with background and recommendations in section D2 of L2/21-012 UTC #166 properties feedback & recommendation, pp. 9-11., which takes Charlotte Buff's feedback into account).

**Recommendation:** We recommend the UTC make the following disposition:  
Notes the comments above and defers to the recommendations in L2/21-012, section D2.  
(Reference: Section 27 of L2/21-016 Script Ad Hoc Recommendations).

---

### 28 CCC of Znamenny Notation Characters

**Document:** L2/21-033 Canonical Combining Classes of Znamenny Notation Characters – Buff

**Comments:** We reviewed this document, which requested the Znamenny musical notation combining marks be given non-zero CCC values, based on their placement on and around base characters.

The Znamenny characters were approved for Unicode version 14.0, based on the proposal L2/19-053, which assigned the combining marks be CCC=0. Earlier versions of the proposal seen by the Script Ad Hoc had suggested non-zero canonical combining classes, but the Script Ad Hoc advocated CCC=0, because non-zero combining classes can cause problems for normalization, if not properly assigned. If

mistakes are made in initial assignments, because of the complexity of the rendering model, they nevertheless cannot be corrected in the future because of stability guarantees.

Buff has done considerable work and proposes CCC values on pages 5-8. A problem case was identified in Buff's analysis, however: #100 and #101 on page 31 of the [L2/19-053](#) has 1CF24, which is not a left-only mark. To assist in review of the CCC values, a list of all the marks by grouping (i.e., a list of all the glyphs of the Attached\_Right marks, all the Below\_Left marks, etc.) is also requested.

In our view, it is safest to stay with CCC=0 at this point. If additional information is provided, the CCC values may be re-visited, though any change would need to come before the April 2021 Unicode Technical Committee meeting, when version 14.0 will be locked down. The Properties & Algorithm group should probably be consulted on any property change issues.

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

Assigns an AI to Deborah Anderson to relay the comments above to the author of L2/21-033.

(Reference: Section 28 of L2/21-016 Script Ad Hoc Recommendations).

---

---

## X. PUBLIC REVIEW FEEDBACK


**Document:** [L2/21-011](#) Public Review Feedback


### ARABIC

#### *a. From Fawaz Ahmed, October 5, 2020, 06E0 Character has wrong description*

*Public Feedback:* U+06E0 should be described as “circular zero”.

*Response from SAH:*

The character's name is “ARABIC SMALL HIGH UPRIGHT RECTANGULAR ZERO” and glyph is . The character is a Quranic annotation sign, and comes from Unicode 1.1, inherited from ISO/IEC 10646.

This character was discussed in [PRI 73](#), which reported that some glyphs were incorrect, provided evidence, and proposed changing them. The proposed glyph change for U+06E0 in PRI 73 is that currently in the code charts, . Additional feedback from Kamal Mansour in [L2/05-231](#), makes it clear the name “rectangular zero” is a translation of the Arabic name. To prevent future confusion, we recommend an annotation be added.

*Recommendation:*

We recommend the UTC make the following disposition:

Assigns an AI to Ken Whistler to add an annotation to U+0630, indicating the name is a translation of the Arabic name. (Reference: Section X of L2/21-016 Script Ad Hoc Recommendations).

***b. From David Corbett, October 12, 2020, Comments on L2/19-306***

*Public Feedback question 1: What does final ARABIC LETTER THIN YEH look like?*

*Response from SAH:*

There are no examples of the final or isolated forms of ARABIC LETTER THIN YEH.

*Recommendation:*

We recommend the UTC make the following dispositions:

Assigns an AI to Ken Whistler to add an annotation U+0886 ARABIC LETTER THIN YEH that no final or isolated forms are attested. (Reference: Section X of L2/21-016 Script Ad Hoc Recommendations).

Assigns an AI to EdComm to add a note to Table 9-8 “Dual-Joining Arabic Characters” of the Core Spec for Unicode 14.0, that U+0886 ARABIC LETTER THIN YEH has no final or isolated forms attested. (Reference: Section X of L2/21-016 Script Ad Hoc Recommendations).

*Public Feedback question 2: If U+0887 ARABIC BASELINE ROUND DOT has gc=Lo, why does U+0888 ARABIC RAISED ROUND DOT have gc=Sk?*

*Response from SAH:*

U+0887 behaves like a letter, whereas U+0888 is a diacritic, modifying a letter. The distinction between Lo (or Lm) versus Sk is that those characters whose general category is Lo (or Lm) can be used in identifiers. Ideally, identifiers should include letters and letter modifiers used in ordinary textual orthographies. On the other hand, those characters whose gc=Sk cannot be in identifiers by default. In theory, Quranic annotations would not be considered an ordinary orthography. It is easier to keep Sk and change it, if needed, to Lo (or Lm), but harder to go from Lo (or Lm) to Sk. In sum, we consider the general category designations to be correct.

*Recommendation:*

We recommend the UTC make the following disposition:

Assigns an AI to Rick McGowan to relay the feedback above to David Corbett.  
(Reference: Section X of L2/21-016 Script Ad Hoc Recommendations).

*Public Feedback question 3: U+08CE ARABIC LARGE ROUND DOT ABOVE behaves like ARABIC HAMZA ABOVE. Add to UTR #53? Add U+08CF ARABIC LARGE ROUND DOT BELOW, too?*

*Response from SAH:*

In our view, this recommendation makes sense.

*Recommendation:*

We recommend the UTC make the following disposition:

Assigns an AI to Lorna Evans to revise the proposed update of UTR #53 Unicode Arabic Mark Rendering, adding U+08CE and U+08CF as Modifier Combining Marks (MCM), and include attestation from the document <https://app.quranflash.com/book/Warsh2?en#/reader> (such as page 560ff.). (Reference: Section X of L2/21-016 Script Ad Hoc Recommendations).

## **Javanese**

**From PANDI ID Registry (Alicia Nabilla), December 14, 2020, PANDI Inquiries**

**Public Feedback:** A query was received from the PANDI .id registry, asking how to change the status of Javanese script from “Limited Use” to “Recommended” for their IDN process to ICANN.

### **Response from SAH:**

Domain names falls outside the purview of the Script Ad Hoc and we defer to the comments in the Properties & Algorithms report (L2/21-012), which requests PANDI work with ICANN. A document with criteria on Identifier\_Type=Recommended in UAX #31 was recently posted (L2/21-030). The Script Ad Hoc can assist on the decision regarding which scripts should be Identifier\_Type=Recommended, based on the criteria in L2/21-030.

### **Recommendation:**

We recommend the UTC make the following disposition:

Assigns an AI to Rick McGowan to relay the feedback above to the PANDI representative.  
(Reference: Section X of L2/21-016 Script Ad Hoc Recommendations).

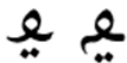
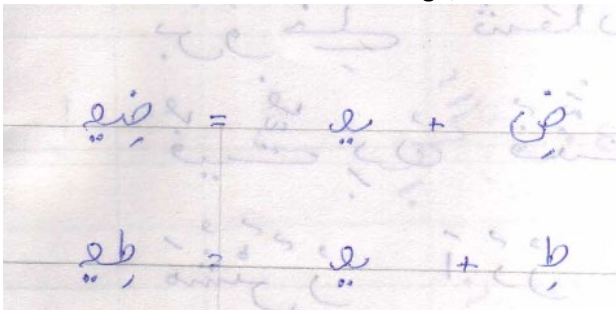
## **Rohingya**

**From David Corbett, Nov. 26, 2020, Isolated U+08AC ARABIC LETTER ROHINGYA YEH**

**Public Feedback:** The Core Spec states there is no isolated form of U+08AC ARABIC LETTER ROHINGYA YEH, but it may exist.

### **Response from SAH:**

Lorna Evans confirmed that the isolated form appeared in teaching materials. The form was discovered after the script was encoded. Below is an image provided by Evans, and the Scheherazade forms below the image, with the isolate on the left, the final on the right.



### **Recommendation:**

We recommend the UTC make the following disposition:

Assigns an AI to Lorna Evans and the EdComm to provide a glyph for the isolated form of ROHINGYA YEH for Table 9-9, and remove the note “Isolated form does not occur.”  
(Reference: Section X of L2/21-016 Script Ad Hoc Recommendations).

## ***SignWriting***

### ***a From David Corbett, Dec. 4, 2020, Rendering U+1D9FF SIGNWRITING HEAD with forehead marks***

#### ***Public Feedback from David Corbett:***

Some SignWriting marks are placed on the forehead. To avoid overlapping the top part of U+1D9FF SIGNWRITING HEAD, the top part of U+1D9FF is omitted. This is encoded as <U+1D9FF, U+1DA9B> (SignWriting ID 04-01-001-01-02-01). The corresponding symbols in ISWA 2010 (<http://www.signbank.org/iswa/>) are drawn with the top part of the head omitted; an example is 04-01-003-01-04-03. Should 04-01-003-01-04-03 be encoded as <1D9FF, 1DA01, 1DA9D, 1DAA2> or as <1D9FF, 1DA9B, 1DA01, 1DA9D, 1DAA2>?

#### ***SAH response:***

Kamal Mansour, who implemented a Noto font for SignWriting, reports that the two different sequences can correctly produce the same visuals in Noto; there is no promise for unique visual results for distinct character sequences.

#### ***Recommendation:***

We recommend the UTC make the following disposition:  
Assigns an AI to Rick McGowan to relay the feedback above to David Corbett.  
(Reference: Section X of L2/21-016 Script Ad Hoc Recommendations).

### ***b From David Corbett, Dec. 7, 2020, Rendering SignWriting symbols without valid fill-1***

#### ***Public Feedback from David Corbett:***

According to chapter 21, in the section for SignWriting, “There are no explicit modifiers encoded for fill-1 or rotation-1, as those values are considered inherent in the base character”. However, there are some characters for which fill-1 is not valid, such as U+1D8F6 SIGNWRITING HAND-FIST THUMB HEEL. How should such characters be rendered when not followed by a valid fill modifier?

#### ***SAH response:***

Kamal Mansour reports this comment reveals an error (i.e., declaring that the nominal encoded character is invalid) and should be corrected in the Noto font and in documentation. He comments:

In all other cases, the default fill variant (the implied F1) of any character represents its nominal encoded form which must be assumed as valid. U+1D8F6 is the only exception to this normal rule. By inspecting the lists of valid and invalid character sequences for Sutton Sign Writing, I’ve verified that there are NO single-character, invalid codes. The existence of F2-variants of U+1D8F6, along with the absence of F1-forms, leads me to think that the proposers made a mistake in their specification. Since there is only one fill-form (plus rotations) of U+1D8F6, we should change all the F2 forms to the corresponding F1 equivalents. We will simultaneously need to add new entries in the list of invalid sequences for all F2–F2 variants of U+1D8F6.

No detailed information on how to handle variations is provided by the Unicode Standard. To address questions such as the question from David Corbett, better documentation is clearly needed to ensure interoperability. The recently released Noto font will be a useful baseline of shapes and variants, and can be used to create a UTN on SignWriting.

It was noted that the valid and invalid sequences for SignWriting are currently located at: <https://www.evertype.com/standards/iso10646/pdf/signwritingsequences.txt>. Putting this large file (4MB) in the UCD was not deemed appropriate. If such a request were made, a strong rationale would be needed, and should be sent to the Properties & Algorithms group. In any event, a concrete proposal to consider the options on how to make the valid and invalid sequences accessible is invited. Before perpetual maintenance of any complex data file can be entertained by the UTC, it would first need approval of a documenting context, preferably in the form of a Unicode Technical Report, specifying the rendering model for SignWriting and how the data file would be used algorithmically in that model.

*Recommendation:*

We recommend the UTC make the following dispositions:

Assigns an AI to Debbie Anderson to relay the feedback above to the authors of the SignWriting proposal, pointing out the error. (Reference: Section X of L2/21-016 Script Ad Hoc Recommendations)

Assigns an AI to Debbie Anderson to relay to comments above to Kamal Mansour concerning how to make the valid and invalid SignWriting sequences accessible. (Reference: Section X of L2/21-016 Script Ad Hoc Recommendations)

Assigns an AI to Kamal Mansour to provide tables of glyphs from the font for review by Peter Constable and other interested parties and can be used as a basis for a future UTN. (Reference: Section X of L2/21-016 Script Ad Hoc Recommendations)

***Todhri/Vithkuqi***

***From David Corbett, Oct 1 2020, Comments on L2/20-187R and L2/20-188R***

*Public Feedback:* Consider handling Albanian “e” and “ë” in Todhri and Vithkuqi in the same manner as in Elbasan (i.e., ASCIIfy them as EI and E).

*SAH response:*

The feedback was forwarded to the proposal author (and the change has been incorporated in the latest version of the proposals).

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

Assigns an AI to Rick McGowan to thank David Corbett for his feedback. (Reference: Section X of L2/21-016 Script Ad Hoc Recommendations).



### ***Other (Currency symbol)***

***From Eduardo Marín Silva, Oct 2 2020, Kyrgyz som observation***

*Public Feedback:* Consider changing the name of the KYRGYZ SOM character to SOM SIGN.

*SAH response:*

The following feedback has been taken into account in the October 2020 SAH Recommendations (L2/20-250, p. 16).

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

Assigns an AI to Rick McGowan to thank Eduardo Marín Silva for his feedback.

(Reference: Section X of L2/21-016 Script Ad Hoc Recommendations).

### ***Other (Variation Sequences L2/20-244, L2/20-247)***

***From David Corbett, Sept 30 2020, Comment on L2/20-247***

*Public Feedback:* A different option regarding a composed code point is suggested.

*SAH response:*

This comment was taken into account by the Properties and Algorithm group (see [L2/21-012](#)).

### **Public Review Feedback to be carried over to next set of recommendations:**

Mandaic (Dec. 20, 2020 from David Corbett) and Old Hungarian (Nov. 26 2020 from David Corbett)

## **XI. RECOMMENDATIONS FOR UNICODE 14.0**

### **Script and Character Additions**

Of the various approval recommendations made in the above sections of this report, we recommend that the UTC specifically target the following subset of recommended approvals for immediate processing and eventual publication in Version 14.0 of the Unicode Standard. In the identifying lists provided below, links are provided to the relevant proposal documents, but note that in some cases, code points and/or character names may have been adjusted from those in the proposals. Recommended code points, character names, glyphs, sequences, and properties are as documented in the case-by-case recommendations above in this report.

Cypro-Minoan (1 character) ([L2/20-156R](#))

Ethiopic (Gurage additions) ([L2/21-037](#))

Kana (additions [L2/20-209R](#))

Kannada (1 character [L2/20-228R](#))

Latin

Phonetic characters ([L2/20-266=L2/21-021](#))

IPA Modifier Letters – Pulmonic ([L2/20-252R](#))

IPA Modifier Letter – Non-Pulmonic ([L2/20-253R](#))

Modifier Latin capital letters ([L2/20-251](#))

Oblique hyphen ([L2/21-036](#))

Old Polish O ([L2/21-039](#))

Phonetic punctuation and diacritics ([L2/21-042](#), minus two characters)

Additional Para-IPA letters (L2/21-041)  
Dezh with retroflex hook (L2/21-004)  
Medieval punctuation (L2/20-270R)  
Sigmoid S (L2/20-269)  
Two characters for Middle Scots (L2/20-267)  
Kawi (L2/20-284R)  
Mundari Bani (L2/21-031)  
Old Uyghur (L2/20-191)  
Tangsa (L2/21-027)  
Vithkuqi (L2/20-187R2)

We invite the UTC to discuss whether there are any reservations about the inclusion of any of the specified scripts or collections in 14.0, given the results of the item-by-item discussions.

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

Assigns an AI to Ken Whistler to adjust the Pipeline to reflect the decision about which repertoire to include in 14.0 (and which will be in the bucket for publication in a future release).

### **Glyph Changes**

We recommend that all the glyph changes recommended for approval in this report be targeted for publication in Unicode 14.0. This includes changes for the following:

Arabic (for Kazakh, Kyrgyz, and Uyghur L2/20-289; for one Sindhi character L2/20-288)

Egyptian Hieroglyphs (L2/21-028)

Tangut (L2/20-166)

### **Variation Sequences**

We recommend that the following variation sequences recommended for approval in this report be targeted for publication in Unicode 14.0:

Math calligraphic alphabets (L2/20-275R).