L2/19-184

TO: **UTC**
FROM: **Deborah Anderson, Ken Whistler, Roozbeh Pournader, Lisa Moore, Liang Hai, Ben Yang, Craig Cornelius, Richard Cook, Norbert Lindenberg, Manish Goregaokar, Marek Jeziorek, Chris Chapman, and Steven Loomis**
SUBJECT: **Recommendations to UTC #159 April-May 2019 on Script Proposals**
DATE: **April 29, 2019**

The Script Ad Hoc group met on 15 February, 22 March, and 16 April 2019 in order to review proposals. The following represents feedback on proposals that were posted in the Unicode document registry at the time the group met. A table of contents is provided below.

	page
AMERICAS	
Mayan Hieroglyphs	2
EUROPE	
Cypro-Minoan	2
Latin	3
Palaeohispanic	4
AFRICA	
Adlam	4
Garay	6
MIDDLE EAST	
Arabic	8
Hebrew	9
Yezidi	9
SOUTH AND CENTRAL ASIA	
Bengali	10
Chakma	12
Gurmukhi	12
Malayalam	13
Mongolian	13
Newa	15
Old Turkic	16
INDONESIA AND OCEANIA	
Javanese	16
EAST ASIA	
Bopomofo	17
Naxi Dongba	17
Tangut	17
SYMBOLS AND NUMERICAL NOTATION SYSTEMS	
Ascia symbols	18
Block Elements and Symbols	18
Breath Punctuation Marks	19
Christian symbols	19
Cross Symbols	20
Half-length Hyphen	21
Znamenny Notation	21
Zoroastrian Symbol	22

OTHER

Italics in Plain text 22

AMERICA

1. Mayan

Document: L2/19-171 Final list of characters and Quadrats for Mayan Codices – Pallán

Comments: We briefly reviewed these documents, which contain final versions of the glyph blocks and list of characters contained in L2/18-038 Preliminary Proposal for Encoding Mayan Hieroglyphic Text.

We recommend the author seek review of these lists by scholars. Also, how many controls are needed, based on quadrats?

Recommendations: We recommend UTC members review these documents at their leisure.

EUROPE

2. Cypro-Minoan

Document: L2/19-166 Brief report from meeting 13-14 March with experts – Pournader and Anderson
Background documents: L2/16-179 Revised proposal to encode the Cypro-Minoan script in the SMP

Comments: We reviewed the written summary.

As relayed by Roozbeh Pournader, Cypro-Minoan has been divided into four eras:

- CM0 is the earliest set of signs, but is based on one document, which showed no repetition of signs.
- CM1 includes the largest number of artifacts, though some inscriptions are quite short.
- CM2 comprises just four tablets, but the four are lengthy, and hence are an important source.
- CM3 are artifacts from the coast of Syria. CM3 could be a branch of CM1.

CM1, CM2, and CM3 glyphs will be represented by fonts

According to the experts at the meeting, Olivier is the major reference used today, although it was acknowledged Oliver's repertoire very likely includes some variants – but exactly which signs are variants is still a matter of dispute. It was decided to use Olivier as the basis of the repertoire, and to not rely on Masson's earlier work. If signs from Masson are needed, they can be proposed and added later.

Other points raised include:

- The punctuation marks may be unifiable to Aegean punctuation.
- There are only ten examples of numbers, and they may need to be encoded visually, since their value is uncertain.

Additional feedback from the two participating experts is expected; that information can be incorporated into a revised proposal.

Recommendations: We recommend the UTC encourage continued discussion with experts.

3. Latin

a. Latin Letter Turned Capital Y

Document: L2/19-072 LATIN LETTER TURNED CAPITAL Y – Schüler

Comments: We reviewed this one-page request for LATIN LETTER TURNED Y.

The rationale given in this request is: "An increasing number of people create turned or flipped text using Unicode's turned or flipped versions of letters." However, the document does not provide any evidence to support the claim. What would the character be used for? If the letter were shown to be used in a well-established orthography (with evidence and a full proposal as outlined on the Unicode website), we could re-consider the request.

Recommendations: We recommend the UTC forward the above comments to the author.

b. Latin Letter Reversed Half H

Document L2/19-092 Proposal to encode Latin Letter Reversed Half H (WG2 N5039) - West and Everson

Comments: We reviewed this proposal to encode a casing pair of Latin letters for reversed half h that appear in Latin inscriptions from Gaul.

Only the uppercase appears in inscriptions and printed text, but the lowercase is proposed here, as was the case for U+2C75/U+2C76 LATIN LETTER HALF H (L2/05-193).

Recommendations: We recommend the UTC approve the following two characters in the Latin Extended-D block:

A7F5 LATIN CAPITAL LETTER REVERSED HALF H
A7F6 LATIN SMALL LETTER REVERSED HALF H

c. Phonetic characters for Scots

Document: L2/19-075 Proposal to add six phonetic characters for Scots – Everson

Comments: We reviewed this proposal for six characters used in linguistic transcriptions of Scots found in the *Scottish National Dictionary* and the *Linguistic Atlas of Scotland*. Published evidence is provided for the characters.

During discussion, the following comments were made:

- The caption in figure 1 needs to be fixed, as the glyphs for the combining characters are incorrect.
- Use the next two available code points, U+1ABF and U+1AC0 for the two combining characters, instead of U+1AC1 and U+1AC2.

The author may want to contact those involved in the Dictionary of the Scots Language (DSL) online project (www.dsl.ac.uk), which has digitized and makes publicly available the *Scottish National Dictionary*, as well as *A Dictionary of the Older Scottish Tongue*. In DSL, the MODIFIER LETTER SMALL TURNED W appears to be represented by U+028D LATIN SMALL LETTER TURNED W, presumably superscripted.

Recommendations: We recommend the UTC accept the following six characters, with adjusted code points suggested for the two combining characters as listed below:

[Combining Diacritical Marks Extended]
1ABF COMBINING LATIN SMALL LETTER W BELOW
1AC0 COMBINING LATIN SMALL LETTER TURNED W BELOW

[Latin Extended-E]
AB68 LATIN SMALL LETTER TURNED R WITH MIDDLE TILDE
AB69 MODIFIER LETTER SMALL TURNED W
AB6A MODIFIER LETTER LEFT TACK
AB6B MODIFIER LETTER RIGHT TACK

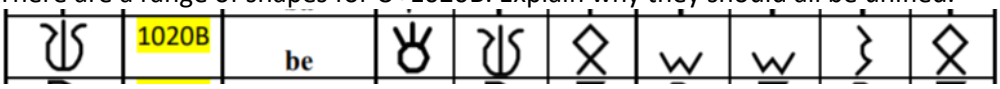
4. Palaeohispanic

Document: L2/19-045 New charts for Northern and Southern Palaeohispanic – Ferrer et al.
Background document:
L2/18-283 Proposal to encode the Palaeohispanic script – Ferrer et al.

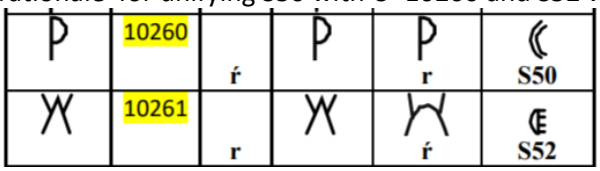
Comments: We reviewed the new Northern and Southern Palaeohispanic charts, which now include variants. The overall encoding model – separating a single, unified script into Northern Palaeohispanic and Southern Palaeohispanic – seems to be the correct approach, in our opinion.

The following comments were made:

- There are a range of shapes for U+1020B. Explain why they should all be unified.



- The chart for Southern Palaeohispanic will need careful review, including providing a strong rationale for unifying S50 with U+10260 and S52 with U+10261.



- We recommend a review be done comparing the earlier versions of the proposal versus the latest charts to ensure that no characters or significant glyphs have been removed.

Recommendations: We recommend the UTC members send feedback to the authors and encourage ongoing discussion with experts.

AFRICA

5. Adlam

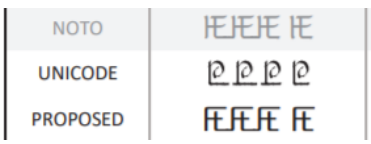
Document: L2/19-119 Replacement of Adlam Reference Font in Codesheet to Updated Design – Patel, Jamra, Cornelius, I. Barry and A. Barry
Reference document: *L2/14-219 Revised Proposal for encoding the Adlam script in the SMP – Everson*

Comments: We reviewed this document which requests a change of the current code chart font for Adlam to the set of proposed glyphs (which are found in the Ebrima / Kigelia fonts). The comparison

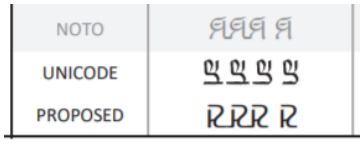
chart shows the glyphs from the 2007 non-Unicode font Pulaar, Noto Sans, the glyphs as documented in the Unicode code chart and L2/14-219, and the proposed shapes, with comments describing the changes and any rationale for the changes.

The following highlights the discussion:

- Since the publication in Unicode 9.0 in 2016, there have been several changes to the glyphs, notably in the supplementary letters (see examples below), which are used in loanwords. The document states although such words are common, their use varies by region, and users can identify the characters through context.



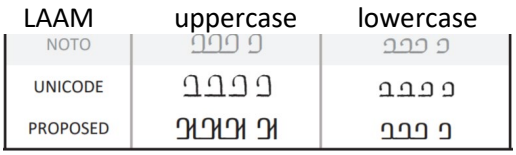
CAPITAL LETTER KHA U+1E91D

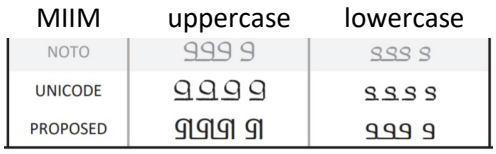


CAPITAL LETTER GBE U+1E91E

CAPITAL LETTER SHA U+1E921

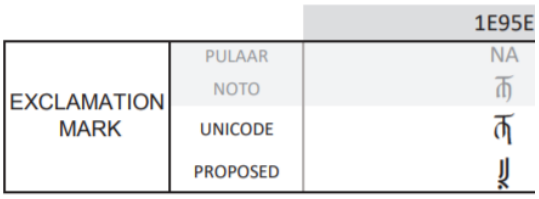
Another set of significant glyph changes involve those made to uppercase letters, to differentiate them from lowercase forms. Examples include LAAM and MIIM (below).

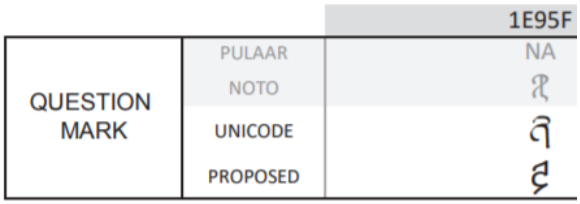




The marked difference in glyph shape for these characters raised the question whether the new shapes are new graphemes and hence should be considered new characters. However, Adlam is a relatively young script that does not have much digital data yet, so it would be more prudent to change the glyphs to the new shapes, in our opinion, rather than maintain the early graphic history in the code charts. If, in the future, there is a project to digitize content from earlier materials, the earlier shapes can be separately proposed, as was done for Bamum.

- The two punctuation characters, exclamation mark and question mark, also are quite different from the current code chart (and Noto Sans):





Ibrahim Barry, one of the script's creators, stated that the characters are reported not to be in wide use. These marks are used at the beginning of sentences.

- A question was raised about the two quite different shapes for U+1E945 VOWEL LENGTHENER:

The original proposal (L2/14-219R) explained the usage:

> **3.2. The ADLAM VOWEL LENGTHENER** is used with other vowels: 𞤕 *du* is lengthened to 𞤕𞥄 *dū*. This typically has a different glyph shape when used with the capital letters than when it is used with the small letters: compare Ú *Ī* and ʊ̄ *ī*.

The right-hand shape (◌̄) should appear in the code chart.

- The digit "6" is correct in Noto Sans, but incorrect in the Ebrima and Kigelia fonts. The Kigelia font has been fixed and an updated version of Ebrima with the correct "6" glyph is slated for release soon.
- Because the script has been constantly evolving and its glyphic repertoire has changed, the question was raised how much more will the script change? Ibrahim Barry stated this is the last set of changes; the script now will be easy for users to read. (It was noted that even if the new glyph changes are incorporated in fonts, old Noto Sans implementations will still be circulating for ca. 5 years on mobile devices.)
- Today the Noto Sans and Ebrima/Kigelia fonts are used. (Kigelia is a different typeface from Ebrima but has the same underlying structure.) The Ebrima font is part of the Windows 7(+) operating system. Kigelia is available in an app from JamraPatel, but is slated to be released in Office later this year. Based on the script creators' view, the Noto Sans and Ebrima (/Kigelia) fonts are both usable, but the glyphs in the Ebrima (/Kigelia) are preferred.

In sum, the Script Ad Hoc participants see the merits of making the glyph changes as proposed.
(Note: After the Ad Hoc met, Kamal Mansour of Monotype communicated that he was satisfied with glyph changes and the rationale, based on L2/19-119.)

Recommendations: We recommend the UTC discuss this document, and decide whether to change the font in the code chart.

6. Garay
a. Feedback (to Script Ad Hoc)
Document: L2/19-163 Feedback on Garay – Rovenchak
Background documents:
L2/16-069 Proposal for encoding the Garay script in the SMP – Everson
L2/18-168 Script Recommendations (comments on Garay are on pp. 28-29)

Comments: We reviewed this feedback document, which contained responses from the user community on outstanding questions posed by the Script Ad Hoc in L2/18-168.

The user community agreed to follow the basic set of recommendations made by the Script Ad Hoc (covered by #1-#3 in this document), resulting in:

- Removal of four letters from the repertoire (MBA, NGGA, NDA and NJA) and documentation of they should be represented (i.e., base letter and VOWEL SIGN E).
- Separately encoding a combining dot above character and removal of NJA and NGA.
- Changing the canonical combining class for VOWEL SIGN E and COMBINING GEMINATION MARK to 230.

The authors were uncertain about the request to change the Bidi_Class for the digits (#4). We recommend letters be made AL and numbers AN, but authors provide examples of numbers in context (i.e., where they occur next to punctuation, a plus or minus, section sign, or percentage). The examples

will help in determining the appropriate Bidi_Class for the characters. (Note: Additional information on math notation is contained in a separate document, L2/19-162, discussed below.)

On the remaining topics (#5-#6), we recommend the information on the macron and the use of the squiggle be incorporated into a revised proposal. The Script Ad Hoc and the UTC can then review the entire proposal and provide additional comments or feedback.

Recommendations: We recommend the UTC forward the comments above to the author.

b. Garay math notation
Document: L2/19-162 Garay math notation– Rovenchak and Riley
Background documents:
L2/16-069 Proposal for encoding the Garay script in the SMP – Everson
L2/18-168 Script Recommendations (comments on Garay are on pp. 28-29)

Comments: We reviewed this document, which provided useful examples and details on Garay mathematical notation.

The following is a summary of the discussion:

- In our view, U+002B PLUS SIGN can be used for the Garay division mark. To represent the Garay subtraction mark, U+22A5 ⊥ UP TACK can be used, and for the Garay addition mark, U+22A4 ⊤ DOWN TACK. Both characters have the general category "Sm", are Other_Neutral for their Bidi_Class, and are located in the Mathematical Operators block.
- Use of U+10D62 GARAY CAPITAL LETTER YA is acceptable for the Garay multiplication mark.
- In figure 1 the "+" (indicating division) appears upright in an equation whose numbers are also upright. In figure 2, the addition mark and numbers look italicized (i.e., slanted to the right). In figure 3 the numbers are upright, but the subtraction and addition mark appear again italicized.

  Figure 1:

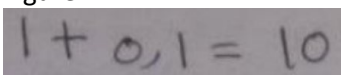
  

  Figure 2:

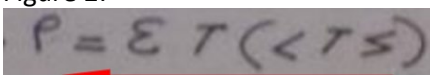
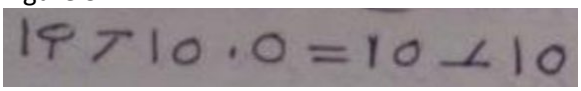
  

  Figure 3:

  Is the italicization part of the identity of the characters? Are there upright versions of the addition and subtract signs? To help answer this question, provide additional examples showing the mathematical signs. (Compare section of 22.2 The Unicode Standard 12.0 ["Mathematical Alphabets" p. 812], which explains that in mathematical notation italics have distinct semantics.)
- It was noted that the change of direction for math is found elsewhere, including in the Arab world.

Recommendations: We recommend the UTC members review this document at their leisure and forward the comments above to the document authors.

MIDDLE EAST

7. Arabic

a. ARABIC LETTER JEEM WITH THREE DOTS

Document: L2/19-118 Proposal to add Arabic letter JEEM WITH THREE DOTS (revised) – Patel, Riley, and MacLean

Comments: We reviewed this proposal for two characters used in the Arabic-based script called Ajami. The characters JEEM WITH THREE DOTS BELOW and JEEM WITH THREE DOTS ABOVE are used to write Wolof, typically to represent nasalized sounds. Separately encoding the characters with dots above and below is warranted, in our view.

The names and code points appear to be acceptable. The ArabicShaping data will need to be reviewed.

Recommendations: We recommend the UTC approve the following two characters in the Arabic Extended-A block:

08C6 JEEM WITH THREE DOTS ABOVE
08C7 JEEM WITH THREE DOTS BELOW

We further recommend the UTC assign an Action Item to Roozbeh Pournader to check the ArabicShaping data for the two characters.

b. ARABIC LETTER LAM WITH SMALL ARABIC LETTER TAH ABOVE

Document: L2/19-111 Proposal to encode ARABIC LETTER LAM WITH SMALL ARABIC LETTER TAH ABOVE in the UCS – Evans and Malik

Comments: We reviewed this proposal for one modern-use Arabic character. The character is used to represent a sound in the Western Punjabi language of Pakistan. The sound is not representable with an Arabic character today, although it is represented in the Gurmukhi script. (It was noted that Western Punjabi in Pakistan is not yet used as a medium of instruction and has no standardized alphabet.)

Examples are provided, though they date to 1792. More recently, a number of users have experimented with different ways to represent this sound, as described on the bottom of page 1 of the proposal.

The name and code point are acceptable. The authors have provided letters of support for the encoding of this character.

Recommendation: We recommend the UTC approve U+08C5 ARABIC LETTER LAM WITH SMALL ARABIC LETTER TAH ABOVE in the Arabic Extended-A block.

c. Balti

Document: L2/19-077 Proposal to include Balti alphabet – Lateef Shaikh

Comments: We reviewed this proposal for one letter, ARABIC LETTER KEHEH WITH HAMZA ABOVE, used to write the Balti language (629-3: bft) in Pakistan and northern India. The Balti language is written in both the Arabic and Tibetan scripts (although use of Tibetan is part of a revival effort according to *Ethnologue*). Lorna Evans of SIL has confirmed that there is a keyboard using this character.

The following were comments made during discussion:

- In order to evaluate the proposal, we suggest the author include the IPA value of the letter when used to write Balti.
- Provide additional examples of the letter in running text (with the letter circled), citing the source of the examples (with full bibliographic information).
- Provide the full bibliographic information on the two references cited in the proposal (page 3).
- Note that the code point U+08C3 is already taken (see Pipeline of proposed characters).
- Explain what the discussion on pp. 322-323 of *Tareekh-e-Baltistan* is about. (Page 323 includes the proposed character.)

Recommendation: We recommend the UTC review this document and send the comments above to the author.

d. Shughni/Roshani

Document: L2/19-074 Proposal to include Shughni/Roshani Alphabet – Zahoori

Comments: We reviewed this proposal for one character used to write Shughni (/Roshani, ISO 639-3 shg), an Indo-Iranian language spoken in the Kuhistoni Badakhshon region of Afghaistan and Tajikistan.

The proposal needs to include more examples, preferably from various printed materials with full bibliographic citations. A translation of page 4 of the second PDF is also needed, along with an explanation.

Recommendation: We recommend the UTC review this document and send the author the comments above.

8. Hebrew

Document: L2/19-183 SBL Hebrew Font User Manual (Font version 1.50) – Hudson

Comments: We briefly reviewed this document, which addresses problems that Hebrew users have faced when two different combining marks can occur in any order, but have different CCC values. It was noted that a similar problem was faced in Arabic, which resulted in the creation of UTR #53 Unicode Arabic Mark Rendering.

In the case of Hebrew, a customized normalization routine has been devised, and is documented on pp. 21-22 of this document. In order to evaluate the usage of this customized algorithm, it would be useful to know if any browsers use it, and if Microsoft is using it when rendering? Since this is not conformant to Unicode we recommend this document be discussed.

Recommendations: We recommend the UTC discuss this document.

9. Yezidi

a. Yezidi script proposal

Document: L2/19-051R Proposal for encoding the Yezidi script in the SMP of the UCS – Rovenchak et al.

Comments: We reviewed this revised proposal, which removed the UUM character, but left space for a stand-alone hamza, should evidence be provided in the future. In an email to Debbie Anderson, the author stated he contacted the user community, and verified that any word with UUM can be written

either with a ligatured or non-ligatured form, with no distinction between the two forms. (Figure 10 shows both forms.)

Recommendations: We recommend the UTC review this proposal and approve the 47 characters for encoding in a new Yezidi block that extends from U+10E80…U+10EBF.

b. Yezidi UUM and hamza
Document: L2/19-164 Information on Yezidi UUM and hamza – Rovenchak

Comments: We reviewed this document which provides information on Yezidi UUM/UM and the stand-alone hamza.

Page one is a response to the question from David Corbett (from feedback submitted in January 2019), who asked whether UUM ever contrasted with UM+UM. The new information in this document confirms that UUM (𐺻) could be considered as a ligature of two UM's (𐺹) and both can appear in the same text (see examples on page 1). Can the user community confirm that for any word that has UUM, it could be written with either ligatured or non-ligatured form, and there is no distinction between the two? If there is no distinction, we recommend UUM be removed and the hole in the chart be closed up. Also, a note should be included in the proposal describing how UUM should be represented.

On page two, examples of the hamza glyph appearing beside ELIF and alongside MIM, not on top of them, as one would expect for a combining hamza (which is a proposed character). The author of the document mentions the placement was due to a technical limitation, since the hamza could not be displayed above these letters. In our view, the examples do not support the encoding of a stand-alone hamza, which is typically rare. We recommend the hole for a stand-alone hamza be retained in the code chart and if evidence is provided, the character can be proposed and encoded later.

Recommendations: We recommend the UTC forward the comments above to the author.

SOUTH AND CENTRAL ASIA
10. Bengali
Document: L2/19-079 Proposal for revision of Unicode Bangla Block – ICT Div, Gov't of Bangladesh

Note: We reviewed this set of 5 requests. In the following, each request is followed by comments and recommendations from the Script Ad Hoc.

1. Remove 09BC ◌় BENGALI SIGN NUKTA

> **Comments**: Due to the Character Stability policy, characters published in the Unicode Standard cannot be removed. However, users are not required to use all the characters in the code charts, nor are fonts required to include all the characters and their glyphs found in the code chart.
>
> There may be other, non-Bangla communities that use the nukta to extend the alphabet to represent sounds that are not found in Bangla. The nukta could also be used in transcription or transliteration.

Recommendation: For those non-Bangla communities that use the nukta in their writing systems, we welcome information so we can document it in the Core Spec and/or CLDR.

2. Recognize Bengali Letter RRA (ড়), Bengali Letter RHA (ঢ়) and Bengali Letter YYA as atomic characters, not conjunct characters

Comments: In Unicode, normalization makes it possible so the two ways of representing characters will be treated as equivalent. Normalization allows one to perform searches, and find, for example, ড় whether it has been encoded as U+09DC or the sequence <U+09A1, U+09BC>. (See further the FAQ on Normalization, UAX #15 "Unicode Normalization Forms" and Section 3.11 of Chapter 3 "Conformance" in the Unicode Standard.)

Keyboard input is different from the encoding layer: keyboard designs have the capability of emitting more than one character with a single keystroke.

Recommendation: We recommend Lisa Moore be assigned an Action Item to review the current Bengali block introduction for possible improvement in the wording, including discussion of the normalized forms of U+09DC, U+09DD, and U+09DF.

3. Rename the Bengali Currency signs as 'Bangla Taka Mark' instead of 'Bengali Rupee Mark'.

Comments: The BENGALI RUPEE MARK and BENGALI RUPEE SIGN already have an annotation "*taka*" (see below). The Name Stability policy does not allow character names to be changed, once they have been encoded. Do the authors have suggestions for an improvement to the annotation?

Currency symbols

09F2 ৲ BENGALI RUPEE MARK
= taka
• historic currency sign

09F3 ৳ BENGALI RUPEE SIGN
= Bangladeshi taka

4a. Allow printing of the dependent vowel signs independently

Comments: We agree with the document on the need to specify how to represent a spacing combining mark as a standalone character, such as for pedagogical purposes.

Recommendations: We recommend the UTC discuss the representation of a spacing combining mark as a standalone character for Indic scripts, such as NBSP + the mark – but without the extra space. Based on the UTC discussion, we recommend text be added to the Core Spec.

4b. Allow changes in typing sequence of dependent vowels with consonants

Comments: The request to change the encoding order to visual order for left-side dependent vowel signs is not possible, as it would de-stabilize millions of existing documents in Bangla. An

input sequence could, in principle, be accommodated via an input method, but that is different from the underlying order, which is fixed by the standard.

5. Use “Bangla” instead of “Bengali” in Unicode Standard docs

Comments: The name “Bengali” cannot be changed in character names, as it would break the Name Stability policy. CLDR uses “Bangla” and “Bangla” has been incorporated in the block introduction (12.2 of the Unicode Standard), where it states that “[i]n the Indian state of West Bengal and the People’s Republic of Bangladesh, the preferred name for the Bengali script and language is *Bangla*.” It also appears at the top of the names list of the code chart.

11. Chakma
Document: L2/19-143 Proposal to encode CHAKMA LETTER VAA for Pali – Scheuren

Comments: We reviewed this proposal for one Chakma character that is used in Pali texts. The letter was created in 2013 to print the Pali Tipiṭaka.

The proposal includes evidence for the character and seems justified. It also states that there are no conjoined or subjoined forms of *vaa*.

Recommendations: We recommend the UTC accept U+11147 CHAKMA LETTER VAA.

12. Gurmukhi
Documents:
L2/19-167 Response to Script Ad Hoc (L2/19-047) re: L2/18-319 on Bindi before Bihari in Gurmukhi – Singh
L2/18-319 Proposal for Bindi before Bihari in Gurmukh – Singh
Background Documents:
L2/05-088 Proposed Changes to Gurmukhi – Sidhu
L2/06-030 Proposed Changes to Gurmukhi 4 – Sidhu

Comments: We reviewed this response by the author of “Proposal for Bindi before Bihari in Gurmukhi” to the Script Ad Hoc recommendations in L2/19-047.

The following comments were made during discussion:

- Examples showing contrastive usage of BINDI before vs. after Bihari in the same document are still required. It was noted that Unicode focuses on the representation of the graphic elements on a page and not the pronunciation.
- Identify whether the examples come from a manuscript or pre-computer typesetting.
- Is the following the intended shape for SGGS, or is it due to font failure?

  (page 403) 

- Without examples of contrastive usage, the placement of the BINDI appears to be stylistic. In that case, the encoding order would be the same for BINDI before or after Bihari, and positioning of the BINDI should be handled by the font. The Core Spec should note the variation in placement (which does not affect underlying encoding order).

Recommendations: We recommend the UTC review this document and forward the comments above to the author. If it is confirmed that the placement of the BINDI before or after Bihari is stylistic, then the EdComm should be assigned an Action Item to update the Core Spec noting that the placement of the BINDI can vary, when occurring with Bihari.

13. Malayalam
Document: L2/19-086 Encoding Model for Conjuncts with Chillus – Cibu
Related document: L2/19-047 Jan 2019 Script Ad Hoc recommendations (see page 7)

(Note: The author of L2/19-086 has since submitted a revised version of his proposal, based on feedback in L2/19-125 Comments on L2/19-086: Archaic Malayalam Sequences. The new version of L2/19-086 was not seen by the Script Ad Hoc.)

Comments: We reviewed this revised document, which represents a much improved revision of L2/18-346. The document now clarifies some of the points that were raised in the January 2019 Script Ad Hoc Recommendations (L2/19-047), including how different conjuncts should be encoded and use of the vertical bar virama. The author has also provided normalized glyphs in the table in section 2.1 ("Table 12-38" on page 3) and in attestations 1, 4 and 5. The document also provides an "Alternate Options" section 3 (pp. 4-5), with evidence (pp. 5-11) and a list of those chillus which participate in conjuncts and those that don't (page 12).

In our opinion, the revised text answers the outstanding questions and is ready to be incorporated in the Core Spec.

Recommendations: We recommend the UTC forward L2/19-086 to the Editorial Committee, after first reviewing the document. We recommend Liang Hai be assigned an Action Item to incorporate new text in the Malayalam block intro based on this document, and to make the necessary changes.

14. Mongolian
a. Working Group Meeting Report
Document: L2/19-139 Mongolian Working Group Meeting 3 (MWG3) Report– Moore

Comments: We reviewed this report from the meeting in Ulaanbaatar, Mongolia, that took place from April 3-5, 2019. The summary drew on proposals presented at the meeting (chronicled in the summary in L2/19-141 Summary of proposals made during Mongolian Working Group Meeting 3).

The following highlights comments made during discussion.

- Issues identified under the rubric "Areas of Significant Interest" were topics that Mongolia and Inner Mongolia both agreed to, but which need UTC input and will require additional details before implementation:
    1. NNBSP remains problematic for Inner Mongolia and Mongolia; both want the functions handled by a different character. NNBSP has been poorly implemented in various environments.
    2. Mongolian Vowel Separator (MVS) has not been clearly specified and has been implemented in different ways. The suggestion to merge the behavior of NNBSP into

MVS was made, but a detailed specification of the exact behavior of both NNBSP and MVS is required first and the various options studied.

3. Open source shaping rules and test sets was an idea generally agreed to. It would be based on Arabic shaping engine with complex shaping rules.
4. The effect of FVSes, i.e., whether a certain shape always requires use of a particular FVS ("context-independent") or whether the certain shape is affected by what precedes or following the character ("context-dependent") was discussed at length. Whichever choice is made, we should seek consensus from advocates for both approaches, before making any changes.

- The second part of the summary listed the endorsed meeting resolutions:
    1. Better specifications of format control characters in needed.
    2. Rather than using the Unicode code charts to capture the complicated shaping rules and text representation of Mongolian, a much better mechanism would be a Unicode Technical Note. A decision regarding this approach is time-critical, as it will affect Unicode 13.0 code charts and the chart for the CD of the 6th edition of ISO 10646.
    3. FVSes must be visible by default, but if they have an effect on the display of text, they would be invisible. Further discussion by the UTC is needed on this topic. How will this affect existing documents?
    4. To display stand-alone form and abbreviations, either the NIRUGU or ZWJ may be used. This was a request by EAC, but doesn't appear to imply any changes on the UTC's part.
    5. Specifications need to clarify the behavior of NIRUGU and ZWJ inside words.
    6. Positional mismatches in Mongolian variant documentation needs to be corrected. This should be resolvable once the variant details are removed from the code charts (item 2, above) and consistent, proper terminology applied to characters in a UTN.

- The meeting report concluded with two topics which will require further study:
    - Compound words, which will be studied by Liang Jinbao; he will propose a solution for the next working group meeting.
    - New rules for the MVS (namely, if an MVS can be used in place of NNBSP), which will be investigated by Liang Hai and Liang Jinbao.

Recommendations: We recommend the UTC review this document and discuss it.

b. Mongolian Punctuation
Document: L2/19-132 On Mongolian punctuation marks – Shen Yilei

Comments: We reviewed this document that discusses features of Mongolian punctuation that deserve attention. The discussion focused on those parts of the document that may be actionable by the UTC.

Section 1 Whitespace (non-)incorporation
The author demonstrates how the incorporation of whitespace around Mongolian punctuation marks is not consistently implemented across all vendors (Table 1), and has identified tests to help decide whether whitespaces should be incorporated or not. He concludes it is not a good idea for Mongolian punctuation. He further notes that punctuation borrowed from CJK blocks used in Mongolian text can present a problem.

Section 2 Character orientation
In this section, the author requests U+FF1A FULLWIDTH COLON and U+FF1B FULLWIDTH SEMICOLON be changed from “Tr” to “Tu”. We recommend a separate analysis be done in a new document, explaining why the change is needed (with evidence). CJK experts can then invited to comment on the change, based on the new document.

Section 3: Line breaking
The author classifies LB for Mongolian punctuation in Table 3 and proposes LB changes to 12 Mongolian punctuation marks in Table 5. We recommend inviting the author to provide further information on the proposed changes (with examples) in a separate proposal.

Recommendations: We recommend the UTC review the document and forward comments, including those above, to the author.

c. Mongolian Specification
Document: L2/19-130 Towards a well-formed Mongolian specification that allows interoperable implementations – Liang Hai

Comments: We briefly reviewed this document, which lays out clearly the detailed information on Mongolian letter forms and shaping, information typically needed for implementers and font developers. Such a document could be used as the basis of the future UTN on Mongolian. We recommend it include documentation of the EAC implementation in its next version. We further suggest scheduling a review of the next version at a future Script Ad Hoc meeting.

Recommendations: We recommend the UTC review this document.

15. Newa
Document: L2/19-170 Request to change the default behavior of Newa Initial Ra – Scheuren

Comments: We reviewed this request to change the default behavior for initial *ra* in Newa, so it matches that of Devanagari (Devanagari rendering rules R2 and R5a, TUS 12.1). The requested change would make the default behavior be:

Repha (hook ra):	<RA, VIRAMA, Consonant>
Eyelash ra:	<RA, VIRAMA, ZWJ, Consonant>

The following comments were raised during discussion:

- What do other Unicode-compliant Newa fonts do? (Zachary Scheuren has subsequently reported that to his knowledge the only Unicode-compliant font for Newa is Noto Sans. Other fonts overlay the Newa glyphs onto Devanagari code points.)
- The current text in the Newa block intro does not specify how to encode initial *ra.*
- There are three options:
    1. If we were first encoding Newa today, we would recommend encoding a separate character for either the *repha* or *eyelash ra*. This approach is a simpler model than one relying on ZWJ, and is now being recommended for new Brahmi-derived scripts.

2. Use a joiner for *hook ra*, with the default being the *eyelash ra* (the current situation, recorded in Anshu Pandey's original proposal, L2/12-003r =WG2 N4184).
3. Make the change as recommended in this document.

The decision between options 2 and 3 could be based on frequency. If there is no support for option 1 and the hook ra is more common, making the change to <RA, VIRAMA, Consonant> would be our preference for initial *repha*.

Note: According to Ananda Maharjan, *eyelash ra* is rarely used in Newa, and most native Newa-language non-Sanskrit words actually use *repha*. It does appear in initial *ry-* (cf. similar situation in Marathi), especially in loanwords, but this is rare. He also confirmed that users are accustomed to typing Devanagari (which is used for Nepali),

Recommendation: We recommend further discussion of this document and the UTC should decide how *repha* and *eyelash ra* should be represented in Newa. Once a decision is made, we recommend the UTC assign an Action Item to Liang Hai to make the appropriate changes in the Newa block introduction.

16. Old Turkic
Document: L2/19-069 Proposal to encode the Old Turkic ligature ORKHON CI – Pandey

Comments: We reviewed this proposal for one Old Turkic ligature, which appears only once.

The following points were raised in discussion:

- Why couldn't a joiner be used? Such an approach would mean users don't need to wait years for a new character. Use of a joiner for the ligature could be documented in the Core Spec.
- Will having a ligature provide long-term benefit to the community? Discuss the options, and give a strong rationale for separately encoding this ligature.
- Provide collation information on the character.

Recommendations: We recommend the UTC forward the comments above to the proposal author.

INDONESIA AND OCEANIA
17. Javanese
Document: L2/19-083 Positional category of Javanese *pengkal* – Lindenberg and Bayu Perdana

Comments: We reviewed this request to change the Indic positional category U+A9BE JAVANESE CONSONANT SIGN PENGKAL from Right to Bottom_And_Right. The request was based on an error report from 2015 (Public Review Issue #297) by R.S. Wihananto. The document provides clear evidence for making the change. The implications in Universal Shaping Engine for this change should be handled offline.

Recommendation: We recommend the UTC review this document and approve the change of Indic positional category for U+A9BE JAVANESE CONSONANT SIGN PENGKAL from Right to Bottom_And_Right, and assign an AI to Roozbeh Pournader to make the change in IndicPositionalCategory.

EAST ASIA

18. Bopomofo

Document: L2/19-100 Preliminary proposal to encode four extended Bopomofo letters for Cantonese – Eiso Chan

(Note: A revised version of this proposal with information from Ben Yang has been posted as L2/19-177, but the Script Ad Hoc did not review this document.)

Comments: We reviewed this preliminary proposal for four Bopomofo characters for Cantonese. The four characters are proposed for inclusion in the Bopomofo Extended block, which has five open spots.

In our opinion, the proposed characters are acceptable and the proposal includes all the necessary property information. Ben Yang reports he has found additional evidence for the proposed characters.

Recommendations: We recommend the UTC accept the following four characters in the Bopomofo Extended block:
31BC BOPOMOFO LETTER GW
31BD BOPOMOFO LETTER KW
31BE BOPOMOFO LETTER OE
31BF BOPOMOFO LETTER AH
We further recommend the proposal be revised to include additional source references.

19. Naxi Dongba

Document: L2/19-071 Preliminary report on L2/17-337 Revised chart of Naxi Dongba characters (follow-up to L2/18-321) – Poupard
Background documents:
L2/18-321 Comments and initial review of L2/17-337– Poupard (via Anderson)
L2/17-337 Results of the ad-hoc meeting on Naxi Dongba in Hohhot (WG2 N4895) – Everson et al.

Comments: We briefly reviewed this document, which supplemented comments on Naxi Dongba by the same author in L2/18-321. Both provide a review of the repertoire in L2/17-337. From the standpoint of the encoding model, a key question is what two-dimensional model will be used to represent the script?

Recommendations: We recommend the UTC note this document and forward it to the Naxi Dongba experts in China.

20. Tangut

Document: L2/19-064 Investigation of Tangut unification issues – West and Zaytsev

Comments: We reviewed this in-depth investigation, which studied seven Tangut ideographs that are currently unified in Unicode, but for which there may be a semantic difference. The work undertaken was based on a request in L2/16-243 (pp. 2-3).

The document concludes that there is evidence for “subtle but systematic glyph differences”, warranting separate encoding. In addition, two other Tangut ideographs should be disunified, pending agreement from other Tangut experts.

This document also recommends disunification of five Tangut components, a new key with radical and stroke count be added to the Tangut data file (from Jia and Jing's *Table of Xixia Characters and Annotated Properties*), and a new font be provided for the code charts. If new characters are approved after feedback from experts, a new Tangut Extended block will be needed (with proposed allocation from 18D00..18D3F).

On the disunifications, we defer to Tangut experts in China who will review this document, which is informational only.

Recommendations: We recommend the UTC note this document and await review by Tangut experts from China.

SYMBOLS AND NUMERICAL NOTATION SYSTEMS

21. Ascia Symbols

Document: L2/19-091 Proposal to encode a pair of Ascia symbols for Roman epigraphy (WG2 N5038) – West and Everson

Comments: We reviewed this proposal for a right- and left-facing pair symbols for an *ascia*, which is a tool defined as 'a hoe, axe, or adze'. It is found on funerary inscriptions from the first few centuries AD in Gaul, where the sign(s) often appear on the top of the inscription, between *D[iis]* and *M[anibus]* ('for the Manes [departed spirits]').

While examples are provided in running printed text, there is no contrastive use showing distinct semantics of the two. In our view, only one symbol needs to be encoded, until contrastive use in text demonstrating the need for two symbols is provided.

The left-tilted *ascia* appears to be more common. Glyph A would be our preference for the glyph.

Recommendations: We recommend the UTC accept one *ascia* symbol, and relay to the authors that additional evidence is needed for a second right-facing *ascia* symbol.

22. Block Elements and Symbols

Document: L2/19-068 Proposal to add Block Elements and Symbols for Software Developing – Renzhi Li

Comments: We reviewed this proposal to add 8 characters that are used in software development. The following points were raised:

- The Adobe representative confirmed the characters are currently in an Adobe font, located in the PUA.
- The proposed location (U+2FE0..U+2FEF), next to Kangxi Radicals and IDC in the BMP, is not a good location. (As noted in the proposal, the 214 newly approved characters for legacy computers are in the SMP.)
- Since an additional set of characters for computers will be proposed by Doug Ewell, we recommend the author work with Doug Ewell, so an omnibus proposal with characters for computers can be created.

Recommendations: We recommend the UTC send the comments above to the proposal author.

23. Breath Punctuation Marks
Document: L2/19-073 Proposal to add BREATH punctuation marks – Sanhueza

Comments: We reviewed this proposal for two BREATH marks. The figures 2, 3 and 4 provide examples that are close to use plain text usage, though figures 5 and 6 appear to be different characters (cf. "CENTERLESS SPLASH" in L2/11-288).

In our view, U+1F5E6 THREE RAYS LEFT and U+1F5E7 THREE RAYS RIGHT should be used. While they derive from Webdings, an annotation could be added to indicate their use as breath marks. The proposal discounts the use of these characters (and the other set of similar-looking characters, U+269E THREE LINES CONVERING RIGHT and U+269F THREE LINES CONVERGING LEFT) because they do not have the general category properties Open_Punctuation (Ps) or Close_Punctuation (Pe). However, the THREE RAYS characters can be used in different ways, and need not be restricted by having the open and close punctuation properties, like parentheses.

Recommendations: We recommend the UTC reply to the author with the comments above.

24. Christian Symbols
Document: L2/19-093 **Proposal to encode three Christian symbols** (WG2 N5040) – West and Everson

Comments: We reviewed this proposal for three Christian symbols: RHO CROSS, RHO CROSS WITH ALPHA AND OMEGA, and CHI RHO WITH ALPHA AND OMEGA.

The request for encoding a RHO CROSS character that is distinct from U+101A0 GREEK SYMBOL TAU-RHO is based on the following:

a. the shape of the crossbar's serifs (vertical across the RHO CROSS, but purportedly only pointing downwards in TAU-RHO), though the proposal admits the glyphs could be identical in sans serif fonts
b. the script property of GREEK SYMBOL TAU RHO is Greek, whereas the RHO CROSS should have script property of Latin or Common (preferring the latter)
c. there is no scholarly consensus that the RHO CROSS derives from the Greek *staurogram* (derived from the Greek word *stauros* 'cross')
d. the RHO CROSS appears as a stand-alone symbol, not as part of a word, unlike the *staurogram*

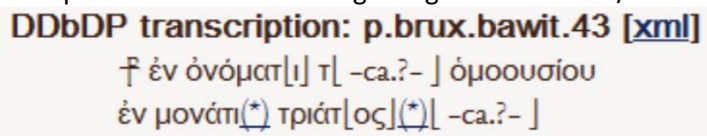
Current:

101A0 ⏀ GREEK SYMBOL TAU RHO
= rho-cross, staurogram

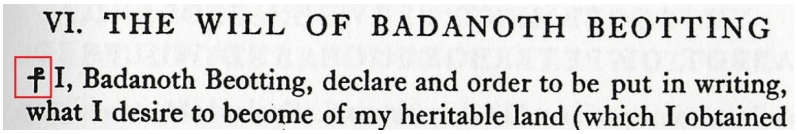
Proposed new character:

U+101A1	⏁	RHO CROSS

According to L2/12-034, TAU-RHO was encoded "as a free-standing character, appearing in Christian inscriptions or other items…signaling 'Christian' and/or as a symbol for Christ" such as in Fig. 33:



This usage seems to match Fig. 33 (above), and is contrary to point d, above:


VI. THE WILL OF BADANOTH BEOTTING

I, Badanoth Beotting, declare and order to be put in writing, what I desire to become of my heritable land (which I obtained

In our view, a stronger case still needs to be made for separately encoding RHO CROSS, since fonts could handle serif-versions of U+101A0 and its script property could be changed to "Common."

The two other proposed characters, RHO CROSS WITH ALPHA AND OMEGA and CHI RHO WITH ALPHA AND OMEGA (below), appear to simply be more iconographic variants of generalized Christian signs.

U+101A2		RHO CROSS WITH ALPHA AND OMEGA
U+101A3		CHI RHO WITH ALPHA AND OMEGA

A separate section on "Early Christian Pictographic Symbols" (p. 11ff.) lists 9 additional typographical signs that appear in two catalogues of Roman inscriptions for discussion, but they are not formally being proposed. In our view, these symbols are part of Christian thematic iconography and may not be appropriate to encode as symbols for use in text. There is nothing special about the two instances picked out in the proposal for 101A2 and 101A3 with alpha and omega, so we do not recommend encoding those.

Recommendations: We recommend the UTC review this proposal and forward any comments, including those above, to the proposal authors.

25. Cross Symbols
Document: L2/19-076 Proposal to add three cross symbols – Everson

Comments: We reviewed this proposal for three cross symbols: one for the "genuine" Maltese cross, and two for the right- and left-half shapes of the cross patty, which the author identifies as the character whose glyph appears in the code charts for U+2720 MALTESE CROSS.

As noted in the proposal (and the current annotation), U+2720 MALTESE CROSS derives from the Zapf Dingbat set. The current annotation also notes that U+2720 MALTESE CROSS took many forms., and the proposal mentions that the shape for the proposed "Cross of Malta" character *can* be covered by fonts today, though apparently few fonts include the "Cross of Malta" glyph. This proposal requests disunifying the two, and additionally requests two half-shapes of cross patty (MALTESE CROSS).


proposed CROSS OF MALTA


current code chart glyph for MALTESE CROSS (U+2720) (=cross patty)

While the two half shapes of the cross patty do appear in the printed transcription of one text (Egerton MS 2880), no textual evidence of the proposed Cross of Malta in running text is provided. Additional examples of the half-shapes would strengthen the argument for encoding the two "CROSS PATTY WITH CROSSBAR" characters.

Recommendations: We recommend the UTC review this proposal, and send the comments above to the author.

26. Half-length Hyphen
Document: L2/19-165 Peoplese half-length hyphen – Moore

Comments: We reviewed this request for a half-length hyphen, which was proposed in 2014 (L2/14-302). The recommendation to the 2014 proposal, contained in the January 2015 Script Ad Hoc recommendations (L2/15-045), was to use U+2010 HYPHEN or U+2043 HYPHEN BULLET.

In our view, a better choice is to use U+2027 HYPHENATION POINT, which is described in the names list as a " visible symbol used to indicate the correct positions for word breaking, as in dic·tion·ar·ies". This character is also likely to be supported in generic system fonts.

It was noted that the Unicode Standard is not intended as a vehicle for any new proposed spelling reform. To indicate low-level morphology, use of existing characters (preferably ASCII) and following standard practice used by linguists today would be a more productive direction than creating new characters.

Recommendations: We recommend the UTC send the comments above to the author.

27. Znamenny Notation
Document: L2/19-053 Proposal to Encode Znamenny Musical Notation in Unicode (rev. April. 2019)–Andreev and Simmons

Comments: We reviewed this proposal for 186 characters used to represent Znamenny musical notation. The UTC saw an earlier version of the proposal at the January 2019 UTC meeting, and the Script Ad Hoc has reviewed several iterations of the proposal.

The following points were raised during discussion:

- There was initially some concern about having two control characters next to one another, but it was pointed out that such a combination occurs in SignWriting (L2/12-321) as well as in compound stacks of Egyptian Hieroglyphs (see L2/17-112, p. 11).

- It would be helpful to have feedback from Monotype and Microsoft, confirming that the model is the most elegant and simple-to-implement approach, given this set of data. A different alternative, for example, would be use of combining marks instead of control characters.

  Kamal Mansour of Monotype has since responded, saying that the model is believable and implementable, but the cost would be on the side of data entry. He later suggested a modified approach to that given in the proposal:

  > If I'm now understanding it correctly, data entry would be much clearer if the control characters (P2, P3, PU…) were typed before the *neume*. The control

characters could appear as visible glyphs until an appropriate successor character is entered, at which time both characters would disappear and be replaced by the resulting “priznaked” form. While both orders (control + *neume*) and (*neume* + control) are equally implementable, the first is more obvious for the user typing the character sequences.

- The proposal already reflects changes made after review by the Script Ad Hoc. In our view, the proposal is very close to being ready for approval, but review and discussion of this complex proposal by UTC members is needed.

Recommendation: We recommend this proposal be reviewed by the UTC, and solicit feedback from other members, including Andrew Glass at Microsoft.

28. Zoroastrian symbol
Document: L2/19-067 Proposal to encode the Zoroastrian symbol 'Fravahar' in Unicode – Pandey

Comments: We reviewed this proposal for the Zoroastrian symbol, which was proposed earlier in L2/15-099. We suggest this proposal be forwarded to the Emoji Subcommittee. (If it were proposed only as a symbol, evidence in-line or as a map symbol would be required.)

Recommendation: We recommend this proposal be forwarded to the Emoji Subcommittee, along with the comments above.

OTHER
29. ITALICS IN PLAIN TEXT
Document: L2/19-063 A proposal for encoding italics in plain text using Variation Selector 14 – Overington

Comments: We reviewed this document which asked that U+FE0D VARIATION SELECTOR-14 be designated so the character preceding VS-14 be represented with an italic glyph.

The following comments were made during discussion:

- For the past 30 years, italicization has been identified as part of the rich text layer, above plain text representation. Departing from this tradition would create a huge amount of confusion and ambiguity in data.
- Italicization of text requires a identifying a span of text, which is a characteristic of text styles (bold, bold-italic, etc.).
- When a need for plain text italics has been shown, such as for math notation, italicized letters have been separately encoded.
- What would italics mean for a script that doesn’t use italics?
- If the use of VS-14 were to be designated in Unicode as proposed, each character to which it is applied would need to be explicitly listed. In other words, a glyphic registration for each combination would be required, which would be very burdensome to maintain.

Recommendations: We recommend the UTC reject this proposal, and use the comments from from the Script Ad Hoc (above) to serve as the basis for a notice of non-approval.