TO: UTC
FROM: Deborah Anderson, Ken Whistler, Roozbeh Pournader, Lisa Moore, and Liang Hai[1]
SUBJECT: Recommendations to UTC #162 January 2020 on Script Proposals
DATE: January 10, 2020

The Script Ad Hoc group met on November 1 and December 6, 2019, and January 6, 2020 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.

EUROPE

1. Latin

1a. Extended IPA

Document: L2/20-039 Unicode request for extIPA support – Miller

Comments: We reviewed this request for 20 IPA extension characters, which are used for transcribing disordered speech. The characters appear in a revised "extIPA" chart, which was approved in 2016 at a meeting of the International Clinical Phonetics and Linguistics Association.

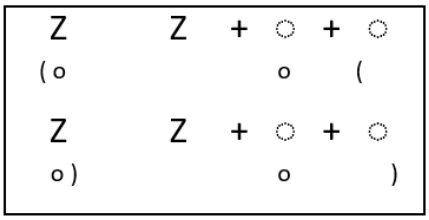
The following were comments made during discussion:

- Add formal names (in all caps) to the proposed characters, following the pattern of character names in the IPA Extensions block (U+0250..U+02AF) and the Spacing Modifier Letters blocks (U+02B0..U+02FF). The hyphen should be removed in in l-ezh (cf. 026E LATIN SMALL LETTER LEZH), l-s, and l-z names.
- Propose code points for the characters. The combining marks should start with U+1AC1 and continue from that point. The first four baseline characters should be assigned U+AB6C,

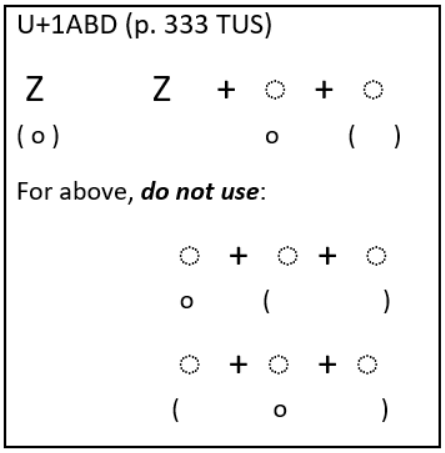
[1] Also participating were Richard Cook, Patrick Chew, Chris Chapman, Ben Yang, Norbert Lindenberg, Craig Cornelius, and (on phone) Ned Holbrook and Andrew Glass.

U+AB6D, U+AB6E, and U+AB6F in the Latin Extended-E block. The remainder should go into Latin Extended-D, starting at U+A7CB, and follow sequentially after that.

- For the combining right/left parentheses, the proposal should discuss the order of the parentheses, since users may intuitively encode the subscript or superscript "(o" as "(" and "o". See below for the correct order, which encodes the under-ring diacritic first, followed by the modification (i.e., parenthesis indicating initial or final partial voicing or devoicing).



- Also, we suggest the authors recommend an update be made to the text on page 333 of the Core Spec, specifying that users who want to represent a paired superscript or subscript parentheses should not use the single right and left parenthesis proposed in this document, but instead use the already encoded paired characters U+1ABB COMBINING PARENTHESES ABOVE or U+1ABD COMBINING PARENTHESES BELOW (for the latter, see below).



- Add a section with character properties, following the general category properties of similar characters (found in https://www.unicode.org/Public/UCD/latest/ucd/UnicodeData.txt). For combining characters, change the combining class to 220 for those below (i.e., CODEPOINT;NAME;Mn;220;NSM;;;;;N;;;;;) and 230 for those marks above (i.e., CODEPOINT;NAME;Mn;230;NSM;;;;;N;;;;;).
- Add page numbers.

Recommendation: We recommend the UTC relay the comments above to the proposal author.

1b. VoQS

Document: L2/20-038 Unicode request for VoQS support – Miller and Ball

Comments: We reviewed this document, which discussed a few topics relating to Voice Quality Symbols (VoQS) and requested one character, MODIFIER SMALL CAPITAL AA LIGATURE.

The following summarizes the comments:

- The authors provide evidence that U+A7F8 MODIFIER LETTER CAPITAL H WITH STROKE (L2/10-161) was incorrectly encoded as a capital; it should have properly been MODIFIER LETTER *SMALL* CAPITAL H WITH STROKE. The character now has a compatibility decomposition to <super> U+0126 LATIN CAPITAL LETTER H WITH STROKE. Because a change would affect normalization, the Script Ad Hoc recommended adding an annotation to U+A7F8 and a cross-reference to U+029C LATIN LETTER SMALL CAPITAL H.
- "Question 1" on page 2 asks about the representation of a modifier letter small capital h, which is used in the transcription of a click consonant in the Khoisan languages. The authors ask if U+2DE9 COMBINING CYRILLIC LETTER EN could be used, since its shape is identical to modifier letter small capital h? In our view, the modifier letters U+1D78 MODIFIER LETTER CYRILLIC EN or perhaps better U+1D34 MODIFIER LETTER CAPITAL H should suffice. If not, justify the disunification from these two modifier letters and provide evidence for encoding a new character. What is the source of the modifier letter small capital h: is it from Cyrillic or Latin?
- "Question 2" on page 2 asks whether the two modifier letter reversed glottal stops (U+02C1 and U+02E4), which appear identical in some fonts, justify having two modifier glottal stops? (That is, whether it is a good idea to add a second modifier glottal stop, beside U+02C0 MODIFIER LETTER GLOTTAL STOP.) In our view, having two would be redundant – just one is needed.
- Create a separate proposal for the "modifier small capital AA ligature" character, with proposed name, code point, and properties. The name should follow the pattern of U+A732 LATIN CAPITAL LETTER AA (i.e., do not use "LIGATURE" in the name).
- Add page numbers to the proposal document.

Recommendation: We recommend the UTC relay the above comments to the proposal author.

1c. Dania
Document: L2/20-040 Unicode request: Danish dialectological symbols – Miller and Schachtenhaufen

Comments: We reviewed this document for Danish dialectal symbols, based on a "Latin Modern Dania" chart from Andreas Stötzner and Ømålsordbogen. Dania is the linguistic transcription system for the Danish language, which is still widely used in Denmark according to Lars Trap-Jensen of the Det Danske Sprog- og Litteraturselskab. Dania is used in ongoing projects, including two large dialectal dictionaries, Ømålsordbogen for the Eastern island dialects and Jysk Ordbog for the Western dialects in Jutland. Table on page 3 of this document has highlighted those symbols that the author reports cannot be represented in Unicode or for which the representation in Unicode is uncertain.

The following comments were noted:
- We agree that italicization of the symbols is a font issue.
- Examples and evidence of any symbols still need to be provided as well as glyphs for a number of characters).
- Provide the document numbers to the cited Unicode Teuthonista proposal(s).
- Add page numbers.

Recommendations: We recommend the author continues his investigation with colleagues in Denmark and submit a proposal after evidence has been collected.

2. Northern Palaeohispanic
Document: L2/20-047 Proposal to encode the Northern Palaeohispanic script -- Ferrer et a*l.*

Comments: We reviewed this revised proposal for Northern Palaeohispanic.
The following comments were made:

- The proposal now includes charts from various publications, which is very helpful.
- This document provides the authors' thinking and justification for not encoding a 3-way (***ti*** and ***to***) or 4-way distinction (for ***o***), highlighted on page 3. In our view this remains a key issue, since the text model for this script is graphetic, that is, it aims to represent the graphic 3-way or 4-way distinction across the entire set of local variants, even though any one abecedary may only have a 2-way distinction. By providing an encoding that maximally captures the distinctions, a naïve user would be able to type the signs found on inscriptions, without requiring them to know that the 3-bar sign from Castellet de Bernabé should be represented with the 2-bar ***o*** character (U+10206). (We noted that the 3- and 4-bar ***o*** are missing from the glyphic variants chart on page 13.) Indeed, a 3-way distinction is made for ***ka*** and ***ke.*** Why not extend the 3- (or 4-) way distinction to ***ti***, ***to***, and ***o***?
- A new script-specific word separator character is proposed for Northern Palaeohispanic (and Southern Palaeohispanic). In our view, this common punctuation mark, found across many scripts, does not need to be separately proposed. Two approaches that can be taken to represent this mark:
    - Use an already encoded character that resembles the glyph (hence U+205A TWO DOT PUNCTUATION, U+205D TRICOLON or U+205E VERTICAL FOUR DOTS). This approach guarantees better presentation, since the glyph will match what appears in the text.
    - Use U+205A (or another common dot punctuation character, such one of those listed above) and modify the Palaeohispanic font so it appears in the specified shape for the particular script. This option may aid processing in simpler implementations, since one can search for all separators, no matter what the glyphic shape.

Recommendations: We recommend the UTC forward the comments above to the proposal authors

3. Southern Palaeohispanic
Document: L2/20-048 Proposal to encode the Southern Palaeohispanic script -- Ferrer et al.

Comments: We reviewed this revised document for Southern Palaeohispanic.

The following comments were made:

- The authors have added informative captions and charts, which are very useful.
- The numbers for figure 18 need to be adjusted: there are three figure 18s.
- On script-specific word-separator, see comments above (Northern Palaeohispanic)
- Additional study by interested Script Ad Hoc members is needed to double-check the new charts versus the proposed characters.

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

MIDDLE EAST
4. Arabic
Comments on Arabic additions for Quranic (L2/19-306)
4a. **Document**: L2/19-393 Comments about "Arabic additions for Quranic orthographies" proposition

(L2/19-306) -- Azzeddine Lazrek

Comments: We reviewed this feedback document, submitted by one of the key people who contributed evidence of various Quranic characters. The author agrees with 16 of the 39 proposed characters in L2/19-306, but has comments on the other characters.

We carefully considered the feedback in this document. The following summarizes the discussion:

- For #2 ARABIC SYMBOL WASLA ABOVE, the author requests a COMBINING WASLA ABOVE be encoded. The combining character was earlier proposed in L2/17-252 by Lazrek and by Miikka-Markus Alhonen in L2/03-166. The combining character fits with Lazrek's encoding model described in L2/18-028.

  While it is true that a combining ARABIC WAVY HAMZA BELOW (U+065F) was encoded in Unicode 6.0 (2010), the case for a combining WASLA ABOVE is less compelling in our opinion. WAVY HAMZA BELOW occurs below various characters and was separately encoded in the Indian standard PASCII (L2/09-176). It was only after careful deliberations about normalization that the UTC agreed to deprecate U+0673 ARABIC LETTER ALEF WITH WAVY HAMZA BELOW, and encode ARABIC WAVY HAMZA BELOW (L2/10-455).

  Unlike WAVY HAMZA BELOW, WASLA ABOVE appears to occur only over one Arabic character, *alef.* The encoding of a WASLA ABOVE would incur canonical equivalencies issues (such as for U+0671 ARABIC LETTER ALEF WASLA). Indeed, the author of L2/19-393 admits there will be two ways to represent U+0671. Such a duplicate encoding would be a serious drawback to encoding a combining WASLA ABOVE, in our view.

- For items #8-26, the author agrees with encoding the *alef* characters, but argues the diacritics should be encoded as combining marks instead of as precomposed characters. He cites the advantages of search, since search engines could find the examples of words with and without the Quranic marks.

  Changing the current model to one with a base and combining mark might work for some combinations with *alef*, but not for others (i.e., those where the diacritics are attached). Because making such a change would complicate the model, we advise against it. Removal of diacritics would be difficult and any attendant normalization implications would cause problems for users and implementers alike.

- The author proposes several name changes:
  - SMALL HIGH FARSI YEH should be changed to SMALL HIGH DOTLESS YEH
  - SMALL FARSI YEH should be changed to SMALL DOTLESS YEH.
  - SUKUN BELOW should be changed to SMALL CIRCLE BELOW. It was noted that the small circle below is graphically the same as a small circle.

    The above name changes are reasonable, but are not necessary, in our view. The names proposed in L2/19-306 were accepted by the UTC, and are consistent with existing character names, such as U+06CC ARABIC LETTER FARSI YEH. However, the names proposed in L2/19-393 do not follow the pattern of other Unicode Arabic character names.

- In set of characters from #8-#26, the author further recommends the following name changes:
  - Remove ATTACHED in those names containing “ALEF WITH ATTACHED…”, since attachment is not always observed. In our opinion, the attachment has meaning. In cases where *alef* is detached from the *kasra* or *fatha*, the text is already representable by the sequence <*alef*, *fatha* > or < *alef*, *kasra*>.
  - Change FATHA and KASRA in the names to STROKE, so they are consistent with MIDDLE STROKE (for *damma*). In our view, the existing *kasra* and *fatha* names appear closer to the conceptual structure of the characters with middle stroke, whose stroke is visually very different from a *damma*. The middle stroke may also be used for other purposes than a *damma*.

In sum, while name changes are deemed reasonable, we believe changing the encoding model at this point would destabilize encoding for Arabic and jeopardize the technical stability policy .

It was also noted that having Saudi experts’ comments would be very helpful.

Recommendations: We recommend the UTC discuss changing the names of SMALL HIGH FARSI YEH to SMALL HIGH DOTLESS YEH, SMALL FARSI YEH to SMALL DOTLESS YEH, and SUKUN BELOW to SMALL CIRCLE BELOW. We also recommend the UTC thank the author for his feedback.

4b. **Document*:*** L2/20-041 Comments on L2/19-306 Arabic additions for Quranic – Haralambous

Comments: We reviewed this set of comments from Yannis Haralambous on L2/19-306 “Arabic additions for Quranic”, which recommends encoding of combining marks, as described by Azzeddine Lazrek in L2/19-393.

Recommendations: We recommend the UTC put the document in the Unicode document registry and thank the author for his comments.

5. Mandaic
5a. **Document**: L2/20-044 The Non-Joining U+0858 in the Mandaic Unicode Standard -- Ardwan Al-Sabti

Comments: We reviewed this proposal to change the Joining_Type of U+0858 MANDAIC LETTER AIN from Non_Joining to Right_Joining.

The proposal provides evidence in support of the change for U+0858 to be Right_Joining. We agree with this request, but ask the author to include figure 8 of L2/08-270R, since it shows that U+0858 is not always consistent: besides its right-joining behavior, the AIN can appear non-joining beside a dual-joining letter (as shown below).

Recommendation: We recommend the UTC review this proposal and assign an Action Item for Roozbeh Pournader to change Joining_Type property for U+0858 MANDAIC LETTER AIN from Non_Joining to Right_Joining in ArabicShaping.txt .

***5b*. Document**: L2/20-043 The Non-Joining U+0856 in the Mandaic Unicode Standard -- Ardwan Al-Sabti
Note: The Script Ad Hoc reviewed an earlier version of this proposal, which included KAD.

Comments: We reviewed this document to change the Joining_Type of two Mandaic characters, U+0856 MANDAIC LETTER DUSHENNA and U+0857 MANDAIC LETTER KAD, from Non_Joining to Right_Joining.

The following captures comments made during discussion:
U+0856 MANDAIC LETTER DUSHENNA

- The proposal provides evidence of U+0856 MANDAIC LETTER DUSHENNA as right-joining. We agree with this request.

U+0857 MANDAIC LETTER KAD

- KAD derives from a ligature of U+084A MANDAIC LETTER AK and U+0856 MANDAIC LETTER DUSHENNA. The original proposal, L2/08-270R, states "KAD never joins with a preceding character." However, this proposal provides evidence of KAD as right-joining in the Neo-Mandaic words *bkḏ* and *lkḏ*.
- For the proposal author:
  - To strengthen the case for right-joining KAD, provide examples (in print or in a manuscript) of the two words below:

ࡁࡗ	bkḏ	U0841+ U0857	alike
ࡋࡗ	lkḏ	U084B+ U0857	till then

  - Does *kaḏ* appear after a non-connecting letter in the same word (i.e., does *kaḏ* ever occur preceded by character that is not joining)?
- Additional information is required to make a decision on U+0857 MANDAIC LETTER KAD. Questions for the UTC to consider:
  - In a case like *ukḏ*, should the encoding be *u + k + ḏ* or *u + kḏ*?
  - If KAD is made right-joining, how disruptive would it be to existing data?
  - Possible options for handling KAD:
    - Make KAD right-joining, and advise users to never use the sequence <*k, ḏ*>
    - Keep KAD non-joining; in cases where it is right-joining, use <*k, ḏ*>*.*

Recommendation: We recommend the UTC review this proposal and assign an Action Item for Roozbeh Pournader to change Joining_Type property for U U+0856 MANDAIC LETTER DUSHENNA from Non_Joining to Right_Joining in ArabicShaping.txt. We further suggest the comments above on KAD be sent to the author.

6. Syriac

Document: L2/20-019 Proposal to encode COMBINING DOT BELOW LEFT for Syriac -- Yang, Kiraz and Kiraz

Comments: We reviewed this proposal to encode one Syriac character, COMBINING DOT BELOW LEFT, which is needed for the representation of Syriac text. As noted in this proposal, Syriac uses many dots, which can be used to disambiguate words that are written identically, but differ in pronunciation and/or semantics.

The following were comments made on the proposal the Script Ad Hoc reviewed:

- The code point, U+1DFA, and properties look acceptable.
- Use a generic round dot for the representative glyph of COMBINING DOT BELOW LEFT.
- Change the glyph for U+1DF8 COMBINING DOT ABOVE LEFT to also be a generic round dot.
- Recommend the names list editor add an annotation for the new character, make changes to the annotation for U+1DF8 (noting that in Typicon the dot may be square), and adjust the subheads above U+1DF8.

Recommendations: We recommend the UTC accept U+1DFA COMBINING DOT BELOW LEFT. We recommend the UTC assign an Action Item to Ken Whistler to make the appropriate annotation and subheading changes to the names list and Roozbeh Pournader add Syriac and Cyrillic to the ScriptExtensions property for U+1DF8. In addition, we recommend the UTC remand the proposed text for Table 9-12 (contained on the top of page 2 of this proposal) to the Editorial Committee.

SOUTH AND CENTRAL ASIA

7. Brahmi

Documents:

L2/19-402 Proposal to Encode 6 Characters in the Brahmi Block -- Vinodh Rajan and Shriramana Sharma
L2/20-037 Comments on L2/19-402 Proposal to Encode 6 Characters in the Brahmi Block – Glass
Note: The Script Ad Hoc did not review the comments from Andrew Glass.

Comments: We reviewed this proposal to add six Brahmi characters to support the Old Tamil orthography. As noted at the beginning of this document and in the 2007 proposal (page 4 of L2/07-342), the Brahmi block was intended to be able to represent Old Tamil. Certain features of Old Tamil appear to warrant separate characters be encoded, and this document proposes six characters. These include Brahmi sign Old Tamil virama, Brahmi Old Tamil short e and short o (letter and vowel sign for each vowel, for a total of four characters), and Brahmi letter Old Tamil LLA.

The following summarizes comments during discussion:

- Provide examples of the proposed characters from inscriptions or handwritten materials.
- Because the Brahmi virama has different behavior from Old Tamil virama (conjunct forming vs. pure killer) and the two have distinct shapes (Brahmi horizontal bar above vs. Old Tamil dot above), we agree there is good reason to separate the two. However, the Brahmi block already has a dot above: U+11001 BRAHMI SIGN ANUSVARA. If the *anusvara* were to be used as the Old Tamil virama, what would be ramifications for shaping or other processes? (It was noted that the *anusvara* does not occur in Old Tamil.)

- Atomically encoding letters and signs for the two short vowels is reasonable in our opinion, since encoding them with the dot decomposed would deviate from the existing, ISCII-based model
- Provide detailed analysis on why the shaping models (such as the USE) fail to render <vowel letter/sign, VIRAMA>.
- Provide evidence in support of the glyph change for LLA, with more discussion. (For example, "This is the character we are disunifying the new LLA from..", "We are proposing this glyph..", "This was the original shape", etc.). Refer to the document trail (L2/12-106, L2/12-165, L2/12-226, L2/12-233, L2/12-236, L2/12-292) and note the glyph change from August 2012 (6.2 Errata) to provide context for the requested change.
- Number the pages.
- Correct name of U+11075 under Character Properties from LLLA to LLA, and indent the second line under U+11073 and U+11074 so it doesn't appear that U+11070 are separate entries.

Recommendations We recommend the UTC relay the comments above to the proposers. Also, we suggest Andrew Glass be invited to participate in discussion of this document at the UTC.

8. Telugu
Documents: L2/19-401 Proposal to Encode Telugu Sign Nukta -- Vinodh Rajan et al
L2/19-405 Additional evidence for the use of Nukta sign in Telugu -- Sridatta and Srinidhi

Comments: We reviewed this proposal to add a *nukta* sign in Telugu. The *nukta* is used in various Indic scripts to extend a script's repertoire so it can represent non-native phonemes.

Within the Unicode document registry, the character appeared in a presentation from 2005 (slide 46) and is mentioned in a document dated from 2004, but no formal proposal was ever made**.** The proposed Telugu *nukta* is attested under three letters, with evidence from the proposal and the ancillary document by Srinidhi and Sridatta.

The code point and properties in sections 7-10 are acceptable.

Recommendations: We recommend the UTC accept for encoding U+0C3C TELUGU SIGN NUKTA.

9. Old Uyghur
Document: L2/20-003 Revised proposal to encode Old Uyghur in Unicode -- Pandey

Comments: We reviewed this revised and very large proposal for Old Uyghur, which expands upon L2/19-016. Page 1 summarizes changes made since the last version.

The following were comments raised during discussion:
- The author explained during a call with the Script Ad Hoc that the block print style will be used for the representative glyphs. He also reported the next revision of the proposal will include a discussion (under "Scope") on encoding rapidly written cursive text, which is very hard for anyone but a seasoned scholar to read.
- The author is seeking input on the best approach for *alef* and *nun* (pp. 15-16) as well as the final-only alternate forms. Also, is the approach for *gimel-heth* valid (p. 17)? *Samekh-shin* (p. 18)?

Recommendations: We recommend the UTC set up an ad hoc on Old Uyghur to be held during the UTC meeting in order to provide specific feedback to the author.

SOUTHEAST ASIA, INDONESIA, AND OCEANI*A*
10. Western Cham
Documents: L2/19-217R3 Proposal to encode Western Cham in the UCS (Revised) – Hosken
L2/20-018 Response to proposal to encode Western Cham in the UCS – Alberto Pérez Pereiro, et al.
(Note: The Script Ad Hoc did not review the Response document above.)

Comments: We reviewed this revised Western Cham proposal, which reflects an update to the September 2019 version of the proposal (L2/19-217). This revision incorporates feedback from the Script Ad Hoc recommendations in L2/19-343.

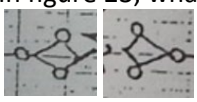
Below are the comments raised during discussion.

- General comments
    - In the proposal, list figure numbers after each character (i.e., in the Repertoire section or another prose section, if it is discussed there). Each proposed character should appear in at least one figure.
    - Move the code charts from the end of the proposal to before the acknowledgements, bibliography and samples. This makes it easier to compare the charts with the properties and the main body of the proposal.
    - Justify the encoding model.
        - Specific comments:
            - "Use of a separate 'finalising' character would not be appropriate given cultural sensibilities. In addition it breaks with the Unicode principle of not using modifiers as productive of other characters. This would make Western Cham unique if it were introduced (cf. New Tai Lue and Cham)."
                - Explain the cultural sensibilities that a finalizing character would incur.
                - What "Unicode principle" does this refer to?
                - What does the comparison of "New Tai Lue and Cham" mean?
            - Justify the phonetic order of left-side combining marks.
    - Provide a chart of the characters, if possible. (One example is found under "ALPHABET ČAM DU CAMBODGE" at https://archive.org/details/dictionnaireamf00cabagoog/page/n24, although the work dates to 1906 and varies from the proposed repertoire.)
    - We recommend removing the following lines, as they refer to deleted characters or code points from an earlier proposal, which is confusing:
        - In this discussion, codes presented in italics are for characters not present in the chart, and therefore having no representative glyph (even if the code is used for a different character). They are used for discussion of deleted characters based on their alignment with Eastern Cham.
        - 3. The 5 characters from N4734, *U+1E23B* (FINAL PH) … We close the gap, thus *U+1E23B* (FINAL PH) from N4734 becomes U+1E23D ([glyph] FINAL PH).

Instead, we suggest using an easy to recognize reference (such as the name or shape. Also advisable is to aim for encoding written structures from scratch, rather than basing it on Eastern Cham.

- Comments on repertoire
  - Move U+1E263 SIGN TANA TAMAT AYET to the Supplemental Punctuation block, since it occurs at the boundary between Arabic and Western Cham.
  - Are there examples of U+1E264 SIGN TANA TAMAT TAKUE beside Arabic text? If so, it should also be moved to Supplemental Punctuation block.
  - The ARABIC END OF TEXT appears to be a sequence of U+1E263 SIGN TANA TAMAT AYET characters repeated, and getting smaller, in a manner similar to the way the dandas extend out to fill the end of the line in figure 22. Based on the evidence, we recommend this character be removed.
  - In our opinion, the "TRIPLE DANDA" as shown in figure 20 is actually a sequence of double dandas and single dandas. In our view, the TRIPLE DANDA should be removed as a separate character.
  - Explain the difference between WESTERN CHAM PUNCTUATION DANDA and DEVANAGARI DANDA. Provide an example of a single danda.
  - Change the lunar symbol names from ARABIC SYMBOL LUNAR 1 DOT to either ARABIC SYMBOL ONE-DOTTED LUNAR MONTH or ARABIC SYMBOL ONE DOT LUNAR MONTH (etc.)
  - Page 3 calls U+1E237 a consonant sign, but says it "acts like a vowel", and page 8 gives its Indic syllabic category Vowel_Dependent. Its name is CONSONANT SIGN TKAJ KA. Is it a vowel or a consonant?
  - More generally, what does "vowel" or "consonant" refer to in the proposal: its phonological analysis, Unicode Indic properties, or the OpenType USE shaping model?
  - In the names list on page 28, explain why WESTERN CHAM LETTER NAN is in the middle of CONSONANT SIGNs.
- Properties
  - Lunar symbols should be "So" rather than "Po", which is the General Category for the Khmer lunar symbols (U+19E0 ff.).
  - Remove "Ranges may include unassigned characters for convenience."
  - Besides the order of characters in a syllable, it would be useful to know which characters can be combined with each other, and which ones can be repeated. A regular expression would help.
  - Revise the section on "Combining Orders" and write about "cluster pattern" generally.
  - The characters whose properties are incompatible with the Universal Shaping Engine are U+1E22E SIGN VOWEL OE and U+1E237 CONSONANT SIGN TKAJ KA. Provide more information on the identity of U+1E237 CONSONANT SIGN TKAJ KA.
  - Note that "Final vowels" in the "Combining Orders" chart don't work with USE's cluster pattern when U+1E232 VOWEL SIGN U precedes. "Medial ya" + "Medial wa" are also not allowed by USE. These cases may not currently be used actively, but they should be discussed if the author is trying to specify the "combining order".
  - The position of several marks as rendered by the font used in the document doesn't seem to match their Indic Positional Categories. Which one is right?
  - "Indic Properties"

        - On page 9, U+1E234 CONSONANT SIGN RA is given Indic Positional Category Left, but looks Left-And-Bottom, according to a Script Ad Hoc member. Is it true Western Cham has no signs interacting with the bottom sign?
        - On page 9, U+1E235 CONSONANT SIGN LA and U+1E232 VOWEL SIGN U are given Indic Positional Category Bottom, but look Right. If Bottom is correct, perhaps adjust the representative glyphs to better reflect this behavior.
        - The final vowels pages 8-9 are given Indic Positional Category Top, but to one Script Ad Hoc member they look Top-And-Right. If they don't increase inline spacing, then Top would be fine, if confirmed.
- Figures
    - In figure 23, what are the following? Which scripts do they occur beside?



    - What are figs. 24-25? Where do they come from and what is the context?
- Typos or other errors:
    - On page 3 "PUCTUATION"
    - On page 2, change the name of U+1E22E from VOWEL OE OR FINAL NG (4X under Repertoire 2) to VOWEL SIGN OE. Also, change the name of U+1E22E SIGN VOWEL OE on pages 5, 6 and 28.
    - Discrepancy between code points in chart on page 25 and the list on page 26.

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

EAST ASIA
11. Kana
***11a.* Document:** L2/19-381 Proposal to Encode Missing Japanese Kana - Abraham Gross

Comments*:* We reviewed this proposal for several Japanese Kana.

The following captures the comments made during discussion:

- Based on feedback from Japan (provided by Ken Lunde), we agree that the four characters identified by the proposal author as most important are good candidates for encoding:
  HIRAGANA LETTER ARCHAIC WU
  KATAKANA LETTER ARCHAIC YI
  KATAKANA LETTER ARCHAIC YE
  KATAKANA LETTER ARCHAIC WU.
  However, the proposed code points should be shifted from U+1B12A..U+1B12E to U+1B11F..U+1B122.
- The request from Japan for evidence showing the base character of HIRAGANA LETTER ARCHAIC WU is U+7D06 need not hold up encoding the character; that information can be added as an annotation later.
- The other characters will require more discussion.

Recommendations: We recommend the UTC approve the four characters at the following code points:
U+1B11F HIRAGANA LETTER ARCHAIC WU
U+1B120 KATAKANA LETTER ARCHAIC YI

U+1B121 KATAKANA LETTER ARCHAIC YE
U+1B122 KATAKANA LETTER ARCHAIC WU.

11b. **Document:** L2/19-382 Proposal to Encode a Kana Character for Transcription of Late Middle Japanese - Alexander Zapryagaev

Comments: We reviewed this proposal to encode one character, HENTAIGANA LETTER TU-2, which is a small form of U+1B06A and would be included in the Small Kana Extension block. As recommended by Japan, we recommend additional evidence be provided.

Recommendations: We recommend the UTC review this document.

12. Tangut
Document: L2/19-403 Additional Tangut Glyph Corrections (WG2 N5126) -- West and Zaytsev

Comments: We briefly reviewed this request to make 46 Tangut glyph corrections (42 to ideographs and 4 Tangut components). As noted in the document, detailed evidence is provided in the newly revised version of UTN 42. In our opinion, the document is well-formed and nothing problematical was identified.

Recommendations: We recommend the UTC review this document and decide how to proceed.

13. Other – Glyph Changes
***13a*. Document:** L2/20-011 Proposal to synchronize two glyphs in the Core Specification – Schneider

Comments: We reviewed this proposal which requests two glyphs be updated in the Core Spec to those in the current code charts:

- U+2D7F TIFINAGH CONSONANT JOINER in figure 19-2 on page 759
- U+10A3F KHAROSHTHI VIRAMA on pages 571-574

We agree these changes are warranted.

Recommendations: We recommend the UTC remand L2/20-011 to the Editorial Committee.

***13b*. Document:** L2/20-012 Proposal to synchronize seven glyphs in the Code Charts – Schneider

Comments: We reviewed this proposal to change seven glyphs in the code charts. Of the seven, we agree that changing the glyph for U+1BAB SUNDANESE SIGN VIRAMA (#1 on page 1) is a good idea. Five of the other proposed glyph changes come from the 2000 block (General Punctuation), where the author is trying to ensure the abbreviation in a glyph's dotted box matches the character's name or formal name alias abbreviation. We don't agree that making the change is needed.

Recommendations: We recommend the UTC assign an Action Item to the code chart editors to add a dotted circle in the glyph for U+1BAB SUNDANESE SIGN VIRAMA. We further recommend the UTC review and discuss the other requests.