TO: UTC
FROM: Deborah Anderson, Ken Whistler, Roozbeh Pournader, Lisa Moore, and Liang Hai
SUBJECT: Recommendations to UTC #152 July-August 2017 on Script Proposals
DATE: 28 July 2017

EUROPE
1. Ottoman Siyaq
Document: L2/15-072 Proposal to Encode Ottoman Siyaq Numbers (revision 2) – Pandey

Comments: We reviewed this document, which dates to 2015.

The following comments were made during discussion:
- In section 2.9, refer to figure 21 (page 40) as evidence for the comment, "Variant forms of the ten thousands are attested. Some of these are shown below. The first row contains representative glyphs for the proposed characters, the rest are variant forms."
- Provide evidence for ALTERNATE NUMBER TWENTY THOUSAND.
- Add characters for 1/2 and 1/6 and place the fractions together (without empty character slots between them) in the code chart; additional fractions can be added later.
- State that the NUMBER MARK is not needed, as its actual attested usage is not strong
- Consider adding a second ALTERNATE FORM OF SIX.

Recommendations: We recommend the UTC review this proposal, and if the above changes are made, approve Ottoman Siyaq.

AFRICA
2. Loma
Document: L2/17-233 Cumulative chart of the Loma script (WG2 N4837) – Everson

Comments: We reviewed this set of charts for Loma, which pulled together glyphs and Latin transcriptions from a variety of sources and documents.

These charts will be helpful when the user community standardizes their orthography.

We suggest the user community consider the letter shapes in *printed* material versus how the letters are written in *handwriting*. When the script is written by hand, the user community may want to extend certain modifications into typographic practice (rather than feel obliged to follow what appears in printed publications).

Recommendations: We recommend UTC members review this document at their leisure.

MIDDLE EAST
3. Arabic
a. Hamza Above
Document: L2/17-149 Proposal to Encode Arabic Hamza Above Isolated Form – Esfabod

Comments: We reviewed this proposal to encode one character in the Arabic Presentation Forms-B block.

The proposed character is requested based on the perceived need to provide a mapping from the isolated form of HAMZA ABOVE found in the desktop publishing software Zarnegar1 to a Unicode character that is located in the Presentation Forms-B block (alongside other spacing-form Arabic diacritic marks located there).

The following comments arose during discussion:
- It was noted that there are several old vendor sets in existence, of which Zarnegar1 is one. Zarnegar1 dates to the 1990s.
- Characters contained in the Arabic Presentation Forms blocks were being interchanged as plain text in the 1990s. However, Zarnegar1 is an advanced word publishing program with its own internal character set. Hence the request to provided 1-1 mapping for the proposed Zarnegar1 character is not considered a plain-text interchange issue. The character can already be represented by <00A0, 0654>.
- The encoding of further characters in the Arabic Presentation Forms blocks is not recommended.

Recommendations: We recommend the UTC review this document, and respond to the author that <00A0, 0654> is the recommended way to represent this character.

b. Wasla Characters
Document: L2/17-215 Proposal to Encode Additional Wasla Characters for the Holy Quran - Murodulla Begmatov

Comments: We reviewed this proposal, which requested 6 *wasla* Arabic characters in the Arabic Extended-A block.

The author has invented new *wasla* characters to aid non-Arabic speakers in reading the Quran. However, new characters such as these need to be shown as being in widespread, demonstrated use, preferably with official support.

Recommendations: We recommend the UTC respond to the author that published examples are required. In particular, encoding new contextual presentation forms for any Arabic characters is not recommended

c. Algorithm for the order of Arabic combining marks
Document: L2/17-253 Algorithm for order of Arabic combining marks – Pournader, Hallissey, and Evans

Comments: We discussed an early draft of this document, which describes an algorithm to determine the correct rendering of canonically equivalent sequences, so they render in the same way. This revised version builds upon an earlier document by Roozbeh Pournader (L2/14-127), but includes more clarification and examples.

Lorna Evans and Roozbeh Pournader will update this document, which will be proposed as a UTR.

Recommendations: We recommend the UTC discuss the proposed draft UTR.

4. Proto-Cuneiform
Document: L2/17-157 Proposal to encode Proto-Cuneiform in the SMP of the UCS – Everson

Comments: This proposal is an updated version of the 8 May 2017 proposal. In this revised version, more information is included about compound/complex forms, documentation has been added about the numbers (pp. 7-18), the code points have been changed and now align with the latest Roadmap draft, and glyphs for some of numbers have been slightly modified.

The following comments were made during discussion:
- The proposal needs to provide a solid rationale for encoding. Discuss the need for the characters in plain text, vs. the signs used as images in catalogs, vs. higher level protocol.
- Do any scholarly materials use the characters in running text, outside of occasional words, set of symbols, or figures? (If so, such information could provide information on line-breaking property for the characters.) Provide examples. Does line spacing need to be increased? Is there any stacking or grouping behavior for the signs?
- Address the relationship between the existing cuneiform characters and Proto-Cuneiform. (For example, U+122B9 CUNEIFORM SIGN SHAR2 is a single round dot, made by inserting the end of a stylus into clay. Is there any relation between this character and Proto-Cuneiform dots?)
- Glyphs need to be improved for dots and black circles with lines through them (i.e., U+1268F and U+126BF, U+125FF, U+126C0, U+126C1)
- Make the size distinction larger between U+1268E (N14) and U+126BE (N45), so it is comparable to the size difference shown on page 8.
- What does TIMES ONE (U+1262C) indicate in the names?
- Show examples of touching behavior, and demonstrate the difference between a compound and complex character. In some cases, it looks like the sign may be a compound made up of two characters, not one, cf. U+12682.
- Reconsider the "Gestalt" of the imprinted signs in clay, and how they should best be represented digitally in fonts. For example, the fine lines shown in the glyphs on the code chart on page 19 may not be the best way to represent the signs.
- Re-analyze the counting system into its encodable pieces. For example, pp. 10-11 suggests the underlying system may be a sequence, composed of a number with a "diacritic" (which acts as a determinative):

"60" used to count most object items

"60" used to count rations of certain items, perhaps a fish
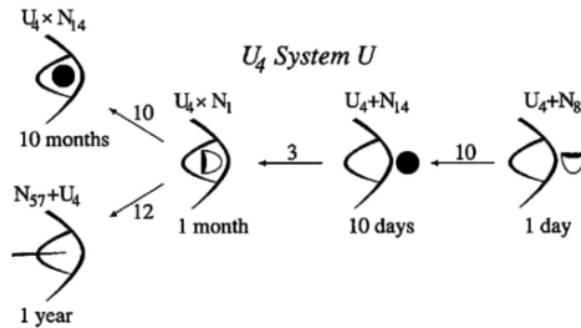

number used to indicate capacity of grain

number used to indicate capacity of certain grain

- Analyze the figures, and determine how they will be represented. In the following, will one character go inside of another? Do the characters have combining behavior?

Recommendations: We recommend the UTC members review the proposal at their leisure, and forward comments, including those above, to the proposal author.

5. Elymaean
Document: L2/17-226 Proposal to encode the Elymaean script in Unicode – Pandey

Comments: We reviewed this proposal for a non-joining script, derived from Aramaic. This proposal builds off the earlier preliminary proposal L2/17-055, which was not reviewed by the script ad hoc.

The following comments were raised during discussion:
- Does the script on coinage belong to a different script? Investigate more fully.
- Is "Elymaic" the more widespread name for this script today? Cf. "Elymaic" in figure 3 (1997) and figure 7 (2006), vs "Elymaean" in figure 12 (1952) and mentioned in a quote in §3.3 by Bivar and Shaken (1964). *Encyclopaedia Iranica* uses "Elymaic" in its article "Epigraphy i. Old Persian and Middle Iranian Epigraphy" (http://www.iranicaonline.org/articles/epigraphy-i) when referring to the inscriptions, but "Elymaean" in referring to the people and the coinage in the "Elymais" article (http://www.iranicaonline.org/articles/elymais).
- Based on glyph shape, it is unclear whether AYIN, KOPH, and RESH are different characters, or could be unified. Are they distinct in source materials?

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

SOUTHEAST ASIA
6. Lao
Document: L2/17-106 Revised Proposal to Encode Lao Characters for Pali – Rajan et al.

Comments: We reviewed this revised proposal, which has taken into account comments from the ad hoc. The author has provided additional property information for Indic Syllabic Categories and Indic Positional Categories, modified the collation tables, and moved LAO LETTER PALI NYA to the position we recommended, U+0E8E. In order to make the order of the new characters stand out, we recommend the author highlight all the letters that differ from the default in 10.2 Tailored Collation for Pali. The authors have incorporated this change.

Recommendations: We recommend the UTC approve 15 Lao characters for Pali, after first discussing the location of LAO LETTER PALI NYA.

SOUTH ASIA
7. Indic Editorial Updates
Document: L2/17-098 Request for editorial updates to Indic scripts -- Srinidhi, Sridatta

Comments: We reviewed this document of corrections and changes for Indic, most of which are editorial in nature.

> *§1 Devanagari*
> *§1.1 A8F8 DEVANAGARI SIGN PUSHPIKA*
> The authors have two requests:
>   a) Remove the alias "vaidika pushpikaa" to U+A8F8, since the character is not limited to Vedic texts and Sanskrit.
>   b) Add Kannada and Newa to the set of scripts in the ScriptExtensions property for DEVANAGARI SIGN PUSHPIKA, and add Tigalari and Nandinagari after they are encoded.
>
> In our view, the alias "vaidika pushpikaa" to U+A8F8 should be retained, as the information is not incorrect, and reflects the agreed-upon pattern found elsewhere in the names list in the Devanagari Extended and Vedic Extensions blocks. We note that the names list annotations should not be considered complete or encyclopedic; annotations are included if the information is considered useful.
>
> Based on the solid evidence provided, we agree Newa and Kannada should be added to the set of scripts in the ScriptExtensions property for U+A8F8.
>
> **Recommendations:** We recommend the UTC remand these items to the Editorial Committee, so an annotation "attested in Sanskrit and other scripts" can be added to the names list. We also recommend Newa and Kannada be added to set of scripts in the ScriptExtensions property for U+A8F8. Note that when Nandinagari and Tigalari are encoded, similar adjustments should be made to the set of scripts in the ScriptExtensions property for U+A8F8 to account for usage by those scripts
>
> *§1.2. A8F8 DEVANAGARI CARET*
> The authors request the removal of "vaidika" from the names list, since the character appears in other Indic scripts.

As noted above for PUSHPIKA, the annotation is not strictly incorrect and follows the general pattern elsewhere, so we do not suggest it be removed. The document does not provide evidence of the CARET in other scripts. If a list of such scripts using CARET (with examples) was provided, the Scripts.txt property could be changed from Devanagari to Common, and the various scripts could be added to set of scripts in the ScriptExtensions property, if deemed justifed.

Recommendations: We recommend the UTC request the authors submit a list of scripts using the CARET, with evidence.

§1.3 Representation of rya in North Indian languages.
The authors provide evidence of special rendering of *rya* in a number of North Indian languages, and request the Devanagari chapter of the core specification document the special rendering.

This information is valuable input and requires feedback from UTC members.

Recommendations: We recommend the UTC request the authors submit a separate document on this topic, since it affects the Devanagari model and should be reviewed by all stakeholders. The document should also mention the rendering behaviors of RA and YA in Bengali.

§1.4 U+0904 DEVANAGARI LETTER SHORT A
The authors request the names list include an annotation about the use of U+0904 as a short *e*.

The document provides useful evidence for the use of DEVANAGARI LETTER SHORT A, a character requested by the Government of India in 2001 (L2/01-304), but which has been missing attested usage (cf. L2/09-321). The attestations demonstrate that U+0904 is used for short *e* in the Awadi language, as well as Hindi translations and Devanagari transliterations of the Kannada, Telugu, Tamil, Malayalam and Kashmiri languages by a publisher in Lucknow.

Recommendations: We recommend the UTC remand this item to the Editorial Committee, so the core specification can note the use of U+0904 and an annotation can be made to the names list.

§1.5 Tibetan in Devanagari
The authors request that Devanagari be added to the set of scripts in the ScriptExtensions property for 6 Tibetan punctuation characters. The evidence shows examples of the Tibetan language written in Devanagari, which is apparently done infrequently.

We do not recommend the characters be added to set of scripts in the ScriptExtensions property. Instead of using the Tibetan SHAD characters, we would recommend the use of Devanagari dandas, which appear to be used stylistically in the examples. The TSHEG characters appear to be borrowed from Tibetan, and should be used in this context.

If the use of Tibetan characters in Devanagari reflects a true orthography for writing Tibetan, there may be a stronger case, and the authors are invited to submit another proposal.

**Recommendations:** We recommend the UTC note this item, but take no action.

§2 Sharada and Kannada
Sharada
The authors have two requests:
   a) Correct the glyphs of *jihvamuliya* + KA and *jihvamuliya* + KHA in §15.3 of *TUS*.
   b) Change the Indic_Syllabic_Category for Sharada *jihvamuliya* from Consonant_Prefixed to Consonant_With_Stacker.

The request to correct the glyphs of U+111C2 *jihvamuliya* + KA/KHA i in §15.3 of *TUS* is well-justified. As shown below, the *jihvamuliya* should appear *below* the headline.

Current glyph of *jihvamuliya* + KA in §15.3 of *TUS*:   Correct:

Current glyph of *jihvamuliya* + KHA in §15.3 of *TUS*:   Correct:

The shapes documented by the authors match those in the original proposal (L2/09-074, p. 20) for *jihvamuliya* + KA, and the example of *jihvamuliya* + KHA is helpful, as the original proposal did not specifically call out the shape.

We do not recommend a change be made for the Indic_Syllabic_Category of Sharada *jihvamuliya*, unless evidence is provided showing *jihvamuliya* occurring by itself. If the examples only show the *jihvamuliya* on top of another consonant, then the current syllabic category should remain. The authors are requested to provide more information and other examples. Does the change of shape for KA take place because of the *jihvamuliya* or because KA is subjoined below another consonant?

Recommendations: We recommend the UTC remand the Sharada *jihvamuliya* glyph correction to the Editorial Committee, and forward the comments above to the author regarding the request to change the Indic_Syllabic_Category for Sharada *jihvamuliya* .

Kannada
The authors request the glyph for Kannada *ḥka* be corrected in §12.8 of *TUS*.

The example provided and further attestations in L2/13-242 by Srinidhi provide ample justification for the correction.

Current glyph of *jihvamuliya* + KA in §12.8 of *TUS*:   Correct:

Recommendations: We recommend the UTC remand the glyph correction of Kannada *jihvamuliya* + KA in §12.8 of *TUS* to the Editorial Committee,

§3 Sora Sompeng
The authors request the description of Sora Sompeng be changed to "alphabet" from "abugida" in §15.14, and Table 6-1, based on recent examples provided.
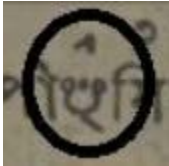
The examples support the request.

Recommendations: We recommend this item be remanded to the Editorial Committee, correcting the references to Sora Sompeng as an alphabet (not an abugida).

§4 Vedic Extensions
§4.1 Veda in Bengali
The authors request Bengali be added to the set of scripts in the ScriptExtensions property for the following Vedic characters U+1CD5, U+1CD6, U+1CD8, U+1CE1, U+1CD0, U+1CD2, U+1CEA, and U+1CED.

The attestations provide support for U+1CD5, U+1CD6, U+1CD8, U+1CE1, U+1CD0, U+1CD2, and U+1CED. For U+1CEA VEDIC SIGN ANUSVARA BAHIRGOMUKHA, the examples provided in 4.1.4 on page 12 show a dot between the two loops.

Is it a combining mark? Part of the character or a random dot?

Recommendation: We recommend Bengali be added to ScriptExtensions.txt for the following 7 characters: U+1CD5, U+1CD6, U+1CD8, U+1CE1, U+1CD0, U+1CD2, and U+1CED. For U+1CEA, we recommend the authors provide further information on the dot in the glyph, at which time the character can be reconsidered as evidence for adding Bengali to the ScriptExtensions property for U+1CEA.

§4.2 Veda in Tirhuta
The authors request Tirhuta be added to the set of scripts in the ScriptExtensions property for two characters in the Devanagari block, U+0951 and U+0952.

Examples confirm use of both characters. We noted that both U+0951 and U+0952 have a long list of scripts (12 and 11, respectively) in ScriptExtensions.

Recommendation: We recommend Tirhuta be added to the scripts in ScriptExtensions.txt for U+0951, and U+0952.

7. Indic Editorial Updates (continued)

§4.3 Samavedic characters in Kannada
Based on the evidence on pp. 4-5 of L2/15-164, the authors ask Kannada be added to the set of scripts in the ScriptExtensions property for the following Vedic Extensions characters: U+1CD0, U+ 1CD2, U+ 1CD3, and U+1CD4.

We note that a preferable way to present the evidence is to include the examples (circled) as part of the document requesting them, rather than referring to another document.

The examples for U+1CD0, U+ 1CD2, and U+1CD4 appear to have solid evidence, but we request more evidence be provided for U+1CD3, which is a spacing character and seems to be represented in L2/15-164 by the following:

Recommendation: We recommend Kannada be added to the scripts in ScriptExtensions.txt for U+1CD0, U+1CD2, and U+1CD4, and ask for more evidence on U+1CD3.

§4.4 Veda in Odia
The authors request that U+1CDA VEDIC TONE DOUBLE SVARITA in the Vedic Extensions block be extended for Odia (Oriya) in ScriptExtensions.txt.

The evidence provided shows the request is reasonable. The examples show this combining mark to the right of the character instead of top of a character, presumably because of limited technology (cf. U+0951, which also appears to the right):

If there is evidence that the character *regularly* occurs on the right, the authors should submit a document so a note can be added to the core specification on this usage.

Recommendation: We recommend Oriya be added to the scripts in ScriptExtensions.txt for U+1CDA.

§4.5 U+1CF2 VEDIC SIGN ARDHAVISARGA
The authors request:
a) U+1CF2 in the Vedic Extensions block have Telugu and Tirhuta be added to its ScriptExtensions property, with evidence for Telugu provided in this document, and for Tirhuta usage, a reference to §4.11 of L2/11-175R, the Tirhuta proposal.
b) the annotation for U+1CF2, *vaidika jihvaamuuliiya upadhmaaniiya*, be removed, since the character appears outside of Vedic and Vyakarana texts.

The evidence presented for usage of *ardhavisarga* in Telugu and Tirhuta is strong, although it

9

would help the script ad hoc if the authors would include the text and examples from referenced documents.

Recommendation: We recommend Telugu and Tirhuta be added to the scripts in ScriptExtensions.txt for U+1CF2. As above for DEVANAGARI SIGN PUSHPIKA (§1.1) and DEVANAGARI CARET (§1.2), we do not recommend the annotation *vaidika jihvaamuuliiya upadhmaaniiya* be removed, as it is not strictly incorrect and annotations are not meant to cover all usages of a given character.

§4.6 VEDIC SIGN NIHSHVASA
The authors note a discrepancy between the names list and the core specification on the use of U+1CD3 VEDIC SIGN NIHSHVASA. The names list annotation ("separates sections between which a pause is disallowed") is corroborated by L2/09-372 (§4.5.2) and L2/09-298. However, the text of *TUS* (§12.1, p. 467) differs: "U+1CD3 VEDIC SIGN NIHSHVASA indicates where a breath may be taken." The authors suggest the wording be modified to: "Separates sections of Sama Vedic singing between which a pause is disallowed."

The point made by the authors is well-taken.

Recommendation: We recommend this item be remanded to the Editorial Committee.

§5 Brahmi
Based on early sources, the authors ask that the glyphs for Brahmi '30' U+1105D and Brahmi '40' U+1105E be modified so they do not have a headstroke. They note that in the Segoe UI Historic font , contains the proposed glyph for '40'

Recommendation: We recommend this request be forwarded to Andrew Glass, one of the authors of the original Brahmi proposal, to get confirmation on the proposed glyphs.

8. Malayalam
Documents:
L2/17-207 On the Origin of Malayalam Candrakkala – Srinidhi, Sridatta
L2/17-230 Feedback on L2/17-207 on Malayalam Chandrakala – Sharma
(Note: Additional feedback has been submitted via the Feedback form from Cibu Johny)

Comments: We reviewed the two documents, both of which comment on the history of Malayalam *candrakkala* and recommend a minor modification of wording in the core spec (§12.9, pp. 501-2). We noted that additional feedback has been submitted from Cibu Johny in L2/17-224, the author of the two virama proposals (L2/14-014 and L2/14-015).

Recommendation: Since the proposed changes are editorial, we recommend the UTC remand this item to the Editorial Committee, and refer to the two documents L2/17-207 and L2/17-230 and feedback from Cibu Johny.

9. Takri

Documents:
L2/17-209 Proposal to encode the TAKRI LETTER SSA – Srinidhi, Sridatta
L2/17-231 Feedback on L2/17-209 on Takri SSA - Sharma
Background doc: L2/09-111 Proposal to Encode the Takri Script – Pandey

Comments: We reviewed these documents. The proposal L2/17-209 provides ample evidence that the original shapes for Takri KHA and SSA differed:

KHA ㅋㅋ   SSA ㅂ.

As noted in the proposal, KHA was later replaced in writing by ㅂ. SSA. The current character in the code chart is: U+1168B ㅂ. TAKRI LETTER KHA.

The proposal from Srinidhi and Sridatta proposes encoding one new character, TAKRI LETTER SSA , ㅂ. and requests a glyph change for the current character TAKRI LETTER KHA to ㅋㅋ. Sharma instead suggests encoding a new character U+116B8 TAKRI LETTER TRADITIONAL KHA ㅂ, and recommends retaining the current glyph for KHA, but adding an annotation that it is also used for SSA.

In our opinion, the approach of Srinidhi and Sridatta focuses too heavily on the phonetic value reflected in the name. In our view, the character's identity should instead be based on the name, shape, and usage of a character, and Sharma's approach is more in line with the Unicode perspective. We would recommend modifying the name Sharma suggests, however, from TRADITIONAL KHA to ARCHAIC KHA, as this is the preferred name (ARCHAIC appears in 33 names in Unicode, TRADITIONAL in 0).

Recommendations: We recommend the UTC approve U+116B8 TAKRI LETTER ARCHAIC KHA with shape

ㅋㅋ ,and add an annotations that U+1168B TAKRI LETTER KHA is also used for SSA.

10. Nandinagari
Document: L2/17-213 Proposal to encode the Prishthamatra for Nandinagari – Srinidhi, Sridatta

Comments: We reviewed this proposal to encode a new character NANDINAGARI VOWEL SIGN PRISHTHAMATRA E at U+11BD4.

The *prishthamatra* is a left-side mark that appears alone or with other vowel signs in the earlier orthography of Nandinagari. It has correspondences in Devanagari (and Sharada, see L2/17-214, discussed below). Attestation is provided in the proposal, though it was noted that in the examples, the glyph may appear somewhat shorter than the proposed representative glyph. The proposal is missing the character property values for Indic_Syllabic_Category and Indic_Positional_Category.

Recommendations: We recommend the UTC approve U+11BD4 NANDINAGARI VOWEL SIGN PRISHTHAMATRA E, but suggest the UTC request the authors revise the proposal to include the character property values for Indic_Syllabic_Category and Indic_Positional_CategoryIndic properties.

11. Sharada
Document: L2/17-214 Proposal to encode the Prishthamatra for Sharada – Srinidhi, Sridatta

Comments: We reviewed this proposal to encode a new character SHARADA VOWEL SIGN PRISHTHAMATRA E at U+111CE. The proposed glyph and code point location are acceptable and the evidence solid, but the authors need to provide Indic properties.

Recommendations: We recommend the UTC approve U+111CE SHARADA VOWEL SIGN PRISHTHAMATRA E, but suggest the UTC request the authors revise the proposal so to include the Indic properties.

12. Telugu
Document: L2/17-218 Proposal to encode the TELUGU SIGN SIDDHAM – Srinidhi, Sridatta

Comments: We reviewed this proposal for one character, TELUGU SIGN SIDDHAM, with the proposed location U+0C77. This character is well justified, and its name and code point acceptable.

Recommendations: We recommend the UTC approve U+0C77 TELUGU SIGN SIDDHAM.

13. Wancho
Document: L2/17-067 Proposal to encode the Wancho script in the UCS – Everson

Comments: We reviewed this revised proposal, which now includes four tone marks and two short vowels (U+1E2C1 WANCHO LETTER A and U+112E4 WANCHO LETTER ANG).

The creator of the script is a speaker of Northern Wancho, but there are other varieties of Wancho, including Southern Wancho, where the tones vary from Northern Wancho. As a result, the proposed tones do not mark absolute tone. (Note: In the earlier version of the proposal, tones were marked by doubling or tripling letters.)

The current repertoire – minus the four tones and the two new short vowels – is well attested and has clear evidence of use, and could be approved. In our opinion, holes could be left for the 6 characters that have no published examples. Once printed examples are provided, then they can be proposed.

Recommendations: We recommend the UTC discuss this proposal.

14. Newa
Document: L2/17-093 Proposal to encode the NEWA LETTER VEDIC ANUSVARA-- Srinidhi, Sridatta

Comments: We reviewed this proposal for one Newa *anusvara* character, which includes two examples of its use. We noted that the glyph should be improved.

Recommendations: We recommend the UTC approve U+1145F NEWA LETTER VEDIC ANUSVARA, but request the author provide a better glyph.

15. Vedic
Document: L2/17-095 Request to change the glyphs of Vedic signs *Jihvamuliya* and *Upadhmaniya* -- Srinidhi, Sridatta

Comments: We reviewed this document which had three main requests:

a) Change the representative glyphs for U+1CF5 VEDIC SIGN JIHVAMULIYA and U+1CF6 VEDIC SIGN UPADHMANIYA to reflect the shapes in the original sources.
b) Assign the Indic_Syllabic_Category for the 2 characters to be Consonant_With_Stacker
c) Remove Kannada from ScriptExtensions.txt for U+1CF5 VEDIC SIGN JIHVAMULIYA

The following points were raised during discussion:
- We consider the change of glyphs to be well-justified, but the glyphs should be enclosed in a dotted box, as is done for the *jihvamuliya* and *upadhmaniya* in Sharada and Kannada. The glyph for *upadhmaniya* should be changed to so it varies from U+1CE9 VEDIC SIGN ANUSVARA ANTARGOMUKHA.

- The text of *TUS* should be modified to state that the characters participate in syllables. Because they also appear in Devanagari outside Vedic materials, we agree it makes sense to mention their usage in Devanagari in §12.1 of *TUS*.

- In order to correctly identify the Indic_Syllabic_Category for the characters , the authors should provide a list of characters with which the *jihvamuliya* and *upadhmaniya* appear in clusters.

- A few other questions:
    o What is the relation between the *upadhmaniya* here and U+1CE9 VEDIC SIGN ANUSVARA ANTARGOMUKHA?

    o In figure 4, how would the following be represented (by encoded characters), and why does it not ligate to the following consonant?

- Removing Kannada from ScriptExtensions.txt U+1CF5 VEDIC SIGN JIHVAMULIYA seems reasonable.

Recommendations: We recommend the UTC change the glyphs for U+1CF5 VEDIC SIGN JIHVAMULIYA and U+1CF6 VEDIC SIGN UPADHMANIYA, but enclose each in a dashed box. We further recommend the suggested changes to §12.1 of the core spec be remanded to the Editorial Committee. We recommend the other questions listed above be forwarded to the proposal authors, in order to get full information on the characters and to identify their Indic_Syllabic_Category. Lastly, we recommend the UTC remove Kannada from U+1CF5 VEDIC SIGN JIHVAMULIYA in ScriptExtensions.txt.

16. Dogra
Document: L2/17-201 Proposal to encode the DOGRA VOWEL SIGN VOCALIC RR – Srinidhi and Sridatta

Comments: We reviewed this proposal for the addition of one character, DOGRA VOWEL SIGN VOCALIC RR, which was used in New Dogra.

This proposal states that New Dogra used the independent letter R (or L) and VOWEL SIGN VOCALIC RR to represent the long vocalic *r* and *l*. (Cf. Devanagari, which can represent the long vocalic *r* by the full letter RA U+0930 and the dependent form of the vowel sign vocalic *rr* [see Table 12-4 in *TUS*].)

If accepted, the authors request a change of code points for 9 characters already approved by the UTC, so the new VOCALIC RR would replace the current DOGRA VOWEL SIGN E, and the following characters be shifted down one position. As a result of this change, VOCALIC RR would coincide with the position of the VOWEL SIGN RR In other Indic scripts (i.e., be with the other dependent vowel signs).

The evidence in the proposal provides solid justification for the character.

Dogra is currently in PDAM 1.3. Because this PDAM is advanced in the encoding process, we don't recommend changing the code points to insert DOGRA VOWEL SIGN RR amongst the other vowel signs, but instead suggest the character be placed at U+1183B, after ABBREVIATION SIGN. If located in this position, the proposal should include a note about collation, specifying that DOGRA VOWEL SIGN VOCALIC RR should be collated after U+11831 DOGRA VOWEL SIGN VOCALIC R, but before DOGRA VOWEL SIGN E.

Recommendations: We recommend the UTC approve DOGRA VOWEL SIGN VOCALIC RR, but discuss the code point location. If the character is placed at U+1183B, the proposal authors should include text on how it would collate. We also suggest that the UTC remand to the Editorial Committee an action that when the script is published, the block introduction mention how the independent liquids are represented in Dogra.

17. Tamil
Document: L2/17-158 Reconsidering ScriptExtensions added for supporting Tamil fractions –Sharma

Comments: We reviewed this document, which requested the removal of Grantha from ScriptExtensions.txt for 0BAA TAMIL LETTER PA and 0BB5 TAMIL LETTER VA. The document provided the background on the two characters, and the reason they had been included in ScriptExtensions.txt.

The request is based on the outcome of the 2016 meeting on Tamil fractions, as documented in L2/17-069 (p. 6, §3, items 1-2), which identified these two Tamil fractions as distinct from the similar-looking Grantha characters.

Recommendations: We recommend the UTC remove the two Tamil characters U+0BAA and U+0BB5 from ScriptExtensions.txt.

CENTRAL ASIA
18. Soyombo
Document: L2/17-235 Proposal to encode JIHVAMULIYA and UPADHMANIYA for Soyombo – Pandey

Comments: We reviewed an earlier version of this proposal, which had recommended the Indic Syllabic Category be changed to Consonant_Prefixed (based on the single example provided). The author has made this modification.

Recommendations: We recommend the UTC accept the two characters U+11A84 SOYOMBO SIGN JIHVAMULIYA and U+11A85 SOYOMBO SIGN UPADHMANIYA.

19. Khwarezmian
Document: L2/17-054 Proposal to encode the Khwarezmian script (revised) – Pandey

Comments: We reviewed this proposal for an abjad derived from Imperial Aramaic. The following comments were made during the discussion:
- Remove the holes from the chart
- Provide references to the following (from §3.1.1): "Variant forms are attested for some letters, eg. [..]. In some inscriptions the letters waw, zayin, yodh are represented using a highly similar if not identi[c]al form, in others they are distinguished."
- Turn figure 1 180 degrees (to aid in readability)
- Provide other sources with charts, if available
- Fill out §3.4 Ligatures more fully. It appears that words beginning with BETH and and those with initial TAW and FINAL NUN also connect. Provide a more in-depth analysis of shapes when they touch. Note: We don't necessarily agree that these are ligatures; the script, or a later version of it, could be joining.

    Compare the following from Figure 22:

    Also, in figure 15 (p. 24), it was noted that the engraver must intentionally have had the letters touch one another, which suggests intentional ligatures:

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

EAST ASIA
20. Hangul
Document: L2/17-241 Response to L2/17-153 Script ad hoc report and L2/17-126 Comments on Preliminary Proposal for Encoding New Hangul – Oh

Comments: We reviewed this response to the script ad hoc recommendations (L2/17-153), and the comments from Jaemin Chung (L2/17-126)

We recommend the authors provide evidence from published books. The core business of the UTC is not to develop and extend the Hangul writing system. If, however, the proposed characters are used by Korean speakers and the characters' use is demonstrated in published materials, then it would be appropriate to re-consider a proposal.

Recommendations: We recommend the UTC note this document.

21. Shuishu

Document: L2/17-239 Towards the ordering of the Shuishu script (WG2 N4839) - Everson

Comments: We reviewed an earlier version of this document, which proposed two different radical orderings, one from the Shui community (36 classes, for a total of 86 radicals), and one from Michael Everson (67 radicals).

Comments that arose during discussion:
- Is there a native ordering?
- Is there any semantic basis for radical choices made by the Shui community?
- The goal should be to organize a large list of arbitrarily shaped characters in a way that makes sense to users, and will aid them in finding characters. Too many radicals may present too great a burden for user.

Recommendations: We recommend the UTC discuss this document.

SYMBOLS AND NUMBERS
22. Hell pause character
Document: L2/17-151 Character Proposal: Hell pause character – Spix

Comments: We reviewed this proposal for one character, which is well-attested in the evidence provided.

A few comments:
- We recommend the proposed name HELL PAUSE CHARACTER be changed to HELLSCHREIBER PAUSE SYMBOL, which is less ambiguous and misleading than "HELL PAUSE CHARACTER".
- The Bidi_Mirrored property should be "N[o]", unless text examples can be provided.
- The proposed codepoint, U+23FF, is not available. Instead, we recommend the code position U+2BFF in the Miscellaneous Symbols and Arrows block (or another hole in the same block).

Recommendations: We recommend the UTC accept this character, after discussing the name, code point, and property change.

23. Neptune
Document: L2/17-191 Proposal to encode NEPTUNE FORM TWO – Marin Silva

Comments: We reviewed this proposal, which provides evidence for the proposed character. The proposed location is acceptable.

Recommendations: We recommend the UTC accept the character U+2BC9 NEPTUNE FORM TWO, and request the author fill out a summary form.

24. Block Elements
Document: L2/17-194 Proposal to create a new block for missing Block Element characters – Marin Silva

Comments: We reviewed this proposal to create a new block for 5 characters.

The following summarizes the comments made:

- We note that, in general, documents should not propose new blocks, but instead should use the format U+XXXX0, U+XXXX1, etc. for characters whose location they are unsure about, and let the UTC determine the best location.

- Because this proposal is requesting characters from an old character set, we suggest the author collaborate with those working on other old character set proposals (such as anyone working on a PETSCII proposal, as mentioned in early April 2017 on the Unicode email list), and provide solid justification for the request.

- The requirement to encode the characters in BMP is not justified, in our view. A better location would be in the Geometric Shapes Extended block, which has 39 spots available.

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

25. Clock faces
Document: L2/17-179 On the exchange of clock face information in plain text and implications for encoding – Marin Silva

Comments: We reviewed this document, which responded to script ad hoc recommendations (L2/17-153) and UTC feedback to an earlier version of this proposal (L2/17-092). The earlier document had proposed two characters, ROMAN NUMERAL ALTERNATE FOUR and SMALL ROMAN NUMERAL ALTERNATE FOUR. In the current proposal, the latter character was removed, but three new precomposed characters are proposed (a precomposed number ten, eleven, and twelve).

The following points were raised in discussion:
- The Roman numerals in Unicode (U+2160..U+2182) were part of Unicode 1.0.0, and derive from an East Asian legacy set. While the ARIB set did contain Roman numerals (cf. L2/07-391), they did not originally come from the ARIB standard.
- The Unicode Standard does not encode precomposed sequences for letters or numbers for symbolic concepts, when a sequence of existing characters is available for that concept. Note that Unicode Standard does not encode *sin*, *tan*, and *cos*, despite the fact that those are well-understood symbolic math concepts.
- To aid in quickly locating the history of a document, we suggest the proposal include in the heading of the document "Replaces L2/17-092".

Recommendations: We recommend UTC members review this document, and forward their feedback (including that above) to the author.

26. Counting rods
Document: L2/17-187 Proposal to add Southern Song forms of counting rods as separate characters – Marin Silva

Comments: We reviewed this proposal, which was a revision of L2/17-085 "Proposal to add 6 standardized variation sequences for counting rods." The author has revised the proposal in light of the May 2017 script ad hoc recommendations (L2/17-153).

The new proposal is a marked improvement over the earlier version.

The following are comments that arose during discussion:
- The left-hand figure on page 2 should be removed (it is a mirror image of the right-hand figure)
- Change the character names from "COUNTING ROD SOUTHERN SONG" to "SOUTHERN SONG COUNTING ROD..."
- Make the name of the first character consistent with those for FIVE and NINE, hence: "SOUTHERN SONG COUNTING ROD UNIT DIGIT FOUR"
- Modify the glyph for UNIT FIVE, as the head is too short, and discuss the glyph shape. The shape should better match the example in the manuscript on page 1.
- To assist reviewers, it would be helpful to list the entire set of counting rods characters and glyphs (i.e., 1-9 and 10-90 in Southern Song system) so they all appear in one document.
- We recommend an annotation be added for SOUTHERN SONG COUNTING ROD UNIT DIGIT FOUR, "used also for tens".
- The code points are acceptable.
- To aid in quickly locating the history of a document, we suggest the proposal include in the heading of the document "Replaces L2/17-085".

Recommendations: We recommend the UTC review this proposal, and accept the five Southern Song Counting Rod characters from U+1D379..U+1D37D, after discussing the names and other points raised above.

27. Tally marks
Document: L2/17-188 Proposal to change the name of the accepted tally marks and add named character sequences for them – Marin Silva

Comments: We reviewed this document which requested three actions:

a) Change the names for TALLY MARK ONE and TALLY MARK FIVE to a name based on their shape (FENCE TALLY MARK ONE and FENCE TALLY MARK FIVE). The rationale given is that the name TALLY MARK does not take into consideration that there are other tally mark systems.
b) Add three named sequences for TWO, THREE, and FOUR.
c) Give them the property vo=R[otated], to account for display in vertical text

The two characters are currently in PDAM 1.3 ballot. In our view, the name changes are not necessary. The current header for the two characters is "Western tally marks", but could be changed if a more appropriate header was proposed.

The named sequences are not, in our view, needed.

We recommend the author provide evidence of the characters in vertical orientation, at which time the Vertical_Orientation property will be considered.

Recommendations: We recommend the UTC review this document, and send the author feedback, including the comments above.

28. Ancient Chinese Mathematical Symbols
Document: L2/17-219 Proposal to encode Ancient Chinese Mathematical Symbols -- Kushim Jiang

Comments: We reviewed an earlier version of this proposal for ancient Chinese mathematical symbols. The proposal requests 6 combining diacritical marks, 11 mathematical symbols, and 87 enclosed ideographs. In the most recent version, the author proposes 87 sequences for enclosed ideographs (made up of CJK characters, CGJ, and U+25EF LARGE CIRCLE). Instead of such sequences, which are neither correct nor likely to be supported, any such circled symbols, if justified, should instead be proposed as atomic (non-decomposed) circled symbols, to be included in the Enclosed Ideographic Supplement block.

A number of good candidates are proposed for encoding in our opinion, but additional information is needed, including expert review.
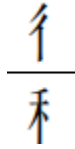
The following comments arose during discussion:
- The 8 Celestial Stem characters below are a coherent, limited set that are good candidates for encoding, but where are '9' and '10'? (cf. https://en.wikipedia.org/wiki/Celestial_stem)

⊞	U+7532,U+034F,U+25EF		戉	U+620A,U+034F,U+25EF
⊿	U+4E59,U+034F,U+25EF		己	U+5DF1,U+034F,U+25EF
丙	U+4E19,U+034F,U+25EF		庚	U+5E9A,U+034F,U+25EF
⊤	U+4E01,U+034F,U+25EF		辛	U+8F9B,U+034F,U+25EF

- The COMBINING LONG SOLIDUS OVERLAY is already encoded as U+0338 (with the same name), so it should be removed.
- The COMBINING ENCLOSING OCTAGON is shown in figure 3, but its use and glyph are not clear. The glyph appears to be a rectangle with cut corners. Provide additional examples.
- The COMBINING FACTORIAL SYMBOL ⌑ is a good candidate. Apparently factorial has a similar shape found in Western European literature, but with the square corner of the box on the left (i.e., $\underline{n}$, see http://mathworld.wolfram.com/Factorial.html). (Note: The webpage https://groups.google.com/forum/#!topic/sci.math/P20zFu2eebs mentions that both were in vogue in England). A graphically similar symbol is encoded at U+20E7, but with different position of the lines, so there is precedent for encoding such a symbol.
- For the CHINESE DIFFERENTIAL SIGN and the CHINESE INTEGRAL SIGN characters, the Kangxi radicals U+2F3B and U+2F72 could theoretically be used (below is an example of the characters from figure 28). However, if the characters came to be used symbolically, or appear in scoping
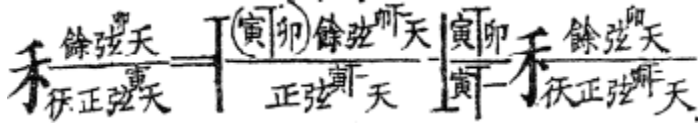
contexts (that is, they grow larger or smaller than surrounding text), then they may be eligible for separate encoding. Provide additional examples of usage.

- The following seems to be reasonable, but are the other pieces of text in the examples (such as figures 25 and 26) capable of being represented?

CHINESE POSITIVE DIFFERENCE SIGN

- Some of the text examples reflect complex notation, which will need higher level representation, such as table in figure 6:
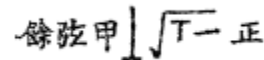
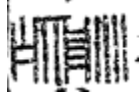

For the text in these examples, what are the pieces?

- What do the dots indicate?

- In figure 15, what does the "T" represent?

- How will the following be represented (from figure 8)?

Recommendations: We recommend the UTC discuss this document, and encourage members to solicit feedback from experts, and send feedback to the author.

NOTATIONAL SYSTEMS
29. SignWriting
Document: L2/17-220 Design Options for Sutton SignWriting with examples and fonts – Slevinski

Comments: We reviewed this proposal, which proposed two options of encoding for Sutton SignWriting:

1. The first option uses a scheme of markers and numbers that overwrite the current Unicode code points for SignWriting, and uses Plane 4 code points for the SignWriting symbols.

2. The second option uses the Unicode code points for SignWriting, but adds additional modifier, number, and marker characters located at two currently unassigned spots in the SignWriting block (1DA9A, 1DAA0) and 15 code points in an unassigned column (1DAB0..1DAB9, 1DABA, 1DABB..1DABE).

Neither option conforms to the Unicode Standard.

The model advocated in this document relies on a coordinate-based system, which requires the characters to be located in relation to one another. Such an approach would require a font-dependent system: if a different font were used, the relationships could be lost, thereby jeopardizing reliable text interchange.

In our view, a well-designed mark-up solution should be used, along with a custom-rendering engine, since plain text won't capture the relation of one character to another.

An instructive model, in our opinion, is musical scoring, which requires specialized rendering of the basic set of musical note characters within their complex data structures. Cf. MusicXML, an XML-based file format used to represent Western musical notation.

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

OTHER PROPOSALS AND DOCUMENTS
Indic Properties
30. Indic Syllabic Categories
Document: L2/17-121 Indic characters without syllabic category assignments -- Pandey

Comments: We reviewed this document which provides a list of all Indic characters that are missing Indic category assignments. While this is a helpful list, it was noted that not all characters necessarily require categorization, such as those that do not take part in graphic syllables (i.e., some punctuation marks or Balinese musical symbols).

Of key interest are those characters that take subscripts or vowels. We suggest the author refine this list, separating out those that do not participate in syllables from those where there is some uncertainty, so evidence can be gathered. We recommend those characters that do not participate in syllables be removed from the list.

Recommendations: We recommend the UTC review this document, and provide the author with feedback.

31. New Indic Categories
Document: L2/17-148 New categories for Indic characters – Pandey

Comments: We reviewed this document, which attempted to classify the uncategorized characters listed in L2/17-121. We recommend the author review UTN 36 "A Categorization of Unicode Characters" and its history.

Suggestions for new categories for characters in Indic scripts are interesting, and the author is welcome to propose such categories to the UTC, but the main function of IndicSyllabicCategories.txt is to provide categories relevant to the participation of characters in Indic syllabic structure. The intent is not to provide a categorization of *all* characters used in Indic scripts.

The history of the withdrawn draft UTR #49, "Unicode Character Categories" is an illustrative cautionary study of the reluctance of the UTC to take on further general categorization of *all* characters, or even all characters used in some limited set of scripts. UTN #36 was the frozen outcome of the failure of UTR #49 to reach consensus in the UTC

Recommendations: We recommend the UTC review this document, and send the author feedback.

32. Indic category for TAMIL NUMBER SIGN
Document: L2/17-178 Feedback on New Indic category for 0BFA Tamil Number Sign - Sharma

Comments: We reviewed this feedback on U+0BFA TAMIL NUMBER SIGN, which was categorized as "Number_Other" in L2/17-148 New categories for Indic characters.

We agree with Sharma that U+0BFA is not a number, and that Pandey should not try to categorize all Indic characters. We recommend Pandey check if any assigned properties are wrong and identify those unassigned characters that are part of an Indic graphic syllable.

Recommendations: We recommend the UTC review this document, and send the author feedback.

NOT COVERED
Small Seal: L2/17-250 Shuowen Seal Encoding Design Issues – Suzuki and Cook
Tigalari: L2/17-182 Comments on encoding the Tigalari script – Srinidhi and Sridatta
Amaragannada: L2/17-186 Introducing the Amaragannada scripts – Srinidhi and Sridatta
Pungchen: L2/17-181 Preliminary proposal to encode the Pungchen script – Pandey
Brusha: L2/17-183 Preliminary proposal to encode the Brusha script – Pandey