

TO: UTC
FROM: Deborah Anderson, Ken Whistler, Roozbeh Pournader, Lisa Moore, Liang Hai, and Richard Cook¹
SUBJECT: Recommendations to UTC #154 January 2018 on Script Proposals
DATE: 19 January 2018

The Script Ad Hoc group met on 17 November 2017, 8 December 2017, and 5 January 2018 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.

EUROPE

1. Palaeohispanic

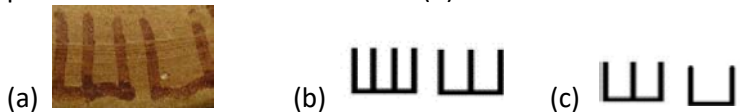
Document: L2/18-030 Proposal to encode the Palaeohispanic script – Ferrer et al.

Comments: We reviewed this preliminary proposal, which has incorporated comments from the November 2017 script ad hoc meeting. (Changes to previous versions are highlighted.)

The latest proposal includes a glyph variation chart (pp.23-25) and addendum with an inscription and code points, and the shapes users would expect in online and print publications (p. 26).

The following summarizes the comments:

- The Palaeohispanic script appears to reflect a 3-way distinction for certain characters, but this distinction is not represented in the latest version of the proposal. In the addendum (p. 26), for example, the top inscription has the shapes on the left (a), but the expected appearance in publication is shown in the middle (b):



We think the appropriate representation for transcribing this inscription would be on the right (c).

- Old Italic is cited as a model for Palaeohispanic (§2 Background), since it unifies various historic alphabets into one script, like Palaeohispanic. Old Italic, however, is probably not a good model for the encoding, since it relies on language-specific fonts and can cause confusion for the user when trying to determine which character to use.
- A source of confusion for the proposers seems to be distinguishing the transcription layer ([c] above, reflecting the text on the artefact) from the normalized transliteration layer (b). To help clarify the different levels of encoding, we recommend the proposers provide samples of text from different eras, and work with the script ad hoc to go through the examples. Unicode provides an encoding for the transcription, but enables transliteration (and the Romanized transliteration).

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

¹ Also calling in or participating in at least one meeting were: Yifán Wáng, Jan Kucera, Mohamed Elsharkawy, and Ziad Khalidi.

2. Latin

Document: [L2/17-437](#) Feedback on proposed Unifon letters – Marin Silva

Comments: We reviewed this document, which provided comments on a proposal from 2014, [L2/14-070](#), which was a revised proposal for Unifon letter additions by Michael Everson.

Two items from the original proposal ([L2/14-070](#)) were deemed “more urgent” in this document:
LATIN CAPITAL LETTER SMALL CAPITAL I WITH STROKE
LATIN SMALL LETTER CLOSED U.

In our opinion, the first, adding an uppercase version to U+1D7B LATIN SMALL CAPITAL LETTER I WITH STROKE, is not urgent. Are there any users actively asking for this character now? For the second character, LATIN SMALL LETTER CLOSED U, which is said to be used in Swedish dialectology, we recommend the author demonstrate its use in a full proposal, with examples.

Note also that encoding the second member of a case pair – just because one of the pair appears in Unifon – is not a strong argument, in our opinion.

This document also has comments on the name and glyph of LATIN LETTER TURNED-E R, suggestions on names for some letters, and an opinion on the unification of certain letters. However, unless a proposal is made for Unifon letters, preferably with support from users, no action needs to be taken.

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

AFRICA

3. Garay

a. Proposal

Document: [L2/16-069](#) Proposal for encoding the Garay script in the SMP – Everson

Comments: We reviewed this proposal for Garay. The following summarizes points raised in the discussion:

- Based upon the evidence presented, such as in figure 7 (below), the VOWEL SIGN E and above-comma diacritical mark on MBA, NGGA, and NDA are graphically the same. As a result, we recommend the characters MBA, NGGA, and NDA be removed, and instead they be represented as a base letter and VOWEL SIGN E.

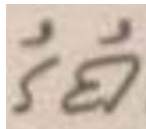


Figure 7:

- Based on the evidence, the dot above should be separately encoded. Besides NJA and NGA, the examples show other letters with a dot above.

Figure 24: $\text{ñ} \text{š}$ and $\text{ñ} \text{š}$ (in caption) and in the image, above the left leg of GA:

Hence, we recommend the removal of NJA and NGA, and instead they be represented as a base letter and combining dot above.

- In §2.2.3 Gemination, the text reads:
 “The gemination mark (◌◌) is written above a consonant letter and above the VOWEL SIGN E where that applies to a letter (◌◌). In encoded text, for searching and matching, perhaps the best practice would be to encode the gemination mark (bound to the consonant) first and the VOWEL SIGN E afterward, even though the presentation form appears to be the opposite.”

The wording in the second sentence should be revised, as the vowel mark should be encoded first, followed by the COMBINING GEMINATION MARK.

- In §9 Unicode Properties, The canonical combining class for VOWEL SIGN E and COMBINING GEMINATION MARK should be 230.
- In §4 Digits (and §9 Unicode Properties), explain why the Bidi_Class is AN (Arabic_Number), instead of EN (European_Number) or L (Left-To_Right).
- Discuss the use of the macron in the following (figure 24)



- Document the various uses of the squiggle in Garay in the proposal (i.e., based on comments from Jack Merrill, in vowel-initial words before the vowel mark; to host the vowel /e/ (or /é/) after voiced stops; to mark the +ATR vowels /ë/, /ó/, /é/ [likely inspired by the use of diacritics in the official orthography]; and used after the sign for /i/, when intended to mark the sound [ü]).

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

b. Report

Document: [L2/17-322](#) Report on the Garay script 2017 (WG2 N4875) - Charles Riley

Comments: We reviewed this report from Charles Riley on his trip to West Africa. Having access to the pages of the Koran in the Garay script would be helpful. (No examples were contained in the latest version of the proposal, L2/16-069.)

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



4. Bété

Document: [L2/17-323](#) Report on the Bété script 2017 (WG2 N4876) – Charles Riley

Comments: We reviewed this report on the Bété script, which contains a preliminary code chart.

The following comments were made:

- A clear case needs to be made for encoding this script. Are documents available?
- It appears that the same signs are used to indicate something else by doubling:

1E600		ba	ba	BETE SYLLABLE	BA
1E603		beu	bə	BETE SYLLABLE	BEU

- The proposal should distinguish the graphical elements of a syllabary, as opposed to a sequence that is mapped onto syllables.

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

5. Egyptian Hieroglyphs

Document: [L2/17-415](#) Revised draft for the encoding of an extended Egyptian Hieroglyphs repertoire – Suignard

Comments: We briefly reviewed this document, which is a revision of L2/17-073, but with the addition of references from Hornung and Schenkel. The additional repertoire in this proposal totals 4504 characters.

In our view, including the rest of the Hieroglyphica characters – even those without references, after removing obvious duplicates – would be worthwhile, and we recommend they follow the IFAO taxonomic order.

Recommendations: We recommend the UTC review this proposal.

SOUTH AND CENTRAL ASIA

6. General -- Indic scripts

NOTE: [L2/17-098](#) Request for editorial updates to Indic scripts (Srinidhi and Sridatta) was an omnibus proposal that was not reviewed by the UTC at either the July/August or October 2017 meetings. However, the script ad hoc did review it carefully in July 2017 and provided comments in [L2/17-255](#). Changes were made by the authors. However, the script ad hoc at its November 2017 meeting recommended the requests be broken up into separate documents, based on the requested action. The resulting documents are: [L2/17-423](#), [L2/17-424](#), [L2/17-425](#) and [L2/17-426](#), which are discussed on the following pages.



a. Core Spec Changes for Indic Scripts



Document: [L2/17-423](#) Changes to Core Specification for Indic scripts for Unicode 11.0 – Srinidhi and Sridatta

Comments: We reviewed this document, which requested changes to the Core Specification for various Indic scripts, and provide below a synopsis of the request, with our recommendations.

§1 Sharada

The authors request the glyphs for U+111C2 *jihvamuliya* + KA/KHA i in §15.3 of *TUS* be corrected (see below). The authors provide additional examples (beyond those originally shown in L2/17-098). We agree with the request; it is well-justified.

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



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

Based on the information provided, the current Indic_Syllabic_Category of Sharada *jihvamuliya* as Consonant_Prefixed should be retained. (If, at a later point, evidence is provided showing *jihvamuliya* occurring by itself, then the property should be changed.)

Recommendations: We recommend the UTC remand the Sharada *jihvamuliya* glyph correction to the Editorial Committee. We further recommend the Editorial Committee ask the authors or Anshuman Pandey to provide the appropriate glyphs.

§2 Kannada

The authors request the glyph for Kannada *hka* be corrected in §12.8 of *TUS*:

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

We agree with this proposed minor glyph fix. It was noted that all subjoined KAs appear to lose the head stroke.

Recommendations: We recommend the UTC remand the glyph correction for Kannada *hka* (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 SIGN NIHSHVASA

The authors request the description of U+1CD3 VEDIC SIGN NIHSHVASA in §12.1 of the Core Specification be aligned with the code chart annotation, which specifies that U+1CD3 “separates sections between which a pause is disallowed”. The current description in §12.1 says the character “indicates where a breath may be taken.” The authors recommend the wording “Separates sections of Sama Vedic singing between which a pause is disallowed.” The request is corroborated with information in L2/09-372 (§4.5.2) and L2/09-298. We agree with the authors.

Recommendations: We recommend the correction to §12.1 of the Core Spec be remanded to the Editorial Committee.

§5 Ligature Forms for Ra + Vocalic Liquids

The authors recommend the Core Specification be revised to mention that the sequence */r vocalic_r/* can appear graphically in two forms in scripts besides Devanagari: the full independent vowel form of *vocalic_r*, with a superscript *repha*, and the full letter form of *ra* with the subscript, dependent form of *vocalic_r*. (The Devanagari *ra* + vocalic letter ligature forms are shown in Table 12-4 of *TUS*.)

Examples from scripts besides Devanagari are provided by the authors, although during discussion it was mentioned that it is not clear the two forms are necessarily a choice in a particular script. (In Telugu, for example, the two forms only appear in older texts, as noted in the document.)

Recommendations: We recommend the remand this item to the Editorial Committee, and give an action item to Liang Hai to add a short paragraph after Table 12-4, indicating the pattern of behavior for *ra* and *vocali_r* is observed in other scripts, such as Kannada, etc.

§6 Brahmi

[Note: The following section was included in a late revision of the document which the script ad hoc did not see. The basic request is summarized below.]

The authors recommend the glyph for *jñā* be changed in figure 14-1 from

current:  proposed: 

Evidence is provided.

The authors also request the text be modified in the “Vowel Modifiers” section of §14.1 (current text on top, with text highlighted by the authors):

“U+11002 brahmi sign visarga is used to write syllable-final voiceless /h/; that is, [x] and [f]. The velar and labial allophones of /h/, followed by voiceless velar and labial stops respectively, are sometimes written with separate signs U+11003 brahmi sign jihvamuliya and U+11004 brahmi sign upadhmaniya.”

Should be corrected as “U+11002 brahmi sign visarga is used to write syllable-final voiceless /h/. The velar and labial allophones of /h/; that is, [x] and [ϕ], followed by voiceless velar and labial stops respectively, are sometimes written with separate signs U+11003 brahmi sign jihvamuliya and U+11004 brahmi sign upadhmaniya.”

Recommendations: This topic should be discussed by the UTC, with input from Andrew Glass.

b. ScriptExtensions.txt changes for Indic

Document: [L2/17-424](#) Changes to ScriptExtensions.txt for Indic characters for Unicode 11.0 – Srinidhi and Sridatta

Comments: We reviewed this document, which recommends a series of additions to ScriptExtensions.txt. A short summary of each request and our recommendations to the UTC follow below.

§1 Devanagari

DEVANAGARI SIGN PUSHPIKA

The authors request Newa and Kannada be added to the set of scripts in the ScriptExtensions property for U+A8F8 DEVANAGARI SIGN PUSHPIKA, and when Nandinagari and Tigalari are encoded, similar adjustments be made the set of scripts in the ScriptExtensions property for U+A8F8 to account for usage by those scripts.

It was noted by the script ad hoc that the character U+A8F8 DEVANAGARI SIGN PUSHPIKA does not currently have a ScriptExtensions property. The shape of the *pushpika* represents a flower, but it does not interact with neighboring characters. The *pushpika* might be compared to OM or the dandas, whose shapes tend to get harmonized on a per script basis.

The question was raised whether it is appropriate to use the Devanagari *pushpika* in other scripts, especially if the scripts are not closely related. Or would it be better to separately encode the *pushpika* in different scripts?

Compare:

Northern scripts: Devanagari  Newa  Nandinagari  (a southern development of the northern Nagari style)

Southern scripts: Kannada  Tigalari 

Recommendations: We recommend the UTC discuss this, particularly regarding use of the *pushpika* character in the southern Brahmic scripts.

§2 Vedic characters

General note: We recommend the authors consider writing a Unicode Technical Note on the use of Vedic signs across the Indic scripts.

§2.1 Veda in Bengali

The authors provided evidence of Vedic characters in Bengali, and satisfactorily answered a question posed earlier by the script ad hoc in [L2/17-255](#). All the characters have Script_Extensions=Deva, except U+1CD0 and U+1CD2, which have Script_Extensions=Deva Gran.

Recommendations: We recommend Bengali be added to ScriptExtensions.txt for the following 8 characters: U+1CD5, U+1CD6, U+1CD8, U+1CE1, U+1CD0, U+1CD2, U+1CEA, and U+1CED.

§2.2 Veda in Tirhuta

The authors provide solid evidence of U+0951 DEVANAGARI STRESS SIGN UDATTA and U+0952 DEVANAGARI STRESS SIGN ANUDATTA in Tirhuta. We noted that both U+0951 and U+0952 have a long list of scripts (12 and 11, respectively) in ScriptExtensions.

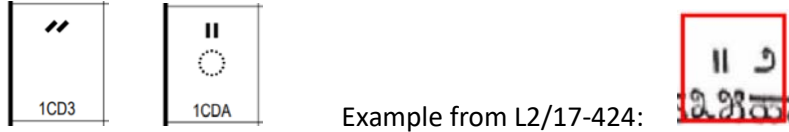
Recommendations: We recommend Tirhuta be added to the scripts in ScriptExtensions.txt for U+0951 and U+0952.

§2.3 Samavedic characters in Kannada

The authors give solid evidence of the use of U+1CD0, U+1CD2, and U+1CF4 in Kannada.

For U+1CD3 VEDIC SIGN NIHSHVASA, the example shown is a combining character above, which is noted in the text (“[t]he combining character written above digit 5 is U+1CD3”). However, U+1CD3 is a spacing character.

The example (below right) looks more like U+1CDA VEDIC TONE DOUBLE SVARITA:



- Explain how U+1CD3 is a combining character.
- Provide a separate document that focuses solely on U+1CD3, and provide evidence from other scripts. Depending upon the evidence provided, a new combining version of VEDIC SIGN NIHSHVASA may be needed.

Recommendations: We recommend the UTC add Kannada to the set of scripts in the ScriptExtensions property for U+1CD0, U+1CD2, and U+1CF4. For U+1CD3, we recommend the UTC forward the questions above to the proposal authors.

§2.4 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, but the authors confirm this was due to typographic limitations. (The latter was in answer to a question posed by the script ad hoc comments in in [L2/17-255](#).)

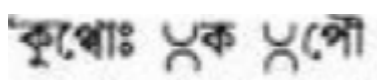
Recommendation: We recommend Oriya be added to the scripts in ScriptExtensions.txt for U+1CDA.

§2.5 U+1CF2 VEDIC SIGN ARDHAVISARGA

The authors request Bengali, Kannada, Telugu and Tirhuta be added to the scripts in ScriptExtensions.txt for U+1CF2 VEDIC SIGN ARDHAVISARGA.

As defined in §12.1 of *TUS*, the *ardhavisarga* marks either a *jihvamuliya*, which is a velar fricative occurring before *ka* or *kha*, or an *upadhmaniya*, which is a bilabial fricative appearing before *pa* and *pha*. It is represented in text in visual order before the consonant it modifies. In the

examples provided, it is not clear that VEDIC SIGN ARDHAVISARGA  as a spacing combining mark (gc=Mc) was correctly analyzed by the standards committees, at least for Bengali. In the Bengali examples, the *ardhavisarga* precedes the character it modifies (left, below), or appears surrounded by spaces (right, below)



Should *ardhavisarga* be changed to a non-combining spacing character? Would that change affect texts or how syllables are rendered? Would it adversely affect line breaking in Bengali? (Sharma had argued for *gc=Lo* for *visarga*, *anusvara*, and related characters including *ardhavisarga* in [L2/09-343](#).)

The Tirhuta example (below, left) and Telugu (below, right) similarly precede the consonant.



Recommendations: We recommend the proposal authors prepare a separate document on U+1CF2 VEDIC SIGN ARDHAVISARGA, and explore the interaction of this sign with characters before and after it. The Unicode Standard character only reflects interaction with the character *before* it, which may be incorrect. Is the character in the Bengali samples U+1CF6 or a variant of U+1CF5 VEDIC SIGN JIHVAMULIA?

§3 Khojki

The proposal authors recommend Khojki be added to the script(s) in ScriptExtensions for 10 characters U+A830..U+A839 in the Common Indic Number Forms block, based on the Khojki script proposal ([L2/11-021](#)). We agree with this suggestion.

Recommendations: We recommend Khojki be added to the scripts in ScriptExtensions.txt for U+A830..U+A839 .

§4. Grantha

The authors provide evidence of the character U+0BF3 TAMIL DAY SIGN in Grantha, and hence the authors recommend Grantha be added to ScriptExtensions. This request is acceptable.

Recommendations: We recommend Grantha be added to the script(s) in ScriptExtensions for U+0BF3 TAMIL DAY SIGN.

Note: *The following South and Central Asian-related documents are ordered based on the script name (in alphabetical order).*

7. Amaragannada scripts

Document: [L2/17-186](#) Introducing the Amaragannada scripts – Srinidhi and Sridatta

Comments: We reviewed this document, which introduces a set of approximately 10 historic scripts for possible encoding. The scripts, used primarily in Northern Karnataka in SW India, were devised to protect the religious literature of Kodekal Basavanna from destruction by Islamic rulers. The major language used was Kannada.

A review of this document raised the following questions:

- How widespread are these scripts?
- Are the scripts named?
- If there is a one-to-one mapping between these scripts and Kannada (such as in figures 5, 6, 7 and 8), why should these be encoded? Why wouldn't a font-change suffice? Provide a

convincing use-case. It was noted that while a cipher *could* become a full-fledged scripts (i.e., Thaana), many do not. Separately encoding a cipher for Kannada may end up obscuring the true content of the Kannada text.

Recommendations: We recommend the UTC members review this proposal, and send comments to the proposal authors.

8. Brahmi

Document: [L2/17-426](#) Request to change the glyphs of THIRTY and FORTY of Brahmi – Srinidhi and Sridatta

Comments: We reviewed this proposal to change the glyphs for Brahmi ‘30’ (U+1105D) and ‘40’ (U+1105E), so the representative glyphs reflect the shapes in early sources. The authors noted that Segoe UI Historic font already contains the proposed glyph for ‘40’. Andrew Glass, the co-author of the Brahmi proposal, agrees with the change.

Recommendations: We recommend the UTC accept the glyph change, and consider issuing a glyph erratum.

9. Devanagari

a. SRI SYMBOL

Document: [L2/17-410](#) Proposal to add very important "sri" symbol in Devanagari Unicode block - Dhruv Pathak

Comments: We reviewed this proposal which requested a character for Devanagari *śri*, since users are reportedly having difficulty in writing it.

Devanagari *śri* is a ligature of <SHA, Virama, RA, VOWEL SIGN I>. The sequence for the consonants is spelled out in Table 12-3 (p. 461 in Chapter 12 of *TUS*):

श + ष् + र → श्र

Note: The author mentions that it exists in Gujarati. However, the Unicode character is not contained in the Gujarati code block (<http://www.unicode.org/charts/PDF/U0A80.pdf>).

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

b. SHORT A

Document: [L2/17-425](#) Request to annotate DEVANAGARI LETTER SHORT A – Srinidhi and Sridatta

Comments: We reviewed this short document, which requested an annotation be added to U+0904 DEVANAGARI LETTER SHORT A, noting that it was used to denote short *e*.

The document provides useful evidence for this character, which was requested by the Government of India in 2001 ([L2/01-304](#)), but which had been missing attested usage (cf. [L2/09-321](#)). The attestations demonstrate its usage 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 in the Devanagari names list.

10. Dives Akuru

Document: [L2/17-417r](#) Dives Akuru (revised) - Pandey

[Note: The script ad hoc reviewed the 31 December 2017 version of the proposal.]

Comments: We reviewed this very well-researched, revised proposal for Dives Akuru. This version made changes to the December 7 2017 version of the proposal.

The following summarizes the recommendations based on the discussion:

- §4.3.2. (top of page 21): Fix the minor typo: “In *lōmāfānu* syllables are separated by spaces and these spaces provide for distinctions between clusters and sequences of consonant syllables, eg...*tka* and...*kata*” (instead of *taka*).

- In §4.3.3, when YA is C2, we recommend it be represented with a new MEDIAL YA character, ೀ.

I.e., *tya* ೀ <TA, MEDIAL YA>.

- In §4.3.4. consider having new, separate characters for *ra-kara* and *repha* so no VIRAMA would be required (i.e., MEDIAL RA and REPHA). If the name MEDIAL RA is proposed, include an annotation in the names list, identifying it as *ra-kara*.
- In §4.4 ‘False Conjuncts’ (top), for both $n=i$ and $n='i$, we recommend using VIRAMA with no joiner, since the joiners are often confusing for users.

I.e., $n=i$ ು and $n='i$ ೂ <NA, VIRAMA, I> and $n='i$ ೃ <NA, VIRAMA, YA, VOWEL SIGN I>

- §4.4. To represent the third example in §4.4 (using a new MEDIAL YA) the sequence would be: *nyi* ೄ <NA, MEDIAL YA, VOWEL SIGN I>

- §4.11 Collation: Insert the new characters next to “normal” RA, and “normal” YA

- Code chart:

- Use the new code point range, starting at U+11900.
- If the new medial characters and *repha* are proposed, insert them in the code chart near DIVES AKURU PREFIXED NASAL SIGN.
- Adjust annotation for VIRAMA, noting its usage (if the new characters are proposed)
- Verify the code points are correct throughout the proposal: the code points need to be corrected for *anusvara*, *candrabindu*, and *virama* in *Indic_Syllabic_Category*, for example.
- Submit the next revised version as a new document in the 2018 registry, noting it replaces L2/17-417R.

In general, we liked the model presented for Dives Akuru. With the proposed changes recommended above and after a final review, a revised proposal may be, in our opinion, ready for the UTC to consider accepting.

Recommendations: We recommend the UTC discuss the final, revised proposal.

11. Kharoshthi

Document: [L2/17-416](#) On Indic Syllabic Category of KHAROSHTHI VIRAMA– Srinidhi and Sridatta

Comments: We reviewed this document, which recommends the UTC re-examine the Indic_Syllabic_Category for U+10A3F KHAROSHTHI VIRAMA, based on three possibilities:

- Virama (only characters that both are killer and consonant stackers)
- Pure Killer (only kills vowels in consonant cluster, but no stacking ability)
- Invisible stacker (invisible stacker)

The current Indic syllabic property value for Kharoshthi Virama is Invisible_Stacker, but the authors recommend it be changed to Virama, since the Kharoshthi virama can act as a visible killer *and* a consonant stacker. The authors point to evidence from Table 14-6 of *TUS* “Examples of Kharoshthi Virama” in support of the change.

The Kharoshthi proposal author Andrew Glass reported that the virama is very rare. When the script was first proposed in 2013, a simple model was put forward. At this point, Andrew Glass would prefer not to change the model, although other approaches would have been possible in hindsight (such as Limbu, which has syllable-final subjoined consonants, or the recently approved Masaram Gondi, which has a separate invisible virama and visible killer). Andrew Glass mentioned that the two categories “Virama” and “Invisible_Stacker” get merged into the same class [for the Universal Shaping Engine], so any change would not have any implications for the shaping engine.

Glass also noted that the example for “pure virama” in Table 14-6 (below) is not a good illustration.

Pure virama	dha + i + k + virama → dhik 𑀢 + 𑀣 + 𑀤 + 𑀥 → 𑀦
-------------	--

It was mentioned during discussion that viramas often tend to have their own shape – but not always, cf. Masaram Gondi virama (U+11D45). The Kharoshthi virama kills vowels, but does not have a visible shape. There is precedence for an invisible virama to have a visible effect: the Masaram Gondi virama removes the tick on the consonant that precedes it, for example.

Recommendations: We request UTC to discuss and decide the appropriate Indic_Syllabic_Category for U+10A3F KHAROSHTHI VIRAMA. We also recommend Roozbeh Pournader and the EdCommittee take an action to document the invisible stacker in Indic_Syllabic_Category.

12. Khojki

Document: [L2/17-307](#) Proposal to encode two characters in Khojki – Srinidhi and Sridatta

Comments: We reviewed this document, which requested two Khojki characters. The following were comments raised during the discussion:

- Is the evidence based on just one source?
- Provide a translation of the text in figure 2.
- How is the independent vowel represented?

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

13. Khwarezmian

Document: [L2/18-010](#) Proposal to encode the Khwarezmian script in Unicode – Pandey

Comments: We reviewed this proposal. A revised version based on the comments below will be circulated to experts on the script. Based on more recent research, the author reported that the script is a full cursive joining script.

The following were comments that arose during discussion:

- §4.1 In the list of letters and their positional forms, we recommend all the red dashes be physically connected to the letter. In addition, explain the vertical line beside TAW forms. Lastly, we recommend the isolated forms be moved closer to the other forms in the listing.
- In cases where the letter occasionally is dual joining (such as *zayin*), we recommend making the letter dual-joining, using non-joiner when the letter does not join. This approach was used for Manichaean ([L2/11-123r](#)). Another approach would be to use ZWJ to modify the default joining behavior, as the author proposes in §5.2. A third option would be to encode both a dual-joining and a right-joining character (cf. Pashto). However, this approach should be avoided, in our opinion, as it leads to encoding characters that look alike for all orthographies. The author may wish to present the different options, and make a case for his choice.
- §6.3 Shaping properties - ArabicShaping.txt: For those characters that contain a single Joining_Group property value, replace the last column with “No_Joining_Group” (Note: Singleton Joining_Group values will be discussed at the January 2018 UTC, see: [L2/17-427](#))

Recommendations: We recommend the UTC and its members review this proposal, and forward comments to the author.

14. Malayalam

a. North Indian Quarter Signs

[L2/17-340](#) Request to Annotate North Indian Quarter Signs for Malayalam Usage – Cibu Johny

Comments: We reviewed this document which demonstrated use of NORTH INDIC FRACTIONS one quarter, one half, and three quarters in 19c Malayalam text, and requested attendant changes to the names list and Core Spec. The evidence was compelling, and we agree with suggested changes.

Recommendations: We recommend the UTC forward to the Editorial Committee the requests for annotations, and the update to §12.9 of the Core Spec, as documented on page 4 of [L2/17-340](#). In addition, we recommend Malayalam be added to the set of scripts in ScriptExtensions.txt for U+A830..U+A832.

b. Malayalam Vedic Anusvara

Document: [L2/17-276](#) Proposal to encode MALAYALAM LETTER VEDIC ANUSVARA– Srinidhi and Sridatta
Feedback document: [L2/17-419](#) Feedback on Malayalam Vedic Anusvara – Shriramana Sharma

Comments: We reviewed [L2/17-276](#), the proposal document for MALAYALAM LETTER VEDIC ANUSVARA, with evidence of its use.

The evidence is solid for this character. We noted that the term “VEDIC” is contained in two other characters: U+09FC BENGALI LETTER VEDIC ANUSVARA and U+1135EGRANTHA LETTER VEDIC ANUSVARA, so there is precedence for the proposed name.

We also reviewed the feedback document by Sharma, which provides rationale to move MALAYALAM LETTER VEDIC ANUSVARA from U+0D50 to U+0D04. We agreed with Sharma’s rationale for the move. The new location (U+0D04) would fill a hole near the beginning of the Malayalam block, and place the character near other similar phonetic characters. As noted by Sharma, it also leaves U+0D50 open in the event an OM character is found. In addition, it may be important to maintain a correspondence with other Indic scripts that typically have OM in that position.

Recommendations: We recommend the UTC approve MALAYALAM LETTER VEDIC ANUSVARA, with the glyph as given in [L2/17-276](#), but decide on the code point location.

15. Nandinagari

Document: [L2/17-213](#) Proposal to encode the Prishthamatra for Nandinagari – Srinidhi and Sridatta

Comments: We reviewed this proposal for one *prishthamatra* character for Nandinagari, a script that is in PDAM 2.2. An earlier version of this proposal was reviewed at the July script ad hoc, ([L2/17-255](#)), which recommended the proposers revise the document and include the Indic properties. The UTC did not take up this proposal at either the July/August or October 2017 meetings.

The revised proposal now includes the Indic properties.

Recommendations: We recommend the UTC approve U+11BD4 NANDINAGARI VOWEL SIGN PRISHTHAMATRA E with glyph and properties as in [L2/17-213](#). Since Nandinagari is in PDAM 2.2 ballot, we recommend a ballot comment be added, requesting this character.

16. Newa

a. JHVAMULIYA and UPADHMANIYA for Newa

Document: [L2/17-369](#) Proposal to encode JHVAMULIYA and UPADHMANIYA for Newa – Srinidhi and Sridatta

Feedback: [L2/17-364](#) Feedback on an Encoding Proposal (L2/17-369) Subject: Indic_Syllabic_Category of Newa jihvamuliya and upadhmaniya

Comments: We reviewed the proposal, which requested two characters for Newa, NEWA SIGN JHVAMULIYA and NEWA SIGN UPADHMANIYA. We agree with David Corbett’s feedback in L2/17-364, which recommends the Indic_Syllabic_Category be Consonant_With_Stacker, instead of Consonant_Prefixed. (The rule-of-thumb seems to be if the *upadhmaniya/jihvamuliya* becomes a superscript form of the consonant, then Consonant_Prefixed seem appropriate, but if the *upadhmaniya/jihvamuliya* is the main consonant, and the following consonant becomes subjoined, then Consonant_With_Stacker is the correct property.)

Other than the Indic_Syllabic_Category, the proposal appears sound.

Recommendations. We recommend the UTC approve the two characters 11460 NEWA SIGN JHVAMULIYA and 11461 NEWA SIGN UPADHMANIYA with glyphs and properties in

L2/17-369, but request the authors change the `Indic_Syllabic_Category` to `Consonant_With_Stacker`. Further, we recommend the UTC record an Action Item for Roozbeh Pournader and Liang Hai to investigate all the currently encoded *upadhmaniyas* and *jihvamuliyas* in `IndicSyllabicCategory.txt` to see if any need to be re-categorized, and recommend any new text for the Core Spec on this topic.

b. DOUBLE COMMA for Newa

Document: [L2/17-440](#) Proposal to encode the DOUBLE COMMA for Newa – Srinidhi and Sridatta

Comments: We reviewed this proposal for one character, NEWA DOUBLE COMMA. Clear examples are provided, and the code point is acceptable.

One outstanding question is: What is the preferred shape of the character in modern typographical usage (which is also consistent with the current code chart glyphs)? Can the authors provide any additional source material in modern typography?

Recommendations: We recommend the UTC approve the character U+1145A NEWA DOUBLE COMMA as documented in L2/17-440. We also suggest an Action Item be given to Deborah Anderson to follow up on the preferred shape of the glyph with contacts in Nepal.

17. Sharada

a. PRISHTHAMATRA for Sharada

Document: [L2/17-214](#) Proposal to encode the Prishthamatra for Sharada – Srinidhi and Sridatta

Comments: We reviewed this proposal, which was a minor modification of the version seen earlier by the script ad hoc in July 2017 ([L2/17-255](#)). The revised proposal now includes the appropriate Indic properties. This proposal was not discussed at either the July/August or October 2017 meetings.

Recommendation: We recommend the UTC approve U+111CE SHARADA VOWEL SIGN PRISHTHAMATRA E as proposed in L/17-214. In addition we recommend this character be added to the “Future Additions” document for WG2.

b. INVERTED CANDRABINDU for Sharada

Document: [L2/17-428](#) Proposal to encode the INVERTED CANDRABINDU for Sharada – Pandey

Comments: We reviewed this proposal for one character, SHARADA SIGN INVERTED CANDRABINDU. The proposal notes that Sharada is alone amongst the northern Brahmi scripts that has the conventional *candrabindu* as an inverted form of the usual shape (cf. (◌̣ U+11180 SHARADA SIGN CANDRABINDU vs. ◌̣ U+0900 DEVANAGARI SIGN CANDRABINDU). The proposal includes several examples from manuscripts showing the inverted and the conventional shapes together in the same document, and states they “are not glyphic variants but alternate pairs whose concurrent usage is intentional.”

While the two forms of *candrabindu* are indeed distinctive, the “intentional use” of the two forms raised some questions. In order to establish whether one form may be a stylistic variant of the other, we recommend the author provide more explanation and information, including:

- Why are the two forms different?
- Can the two be interchanged freely in transcription? Or is there an orthographic rule specifying which should be used (for example, they are used in alternation)? Is it important to scholars to distinguish the two forms?

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

18. Syloti Nagri

Document: [L2/17-418](#) Encoding model to represent conjuncts in Syloti Nagri – Srinidhi and Sridatta

Background document: [L2/05-130](#) Encoding Model for Syloti Nagri Conjoining Behaviour - Constable

Comments: We reviewed this document, which reconsiders the model for rare cross-cluster ligatures and false conjuncts in Syloti Nagri. The authors recommend the UTC (1) discuss the plain-text representation of cross-cluster ligatures and false conjuncts and make changes if necessary to §15.1 Syloti Nagri in the Core Spec, and (2) consider a change in the Indic Syllabic category for U+A806 SYLOTI NAGRI SIGN HASANTA, from Pure_Killer to Virama.

Of the various cases cited in Table 1 and 2, of particular interest are the following:

Table 1 (with Burmese model, with virama)

Case	Example sequence	Conjoining display with specialty font	Bad non-conjoining display with basic font	Correct non-conjoining display
“false” conjunct” C + C	< षा , virama, ञ > (“kot”)	षा	षाञ	षाञ
“false conjunct” with spacing vowel mark	< ब, ी, ब, virama, न, ी > (“bibir” – same sequence as “bibri”)	बीब्री	बीब्री	बीबीन

Table 2 (with ZWJ model)

Case	Example sequence	Conjoining display with specialty font	Conjoining display with basic font
“false” conjunct” C + C	< षा , ZWJ, ञ > (“kot”)	षा	षाञ
“false conjunct” with spacing vowel mark	< ब, ी, ब, ZWJ, ी, न > (“bibir” – distinct sequence from “bibri”)	बीब्री	बीबीन

The following points were raised during discussion:

- The UTC had earlier agreed with Peter Constable’s analysis in [L2/05-130](#), i.e., the model for the script is one with a virama, with no conjoining behavior (like Burmese).
- For special forms, such as rare cross-cluster ligatures or false conjuncts, [L2/05-130](#) recommended OpenType features or ZWJ. The only way to be able display such special conjoining behavior today would be with joiners.
- As noted in the document, the cross-cluster ligatures or false conjuncts are rare, and occur in handwritten documents, but not in modern printed sources. The authors feel use of ZWJ should

not be used in these contexts, but they should only be handled at the font level with OT features.

- Are all the cited conjoining forms orthographically significant?
- Two approaches to the problem could be to either handle this behavior in higher-level protocol or add a new conjunct-forming letter.
- Are conjuncts more common than ligatures? If the script ligates often, then a stacker may not be necessary.

In our view, the present Indic syllabic category, Pure_Killer, is acceptable. If there is a need for representing either conjuncts or false conjuncts (i.e., intrasyllabic ligatures) in plain text, it should be demonstrated. If such a case is justified, then a different solution should be considered, such as a new stacker character.

Recommendations: We recommend the UTC review this proposal, and discuss it with Peter Constable, the author of L2/05-130.

19. Tigalari

Document: [L2/17-378](#) Preliminary proposal to encode Tigalari script -- Vaishnavi Murthy K Y, Vinodh Rajan

Background documents:

[L2/17-182](#) Comments on encoding the Tigalari script – Srinidhi and Sridatta

Related documents:

[L2/17-411](#) Letter in support of preliminary proposal to encode Tigalari - Guru Prasad

[L2/17-422](#) Letter to Vaishnavi Murthy in support of Tigalari encoding proposal - A. V. Nagasampige


[Note: Information has been received that a competing proposal will be submitted.]




Comments: We reviewed this proposal, which addresses some of the points raised by Srinidhi and Sridatta in L2/17-182 and the script ad hoc comments in L2/16-342. The proposal now includes properties.

The following comments were made in a review of this proposal:

- In order to provide a track record of Tigalari proposals, we recommend the authors cite earlier versions of this proposal at the top of the document (on page 1), i.e., “[This document] replaces L2/16-241”. It would also be useful to cite L2/17-182 and the ad hoc comments L2/16-342 within the proposal (such as in §4), so comments from other documents can also be tracked.
- Currently Tigalari is only allocated 6 columns on the [Roadmap](#), but the proposal has 8. The current proposal appears to be based on the Malayalam block, which was itself based on ISCII. A better use of space would be to move the digits left one column, put the letters for vocalic rr and ll and vowel signs for vocalic l and ll, in the usual Sanskrit order (or, for any Dravidian-specific letters, in the appropriate Dravidian relative order).
- §5.1. Include a chart in the body of the proposal showing the representation of independent vowels (which appear to have distinct graphic pieces), similar to the one for Malayalam (in place of a reference in footnote 14):

For	Use	Do Not Use
ഇറ	0D08	<0D07, 0D57>
ഉറ	0D0A	<0D09, 0D57>
ഐ	0D10	<0D0E, 0D46>
ഓ	0D13	<0D12, 0D3E>
ഔ	0D14	<0D12, 0D57>

- §5.1 (pages 6-7) LETTER E and LETTER O:
 - Do research to clarify whether the proposed *e/o* are used currently, and appear in educational materials. If they are in use, then they are eligible for encoding. If not currently being used, then they can be added later.
 - Create a table to show how *e* and *o* appear in different documents in figures 42, 43, and 44, and provide examples of their use in texts by those authors.
- §5.2 (page 8): In the section directly under “For the TIGALARI AI LENGTH MARK...U+113C8 should be used” the text should read “EE-vs + EE-vs” (not E-vs)
Similarly in §9 on page 29, on the top, E-vs should be corrected to EE-vs.
- §5.2 (page 8): The TIGALARI AU LENGTH MARK is an acceptable character to encode; it is a useful graphetic unit for the writing system. However, it should be moved to a position at the end of the range of other dependent vowels
- §5.2 (page 9): The proposal states that the vowel signs U and UU change shape, depending upon the consonant to which they combine. Provide a comprehensive list of the consonant and vowel combinations that attach graphically, noting any exceptions. Provide this information for all the dependent vowels. (Examples appear to be provided on p. 7 of L2/17-182)
- §5.2 (page 9): For the alternate form of the vowel sign U , describe when it occurs. Does it appear consistently in the same manuscript? Is the form interchangeable with other forms of U? This information can help identify whether it is best handled on the font-level (as a font style, for example), or whether it needs to be distinguished in plain text, and hence separately encoded.
- §5.2 (page 10): Under “2. Placed below and right ligating”, the text reads “While...LA + Virama combinations can be used here, it is recommended not to as the characters are not canonically equal.” The wording “not canonically equal” here is unclear; can the authors explain? L2/17-182 states the combination is a conjunct. We recommend the authors incorporate the information from L2/17-182, page 9 (sections 2 and 3)
- §5.4 Virama (page 12ff): For many historical scripts, separate viramas for a killer and a stacker may be advisable, in order to avoid over-reliance on joiners and non-joiners, which can be very confusing for users. It should be noted that using Malayalam as a model is not necessarily advisable, since Malayalam is a different orthography. We recommend two viramas be proposed: a vowel-silencing virama (“killer”, with a visible mark) and a stacker (with no visible mark).
- §5.4 (pages 14-15): The ligature forms of *k*, *t*, *tt*, and *n* resemble Malayalam *chillu*, though the model is simpler than for Malayalam. We recommend separately encoding 4 *chillu* forms.
- §5.5 (page 16, top): Remove the paragraph “The Malayalam model..” with the two lines of examples. If the authors feel strongly that the alternate forms should be represented, demonstrate that the distinction is important, and provide use cases. (Note: This topic could be proposed separately at a later point, with the authors’ suggestions on how to represent the forms in text [i.e., by a stacking virama and a joiner or non-joiner].)

- §5.5 (page 18) 4. Ligating Special Characters: We would recommend a Tigalari *dot reph* be separately encoded (see below, §8.5).

- §5.5 (page 19, top): Represent IRK as <*dot reph, ka, killer*>.
- §6.1 (p. 20): Alternate Glyph Shapes. The third form of Tigalari Letter II, °O°, is separately encoded in Malayalam (U+0D5F MALAYALAM LETTER ARCHAIC II), based on L2/12-225. Consider whether it should be separately encoded (or not).
- §6.1 (p. 21): The shapes of the consonants and alternates in the list on page 21 seem to vary considerably, e.g., JHA  . Could they different be characters? Provide the rationale why the two JHA forms are not proposed as two characters (though pronounced similarly). Provide better discussion for other consonants which have vastly different shapes. Incorporate information from L2/17-182 (pp. 2-4), as applicable.
- §7 Digits (p.22): Is the numbering system as represented in figure 22 (and the chart on p. 22) rare, or is it used widely in historical documents? According to the authors of L2/17-182, the manuscripts in figures 20 and 21 seem to be in the Malayalam script. Is the numbering system as proposed used in schools today? If not, we recommend the authors hold off proposing them until more information on them can be provided.
- §8 (p. 23) *Candra Anunaasaika*:
 - The name *Candra Anunaasaika* varies from the more usual Indic pattern of naming. We recommend it be spelled *anunasaika* (cf. U+0901 with the informative alias *anunasaika*). Is the character ever referred to as *candrabindu*?
 - In order to determine whether the mark is spacing or is more like TELUGU SIGN CANDRABINDU (a right-side mark), demonstrate whether the Tigalari sign functions as part of the syllable or is a spacing mark between syllables.
- §8.2 (p. 24): Add the names and code points to the “Double danda” reference (i.e., U+0965 DEVANAGARI DOUBLE DANDA), if no script-specific dandas are proposed.
- §8.2 (p. 24-25): We suggest the authors put the three characters, OM ALANKAARA, SHRII ALANKAARKA (*shrii* symbol), and PUSHPA ALANKAARA into a chart. In accompanying text, provide justification for PUSHPA ALANKAARA (as the text currently mentions U+2055 – or is that only meant to be a tip to font creators?). The two characters OM ALANKAARA and SHRII ALANKAARKA (*shrii* symbol) seem to be good candidates for encoding.
- §8.3 (p. 25 and figure 25): *Tiddu* mark (Correction mark) is not justified, and we recommend the authors remove this section. The mark is described as an editorial convention used to insert missing text, or to mark imperfections in the text. There are no guidelines as to the mark’s placement, and the *tiddu* was often added later to texts. If a mark is required, the DEVANAGARI CARET at U+A8FA could be employed, or markup could be used. A *de novo* creation of a standardized editorial convention is not the realm of plain text, in our view.
- §8.4 (p. 26) Vedic tone marks:
 - Although Vedic tone marks are not proposed here, we encourage the authors to consider whether existing Vedic marks could be used, since they may appear in related manuscripts in the same area of India.
 - Mark the tone marks in the figure 26.
- §8.5 (p. 26) *Reph*: In order to represent the vertical tick at the beginning of a syllable, we recommend a *dot reph* be separately encoded, which avoids use of joiners/non-joiners. A *stacker* character should be used to represent the post-consonant or sub-based form. Hence, the top example would use a *dot reph*, but a *stacker* character would be used to represent the

bottom example:

ॠ	=	ॠ	+	◌̣	+	ॠ
RVA		RA		virama		VA
ॠ	=	ॠ	+	◌̣	+	ॠ
RVA		RA		virama		VA

- §8.7 (p. 28): If script-specific *dandas* are not proposed, add a firm statement that users should employ Devanagari *dandas*.
- §9 (p. 29): Fix the typos for E-vs in the second line (needs a dotted circle), and in the fifth line, correct EE-vs to E-vs.
- §10.1 (p. 30) Collation: Remove *tiddu*, add 4 *chillus*, *dot reph*, *stacker* and *killer* characters. Note: LLA appears twice (correct one to LLLA).
- §10.2 (p. 30ff): OM ALANKAARA should be given the General_Category value Lo (as in Devanagari). (As above, remove *tiddu*, add 4 *chillus*, *dot reph*, *stacker* and *killer* characters. Incorporate other suggested moves as noted.)

Recommendations: We recommend the UTC forward the above comments to the authors. We also suggest the authors work with the authors of L2/17-182 (Srinidhi and Sridatta).

MIDDLE EAST

20. Arabic

a. Three TANWEEN Characters

Document: [L2/17-377](#) Proposal for the encoding of three Arabic tanween characters – Muller 1924 text by EM: used text by King F and Oman

Comments: This proposal requested three characters for representing the 1924 text of the Quran, MEEM FATHATAN, MEEM DAMMATAN, and MEEM KASRATAN.

In general, Quranic text has been encoded graphically, with the exception of certain unifications. In cases where there is disagreement in the representation of *meen tanween*, such as by King Fahd Complex For The Printing Of The Holy Quran (Hafs' transmission of Asim's reading), it would be useful to document it.

In our opinion, the characters are already representable with Unicode characters, hence we do not feel encoding new characters will solve the problem. We do agree with the author that the rules for representing the text should be very clearly (and thoroughly) documented, saying this is what Unicode has settled on for Cairo 1924 text.


Recommendation: We recommend the UTC review this document and request the author provide additional documentation for the *meen* above and *meem* below.

b. HAMZA

Document: [L2/18-028](#) More precisions about Hamza – Azzeddine Lazrek





Comments: We reviewed this document, which responded to the script ad hoc recommendations contained in [L2/17-384](#). (In the following, the indented text indicates the script ad hoc responses to comments in [L2/18-028](#).)

Item #1 on  ARABIC HIGH WASLA





Ad hoc comments: We agree ARABIC HIGH WASLA is not the same as the  ARABIC SMALL HIGH WORD SAH character. While the author may not need a pedagogical spacing mark WASLA, others may, and one member of the script ad hoc has seen it in text. We still consider this character eligible for encoding, but an example is required.


The table accompanying item #1 suggests the WASLA be used to represent decompositions with WASLA. Because the model for Arabic has already been established, this approach is no longer feasible, as it would break existing documents and implementations.







Comments on bulleted points:



- The variant  must be encoded by the sequence: | U+0627 and  0654, but not by:  U+0623, even if it exists. It will be recommended to not use  U+0623 anymore.

Ad hoc comments: Either U+0623 or <U+0627, U+0654> can be used today, as they are canonically equivalent. In either normalization form NFC or NFD, the two would always be equivalent.

- The variant  must be encoded, by the sequence: | U+0627 and , but not by:  U+0671, even if it exists. It will be recommended to not use  U+0671 anymore.

Ad hoc comments: We would not recommend decomposing  as the two elements only occur in this combination. (Note that U+0671 has no canonical decomposition in the Arabic names list: <http://www.unicode.org/charts/PDF/U0600.pdf>.) It was noted that Syriac has the same combining character, but it too only occurs over ALAPH. The Syriac character has not yet been encoded, but when it is, it will be encoded as a precomposed character.

- The variant  must be encoded by the sequence: | U+0627 and  and , but not only by one code.
- The variant  must be encoded by the sequence: | U+0627 and  and , but not only by one code, as it is proposed by Mussa Abudena L2/15-329 and L2/16-044.


Ad hoc comments: We agree that both  and  need to be encoded, but not as an ALEF with combining marks. The rule-of-thumb for Arabic is that when elements touch, they should be encoded as a single, precomposed character.


By X, any research engine will be able to find any occurrence of the preceding word at all its situations.

Ad hoc comments: We recognize the desire to find all the different ways of writing words in the different editions of the Quran, but the proposed model – representing the text using different combining marks – will not work. Instead, searches should be tailored so they find the different forms. A well-constructed search protocol would be able to include the different characters in a single search, and find them all.


[Note: The character numbers in the left-hand column below are from [L2/17-252](#).]


- Characters #2 and #10 have similar shape with variant sizes (the first one is smaller than the second one) but differ semantically

2	08C1		ARABIC SMALL DOT ABOVE • Starting Hamza Wasl with Fatha
---	------	---	--


10	08C9		ARABIC BIG DOT ABOVE • Facilitate Hamza with Fatha • Used for Replace Hamza with Fatha
----	------	---	--


- Characters #3 and #11 have similar shape with variant sizes (the first one is smaller than the second one) but differ semantically.

3	08C2		ARABIC SMALL DOT LEFT • Starting Hamza Wasl with Damma → 302E • ○ Hangul single dot tone mark
---	------	---	---

11	08CA		ARABIC BIG DOT LEFT • Facilitate Hamza with Damma → 302E • ○ Hangul single dot tone mark
----	------	---	--


- Characters #4 and #13 have similar shape with variant sizes (the first one is smaller than the second one) but differ semantically.


4	08C3		ARABIC SMALL DOT BELOW • Starting Hamza Wasl with Kasra
---	------	---	--

13	08CC		ARABIC BIG DOT BELOW • Facilitate Hamza with Kasra • Used also for Replace Hamza with Kasra
----	------	---	---


Ad hoc comments: We are uncertain whether the “big dot” characters (#10, #11, #13) need to be encoded. Do #2/#10, #3/#11, and #4/#13 appear in the same *mushaf* or do they appear in different *mushaf* in different sizes?


- Characters #7 and #16 have similar shape with variant sizes (the first one is smaller than the second one) but differ semantically.

7	08C6		ARABIC SMALL STROKE ABOVE • Used for Hamza Wasl after a letter with Fatha
---	------	---	--







16	08CF		ARABIC BIG STROKE ABOVE • Transport Hamza with origin Fatha • Used also for Replace Hamza with Fatha
----	------	---	--

- Characters #8 and #18 have similar shape with variant sizes (the first one is smaller than the second one) but differ semantically.

8	08C7		ARABIC SMALL STROKE RIGHT • Used for Hamza Wasl after a letter with Damma
---	------	---	--



18	08D1		ARABIC BIG STROKE RIGHT • Transport Hamza with origin Damma in some printed Mushaf(s)
----	------	---	--

- Characters #9 and #19 have similar shape with variant sizes (the first one is smaller than the second one) but differ semantically.

9	08C8		ARABIC SMALL STROKE BELOW • Used for Hamza Wasl after a letter with Kasra → U+061A  Arabic small Kasra → U+0650  Arabic Kasra
19	08D2		ARABIC BIG STROKE BELOW • Transport Hamza with origin Kasra • Used also for Replace Hamza with Kasra → U+061A  Arabic small Kasra → U+0650  Arabic Kasra

Ad hoc comments: In our view, #7 and #16 are probably one character, and similarly #8/#18 and #9/#19. Do the pairs of big/small stroke characters appear in the same *mushaf*?


- Characters #17 and #18 have same semantic with different positions, we find both in practice.

17	08D0		ARABIC BIG STROKE LEFT • Transport Hamza with origin Damma • Used also for Replace Hamza with Damma
18	08D1		ARABIC BIG STROKE RIGHT • Transport Hamza with origin Damma in some printed Mushaf(s)

Ad hoc comments: In our view, #17 and #18 are different characters.

Character #5  **ARABIC SMALL RING LEFT** and Character #6  **ARABIC SMALL RING BELOW**
 What is the character in the figure below (from L2/17-252)? Is it #6?



Characters #6  and the one used in the figure above have similar shape with variant sizes (the first one is smaller than the second one) but differ semantically.

The first one is used for Starting Hamza Wasl with Kasra (find at some Mushaf Qaloon). However, the second one is used for Al-Imala (find at some Mushaf).

The first one is used only under Alef. However, the second one is used under some other letters

Ad hoc comments: Provide more examples of how the small ring below is used for Al-Imala. What does Al-Imala mean?

11	08CA	•◌	ARABIC BIG DOT LEFT • Facilitate Hamza with Damma → 302E •◌ Hangul single dot tone mark
12	08CB	◌•	ARABIC BIG DOT RIGHT • Facilitate Hamza with Damma used at some Mushaf(s) Qaloon → 08CA •◌ ARABIC BIG DOT LEFT

Additional ad hoc comments: Regarding #11 and #12, only one of these character needs to be encoded, but it should be a spacing character. The distinction would then be provided by putting it before or after the *alef*.

3	08C2	◌◌	ARABIC SMALL DOT LEFT • Starting Hamza Wasl with Damma → 302E •◌ Hangul single dot tone mark
5	08C4	◌◌◌	ARABIC SMALL RING LEFT • Starting Hamza Wasl with Damma used at some Mushaf Qaloon → 08C2 •◌ ARABIC SMALL DOT LEFT

Additional ad hoc comments: We have previously disunified hollow vs. filled Koranic characters, so it appears that these two are also separate characters, although they should be encoded as spacing, like #11/#12. But it's not clear if #3 is the same as #11/#12 or not.

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

AMERICAS

21. Mayan

[L2/18-038](#) A preliminary proposal for encoding Mayan Hieroglyphic Text in Unicode – Pallan
[Note: The script ad hoc reviewed an earlier version of the proposal currently posted in the document registry.]

Comments: We reviewed a draft preliminary proposal, which draws on work done by Carlos Pallan done in 2017 (and earlier, from the MAAYA project).

The following were comments raised during discussion:

- List the quadrat types in the Appendix in order of complexity, so similar formations are next to one another.
- For glyphary and quadrat types in the Appendix, explain what the columns mean and the syntax. For example, what do capital vs. small letters indicate?
- List all the operators that are being proposed.
- Provide a full list of characters being proposed.
- Make sure the syntax for quadrats is correct; cf. 4.27 (11V2)H))33H4))
- Will mappings to the various catalogs be provided?
- Mention the unification guidelines used
- Do signs come in pairs (some tend to be narrow or more block-y)?

Recommendations: We recommend the UTC review this document and forward any comments to the proposal author.

EAST ASIA

22. CJK

Document: [L2/17-429](#) Request to reserve the code point for square Japanese new era name (SC2 N4577) – Japan NB

Comments: We discussed this short document from the Japanese National Body, which requests a BMP code point be reserved for the new era name. The era name will be two-ideograph character with a compatibility decomposition, similar to the era names already encoded at U+337E, U+337D, U+337C, and U+337B. The new era will start on May 1, 2019.

We agree reserving a code point for the new era would be prudent.

Recommendation: We recommend the UTC reserve U+32FF in the Enclosed CJK Letters and Months block for the new era name. When the official decision from Japan is made, the slot will be available for encoding the new character, in time for Unicode 12.0.

SYMBOLS AND PUNCTUATION

23. Ancient Chinese Math Symbols


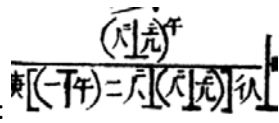
Document: [L2/17-219](#) Proposal to encode Ancient Chinese Mathematical Symbols (revised) - Kushim Jiang

Comments: We reviewed this revised proposal, and had the following comments:

- Provide the figure numbers beside the proposed characters.
- For COMBINING CHINESE PRIME and COMBINING CHINESE DOUBLE PRIME, use of U+0301 COMBINING ACUTE ACCENT and U+030B COMBINING DOUBLE ACUTE.
- For COMBINING CHINESE TRIPLE PRIME, a new character at U+20F1 in the Combining Diacritical Marks for Symbols block would be appropriate.
- Remove COMBINING ENCLOSING CUT-CORNER RECTANGLE. This mark is an editorial convention appearing in a math book. Such a use may be better handled via other means of representation, such as markup. In addition, getting such a mark to work could be difficult for implementers. Is this mark used outside math books? If two neighboring characters were mistakes, would the box enclose both characters? More information is needed.

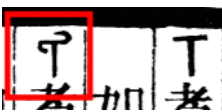
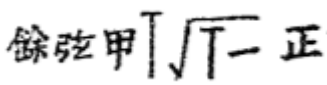


- The CHINESE PLUS SIGN is well-justified by numerous examples. The representative glyph should have the horizontal shortened:

Current glyph:  Example from figure 29: 

Like the other operators (MINUS SIGN, DIFFERENTIAL SIGN, and INTEGRAL SIGN), it is larger than surrounding characters (see figure 29, above). As with the other large math operators, the Unicode Standard does not specify a particular layout, but implementations are expected to follow accepted typographical conventions for layout (see 3.2.3. Large Operators in [UTR 25](#)).

- CHINESE MINUS SIGN similarly appears generally larger than surrounding characters, but in figure 26 it is the same size as POSITIVE DIFFERENCE SIGN, cf.

figure 26  vs. figure 15 

Is this a pair of characters, or just one? Or is the sizing due to style of the headings in this particular work?

As with PLUS SIGN, the horizontal stroke in the representative glyph should be made shorter, so it matches the typical shape.

- The CHINESE PLUS-OR-MINUS SIGN, another large operator, is justified, based on figure 24:




- CHINESE VARIANT PLUS SIGN and CHINESE VARIANT MINUS SIGN require more evidence. The evidence provided is in figure 25 (below). However, the book providing the justification seems to contain a different typographical style than the rest of the text. Provide more explanation on its use. Unless a systematic distinction can be made between the variants and the MINUS SIGN and PLUS SIGN, we recommend they be removed.



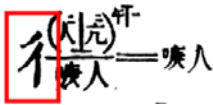
- The CHINESE POSITIVE DIFFERENCE SIGN is shown on a heading in figure 27. The typed example on page 3 (below) suggests it is not a large operator – is that true?

$$5 \text{ 差 } 3 = 3 \text{ 差 } 5 = 2.$$

- For CHINESE FACTORIAL SIGN  we recommend a new character at U+20F2 with the proposed name COMBINING BOTTOM RIGHT CORNER (cf. 23CD/E/F), and an annotation, “used in Chinese.” The mirror image has been used in Western European math literature also for factorial (i.e., $\lfloor n$, 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). (This alternate sign has not been proposed.)

- CHINESE LEFT CURLY BRACKET and CHINESE RIGHT CURLY BRACKET both require more examples showing their usage, beyond what is shown in figure 27. Until more evidence is provided, we recommend these be removed.
- CHINESE ELLIPSIS is shown in figure 27, which is a dictionary entry, but an example showing it in actual usage is required. Remove this character until more examples are provided.
- The CHINESE DIFFERENTIAL SIGN and CHINESE INTEGRAL SIGN appear to be acceptable. It was also noted that these two operators are larger than the neighboring characters. Cf. figure 29:



- This revision of the proposal includes the atomic (non-decomposed) circled symbols for inclusion in the Enclosed Ideographic Supplement block., which is the correct approach in our view.
- We recommend the whole set of Celestial Stems and 12 branches be proposed. Eight of the ten Celestial Stem characters are proposed here (U+1F252..U+1F259) but the listing should show U+1F25A and U+1F25B as “reserved” for the remaining two Celestial Stem characters (for “9” and “10”), when evidence is provided. The 12 branches are represented by U+1F270..1F27B, but a spot at U+1F27A should be reserved, when an example of the circled ideograph 戌 is found.
- The other enclosed operators will require further investigation.

In sum, the following characters were deemed acceptable candidates:

- U+20F1 COMBINING CHINESE TRIPLE PRIME
- U+20F2 COMBINING BOTTOM RIGHT CORNER
- CHINESE PLUS SIGN (with modification of glyph)
- CHINESE MINUS SIGN (with modification of glyph)
- CHINESE PLUS-OR-MINUS SIGN
- CHINESE FACTORIAL SIGN
- CHINESE DIFFERENTIAL SIGN
- CHINESE INTEGRAL SIGN

Recommendation: We recommend the UTC review this document, send feedback to the author, and recommend the proposal be revised to just include the eligible characters (with evidence). For the characters deemed eligible, the UTC should consider whether CHINESE is appropriate in the characters’ names.

24. Fullwidth East Asian Punctuation

Document: [L2/17-436](#) Proposal to add standardized variation sequences for fullwidth East Asian punctuation – Lunde

[Note: The ad hoc did not review [L2/18-013](#) Proposal to add standardized variation sequences for digits and various punctuation, which is the second part of [L2/17-436](#)]

Comments: We reviewed this document, which proposes standardized variation sequences for eight fullwidth punctuation characters. As the proposal states, single-language or single-region fonts can make do with one form of punctuation (i.e., corner-justified or centered), but Pan-CJK fonts need to include both forms to support multiple typographic conventions. Because variation sequences are better supported than language-tagging, and “plain text” environments persist, a plain-text solution with variation sequences is sought. In our view, the proposal is well-structured and provides good justification.

The following comments were raised during discussion:

- Show vertical examples of corner cases and specify what happens in such cases. Do they always move to the top right? Can they be left-justified?
- Mention the option of handling the fullwidth punctuation as halfwidth and space (and explain why this isn't a good option, since the space would change segmentation).

Recommendations: We recommend the UTC discuss this document. If the UTC approves the sequences, we recommend the author and the Editorial Committee take an action to update UTR #50 by adding the sequences and specifying how each sequence with VS would appear in horizontal and vertical text (with examples).

25. Compatibility symbols and punctuation used in DPRK

Document: [L2/18-004](#) Proposal to reconsider compatibility symbols and punctuation used in the DPRK - Marin Silva

Note: The following related documents were not reviewed by the script ad hoc:

[L2/18-011](#) Information on the most recent version of KPS 9566 (KPS 9566-2011?) – J Chung

[L2/18-014](#) Feedback on L2/18-004 (DPRK symbols) – Eiso Chan

[L2/18-017](#) Update on enclosed postal mark – Marin Silva

Comments: We reviewed this document, which discusses a number of DPRK characters that had been proposed in [L2/01-349](#) (drawing also on comments from [L2/02-102](#)). According to one member of the script ad hoc group, the characters that were not encoded were deliberate omissions. It was also noted that there has been no request from vendors to support these characters in 15+ years. If the author can provide the vendor requirement to include these symbols, and make a case for them for conversion or interchangeability, then a full proposal would be welcome. Note: There are historical North Korean *jamos* that still need to be proposed.

Recommendation: We recommend the UTC take no action on this document, until a vendor request and/or interchange requirement can be shown.


26. Legacy characters from computers and teletext

Document: [L2/17-435](#) Proposal to add characters from legacy computers and teletext to the UCS – Ewell et al.

Comments: We reviewed this proposal, which requested 207 new characters and 198 variation sequences to support legacy characters used by home computers that date from mid-1970s to mid-1980s, and the teletext standard that was developed in the 1970s.

The proposed repertoire derives from several different character sets. To strengthen the proposal, additional information for the characters should be provided, so the characters are sourced, have mappings, and appear as a coherent set of characters.

The following summarizes the comments raised during discussion:

- While the block sextant characters seem reasonable, we recommend the holes be closed up, and the cross-references (currently where the reserved spaces are) be put at the top of the names list.
- Provide data tables with sources for all the characters (preferably other than Wikipedia). Data tables should also identify the character sets and machines represented by this repertoire, and should include mapping tables to the character sets, identifying characters that are already in Unicode, and those that are missing.
- Identify who is using these characters. Are the users asking for these characters?
- Why are some of the characters from the figures being proposed, and not others? Cf. figure 4  which is not proposed.
- §7.a. Variation sequences – Reverse Video
 - Show the need for plain text representation for reverse video.
 - “This mechanism is intended specifically for round-tripping” What is being round-tripped?
- 7.b. Variation sequences – Styled digits
 - Make a strong case for the use of VSeS for styled digits.

Recommendations: We recommend the UTC review this proposal and send comments to the authors.

OTHER TOPICS

27. Script Extensions

Document: [L2/17-430](#) Script_Extension values for some combining marks – Davis

Comments: We reviewed this document, which proposes additions to the Script_Extension property values for Unicode 11.0. The details contained on pages 3-9 provide information that is valuable in being able to check the proposed changes.

Recommendation: We recommend the UTC discuss this document, and consider recommending a different format that is more easily maintainable.

Proposals and document not discussed (or mentioned above)

[L2/17-242](#) Proposal to add CC license symbols to UCS [Creative Commons] (rev) – Park

[L2/18-012](#) Proposal of Four IDCs (IRG N 2273; replaces L2/17-386)

[L2/18-015](#) Proposal to encode the END OF TEXT MARK for Malayalam – Srinidhi and Sridatta

[L2/18-020](#) Proposal to define Standardized Variation Sequences for BOPOMOFO LETTER I – Chan and Wei

[L2/18-021](#) Proposal to add informative notes to U+311D, U+3120, and U+3129