TO: **UTC**
FROM: **Deborah Anderson, Ken Whistler, Roozbeh Pournader, Lisa Moore, Liang Hai, Chris Chapman, Richard Cook, Norbert Lindenberg, and Ben Yang**[1]
SUBJECT: **Recommendations to UTC #156 July 2018 on Script Proposals**
DATE: **20 July 2018**

The Script Ad Hoc group met on 1 June and 13 July 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. Because the number of scripts is long, a table of contents is provided below.

[1] Also participating at the June 1, 2018 meeting: Marek Jeziorek, Peter Constable, Michel Suignard, and Yifán Wáng

EUROPE

1. Latin

a. Latin Letters with Overcurl

Document: L2/17-358 Proposal to add one combining character for Medieval Cornish – Everson et al.

WG2 Recommendations (L2/18-209): Ten Latin small letters with overcurl were included in the CD.

Comments in Consent Docket (L2/18-217):

> The original proposal N4907 (L2/17-358) by Michael Everson requested a combining diacritic for the overcurl in Medieval Cornish. WG2 discussion focused on two points: (i) whether base letters with this overcurl should be deemed distinct text elements, or merely presentation forms of existing base-diacritic combinations; and (ii) if distinct encoding were to be approved, whether a productive combining-mark or specific, atomic characters would be preferable. WG2 recommended (and SC2 accepted) the addition of 10 atomic characters to the CD, in the Latin Extended-D block.

Comments from Script Ad Hoc: We reviewed this proposal and the Consent Docket comments. In general, we consider the atomic character option to be an improvement over a combining diacritic, since implementing the combining overcurl is difficult in a font. However, it is not clear that the proposed letters are true characters; they may be stylistic variants or ligatures of a small Latin letter with an inverted breve. In our view, representing the details of hands in manuscripts is better handled by another layer, such as markup or OpenType features (such as stylistic sets).

The proposal needs to be updated to reflect the ten new atomic characters included in the CD added repertoire (L2/18-221). Where there is an orthographic distinction between a letter and a letter with overcurl, evidence of meaningful orthographic distinction between letter with breve and letter with overcurl must be provided, in order to justify that these are not simply ligatures of the letter and inverted breve.

Recommendations: We recommend the UTC discuss this topic, and consider making a ballot comment for the CD.

b. Latin Subscripts, Superscripts, and Small Caps

Document: L2/18-206 1120 more superscripts, subscripts and small capitals – Grochowski

Comments: We reviewed this document, which requests adding 1,120 superscripts, subscripts, and small capitals. In our opinion, adding 1,120 such characters is not a good idea architecturally. A full proposal with orthographic evidence for the specific characters could, however, be considered, if the author provided such a document.

Recommendations: We recommend the UTC note this document, relaying to the author that if orthographic evidence were provided, the request could be re-considered (although a full proposal is required).

c. Latin Letter Thorn with Diagonal Stroke

Documents:

L2/17-236 Proposal to add LATIN LETTER THORN WITH DIAGONAL STROKE, revised – West and Everson

L2/18-242 Feedback on LATIN LETTER THORN WITH DIAGONAL STROKE – Anderson

WG2 Recommendations (L2/18-209): Two characters were added to the CD: A7C0 LATIN CAPITAL LETTER THORN WITH DIAGONAL STROKE and A7C1 LATIN SMALL LETTER THORN WITH DIAGONAL STROKE.

Comments in Consent Docket (L2/18-217)**:**

> The USNB ballot comments for PDAM 2.2 recommended removal of the two Latin thorns with diagonal stroke, since they were considered glyph variants of U+A764 LATIN CAPITAL LETTER THORN WITH STROKE and U+A765 LATIN SMALL LETTER THORN WITH STROKE that could be represented with OpenType features or different fonts. The revised proposal, N 4836R, includes early attestations of the character with a diagonal stroke (found in Anglo-Saxon works), and indicated that this form appeared to be the predominant shape, compared to horizontal stroke used in Nordic traditions.
>
> During discussion at WG2, the proposal authors acknowledged that thorns with diagonal versus horizontal stroke can be considered glyph variants, but contended that disunification was nonetheless warranted. Regarding the alternative — maintaining the current unified encoding — the question of what the representative glyph should be was also discussed: the horizontal stroke has been shown in charts since Unicode 5.1 (2008), and implemented in some commonly-used fonts. The diagonal stroke is more widely used, but a change of representative glyph could lead to changes in fonts that might impact existing documents and, hence, not be welcomed by Nordic scholars.

Comments from Script Ad Hoc: We reviewed the proposal, the Consent Docket comments, and the expert feedback. The evidence provided in the document with expert feedback, L2/18-242, strongly suggests that the user community considers these two shapes as glyph differences and is using the existing character for both shapes. Encoding a new character for one of the forms would be very disruptive to the use community, as a partial uptake of a new character could throw much of their searching and display processes into disarray.

Recommendations: We recommend the UTC discuss this topic and consider making a ballot comment on the CD.

2. Cypro-Minoan
Document: L2/16-179 Revised proposal to encode the Cypro-Minoan script in the SMP – Everson (L2/16-265 – old feedback)

WG2 Recommendations (L2/18-209): Cypro-Minoan was added to the CD.

Comments: We briefly reviewed the proposal, which dates to 2016. No new information has been provided. In our opinion, the proposal is not ready for standardization. We expect extensive feedback from the expert community on the proposal, including issues related to the character model and unification. We don't consider it productive to include the character chart in the CD as is, since we expect changes that would be too hard to do if the repertoire and the model needed to go through several rounds of feedback.

Recommendations: We recommend the UTC discuss Cypro-Minoan and consider a future ballot comment on the CD.

AFRICA

3. Egyptian Hieroglyphs

Document: L2/18-236 A note on the syntax of Ancient Egyptian hieroglyphic control characters – Nederhof

Comments: We reviewed this document which proposes an alternate syntax for the nine Egyptian hieroglyphic control characters that are expected to be published in Unicode 12.0. The document could be useful for implementers.

Recommendations: We recommend the UTC note this document as a FYI to implementers**.**

MIDDLE EAST

4. Arabic

Document: L2/18-061R Proposal to include Luri alphabets – Mogoi and Shaikh

Comments: We reviewed this proposal for three Arabic characters used to represent the Luri language of Iran. This revised proposal removed the characters that can already be represented with Arabic characters (see p. 18 of the April/May Script Ad Hoc recommendations, L2/18-168), and adds new examples.

While there is some hesitation about the characters being used outside a very narrow circle, Figures 3 and 4 demonstrate the characters have been published by different publishers. If the UTC decided to accept the characters, "ABOVE" should be removed from the names. The code points are acceptable.

Recommendation: We recommend the UTC discuss this proposal for the following three characters (with the revised names we recommend):
U+08C5 ARABIC LETTER ALEF WITH INVERTED V
U+08C6 ARABIC LETTER KEHEH WITH INVERTED V
U+08C7 ARABIC LETTER LAM WITH INVERTED V.
If accepted, we recommend the UTC request the authors to revise the proposal with the new names.

5. Yezidi

Document: L2/18-238 Preliminary proposal for encoding the Yezidi script in the SMP – Karaca et al.

Comments: We reviewed this proposal. It was noted that there is almost a one-to-one correspondence with Arabic, although the script is used to write Kurmanji (Northern Kurdish).

The following comments were made:

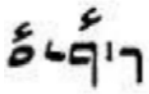
- Provide a better transcription system (cf. XHEYN for GHAIN)
- Provide details on the background of YOT and ET

1ED75 𐺵 YEZIDI LETTER YOT

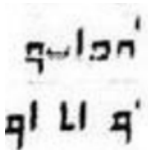
1ED76 𐺶 YEZIDI LETTER ET

- Clarify which letters are a modern invention and which are traditional, with the phonetic value of the newly created letters.
- Explain the differences between the characters in figure 8 and the current repertoire.

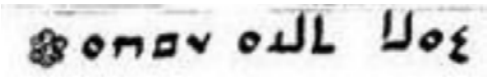
- Provide a chart with the values of the letters with IPA, Kurdish (in Latin script), and Arabic values
- Provide an explanation of the following, describing how they are used and their behavior: hamza, *madda*, hyphenation mark and the *alef-madda*.
    - Is *madda* a spacing or combining mark?
    - The hamza characters in fig. 1 (below) appears to be combining, but the hamza in the chart is not. Should the chart instead be a combining hamza?



- What are the small raised vertical marks below (from figure 2)?



- Give any information known about punctuation. For example, what is the following mark on the left? Does it appear only in this document? What does it indicate?



- Show the historical ligatures, circling them in the examples.

Recommendations: We recommend the UTC members review this preliminary proposal at their leisure and send comments to the Andrij Rovenchak.

SOUTH AND CENTRAL ASIA
6. General Indic* - Script Extensions - *Pushpika
Documents:
L2/17-424 Changes to ScriptExtensions.txt for Indic characters for Unicode 11.0 – Srinidhi and Sridatta
(*§1 Devanagari* A8F8 DEVANAGARI SIGN PUSHPIKA, pp. 1-2)
L2/18-232 Feedback on *pushpika* – Anderson

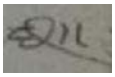
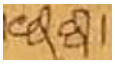
Comments: We reviewed the documents on *pushpika*. In §1 Devanagari of L2/17-424, the authors request Newa and Kannada be added to the set of scripts in the ScriptExtensions property for U+A8F8 DEVANAGARI SIGN PUSHPIKA, and when Nandinagari and Tigalari are encoded, similar adjustments be made the set 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 an explicit property value for Script_Extensions. 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, although the *pushpika* is a rather marginal sign, compared to the dandas.

What is the general approach going forward? Should a *pushpika* be encoded on a per-script basis, or should one *pushpika* to be used across the scripts (using fonts to display the preferred shape)? Or should the approach be based on the grouping of closely related scripts (i.e., one *pushpika* for the northern scripts and one for the southern scripts)?

Compare:

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

Southern scripts: Kannada  Tigalari 

Note a second structure common in the southern scripts:

Tigalari (L2/17-378) has the proposed shape  based on figure 23 

Malayalam (L2/18-015) has the proposed shape  based on examples as:  

In the feedback document on *pushpika* (L2/18-232), Sharma recommends the *pushpika* be unified across Indic scripts "where there is an identifiable unity in the glyph skeleton", and does not feel the north-south dichotomy is necessarily relevant here.

We agree with Sharma that unification for this rather marginal sign may be warranted, unless the signs are structurally different. Unification of similarly shaped signs is not likely to compromise readability. In order to come to a decision on unification of the *pushpika*s across Indic scripts, we recommend an expert (or two) collect examples of *pushpikas* for review, drawing on samples from different scripts as well as samples from the same script.

Recommendations: We recommend the UTC invite experts (such as Srinidhi and Sridatta) collect *pushpikas* from Indic scripts and provide suggestions for unification. Based on this evidence, the UTC can consider making relevant changes to the ScriptExtensions property as needed.

7. Assamese

Documents:
L2/18-181 Proposal for inclusion of Assamese Script – India/BIS
L2/18-205 Assamese Ad Hoc report – Everson

WG2 Recommendations (L2/18-209):

After consideration of the proposal in document N4947 [L2/18-181] to encode the Assamese script, WG2 recommends that SC2 accept the ad hoc report on Assamese script in document N4999 [L2/18-205], which has the following main recommendations:
a. Add Assamese character names in the nameslist as annotations,
b. Change the block header from Bengali to Bengali-Assamese, and,
c. Prepare a revised contribution on new characters to be added.

WG2 encourages the experts on Assamese script to continue the work towards a revised contribution and submit to WG2. WG2 recommends that SC2 invites the national body of India, BIS, to coordinate this effort.

Comments: We reviewed these documents and the WG2 recommendations. In our opinion, it would be useful for the UTC to more fully understand the technical issues.

Recommendations: We recommend the UTC discuss the topic and consider inviting Assamese experts to provide details to the UTC on the technical issues they face.

8. Chorasmian
Document: L2/18-164r2 Proposal to encode the Chorasmian script (revised) – Pandey

Comments: We reviewed this revised proposal for Chorasmian, which is nearing a mature state. The following comments were made.

General comments:

- Provide references to figures, so the examples can be reviewed.
- Specify in the text if "final" is word-final or cursive (/shaping) final.

aleph (p. 8)
1. Comment on wording in bold:

> Moreover, in both of these cases, the disconnected aleph triggers the rendering of the preceding letter using its final form. A form of final ـ that joins to the preceding letter does not exist in the available materials. Therefore***, it may be necessary*** to use a non-joiner control character before a final ALEPH after a dual-joining letter in order break the normal joining behavior and to render the preceding letter using its final or isolated form (see § 6.2).

The wording "it may be necessary" should be changed, as it is (a) either necessary to use a ZWNJ or (b) one should disunify, so there are two characters. If aleph is dual-joining, what is its final form?

2. The following are two options for handling aleph; the choice of which option depends upon scholars' and/or proposal author's view.

> (a) two characters (as in the version of the proposal dated May 31, 2018, https://www.unicode.org/L2/L2018/18164r-chorasmian.pdf); no ZWNJ required for isolated form

	Isolated	*Final*	*Medial*	*Initial*
aleph				

final aleph 
(* indicates an artificial form)

(b) one "aleph" character with a ZWNJ to render the isolated form

Isolated	*Final*	*Medial*	*Initial*
	 *		

(* indicates an artificial form)

3. In the text, clarify that the final form is really an isolated form.

4. Regarding the following text:
"The stylistic form  of medial aleph occurs in some coins and silver vessels instead of  ; this form joins at the baseline on both sides. This is to be treated as a glyphic variant and controlled at the font level."

Does the form ever occur initially or word-initially? What happen when it is not connected to letter on its right? Is it a shaping or cursive initial, or is it an alternate that needs to be separately encoded?

5. Rework the section on *aleph* "Final" to clarify how system works.
Specific comments:

- The wording with "triggers" in the following is confusing: "Moreover, in both of these cases, the disconnected aleph ***triggers*** the rendering of the preceding letter using its final form."

  Suggested revision: "In these cases, the representation of text should contain ZWNJ so that preceding letter would render in a final form, and the isolated form of aleph would be chosen."

- Clarify if "final" is word-final or cursive (/shaping) final.
- Change "a non-joiner control character" to "ZWNJ".

small aleph (p. 8)
Clarify if "final aleph" is word-final or a cursive (/shaping) final, and note in the text that final small aleph is properly an isolated form.

he (p. 8)
Improve the language in the section. In our opinion, the glyph distinction between  and is a stylistic preference, that should not handled by the encoding directly. Remove the reference to the shaping engine.

zayin (p. 9)
The description of the joining behavior of *zayin* is clear, and should be used as a model for *aleph.*
Change "a non-joiner control character" to "ZWNJ".

mem (p. 9)
We feel the distinction being discussed in this section between non-word-final *mem* and word-final *mem* is a stylistic one, not covered by the encoding.

nun (p. 10)
Clarify the text by using “word-final” and “joining final” for the different shapes.

ayin (p. 10)
Change “initial position” to “word-initial position”.
We suggest the following bold text be modified slightly from “***it would be appropriate*** on palaeographical grounds to define the Chorasmian ayin as a right-joining letter” to “it may be suggested…”

pe (p. 10)
Change “a non-joiner control character” and “control character” (last line) to “ZWNJ”.

§6 Joining behavior (p. 11)
In the chart on the bottom of page 11, change the final form of *aleph* to  (with asterisk indicating an artificial form)

(p. 12)
Merge the second, special right-joining table back into the main right-joining table. The *he* and *mem* columns X_n and X_{nf} should be merged and whichever form decided secondary should be either removed or — if really necessary — put in parentheses. Note the general Arabic Cursive Joining algorithm is ignorant of word boundaries, therefore the contrast between X_n and X_{nf} is not directly relevant when talking about joining behavior.

§8.3 Shaping properties (p. 17)
Change all “N” to “U” (non-joining):
I.e, 10FB1; CHORASMIAN SMALL ALEPH; U; No_Joining_Group

9. Khitan Small Script
Documents:
L2/18-121R Cluster Formation Model for Khitan Small Script – West et al.
L2/18-213 KSS Ad Hoc Report – Anderson et al.
Note: the following feedback was not discussed:
Feedback: L2/18-214 Feedback on Additional repertoire for ISO/IEC 10646:2017 (5th ed.) beyond Amendment 2 (L2/18-211) – Marín Silva: The glyph for KHITAN SMALL SCRIPT FILLER should reflect its name, with the initials KSSF.
Background documents:
L2/18-210 Comments on WG2 #67 documents (June 2018) – Anderson et al.
L2/16-245 Final proposal to encode the Small Khitan Script in the SMP

WG2 Recommendations (L2/18-209): (1) Included Khitan Small Script in the CD, (2) accepted one script-specific format character, U+16FE4 KHITAN SMALL SCRIPT FILLER, in the Ideographic Symbols and Punctuation block and (3) changed the name of U+18B00 KHITAN SMALL SCRIPT ITERATION MARK to be KHITAN SMALL SCRIPT CHARACTER-18B00.

Comments in the Consent Docket (L2/18-217):

> The encoding model was the main concern of USNB ballot comments for PDAM 2.2, which prompted the removal of Khitan Small Script in PDAM 2.3. Alternate models were summarized in the WG2 discussion, including the options mentioned by Unicode script experts in N4977 (L2/18-210) along with an additional option [Tibetan-type model], which is mentioned in the Khitan Small Script ad hoc report, N5002 (L2/18-213).
>
> During an ad hoc, representatives of the user community, including experts from China, were favorable toward the model proposed in N4943 (L2/18-121R), which requires use of a single format control character. It was agreed that the behavior of this format control should be completely documented in Unicode, including occurrence in unexpected contexts. Also, the importance of assigning properties defined in UAX #50 was noted.

Comments from Script Ad Hoc: We reviewed the various documents and comments on Khitan Small Script.

The location of the ITERATION MARK in the main Khitan Small Script block is acceptable as long as there is no requirement that the character have any property other than Lo, and the character will require no special treatment in the nameslist. (In blocks where names are algorithmically generated, the characters should all have the same properties.) If the ITERATION MARK were to be moved to the Ideographic Symbols and Punctuation block, there would be more flexibility regarding the character's property and any nameslist annotations.

Before locking down the model for Khitan Small Script, it would be advisable to be sure the KSS experts in China are involved and provide examples of any "unexpected contexts" of KSS characters in order to document them in the revised proposal.

Recommendations: We recommend the UTC discuss this topic and relay the comments above to the proposal authors.

10. Mongolian

a. Comments on Mongolian code charts in DAM1

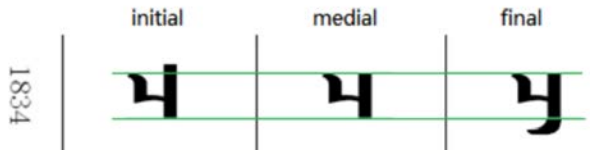
Document: L2/18-243 China's comments on Mongolian code charts in DAM1 – China

WG2 Recommendations (L2/18-209):

> WG2 recommends that SC2 notes the progress made by the Mongolian experts in documents N4989 and N4998, and invites the Chinese national body, working with Mongolian experts, to prepare and submit to WG2, a separate document containing only the proposed additions with rationale and phonetic information.

Comments: We briefly reviewed the document from China and the WG2 recommendations. The document L2/18-243 includes different categories of comments:

- Add new standardized sequences, such as the second form (final) of U+1824
- Replace the "?" in the chart with the nominal forms, such as in U+1829 first form (initial), or variant forms, such as U+1836 first form final (see comments on p. 27)
- Make glyph changes, such as in U+1824 first form (initial)/(medial)/(final), with details on p. 28:



- Delete certain standardized sequences, including rare or old-fashioned forms (i.e., U+1887 U+180D), or sequences where the glyphs are identical in shape, such as U+1889 first form (medial) and first form (final)
- Replace glyphs of certain forms "so they are easy to understand and use"

The Chinese NB reported they would be working on a separate proposal for the new standardized sequences and will include justification and phonetic information.

Recommendation: We recommend UTC members carefully review the requested changes in N4998.

b. **Document:** The Aesthetics of Mongolian Script in 800 Years – Badral

Comments: We reviewed this document which provides a useful reference document and might be worthwhile adding it to the Unicode document registry. (It currently resides in the Mongolian document registry.)

Recommendations: We recommend the UTC note this document.

11. Takri
Document: L2/18-084 Proposal to encode the TAKRI VOWEL SIGN VOCALIC R – Srinidhi and Sridatta
Additional document: L2/18-247 Additional information on L2/18-084 TAKRI VOWEL SIGN VOCALIC R – Anderson
Note: The document "Additional information" was not reviewed by the Script Ad Hoc.

Comments: We reviewed this proposal for one Takri character, TAKRI VOWEL SIGN VOCALIC R, which was not contained in the original Takri script proposal (L2/09-424). It was noted that there is no standard form of the script, and "Takri" includes various regional varieties, with variation in glyph shapes and orthographic styles but with a common underlying structure.

The character being proposed, TAKRI VOWEL SIGN VOCALIC R, appears in some varieties of Takri according to the authors of L2/18-084. The examples provided all come from the Mandeali form of Takri. Although the examples derive from just one source, the character appears to be justified.

Recommendation: We recommend the UTC accept U+116B9 TAKRI VOWEL SIGN VOCALIC R, after discussion. Note: For the glyph, the author reports the glyph can be borrowed from Sharada, U+111B8 SHARADA VOWEL SIGN VOCALIC R.

12. Tamil
Document: L2/18-207 TAMIL homograph sequences for addition to chapter 12 – Freytag

Comments: We reviewed this document, which recommended additions to the Tamil section of Chapter 12 of *The Unicode Standard*, and suggested adding corresponding data to UTS #39 Unicode Security Mechanisms (if not already present).

This document provides information on homographs for TAMIL LETTER LLA that we agree should be noted (or highlighted more clearly) in the Tamil chapter. While closely examining the Tamil chart, other look-alikes or potential confusables were spotted, which should also be noted.

We suggest the following be done:

1. Create a table (or some other vehicle in Chapter 12.6) that identifies which characters should be used (and *not* used) to represent letters that could be visually analyzed as having multiple parts. For example, it should note that for Tamil letter AA, users should employ U+0B86 TAMIL LETTER AA and not the sequence <U+0B85 TAMIL LETTER A, U+0BC2 TAMIL VOWEL SIGN UU>.

   Also, for U+0B94 TAMIL LETTER AU and U+0BCC TAMIL VOWEL SIGN AU, which have canonical decompositions that involve U+0BD7 TAMIL AU LENGTH MARK, users should not use U+0BB3 TAMIL LETTER LLA in the place of U+0BD7.

2. Note that the glyph for U+0B95 TAMIL LETTER KA is identical to U+0BE7 TAMIL DIGIT ONE in the documentation.
3. Check whether any of the above lookalike characters are included in the UTS #39 confusables data file(s).

Also, it should be noted that the actions above only ensure that an intended vowel letter/sign AU should have a well-defined encoding, but haven't addressed the issue of unrelated spellings <… O, LLA, …> and <… vowel sign E, LLA, …> that are also valid in the language and can be confusable with the vowel letter/sign AU.

Recommendations: We recommend the UTC remand the items listed above to the Editorial Committee, assigning Liang Hai and the EdComm for items #1-#2, and Mark Davis for #3.

13. Tigalari

Document: L2/18-175 Replies to Script Ad Hoc Recommendations and Comments on Tigalari Proposal – Murthy Yerkadithaya

Comments: We reviewed this document, which appears to be responding to comments from 2016 and 2017. We recommend the author review the comments in L2/18-039.

Recommendations: We recommend the UTC create an AI for Debbie Anderson to send the document author the comments in L2/18-039.

14. Zanabazar Square

Document: L2/18-132R Proposal to encode two additional Zanabazar Square letters – West

WG2 Recommendations (L2/18-209): Two characters were added to the CD: 11A48 ZANABAZAR SQUARE CLUSTER-INITIAL LETTER LA and 11A49 ZANABAZAR SQUARE CLUSTER-INITIAL LETTER SA.

WG2 Consent Docket comments (L2/18-217):

Two Zanabazar Square characters were proposed by Andrew West (N4945 = L2/18-132) and were discussed by the script experts in the N4977 (L2/18-210). Andrew West agreed to revise the proposal, adding guidance on how to handle cases that fall outside the "typical" examples (such as those shown in figures 3, 5, 7, and 11). He also agreed to fix the representative glyphs (changing the dotted circles to dotted squares).

Comments from Script Ad Hoc: We reviewed the Consent Docket and the revised proposal, which now includes wording on pp. 5-6 that was intended to address earlier comments from the Script Ad Hoc (L2/16-168) regarding how to handle the non-typical combinations of LA and SA (i.e., figure 3 compressed and not ligated, figure 5 halfway compressed and not clearly ligated, figure 7 halfway compressed and ligated, etc.)

The new text recommends handling Tibetan words with the newly proposed cluster-initial letters, and Sanskrit words with normal full-size letters. However, without a well understood graphic distinction, this assumes the user needs to know the etymological origin of the word, and thus promotes undesired etymology-based text encoding.

In our view, there doesn't appear to be any systematic distinction in the examples provided. Hence, evidence of intentional orthographic contrast within a single source is still needed; such evidence would justify having two encodings for the same character.

Recommendations: We recommend the UTC review the document and forward comments to the proposal author. The UTC may consider making ballot comments on the CD.

SOUTHEAST ASIA

15. Tai Tham

Document: L2/18-171 Positioning of Tai Tham Vowels Below – Wordingham

Comments: We reviewed this document, which answers, in part, Action Item 151-A147 ("Create a new document addressing the remaining issues for Tai Tham and Lao, for discussion at the next UTC meeting").

Recommendations: We recommend the UTC assign to Roozbeh Pournader an Action Item to add a comment in IndicPositionalCategories that U+1A69 TAI THAM SIGN U and U+1A6A TAI THAM SIGN UU have contextually variable placement in Tai Tham.

16. Thai

Document: L2/18-216 Canonical Ordering of Marks in Thai Script – Constable

Comments: We reviewed this document, which discusses and analyzes the canonical ordering of marks in Thai and provides tentative rendering rules for use in a Thai rendering engine (page 10). Several relevant problematic cases are presented.

It was noted that similar mark ordering issues occur in Arabic – which ultimately resulted in the creation of UTR #53 Unicode Arabic Mark Rendering – as well as in Tibetan.

The following are comments raised in discussion about the rendering rules:

- Case 2: above vowel mark in class 0 with a below mark, with examples below

  Case 2 is, in our view, an architectural issue in Unicode that should be noted in the core spec or in a document similar to UTR #53. A preferred encoding order should be specified. (Alternatively, if UTR #53 were to be expanded to include Arabic and Thai mark ordering, details on Thai mark ordering could be added there.)

Case 2, phintuu + above vowels, order (i)	<0E01 0E3A 0E34> “กิ̣”	
Case 2, phintuu + above vowels, order (ii)	<0E01 0E34 0E3A> “กิ◌̣”	

- Case 3: two below marks from distinct non-zero classes
  Case 5: two above marks from distinct non-zero classes

  Cases 3 and 5 appear to require a decision by the UTC to identify a preferred visual/rendering order. Note the author’s discussion on page 7 about the implications of ensuring a specific visual order. Also note, on the same page, the author’s opinion, “it would have been better to encode a new Thai-specific consonant diacritic, or new consonant letters.”

Case 3, phintuu + below vowels	<0E01 0E3A 0E38> “กุ̣” ≡ <0E01 0E38 0E3A> “กุ̣”	
Case 3, generic below marks + below vowels	<0E01 0331 0E38> “ก◌̱◌ุ” ≡ <0E01 0E38 0331> “กุ̱”	
Case 3, phintuu + generic below marks	<0E01 0E3A 0331> “ก̱” ≡ <0E01 0331 0E3A> “ก̱◌̣”	
Case 5, generic above marks + tones	<0E01 0303 0E48> “ก◌̃◌่” ≡ <0E01 0E48 0303> “ก่̃”	

Case 4: two above marks, one of them in class 0
Case 6: three above marks in distinct classes, one of them in class 0

These two cases need more discussion. Although case 4 itself is not problematic, the author's tentative rendering rules assume that a preferred visual order – hence also a preferred encoding order – can be specified for case 4.

Case 4, above vowels + tones, order (i)	<0E01 0E34 0E48> "กิ่"	
Case 4, above vowels + tones, order (ii)	<0E01 0E48 0E34> "กิ้"	
Case 4, above vowels + generic above marks, order (i)	<0E01 0E34 0303> "กิ̃"	
Case 4, above vowels + generic above marks, order (ii)	<0E01 0303 0E34> "กี̃ิ"	

Case 6, above vowels + tones + generic above marks, equivalence class "A"	<0E01 0E34 0E48 0303> "กิ่̃" ≡ <0E01 0E34 0303 0E48> "กิ่̃"	

Case 6, above vowels + tones + generic above marks, equivalence class "B"	<0E01 0E48 0E34 0303> "กิ้̃"	
Case 6, above vowels + tones + generic above marks, equivalence class "C"	<0E01 0E48 0303 0E34> "กี้̃ิ" ≡ <0E01 0303 0E48 0E34> "กี้̃ิ"	
Case 6, above vowels + tones + generic above marks, equivalence class "D"	<0E01 0303 0E34 0E48> "กี̃ิ่"	

Recommendations: We recommend the UTC discuss this document.

EAST ASIA

17. CJK - Gongche characters

Document: L2/18-245 Gongche characters for Kunqu Opera – Chan et al.

Background document: L2/17-087 Proposal to encode characters for Gongche Notation – Chan et al.

WG2 Recommendations (L2/18-209): Added the 7 characters in a new block at U+2A6E0.. U+2A6FF named CJK Unified Ideographs Supplement, which could be used for exceptional cases, and to encode these characters within that block.

WG2 Consent Docket comments (L2/18-217):

> WG2 recommended encoding 7 ideographs U+2A6E0.. U+2A6E6 used in musical notation, based on document N4967. The discussion at WG2 focused on appropriate process for these ideographs and on their location. There was general consensus that these should be treated as unified CJK ideographs, but that they are exceptional: they should not be processed as part of Extension G or a future extension, and they are not UNC (urgently needed characters) but are valid candidates for encoding. There was general consensus to process them in SC2 as an independent, non-UNC proposal. Also, the basic elements of these ideographs are commonly-used characters, and there was concern about potential confusion for users if the characters were offered in input methods.

Comments from Script Ad Hoc: We reviewed the proposal and the Consent Docket comments. In our view, creating a new CJK block will only serve to create confusion. It would be preferable to tack the seven ideographs onto an existing CJK block. The proposal should provide the status of all the characters listed in Table 1.

Recommendations: We recommend the UTC discuss this proposal and suggest possible ballot comments for the CD.

SYMBOLS AND NUMERICAL NOTATION SYSTEMS

18. Dozenal notation

Document: L2/18-190 Proposal to properly support American style Dozenal notation – Marín Silva

Comments: We reviewed this document which picks up on the 2013 proposal from Karl Pentzlin (L2/13-054). Although the document is entitled "Proposal…", any formal action requires a full proposal that stands on its own, with justification for encoding the characters, in order for the UTC to properly review it.

Recommendation: We recommend the UTC note this document, which is not actionable. The UTC may wish to invite the author to create a full proposal with characters, glyphs, and justification.

19. Legacy Computer and Teletext characters

Document: L2/17-435R Proposal to add characters from legacy computers and teletext to the UCS – Ewell et al.

The following comments arose during discussion:

- For every sourced character set, visuals are needed with text mapping for those characters not already in the current proposal. This will provide a way to check any unifications and to explain why characters are being proposed (and identify their sources).

  To best accomplish this, we recommend creating a spreadsheet that shows how the symbols are cross-mapped, as was done for emoji (cf. L2/08-080r and the earlier version, L2/07-257 with mapping table [in zipped file, accessible from doc register for 2007]).

  Any new characters that have been found can be added to the spreadsheet.

- The proposal notes that a technical note is being created that will explain the mechanisms and recommended techniques for working with the extensive set of mapping tables. Re-send the mapping tables.

Recommendations: We recommend the UTC discuss the proposal, in conjunction with the authors.

20. Shogi
Document: L2/18-170 Proposal to encode a full set of pieces for the game of shogi – Marín Silva

Comments: We briefly reviewed this proposal for 30 shogi characters. Evidence from 2-dimensional diagrams is not considered adequate; demonstrated use in plain text is needed. A useful guide are the examples in the fairy chess proposal L2/17-034, which includes figures showing the symbols inline. The proposal should also include page numbers.

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

21. Symbol for Type A Electronics
Document: L2/18-184 Proposal to encode: SYMBOL FOR TYPE A ELECTRONICS (replaces L2/18-058) – Marín Silva

Comments: We reviewed this proposal for one character, SYMBOL FOR TYPE A ELECTRONICS, which previously appeared in L2/18-058, L2/18-004 (and additional example in L2/18-017), and the original DPRK proposal, L2/01-349.

The following comments arose during discussion:

- The character is justified but should not have an annotation (“postal mark”).
- The proposal needs a font before it can be officially accepted.
- The proposed code point, U+1F1B0, in the Enclosed Alphanumeric Supplement block is not the most appropriate location.
- Number the pages.

Recommendation: We recommend the UTC review this proposal and accept the character but discuss the location of the character. We further recommend the UTC request the author provide an updated proposal based on the comments above and provide a glyph (from Adobe).

22. Tachograph symbols
Document:L2/18-090 On the encoding of tachograph symbols – Spix
Feedback: L2/18-095 Feedback – Marín Silva
Feedback from David Corbett in L2/18-117 Public Review Issues: Feedback on tachograph bed symbol (received April 13 2018)

Comments: We reviewed this proposal for symbols that appear on tachograph devices (and related documentation), which track the speed of a vehicle, as well as the driver's activities (such as being on-duty, taking a break, etc.).

The proposal included a list of characters the author considered unifiable with existing characters, so the author recommends annotations be added describing their use on tachographs. Four characters were identified as not in Unicode:

- STEERING WHEEL
- WHITE SQUARE WITH HORIZONTAL BISECTING LINE
- VERTICAL BAR WITH BLACK RIGHT-POINTING TRIANGLE
- BLACK RIGHT-POINTING TRIANGLE WITH VERTICAL BAR.

The proposal is a useful start in identifying the core set of tachograph characters that are widely used and well-understood. Further research is needed into the set as a whole, providing comments on whether currently encoded characters should be used (or not).

The following are specific comments on the document:

- Number the figures, and point to the figures in the list of symbols.
- The top figure on page 9 contains two bed symbols, one labeled "PA" and the other "RU" (i.e."Pausenzeiten" and "Ruhezeiten"). The proposal should take the two symbols into account.
- The proposed character WHITE SQUARE WITH HORIZONTAL BISECTING LINE is already representable in Unicode with U+229F SQUARED MINUS.
- The two play buttons, ⏵ VERTICAL BAR WITH BLACK RIGHT-POINTING TRIANGLE and ⏯ BLACK RIGHT-POINTING TRIANGLE WITH VERTICAL BAR appear to be good candidates, and evidence of their use in manuals is provided (i.e., pp. 11-12, 14, 16-18). They are not currently included in the "user interface symbols" from U+23E9..U+23EF or U+1F53A..U+1F53D. (Other symbols are used for "eject" U+23CF and "skip to next track" U+23ED, but the glyphs are different.) The names should be checked to ensure they are compatible.
- The proposed glyph for STEERING WHEEL glyph  is not clear, as there is a range of shapes:  ,  . Which is the most common?
- While the "driver, vehicle, tachograph" symbol looks like a CYRILLIC CAPITAL LETTER DE, it should be disunified, if plain text usage is provided.
- The bed symbol  is not the same as the Dingbat character U+1F6CF BED, so the tachograph symbol should probably be separately proposed.
- The "repair" hammer  (pp. 11 and 15) might be eligible as a new symbol, separate from U+1F528 HAMMER.

- The symbol  'card' may need to be separately encoded, as it is different from U+1F4B3 CREDIT CARD.

Recommendations: We recommend the UTC members review this proposal and send the author feedback.

23. Tally Marks
Document:L2/18-088 Proposal to encode the two remaining tally mark systems proposed in L2/15-328 – Marín Silva

Comments: We reviewed this brief document, which refers to two other tally mark systems that were cited in L2/15-328. The author provides no new information for encoding the characters. If the author wishes to pursue proposing the tally mark systems, he is welcome to submit a full proposal that includes a code chart (with nameslist), evidence (besides Wikipedia), and strong justification, along with a font.

Recommendations: We recommend the UTC note this document.

24. Other
Feedback on Additional Repertoire beyond Amendment 2 of ISO/IEC 10646 5th edition
Document: L2/18-214 Feedback on: Additional repertoire for ISO/IEC 10646:2017 (5th ed.) beyond Amendment 2 (L2/18-211) – Marín Silva

Comments: The following captures responses to the various feedback comments sent in this document.

- MALAYALAM VEDIC ANUSVARA (0D04). This character from both the proposed glyph as well as the attestations provided, is clearly a spacing character; but the form in this code chart is that of a combining mark above (which looks like a *candrabindu*). The glyph should be changed. (Proposal: L2/17-276)
  Response: We agree, the glyph should be fixed in the CD. A ballot comment for the CD is warranted.

- SINHALA SIGN CANDRABINDU (0D81). An annotation should be made to indicate its use in Sanskrit.
  Response: This appears to be an acceptable annotation to the namelist for the CD.

- THORN WITH STROKE. A note should be added to A765 that it is used for old Norse, to fill in the information gap.
  Response: The Script Ad Hoc is opposed to the character, so this comment is not actionable.

- LETTERS WITH OVERCURL. The header above these letters should say: Letters for medieval Cornish.
  Response: The Script Ad Hoc still has questions on the set of Latin letters with overcurl, so this comment is not actionable.

- CIRCLED EQUALS. A note should be added to 229C that it is used to indicate "No derivate works" in the Creative Commons license, to fill in the information gap.
  Response: Adding a note to the nameslist as suggested seems reasonable (but the wording

should be corrected to “derivative”).

- CJK UNIFIED IDEOGRAPHS SUPPLEMENT. The notes for these characters should include a reference to the most similar unified ideographs; and the names should not be like the rest of the ideographs, for instance for 2A6E2, I suggest CJK UNIFIED IDEOGRAPH 4E00 WITH DESCENDER.
  Response: The Script Ad Hoc is opposed to the block, so this comment Is not actionable.

- TANGUT COMPONENTS It would be nice information to have the stroke count for each of these characters in the code charts.
  Response: This comment is moot**.**

Recommendations: We recommend the UTC consider a ballot comment for U+0D04 MALAYALAM VEDIC ANUSVARA for the CD, and remand to the nameslist editors to take under advisement the suggestions in L2/18-214 for U+0D81 SINHALA SIGN CANDRABINDU and U+229C CIRCLED EQUALS.

SCRIPT RECOMMENDATIONS CARRIED OVER
25. Bengali
Document: L2/18-035 Encoding model issues with the Vedic *gomukha* characters – Sharma

Comments: We reviewed this document, which requests guidance from the UTC on the encoding model for *gomukha* characters in Bengali, raised specifically by the Bengali form  , cited in §4.1.4 of L2/17-098. The author noted that the Bengali form  is a variant of  U+A8F3 DEVANAGARI SIGN CANDRABINDU VIRAMA.

The form appears to be the sequence U+1CEA  VEDIC SIGN ANUSVARA BAHIRGOMUKHA and U+1CED  VEDIC SIGN TIRYAK, but which character should be used to represent the dot?
The Vedic Extensions section of Chapter 12 of the Core Spec (p. 468) mentions that the *gomukha* characters may be combined with  DEVANAGARI SIGN CANDRABINDU (U+0901) or  DEVANAGARI SIGN ANUSVARA (U+0902). With evidence now in Bengali for *gomukha* characters, which characters should be used for the combining marks? Should the Devanagari characters be used?

The author provides different options:

1. Use U+0982 BENGALI SIGN ANUSVARA  and U+0981  BENGALI SIGN CANDRABINDU, noting that BENGALI SIGN ANUSVARA is not a “dot above” shape.
2. Use U+0902  DEVANAGARI SIGN ANUSVARA and U+0901  DEVANAGARI SIGN CANDRABINDU. Sharma notes that Bengali has its own *candrabindu* at U+0981: 
3. Encode a separate “dot above” character for Bengali.
4. Use U+0307 COMBINING DOT ABOVE and U+0310 COMBINING CANDRABINDU in the Combining Diacritical Marks block
5. Encode two new characters in the Vedic Extensions block, a script=inherited “dot above” character and a CANDRABINDU.
6. Encode only a new “dot above” character with script=inherited in Vedic Extensions, since *candrabindu* is uniform across the Indic scripts, but the glyph for the *anusvara* may vary.

The following comments were raised during discussion:

- Is the “dot above” only found in Devanagari?
- Some members of the Script Ad Hoc felt option #2 or possibly #6 might work, but there was hesitation in adding more dots (options 3, 5, 6) without careful consideration of the viable alternatives.
- Why couldn’t U+09FC  BENGALI LETTER VEDIC ANUSVARA be used, on the model of  U+A8F3 DEVANAGARI SIGN CANDRABINDU VIRAMA?

Recommendations: We recommend the UTC review this document, seeking input from those members with Indic rendering engines and the Bengali user community. We also suggest the author work with Srinidhi and Sridatta, authors of L2/17-098, and consider why U+09FC  BENGALI LETTER VEDIC ANUSVARA would not be suitable.

26. Malayalam

Document: L2/18-015 Proposal to encode the END OF TEXT MARK for Malayalam – Srinidhi and Sridatta

Feedback:

L2/18-034 Feedback on L2/18-015 proposing Malayalam punctuation mark – Sharma
L2/18-009 Comments on Public Review Issues from Eduardo Marín Silva
L2/18-145 Feedback on name for MALAYALAM END OF TEXT MARK (L2/18-015) - Anderson

Comments: We reviewed the proposal, which requested one character, MALAYALAM END OF TEXT MARK.

The feedback from Sharma (L2/18-034) is in support of the character. He provides transliteration of the examples in the proposal, which clarifies the name “end of text,” but noting that the mark predominantly appears alongside dandas. Sharma prefers the name MALAYALAM SIGN PUSHPIKA (*pushpa*=flower), since its usage is very similar to U+A8F8 DEVANAGARI SIGN PUSHPIKA. The Devanagari character also appears alongside dandas.

In our view, the name MALAYALAM SIGN PUSHPIKA is acceptable, and preferable to END OF TEXT MARK (or MALAYALAM SECTION MARK as suggested by Eduardo Marín Silva). The authors of the proposal are not against the name PUSHPIKA as noted in L2/18-145, although they note that the shape technically does not resemble a “flower.”

Recommendations: We recommend the UTC accept the character U+0D53 MALAYALAM SIGN PUSHPIKA, after first discussing the name. However, we recommend discussion about the overall approach to the *pushpika* character in Indic scripts be addressed first (discussed above, #6. General Indic - Script Extensions).

27. Tibetan

Document: L2/18-078 Deprecation of 3 Tibetan Characters - Élie Roux

Comments: We reviewed this request to deprecate three Tibetan characters: U+0F00 TIBETAN SYLLABLE OM and two head marks (U+0F02 TIBETAN MARK GTER YIG MGO -UM RNAM BCAD MA and U+0F03 TIBETAN MARK GTER YIG MGO -UM GTER TSHEG MA).

The Unicode Standard (D13 in section 3.4) states that deprecated characters typically are those characters that pose significant architectural problems or cause implementation problems. In our view, the three characters do not rise to this level.

The characters all appeared in Unicode 2.0 in 1996, and hence were encoded before normalization was first introduced (Unicode 3.1). All the OM characters in Unicode are "Lo", and none are decomposed. Like the Mongolian *birgas*, the two headmarks are encoded as atomic symbols (gc="So"), which is why collation doesn't decompose them.

The author is invited to propose a short annotation for the names list on the usage of these characters or text for the Core Spec that describes their use or provides guidance on the representation of the characters.

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

28. 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	< क , virama, न > ("kot")	क्न	क्‌न	क्‌न
"false conjunct" with spacing vowel mark	< ब, ◌ी, ब, virama, न, ◌ी > ("bibir" — same sequence as "bibri")	बीब्री	बीबेनी	बीबीन

Table 2 (with ZWJ model)

Case	Example sequence	Conjoining display with specialty font	Conjoining display with basic font
"false" conjunct" C + C	< क , ZWJ, न > ("kot")	क्न	क्‌न

"false conjunct" with spacing vowel mark	< व, ◌ी, व, ZWJ, ◌ी, न > ("bibir" – distinct sequence from "bibri")	वीव्री	वीवीन

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.