TO: UTC
FROM: Deborah Anderson, Ken Whistler, Roozbeh Pournader, Lisa Moore, and Liang Hai[1]
SUBJECT: Recommendations to UTC #161 October 2019 on Script Proposals
DATE: October 4, 2019

The Script Ad Hoc group met on 30 August and 27 September 2019 in order to review proposals. The following represents feedback on proposals that were posted in the Unicode document registry at the time the group met. A table of contents is provided below.

EUROPE
1. Glagolitic
Document: L2/19-288R Proposal to add a character used in printed Glagolitic works (rev.) -- Zoran Vukojević

Comments: We reviewed this proposal to add a case pair for a variant of the Glagolitic letter *črv*. Unicode currently contains the pair U+2C1D/U+2C4D GLAGOLITIC CAPITAL/SMALL LETTER CHRIVI. The proposed case pair is used with a combining tilde or combining titlo to indicate *et cetera*, a use not found with U+2C1D/U+2C4D (see figure 2). The currently encoded characters and the proposed variants

---

[1] Also participating were Richard Cook, Craig Cornelius, Ben Yang, Chris Chapman, Manish Goregaokar, Marek Jeziorek, and (on phone) Ned Holbrook, Lorna Evans and Andrew Glass.

are shown on the same page in figures 1, 3 and 4, and therefore, in our opinion, should be differentiated in plain text.

The proposed characters appear in printed works in the Glagolitic alphabet, which is still used in scholarly works.

Printed evidence is provided, and the proposal has been reviewed by other experts, who agree the characters are justified and approve of the name (GLAGOLITIC LETTER CAUDATE ["having a tail"] CHRIVI). The Script Ad Hoc reviewed an earlier version of the proposal; corrections and modifications have been made and incorporated in this revised version.

Recommendations: We recommend the UTC accept the two characters
U+2C2F GLAGOLITIC CAPITAL LETTER CAUDATE CHRIVI
U+2C5F GLAGOLITIC SMALL LETTER CAUDATE CHRIVI.

2. Latin
a. Latin S with short stroke overlay
Document: L2/19-260 Supplemental proposal on the Latin letter S with short stroke overlay – Pentzlin

Comments: We reviewed this document, which proposed two actions:
- modify the representative glyphs for U+A7C9/U+A7CA LATIN CAPITAL/SMALL LETTER S WITH SHORT STROKE OVERLAY, so the mid-stroke is diagonal (S), instead of the current horizontal position (S), since the diagonal form predominates in the historical orthographies of Komi, Kumyk, and Tsakhur and other languages. The author noted that LATIN LETTER S WITH SHORT STROKE OVERLAY may appear in these orthographies in various orientations (horizontal, diagonal, and vertical), citing examples from L2/12-045.
- add an annotation to the new characters (U+A7C9/U+A7CA), indicating the use of this letter in historical orthographies of the languages listed above.

Discussion on this document raised the following points:
- The two characters were approved at the July 2019 UTC meeting, based on a proposal by Michael Everson to represent Gaulish. In a private email, Michael Everson stated that the proposed glyph change for LATIN LETTER S WITH SHORT STROKE OVERLAY is not suitable for Gaulish, since the Gaulish mid-stroke is long and horizontal. In addition, the mid-stroke can touch an adjoining glyph, a characteristic not shared with the early Soviet orthographies. For the inconsistent shapes of early Soviet typography, U+A7A8/U+A7A9 should be used. (The latter view is shared by the author of the recent Komi proposal, L2/19-224.)

In sum, we consider the current glyph for LATIN LETTER S WITH SHORT STROKE OVERLAY to be acceptable. Adding characters to represent all the orientations of LATIN LETTER S with a mid-stroke in early Soviet orthographies is not necessary in our view, particularly as the historical orthographies of the former Soviet Union were not stable.

Recommendations: We recommend the UTC send a note to the proposal author, declining to make the glyph change.

b. Latin letters for Komi
Document: L2/19-261 Proposal to encode Latin letters for the Komi Latin alphabet
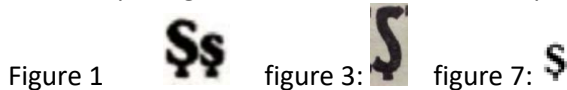
(supplementing [L2/19-224](#)) – Pentzlin

Comments: We reviewed this proposal which offers modifications and additions to the Komi proposal, [L2/19-224](#). The document [L2/19-261](#) is based upon the author's earlier document, [L2/12-045](#) "Revised Proposal to encode Latin letters used in the Former Soviet Union."

The following summarizes comments on specific sections of this document:

- §1.1
    - YERU: Adding another *yeru (=KOMI BACK I)* would further proliferate Cyrillic characters into the Latin script for use in dead orthographies. Instead, users should use Cyrillic characters. (See page 2 of [the July Script Ad Hoc Recommendations](#).)

    - S WITH SHORT STROKE OVERLAY is discussed above ([L2/19-260](#))

- §2.1 Descender vs. cedilla
  In our view, the evidence for the distinct identity of the descender in early Soviet typography isn't compelling, because the evidence is so poor. Compare, for the s-descender:

 Figure 1   figure 3:   figure 7:

  For other letters, cf. figure 7:

  A better analysis is needed to distinguish the descender from the cedilla and comma below.

- §2.2 Character forms similar to Cyrillic ones
  The July 2019 Script Ad Hoc had recommended (in [L2/19-286](#)) and the UTC agreed to create a FAQ stating that it is acceptable to mix scripts for historical, transitional orthographies (including those of the former Soviet Union).

  In our opinion, the Latin letters with cedilla or comma below (where applicable) can be used, with fonts that are specific for Komi (etc.). For Komi CHA, YERU, and REVERSED OPEN E WITH DESCENDER, the already encoded Cyrillic characters are suitable for use.

- §3. Proposed letters
  For KOMI CHA / , why can't the letters be represented with U+0454/U+0404 є/Є CYRILLIC SMALL/CAPITAL LETTER UKRAINIAN IE? In handwriting, the distinction between Ukrainian IE and the Komi forms would be negligible.

  In sum, our view is that a case needs to be made that the proposed letters were adapted into Latin because of systematic shape changes. Until such a case is made, we recommend use of the already encoded characters (listed on the following page).

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

Proposed characters

Ƀ	U+A7C0	LATIN CAPITAL LETTER YERU
ƀ	U+A7C1	LATIN SMALL LETTER YERU
Є	U+A7CB	LATIN CAPITAL LETTER KOMI CHA
є	U+A7CC	LATIN SMALL LETTER KOMI CHA
Ç	U+A7CD	LATIN CAPITAL LETTER C WITH DESCENDER
ç	U+A7CE	LATIN SMALL LETTER C WITH DESCENDER
Ḑ	U+A7CF	LATIN CAPITAL LETTER D WITH DESCENDER
ḍ	U+A7D0	LATIN SMALL LETTER D WITH DESCENDER
Ļ	U+A7D1	LATIN CAPITAL LETTER L WITH DESCENDER
ļ	U+A7D2	LATIN SMALL LETTER L WITH DESCENDER
Ş	U+A7D3	LATIN CAPITAL LETTER S WITH DESCENDER
ş	U+A7D4	LATIN SMALL LETTER S WITH DESCENDER
Ţ	U+A7D5	LATIN CAPITAL LETTER T WITH DESCENDER
ţ	U+A7D6	LATIN SMALL LETTER T WITH DESCENDER
Ӡ	U+A7D7	LATIN CAPITAL LETTER REVERSED OPEN E WITH DESCENDER
ӡ	U+A7D8	LATIN SMALL LETTER REVERSED OPEN E WITH DESCENDER
S	U+A7C9	LATIN CAPITAL LETTER S WITH SHORT STROKE OVERLAY
s	U+A7CA	LATIN SMALL LETTER S WITH SHORT STROKE OVERLAY

Options from encoded characters

042C	Ь	CYRILLIC CAPITAL LETTER SOFT SIGN
044C	ь	CYRILLIC SMALL LETTER SOFT SIGN
0404	Є	CYRILLIC CAPITAL LETTER UKRAINIAN IE
0454	є	CYRILLIC SMALL LETTER UKRAINIAN IE
04AA	Ҫ	CYRILLIC CAPITAL LETTER ES WITH DESCENDER
00C7	Ç	LATIN CAPITAL LETTER C WITH CEDILLA
04AB	ҫ	CYRILLIC SMALL LETTER ES WITH DESCENDER
00E7	ç	LATIN SMALL LETTER C WITH CEDILLA
1E10	Ḑ	LATIN CAPITAL LETTER D WITH CEDILLA
1E11	ḑ	LATIN SMALL LETTER D WITH CEDILLA
013B	Ļ	LATIN CAPITAL LETTER L WITH CEDILLA
013C	ļ	LATIN SMALL LETTER L WITH CEDILLA
015E	Ş	LATIN CAPITAL LETTER S WITH CEDILLA
0218	Ș	LATIN CAPITAL LETTER S WITH COMMA BELOW
015F	ş	LATIN SMALL LETTER S WITH CEDILLA
0219	ș	LATIN SMALL LETTER S WITH COMMA BELOW
04AC	Ҭ	CYRILLIC CAPITAL LETTER TE WITH DESCENDER
0162	Ţ	LATIN CAPITAL LETTER T WITH CEDILLA
021A	Ț	LATIN CAPITAL LETTER T WITH COMMA BELOW
(04AD	ҭ	CYRILLIC SMALL LETTER TE WITH DESCENDER *)
0163	ţ	LATIN SMALL LETTER T WITH CEDILLA
021B	ț	LATIN SMALL LETTER T WITH COMMA BELOW
0498	Ҙ	CYRILLIC CAPITAL LETTER ZE WITH DESCENDER
0499	ҙ	CYRILLIC SMALL LETTER ZE WITH DESCENDER
A7A8	Ꞩ	LATIN CAPITAL LETTER S WITH OBLIQUE STROKE
A7A9	ꞩ	LATIN SMALL LETTER S WITH OBLIQUE STROKE

* Different glyph, so not usable.

c. Middle Scots
Document: L2/19-180R Proposal to add two characters for Middle Scots (WG2 N5045) -- Everson

Comments: We reviewed this revised proposal to add two characters for Middle Scots. An earlier version of the proposal was discussed at the June 2019 WG2 meeting. The latest version of the proposal (L2/19-180R) includes additional discussion about case pairs (§3.1 and §3.2) and two new figures (14 and 15).

The July 2019 Script Ad Hoc recommendations (L2/19-286) suggested that a ligature of two *s*'s would suffice. Alternatively, users could employ U+00DF LATIN SMALL LETTER SHARP S, making minor modifications to the font, if needed. However, later discussions in the Script Ad Hoc led to consensus that SMALL LETTER MIDDLE SCOTS S was an atomic unit of the writing system and was not necessarily a ligature.

Section §3.2 refers to capital use of MIDDLE SCOTS S in figures 14 and 15. The following comments were made about the two figures:

- In making a case for the bicamerality of the MIDDLE SCOTS S, the author points to the bold vertical of MIDDLE SCOTS S (circled) in line 1 of figure 14. Is the long s in "Jeſus" in line 5 also bold? Is there a lowercase MIDDLE SCOTS S in the text? If so, provide a sample.
- It was unclear how figure 15 shows uppercase. The caption in figure 15 only refers to word-initial use. If figure 15 does exemplify uppercase, please explain.
- Can the author point out contrastive use of upper- vs. lowercase of same letter in the figures (or in another figure)? Such evidence would bolster the argument for encoding LATIN CAPITAL LETTER MIDDLE SCOTS S.

Recommendations: We recommend the UTC accept U+A7F1 LATIN SMALL LETTER MIDDLE SCOTS S, after reviewing the proposal, and relay the comments above to the author.

3. Northern Palaeohispanic
Note: The following two script proposals followed the Script Ad Hoc recommendations from July 2019 (L2/19-286), namely, to separate the unified "Palaeohispanic" repertoire (as proposed in L2/18-283) into two separate proposals.

Documents: L2/19-332 Proposal to encode the Northern Palaeohispanic script – Ferrer et al.

Comments: We reviewed the Northern Palaeohispanic proposal and the following comments were noted:
- Remove "ONE" from character names that currently contain "ONE ADDITIONAL STROKE".
- In section 9, Unicode Character Properties, change "No" to "So" for NUMERAL A and NUMERAL B.

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

4. Southern Palaeohispanic
Documents: L2/19-333 Proposal to encode the Southern Palaeohispanic script – Ferrer et al.

Comments: We reviewed this proposal, which now includes variants for the recently discovered Turdetan inscriptions (shown in green on pages 12-15).

The following summarize the comments raised during discussion:
- As with the Northern Palaeohispanic proposal, remove "ONE" from those names that have "ONE ADDITIONAL STROKE".
- Change the character names on pages 6-8 from PALAEOHISPANIC SOUTHERN to SOUTHERN PALAEOHISPANIC.
- Explain the vertical dotted lines in figure 4. Are these word separators, such as those listed in section 7? If so, figure 4 appears to have U+2E3D VERTICAL SIX DOTS, which should be added to the list of encoded vertical dot characters in section 7.
- Add annotations to the figures, explaining what each figure demonstrates and why it is important for encoding the script. If a transcription would help understand the issue better, provide it. If a figure is just decorative, remove it. (Cf. the captions from the Northern Palaeohispanic proposal, such as figure 4: "Lead sheet from Castellet de Bernabé (F.13.75*). Extended dual abecedary with explicit duality for ŕ.")
- For S52, provide an explanation and key parts of the papers referenced in footnote 1 to justify why it doesn't need to be separately encoded.

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

AFRICA
5. Egyptian Hieroglyphs
a. OpenType Implementation
Document: L2/19-331 A note on OpenType implementation of Ancient Egyptian hieroglyphic text – Nederhof

Comments: We reviewed this note which describes an implementation for rendering Egyptian quadrats, using an offline approach. The approach used by Andrew Glass is slightly different; it involves calculating the verticals and horizontals of quadrats within the font. In our opinion, it is encouraging to see two independent implementations that render sequences with similar results.

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

b. September 2019 meeting on Egyptian Hieroglyphs
Documents:
L2/19-342 Summary of Meeting Report on Egyptian Hieroglyphs by Michel Suignard (L2/19-315) and Additional Comments) – Anderson
L2/19-315 Meeting report on Egyptian Hieroglyphs (Paris, Sept. 2019) (=WG2 N5116) – Suignard
Related document:
L2/19-314 Agenda for meeting on Egyptian Hieroglyphs, September 12-13, 2019 (WG2 N5115) – Anderson

Comments: We reviewed the Summary document, which is a précis of the detailed 20-page September 2019 Egyptian Hieroglyph meeting report (L2/19-315).

The Summary identifies changes to glyphs and annotations to be made in the current Egyptian Hieroglyph repertoire and the proposed set of characters in L2/19-220,. It also highlights discussion that took place at the meeting about control characters, and their role in expanding the repertoire of Egyptian hieroglyph characters.

The Script Ad Hoc discussion centered on the question of atomic encoding vs. sequences with OVERLAY MIDDLE, specifically, should Unicode maintain a list of combinations (i.e., Egyptian Hieroglyph characters with the OVERLAY MIDDLE control character), similar to RGI for emoji? If so, should the sequences be documented in a UTN? Further discussion with the UTC and Michel Suignard is needed.

Recommendations: We recommend the UTC hold an ad hoc on Egyptian hieroglyphs with Michel Suignard and interested parties, and use the Summary as a discussion document, bringing the results of the ad hoc back to the UTC for further discussion.

MIDDLE EAST
6. Arabic
a. Additions for Quranic orthographies
Document: L2/19-306 Arabic Additions for Quranic orthographies -- Pournader and Anderson

Comments: We reviewed this proposal, which presents a thorough analysis of 39 Arabic characters used to represent Quranic orthographies. The document builds off proposals and documents from others, and includes additional input from Marijn van Putten of Leiden University. The proposed characters all have evidence, code points, glyphs, and properties provided.

As noted on page 1, the proposed characters follow the current encoding model, which means the new characters can be added to existing fonts more easily. The document also includes suggested annotations for four already encoded characters (pp. 29-30). Ten characters are not proposed (pp. 30-33), as additional research is required.

Because only 11 spots were available in Arabic Extended-A block, a new Arabic Extended-B block is needed. We recommend the new block of three columns be located in the range U+0870…U+089F. (The new allocation has already been incorporated in the latest version of the Roadmap.) One character is proposed for inclusion in Arabic Presentation Forms-A, and 10 characters are proposed for the Arabic Extended-A block.

The proposal has been reviewed by a number of experts.

Recommendations: We recommend the UTC accept 39 Arabic characters:
10 characters in the Arabic Extended-A block (U+08C9..U+08D2);
28 characters in a new Arabic Extended-B block (U+0870..U+0888 and U+089D..U+089F, with the range of the new block extending from U+0870…U+089F);
1 character in Arabic Presentation Forms-A (U+FBC2).
We also recommend the annotations on pages 29-30 be forwarded to the names list editor.

b. Arabic qaf with dot below
Document: L2/19-313 Proposal to encode an Arabic qaf with dot below -- van Putten and Pournader

Comments: We reviewed this proposal for one Arabic character that appears in historic Quranic

manuscripts. The missing character was pointed out to Roozbeh Pournader by Marijn van Putten of Leiden University.

The character is confusable with U+06A2 ARABIC LETTER FEH WITH DOT MOVED BELOW and U+09BB ARABIC LETTER AFRICAN FEH in initial and medial forms; the proposed character differs from the two encoded characters in its final and isolated forms. As noted on page 2, all dots are changed to a diagonal line in some styles.

The proposal includes property information and ample evidence. The proposed code point fills a hole in the Arabic Extended-A block, and is acceptable. The proposed name is based on a new pattern, since there are no character names presently with "no dots above." An alternative name has been suggested by Michael Everson ("dotless qaf with dot below").

Recommendations: We recommend the UTC approve U+08B5 ARABIC LETTER QAF WITH DOT BELOW AND NO DOTS ABOVE in the Arabic Extended-A block, after reviewing and discussing the proposal. We recommend the authors also append a proposal summary form.

c. Arabic honorifics
Document: L2/19-319 Proposal to encode two more Arabic honorifics -- Pournader and Jibaly

Comments: We reviewed this proposal to encode two Arabic honorifics. At the July 2019 UTC meeting, 14 honorifics were proposed (L2/19-289) and approved, but a large set of other honorifics were still lacking evidence in running text. In the meantime, evidence for two honorifics has been found and are now proposed in this document. Published usage of twenty-five forms is still needed (listed on pages 2-3).

Recommendations: We recommend the UTC approve U+FD4C ARABIC LIGATURE SALLALLAHU ALAYHI WA-AALIHEE WA-SALLAM and U+FD4D ARABIC LIGATURE ALAYHAA AS-SALAAM in the Arabic Presentation Forms-A block.

d. Bosnian Arabic characters
Document: L2/19-339 Proposal to encode Bosnian Arabic characters -- Jacquerye

Comments: We reviewed this proposal for two Arabic characters used to write Bosnian. The two letters appeared in publications from 1908 to the 1940s. There has been renewed interest in the orthography using these letters; they appeared in 2013 in the *Šegrt Suljica* comics, for example (see figure 8).

The proposal includes evidence and properties, and the name and code points are acceptable.

Recommendations: We recommend the UTC accept U+0889 ARABIC LETTER NOON WITH INVERTED SMALL V and U+088A ARABIC LETTER HAH WITH INVERTED SMALL V BELOW in the new Arabic Extended-B block. We further recommend the annotations be forwarded to the names list editor.

e. Arabic characters for Balochi
Document: L2/19-320 Proposal to add four new Arabic characters for Balochi language

Comments: This document requests four Arabic characters to write the Balochi language.

Based on the Script Ad Hoc discussion, the four characters can already be represented with Unicode characters, as described below:

1. ـٟ iybalochi / yeh hamza above with Arabic kasra
This character is not a *kasra*. The shape reflects the style used in the Persianate world, where some hamzas get a line underneath them. We believe this is an alternate glyph for U+06D3:

06D3 �‌ـ ARABIC LETTER YEH BARREE WITH HAMZA
ABOVE

However, if the author is able to show how the proposed character is distinguished from U+06D3 (i.e., the proposed character and U+06D3 appear in the same document and need to be differentiated), it may be eligible for separate encoding. If the character is indeed a glyph variant of U+06D3, the author should contact font vendors and request the glyph be included in font for Balochi.

2. ء hamzazabar / hamza fatha above
This character can be represented by the sequence <U+0621 ARABIC LETTER HAMZA, U+064E ARABIC FATHA>.

3. ء hamzapesh / hamza damma above
This character can be represented by the sequence <U+0621 ARABIC LETTER HAMZA, U+064F ARABIC DAMMA>.

4. ء hamzazer / hamza kasra above

This character can be represented by either a sequence of <U+0621 ARABIC LETTER HAMZA, U+0650 ARABIC KASRA> or just a *hamza* in a specific Persianate font style.

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

7. Proto-Sinaitic
Document: L2/19-299 Revisiting the encoding of Proto-Sinaitic (revised) – Pandey

Comments: We reviewed this document which proposes that Proto-Sinaitic, a script that had been listed as rejected on the "Not on the Roadmap" page, be re-visited as worthy of encoding. It was noted that the scripts on the "Not on the Roadmap" page have no formal status, and should not be used as precedent for rejecting a script.

While the history of the decision to not encode Proto-Sinaitic is murky, the following summarizes some background:

- Michael Everson, the author of the earlier proposal for "Sinaitic" (L2/98-035) and a chart for Proto-Sinaitic (L2/99-069), wrote in October 1999 (WG2 N2133) that he had received input from experts about Proto-Sinaitic and stated, "Proto-Sinaitic should not be considered further at this time." This agreed with the September 1999 comments from Prof. Dr. W. Röllig in WG2 N2097,

who stated that scholars use drawings and images and the characters were not normally used in publications. Hence he felt encoding the script was pointless.

- The early document WG2 N2311 considered Phoenician as a catch-all group that should include Proto-Sinaitic. (However, it was noted that the final Phoenician proposal [L2/04-414R2] and §10.3 of the Core Spec do not include Proto-Sinaitic as being covered by Phoenician.)

In contrast to statements in 1999 about the use of Proto-Sinaitic by scholars, the document under review presents evidence of Proto-Sinaitic today in scholarly discussions and publications. The proposal also compares the Proto-Sinaitic repertoire with its descendants, Phoenician (figures 18 and 20), Old South Arabian (figure 18), and other related scripts (figure 18). Indeed, the repertoire of Proto-Sinaitic contains at least 27 characters, whereas Phoenician has 22, and hence should not be unified with it.

We encourage the author to continue his work on the script proposal; it adds considerably more information than the documents from 1998 and 1999. For code points, we recommend 3 columns (U+108B0..U+108DF, displacing Numidian), and moving Numidian to U+10960..U+1097F. We also suggest the author of L2/19-299 remove the line, "Although Proto-Sinaitic was disapproved for encoding in 2000."

Recommendations: We recommend the UTC ask the Roadmap Committee to remove "Proto-Sinaitic" from the "Not on the Roadmap" page and make Roadmap changes outlined above. We recommend those interested in the script contact the proposal author with comments.

8. Syriac
Document: L2/19-338 Expanding the "Syriac Supplement" U+0860 Block for historical glyphs and marks – Kiraz and Kiraz

Comments: We reviewed this proposal for two diacritical marks, a supralinear left dot and a sublinear left dot. According to the proposal, the two dots are differentiated from U+0307 and U+0323 by their position; U+0307 and U+0323 are usually centered above or below characters. In addition, the two proposed marks are used to distinguish two homographs, whereas the centered dots denote a feminine suffix.

The following summarizes points raised during discussion:
- In the original Syriac proposal (L2/98-050), the two characters U+0323 and U+0307 were listed, but without any explanation as to their use:

◌	U+0323	COMBINING DOT BELOW	SYRIAC DOT BELOW	ܙܘܿܘ Bedjan
◌	U+0307	COMBINING DOT ABOVE	SYRIAC DOT ABOVE	ܐܘܿܗܝ Bedjan

U+0323 and U+0307 are described in the Core Spec (page 395) by the following: "These points are used for various purposes—grammatical, phonological, and otherwise. They differ typographically and semantically from the qushshaya, rukkakha points, and the dotted vowel points."

Were the two proposed marks SUPRALINEAR LEFT DOT and SUBLINEAR LEFT DOT originally

intended to be covered by U+0323 and U+0307, as described above?

- How does the proposed SYRIAC SUPERSCRIPT LEFT HOMOGRAPH MARK differ from

 U+0740 SYRIAC FEMININE DOT? Note that Unicode is not encoding function, but graphemes. (From the description in the proposal on page 1, the centered dots denote the feminine suffix.)
- Provide more explanation with examples, comparing the proposed characters vs. U+0740 vs. U+0323 vs. U+0307. If U+0740 is not appropriate for SUPERSCRIPT LEFT HOMOGRAPH MARK, could U+1DF8 COMBINING DOT ABOVE LEFT be used?
- Which style do the proposed characters appear in?
- Update this proposal and refer to the original proposal (L2/98-050), explaining why the currently encoded characters should not be used and answering the questions above.

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

SOUTH AND CENTRAL ASIA
9. Dogra
Document: L2/19-263 Representation of conjuncts involving Ra in Dogra script -- Srinidhi and Sridatta

Comments: We reviewed this request to correct the text in the Dogra block description of the Core Spec, which says, "There is no evidence that special conjunct forms of *ra* occur." The document points out that the phonetic post-base conjunct form of *ra* does occur (and examples in text are provided).

Recommendations: We recommend the UTC assign an Action Item to Liang Hai and the Editorial Committee to draft text for the Dogra block intro in the Core Spec on conjuncts involving *ra*.

10. Manichaean
Document: Error Report (contained in Public Review doc L2/19-272) from David Corbett, July 4, 2019:

> The Manichaean section says "Manichaean has two obligatory ligatures for sadhe followed by yodh or nun", but they are not obligatory and they are not always ligatures
> See https://github.com/googlefonts/noto-fonts/issues/1550 for details and evidence.

Reference document:
L2/11-123 Second revised proposal for encoding the Manichaean script in the SMP – Everson et al.

Comments: We discussed this Error Report on Manichaean obligatory ligatures. We recommend that the author of the Error Report submit a separate document on Manichaean ligatures, highlighting the evidence with embedded images. (If a technical change to the Standard is required, the UTC will need to see a document.)

Recommendations: We recommend the UTC relay the comments above to the submitter of the Error Report.

11. Saurashtra
Document: Error Report (contained in Public Review doc L2/19-272) from David Corbett, May 7, 2019:

> Date/Time: Tue May 7 16:29:23 CDT 2019
> Name: David Corbett
> Report Type: Error Report
> Opt Subject: Saurashtra C2-conjoining forms
>
> The Saurashtra section says "An exception to the non-occurrence of complex
> consonant clusters is the conjunct ksa, formed by the sequence <U+A892 KA,
> U+A8C4 virama, U+200D ZERO WIDTH JOINER, U+A8B0 SSA>. [...] If necessary, U+200D ZERO
> WIDTH JOINER may be used to force conjunct behavior." That implies that the syllable "kra" in
> the old-fashioned style would be formed by the sequence <ka, virama, ZWJ,
> ra>. That is the opposite of the usual Indic practice (see PR #37) where
> a C2-conjoining form (as in Saurashtra) is formed by <ZWJ, virama,
> C2>. I recommend using <ZWJ, virama>.
>
> I know of only two Saurashtra fonts: Pagul and Noto Sans Saurashtra. Neither
> really supports C2-conjoining forms. (Pagul supports one but it doesn't even
> use ZWJ at all. Noto supports only a couple specific syllables.) So I
> wouldn't worry about breaking any existing text.
>
> Since "kṣa" is an atomic conjunct, either order could work. Using <ka,
> ZWJ, virama, ṣa> seems more consistent with other syllables, but using
> <ka, virama, ZWJ, ṣa> would be more compatible with previous versions
> of TUS and would allow <ka, ZWJ, virama, ṣa> to request a non-atomic
> "kṣa," so I recommend keeping "kṣa" as it is.
>
> In any case, I recommend documenting all of this explicitly.

Comments: We reviewed the Error Report from David Corbett on Saurashtra.
In our opinion , the Saurashtra block intro in the Core Spec needs further review. The current text,
which is based on the original proposal L2/05-222R2, implies an encoding model that is indeed
inconsistent with "the usual Indic practice," such as the one for Oriya, but may be more consistent with
that of Sinhala (i.e., the virama is only a vowel killer by default, and becomes a conjoiner when followed
by a ZWJ).

Recommendations: We recommend the UTC assign an Action Item to Liang Hai and the Editorial
Committee to review the entire Saurashtra Core Spec text. If any technical changes are needed, we
recommend he submit a document to the UTC for discussion.

12. Takri
Document: L2/19-264 Proposal to encode the Abbreviation Sign for Takri -- Srinidhi and Sridatta

Comments: We reviewed this proposal to add one character, the TAKRI ABBREVIATION SIGN, to the
Takri block. The abbreviation sign is found in 12 other scripts, and its representative glyph is found in
Sharada and Gurmukhi. The proposal also provides an example in figure 5 of a circle glyph, found in
Devanagari and several other scripts, and notes this should be considered a variant.

(It was noted that the proposal summary form suggests taking the glyph "may be borrowed from the Sharada or Takri blocks", which is a minor error and should read, "may be borrowed from the Sharada or Gurmukhi blocks.")

Recommendations: We recommend the UTC accept U+116B9 TAKRI ABBREVIATION SIGN.

13. Tibetan
Document: L2/19-337 Comments on Tibetan DOUBLE SHAY (response to Error Report by D. Corbett) – West and Lofting

Comments: We reviewed comments from Peter Lofting and Andrew West in response to the Error Report from David Corbett, dated April 2, 2019 (contained in L2/19-124 Public Review Issues).

The Error Report focused on a comment in the punctuation section of Tibetan block intro (§13.4 in the Core Spec), which states that U+0F0E ‖ TIBETAN MARK NYIS SHAD was encoded with the intent that it would have a wider space between the two shays than if two U+0F0D ▏ TIBETAN MARK SHAD were written together. The current text goes on to say that most writers don't use the unusual spacing between the double shay, and the application should allow users to write two U+0F0D characters one after the other, and font designers should decide whether to implement the double shays with a larger than normal gap.

We agree with the comments of Andrew West, who suggests shay + space + shay should be used for the widely spaced double shay, and U+0F0E ‖ TIBETAN MARK NYIS SHAD should only be used when the two shays occur closely together as a unit.

Recommendations: We recommend the UTC assign an Action Item to Lisa Moore and the Editorial Committee to provide text to fix the Core Spec, based on the comments of Andrew West.

14. Toto
Document: L2/19-330 Proposal for encoding the Toto script -- Anderson

Comments: This proposal is for a relatively new script created for the Toto language (txo) by a member of the Toto community. The language is spoken by a very small community in India, near Bhutan. Since the script was created, there has been increased interest in writing the language, although the Toto script is still in limited use.

The proposal builds off the earlier document (L2/19-278 Introducing the Toto Script) and has been reviewed several times by the Script Ad Hoc. The latest version has incorporated comments from the Script Ad Hoc.

In our opinion, the current model is correct: the repertoire includes a separate tone mark but the breathy vowels are separately and atomically encoded (i.e., there is no breathiness combining mark, originally proposed in the earlier proposal, L2/19-278). The proposed character names and locations are acceptable.

Recommendations: We recommend the UTC accept 31 Toto characters, in a new Toto block that extends from U+1E290 to U+1E2BF. We further request the Roadmap Committee allocate a new Toto block in the range U+1E290..U+1E2BF.

SOUTHEAST ASIA, INDONESIA, AND OCEANIA
15. Balinese
Document: L2/19-318 Two Balinese punctuation marks – Ben Yang and Aditya Bayu Perdana

Comments: We reviewed this document, which proposes two punctuation marks found in Balinese texts. The marks are used to indicate the end of text. PANTI LANTANG ("long panti") is used in lower-status works, whereas PAMADA LANTANG ("long pamada") appears in religious or noble works. "Panti" or "long" in the character names refer the tail, which extends out from the side of the character (cf. the two related, but already encoded characters, U+1B5A BALINESE PANTI and U+1B5B BALINESE PAMADA, which have no "tail").

The characters appear to be well-justified, in our opinion.

Recommendations: We recommend the UTC accept the following two characters:
U+1B7D BALINESE PANTI LANTANG
U+1B7E BALINESE PAMADA LANTANG.

16. Western Cham
Document: L2/19-217 Proposal to encode Western Cham in the UCS (Revised) -- Hosken

Comments: We reviewed this revised proposal. The following summarizes the comments:

- Number the pages.
- The chart on the bottom of page 2 identifies three characters as Arabic script, but the accompanying text states "some symbols claiming to be used for Arabic script also get used in Western Cham script" (and, we note, are included in the Western Cham chart). For these symbols, which script are they – Arabic or Western Cham?
- Move all the lunar symbols (U+1E265..U+1E26C) into the Arabic Mathematical Alphabetic Symbols block in the positions U+1EEF8..U+1EEFF, and rename them with Arabic names ("ARABIC SYMBOL …").
- For the character U+1E22E WESTERN CHAM SIGN VOWEL OE OR FINAL NG, change the name to WESTERN CHAM VOWEL SIGN OE, adding an annotation "also used for final ng."
- Include a screen shot of page 3 of L2/16-198 as a figure in the document, and point to it from the Introduction (i.e., after the statement "As stated in N4734, Western Cham while closely related to Eastern Cham, is sufficiently different in style for Eastern Cham characters in plain text to be unintelligible to Western Cham readers and therefore a separate block is required.") It is easier for committee members to review the proposal if everything is contained in one document, rather than flipping between different documents.
- Provide evidence for the WESTERN CHAM PUNCTUATION TRIPLE DANDA and WESTERN CHAM PUNCTUATION DANDA.
- Provide more evidence for ARABIC END OF TEXT beyond figure 8, drawing on examples from other publications. If solid evidence is provided, change the code point to U+0612, filling the last

hole in the Arabic block. (If justification is provided, no Arabic block chart would be necessary to include in the proposal, as long as evidence and the glyph are provided.)

- Provide a chart showing the letters, such as from a primer or academic work listing the characters.
- Be sure at least one example of the proposed characters is given, and the evidence satisfies other parties in Cambodia using the Western Cham script.

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

EAST ASIA
17. CJK - Gongche characters
L2/19-346 On Encoding Policy of Gongche Notations and Upcoming Para-ideographs - Yifán Wáng

Comments: We reviewed this document which raises a number of points about the Gongche characters and other "para-ideographic" characters. Because the Script Ad Hoc reviews proposals for non-CJK characters, we recommend the document be submitted to the UTC and to IRG for discussion.

Recommendations: We recommend the author bring his document to the UTC for discussion and raise issues involving IRG with the IRG.

SYMBOLS, PUNCTUATION, AND NUMERICAL NOTATION SYSTEMS
18. Currency symbols
Document: L2/19-291 Currency signs missing in Unicode -- Silva

Comments: We reviewed this document, which proposed 4 new characters and several edits to current annotations of currency signs.

The following are the comments from the Script Ad Hoc:

- The dollar sign with one or two vertical bars is discussed in *TUS* §22.1:

    > Currency symbols are intended to encode the customary symbolic signs used to indicate certain currencies in general text. These signs vary in shape and are often used for more than one currency…There are therefore many minor variants, such as the U+0024 dollar sign $, with one or two vertical bars…Claims that glyph variants of a certain currency symbol are used consistently to indicate a particular currency could not be substantiated upon further research. Therefore, the Unicode Standard considers these variants to be typographical and provides a single encoding for them.

- The document needs to provide evidence for the characters outside of just Wikipedia.
- For TRANSNISTRIAN RUBLE SIGN, examples showing the sign in plain-text usage, such as on logos or labels from the country, are needed.

Recommendations: We recommend the UTC review this document and send any comments, including those above, to the author. We further recommend the suggested changes to annotations be forwarded to Ken Whistler for review and consideration.

19. Graphics for Legacy Computing

Document: [L2/19-068](#) Proposal to add additional characters into the Graphics for Legacy Computing block -- Renzhi Li

Comments: We reviewed this revised proposal for 8 characters that are used in a block element set called Powerline, employed in certain text editors.

Some Script Ad Hoc members felt these are entities used in a closed system, inappropriate for interchange as characters, and could be handled as PUA characters However, it was noted that discussion at the July 2019 UTC meeting already had signaled that a number of the characters may be eligible for encoding.

Comments made during discussion include the following:

- The document should have page numbers.
- In general, the proposal should include usage, shape, function, and a use case for each of the symbols.
- On the top of page 3, the proposal states: "The UCD properties [will be] similar to other box-drawing characters and block elements." Block elements is the name of a block in Unicode, whose characters have UCD properties. "Block elements" is not a property.
- Regarding specific characters:
    - The name "READ-ONLY SYMBOL" reflects a concept, but Unicode focuses on encoding graphical characters, not concepts. In this case, U+1F512 LOCK should suffice.
    - The VERSION CONTROL BRANCH SYMBOL makes sense and is not otherwise able to be represented. We recommend additional documentation on this symbol be added to the proposal.
    - There is no other way to represent COLUMN NUMBER INDICATOR and LINE NUMBER INDICATOR as units, so these two characters could be useful.
    - In our view, figure 5 does not provide a persuasive case for encoding more triangles and angles and is not necessary. The Powerline triangles and angles can be represented by currently encoded characters with a Powerline-specific font.

Recommendations: We recommend the UTC discuss the proposal and accept the following three characters, with code points in the (new) Graphics for Legacy Computing block:
U+1FBCB VERSION CONTROL BRANCH SYMBOL
U+1FBCC COLUMN NUMBER INDICATOR
U+1FBCD LINE NUMBER INDICATOR.