TO: UTC
FROM: Deborah Anderson, Ken Whistler, Roozbeh Pournader, Lisa Moore, and Liang Hai[1]
SUBJECT: Recommendations to UTC #160 July 2019 on Script Proposals
DATE: 22 July 2019

The Script Ad Hoc group met on 17 May, 4 June, and 5 July 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. Latin
a. LATIN SMALL LETTER TONE SIX
Document: L2/19-201 Proposal to correct LATIN SMALL LETTER TONE SIX (U+0185) -- Juechen Zhong

---

[1] Also participating were Richard Cook, Norbert Lindenberg, Manish Goregaokar, Marek Jeziorek, Craig Cornelius, Ned Holbrook, and Chris Chapman.

Comments: We reviewed this request to change the glyph of U+0185 LATIN SMALL LETTER TONE SIX, a character that dates to Unicode 1.0 and was used to represent the old Zhuang orthography. The examples clearly show the difference between lowercase Latin b and LATIN SMALL LETTER TONE SIX. The author requests the current glyph <b> be modified to be comparable in size to the CYRILLIC SMALL LETTER HARD SIGN (U+044A). The author also asks for a glyph change to U+0184 LATIN CAPITAL LETTER TONE SIX.

In our view, the glyph change for LATIN SMALL LETTER TONE SIX appears justified, but we suggest the glyph be made available for review by the UTC before incorporating the change. A new glyph (with evidence) for U+0184 LATIN CAPITAL LETTER TONE SIX should also be provided.

Recommendations: We recommend the UTC respond to the author and request he provide glyphs for U+0185 LATIN SMALL LETTER TONE SIX and U+0184 LATIN CAPITAL LETTER TONE SIX (with evidence) so the UTC can review the glyphs, before making any changes.

b. Komi Latin Letters
Document: L2/19-224 Komi Latin letters missing in Unicode – Rueter
Related document (not discussed by the Script Ad Hoc):
L2/19-261 Proposal to encode Latin letters for the Komi Latin alphabet (supplementing L2/19-224) -- Pentzlin

Comments: We reviewed L2/19-224, a proposal for 16 letters used in publications of the Komi-Zyrian (ISO 639-3: kpv) and Komi-Permyak (ISO 639-3: koi) languages from the 1930s. The publications are now part of a digitization project in the Komi Republic and Finland. The publications were created during the period when Latin was introduced and briefly used, between periods when the languages were written in Cyrillic. In our opinion, the orthographies shown in this document are transitional, mixing Latin and Cyrillic. In general, it was noted that printed works from this period exhibit poor typography.

The proposal cites the earlier proposal covering the Latin orthographies in the former Soviet Union, which was submitted in 2012 (L2/12-045 = WG2 N4162) and included reasons for not mixing scripts. The Script Ad Hoc comments at the time (L2/13-028) recommended the proposal be re-structured, "identifying which characters are not capable of being represented with Cyrillic characters," and suggested, "the UTC might want to take a position on the practice of mixing of scripts for dead orthographies, particularly in cases, such as those of the FSU, where typography is especially poor and the potential users and scholars have moved on to other orthographies." We recommend a similar approach be taken to this proposal. Note that characters for the other transitional orthographies have, on occasion, been approved, such as four Latin letters for the Sakha (Yakut) language that were proposed in L2/12-044. The four Sakha characters were approved because they were not representable by any characters in Unicode.

We suggest the author revise the proposal to only include those characters that cannot be represented with Cyrillic letters and provide the phonetic values for each letter. Also, the author should add page numbers to the proposal, and move the proposal summary forms to the end of the document.

Recommendations: We recommend the UTC review this document and forward any comments, including those above, to the author. We further recommend the UTC discuss the mixing of scripts in historic, transitional orthographies and depending upon the outcome of the discussion, assign an Action

Item to the Editorial Committee to create a FAQ on the mixing of scripts for historical, transitional orthographies.

c. LATIN LETTER REVERSED HALF H
Document: L2/19-254 Proposal to change the accepted character names for A7F5 and A7F6 – Urupin

Comments: We reviewed this request to change the names for two characters (LATIN CAPITAL LETTER REVERSED HALF H and LATIN SMALL LETTER REVERSED HALF H) that were accepted at the April/May 2019 UTC meeting and accepted at the SC2 meeting in June 2019. The author requests the names be changed to "LATIN [CAPITAL/SMALL] LETTER RIGHT HALF H".

"REVERSED" is found in other names, and the rationale for changing the name is not strong enough, in our opinion, to make the change.

Recommendations: We recommend the UTC thank the author for this input, and forward this document to the author of the proposal for Latin Letter Reversed Half H (L2/19-092), Michael Everson, for consideration.
(Note: Michael Everson has responded that he disagrees with the request to change the characters' names.)

d. Two Characters for Middle Scots
Document: L2/19-180 Proposal to add two characters for Middle Scots – Everson
(Note: The comments below also appear in WG2 Expert Comments N5095.)

Comments: We reviewed this proposal for two Latin characters. In our view, a ligature of <017F LATIN SMALL LETTER LONG S, 0073 LATIN SMALL LETTER S> would suffice, and users need not wait several years. Alternatively, users could employ U+00DF LATIN SMALL LETTER SHARP S (with uppercase 1E9E), making minor modifications to the font, if necessary. No evidence was provided for an uppercase character.

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

e. Tironian letters
Documents: L2/19-172 Proposal to add one or two Latin Tironian letters (WG2 N5042) -- Everson and West
L2/19-199 Response to the proposal to add the Tironinan letters -- Marin Silva
Related documents:
L2/17-359 Proposal to add six Latin Tironian letters
L2/17-384 Recommendations to UTC #153 October 2017 on Script Proposals
(Note: As noted in the Consent Docket L2/19-250, WG2 recommended and SC2 accepted U+2E52 TIRONIAN SIGN CAPITAL ET, but the UTC needs to determine the properties.)

Comments: We reviewed L2/19-172 which contained characters proposed earlier in L2/17-300 and L2/17-359.

The main request of this proposal is to add an uppercase letter TIRONIAN ET, preferably with a casing relationship, either to U+204A TIRONIAN SIGN ET, requiring a change of the General Category property

of U+204A from "Punctuation, other" (Po) to "Letter, lowercase" (Ll) (Option 1), or by adding a new case pair (Option 2). A third alternative, recommended in N4885 (L2/17-367) and L2/17-384 is to add a new non-casing TIRONIAN ET (Option 3).

As noted in this proposal, U+204A TIRONIAN SIGN ET was encoded already in Unicode 3.0 in the Punctuation block, with General Category property Po.
The following are comments from the experts:

- In our opinion, the number of users who would need uppercase is very small, compared to those using U+204A. (Note that U+204A is already described as "an infrequently used character" on page 3 of the proposal.) Hence, the requirement for having automatic uppercasing is not justified. How does not having automatic casing behavior cause a major problem for users? Medievalists would typically be entering a text as it appears on manuscripts or other documents, rather than expecting automatic casing. Search could be tailored to find uppercase characters as needed.
- We agree that adding a new lowercase TIRONIAN ET (Option 2) would cause confusion for users, since the new lowercase character and U+204A would be nearly identical in shape.
- TIRONIAN ET is much like "&": it acts as a mark of punctuation ("hot & cold") and is letter-like (&c.). In Unicode, even though a character may act as a letter in some instances and in other cases as a mark of punctuation, only one value of the partition property General_Category can be assigned to the character, so a decision must be made. (AMPERSAND, like TIRONIAN ET, are both Po.)

In sum, we agree a new character is warranted, and suggest the name TIRONIAN SIGN CAPITAL ET with General_Category=Po and no case mapping. We recommend the code point U+2E52 in the Supplemental Punctuation block, since the earlier U+204A TIRONIAN SIGN ET is already in a punctuation block, and a cross-reference to U+204A TIRONIAN SIGN ET be added.

The feedback from Eduardo Marin was noted.

Recommendations: We recommend the UTC accept TIRONIAN SIGN CAPITAL ET at U+2E52 and discuss the appropriate properties for this character.

2. Palaeohispanic
Document: L2/19-045 New charts for Northern and Southern Palaeohispanic – Ferrer, et al.

Comments: We reviewed the charts for Northern and Southern Palaeohispanic, which were revised after a call between the script expert Noemí Moncunill Martí, Roozbeh Pournader, and Deborah Anderson.

The following summarizes comments made during discussion:
- The revised charts now reflect the graphemes in the writing systems, and our view, the scripts are very close to being ready to encode.
- The authors are encouraged to create a new proposal for the two scripts, with charts and evidence.
- In the revised proposal, provide evidence for S52.

Recommendations: We recommend the UTC review these documents and send comments, including those above, to the authors.
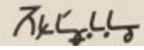

AFRICA
3. Luo Lakeside Script
Document: L2/19-268 Addition of Luo Lakeside Script – Ombewa

Comments: We reviewed this proposal for a recently invented script, which was created between 2009-2012. The script appears to be a limited cursive script with complex shaping behavior: it shows a cursive lowercase, but a disjoint uppercase.

The following comments were raised during discussion:
- Which language(s) is the Luo Lakeside script used for? Provide the ISO-639 codes for the language(s).
- How widespread is its use? Is it currently being taught?
- Change the names for the letters to match the pattern used for Latin, i.e., [script name] CAPITAL LETTER X, [script name] SMALL LETTER X.
- Change the name for the numbers to [script name] DIGIT ONE, etc.
- The preferred name should not be the same as the language ("Luo" in this case), hence "Luo Lakeside" would be acceptable.

- The cover of the book on page 3 shows an initial capital followed by lowercase: How does cursive work in such a case?
- The cursive connections need to be fully described, listing which letters are joined, and whether the baseline connection for the letters is on one side or both.
- Move the Proposal Summary forms to the end of the document.

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

4. Egyptian Hieroglyphs
Document: L2/19-220 Revised draft for the encoding of an extended Egyptian Hieroglyphs repertoire, 1st and 2nd tranche – Suignard

Comments: We reviewed this revised draft proposal, which expands upon the first tranche of Egyptian hieroglyphs in L2/19-052 (containing characters for Humans, Gods, and Goddesses), by adding characters in the categories of Human parts, Mammals, and Mammal parts. The total number of characters in this draft is 3059. The proposal author is seeking expert review.

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

MIDDLE EAST
5. Proto-Cuneiform
Document: L2/19-284 Proposal To Encode Proto-Cuneiform In The SMP Of The UCS -- Hawkins

Comments: We reviewed this draft document, which contains additional statistical information about the repertoire (in §II.1.), not contained in the earlier proposal (L2/17-157).

- The document notes that traditionally the script is thought to include 800 signs, which are published in *Zeichenliste der Archaischen Texte aus Uruk* (*ZATU*), identified as the authoritative list. In our view, it would make sense to encode all of the *ZATU* signs, using a catalog-based naming system, but including the names such as "GI" or "GI6" in the annotations, if this is acceptable to scholars.
- The glyphs should reflect the character's *Gestalt* and be legible in the code chart.
- The numerical system still requires analysis. How do the elements interact? If the numbers have different catalog numbers in *ZATU*, they should be encoded. (More detailed comments are contained in Script Ad Hoc recommendations L2/17-255, pp. 3-4.)
- The goal of encoding should be to make available the sign list for scholars to be able to discuss the signs and to digitally represent texts.
- Provide an image from *ZATU*.
- What font was used to produce the figures 4a and 4b?

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

6. Arabic
Balti
Document: L2/19-252 Proposal to include Balti alphabet ARABIC LETTER GRAF (rev.) -- Lateef Sagar Shaikh

Comments: We reviewed this proposal for one Arabic character used to represent the Balti language in Pakistan. Balti can be written with the Arabic and Tibetan scripts. "GRAF" is the character's name in the Balti language.

This proposal has been seen by the Script Ad Hoc a number of times and was briefly discussed at the Spring 2019 UTC meeting. L2/19-252 answers questions posed by the Script Ad Hoc in L2/19-173.

The author explained offline that the glyph in this latest proposal is based on the Balti Primer that was approved by Pakistan's National Language Authority. The glyph that appeared in the earlier version of the proposal was from a book that was not approved by the NLA.

The codepoint is acceptable. The name follows the same pattern as ARABIC LETTER GAF, and is also acceptable.

Recommendations: We recommend the UTC accept the character U+08C8 ARABIC LETTER GRAF. We also recommend an Action Item be assigned to Lorna Evans to provide a glyph.

SOUTH AND CENTRAL ASIA
7. Chakma
Document: L2/19-218 Proposal to improve alternative names in the Chakma block -- Chakma

Comments: We reviewed this document which provides corrections to the informative aliases in the Chakma code chart. This document provides useful information, although Liang Hai noted that Chakma does not have a standardized transliteration or transcription system.

Because many of the current Chakma aliases are inconsistent and incorrect, we recommend removing them from the code chart and placing the information in another location. One option would be a Unicode Technical Note that would contain the names in the Chakma script and in Latin transliteration. The UTN format would also allow use of combining marks in the names, which is currently restricted in the names lists. A second option would be to move the information to CLDR, whose structure allows for annotations (for both emoji and non-emoji).

Recommendations: We recommend the UTC discuss this document, decide whether to remove the Chakma aliases from the names list, and, if so, then assign an Action Item to Liang Hai to move the information to another location.

8. Gurmukhi
Document: L2/19-283 Feedback on Gurmukhi BINDI and TIPPI – Anderson
Related documents:
L2/19-167 Response to Script Ad Hoc (L2/19-047) re: L2/18-319 on Bindi before Bihari in Gurmukhi – Singh
L2/18-319 Proposal for Bindi before Bihari in Gurmukh – Singh
L2/05-088 Proposed Changes to Gurmukhi – Sidhu
L2/06-030 Proposed Changes to Gurmukhi 4 – Sidhu

Comments: We reviewed this document, which provided more information on Gurmukhi BINDI and TIPPI from two users. The main question was whether the graphical placement of the dot (BINDI) before vowel sign ii on a base (of either a typical consonant letter or the vowel base iri) has a semantic distinction from its placement after vowel sign ii. Not all members of the Script Ad Hoc were convinced the distinction needs to be made in plain text. The placement of the dot could, for example, be represented today using fonts.

Manvir Singh's comment on page 3 that some people pronounce the BINDI differently suggests that this may be a distinction that needs to be retained, and is not just stylistic. More information is required on exactly what the intended pronunciations are and who is making this distinction. In our opinion, a new proposal with additional analysis (including the phonetic values and the conscious orthographical rules for making the phonetic distinction) and manuscript evidence is needed.

Recommendations: We recommend the UTC discuss this topic, and respond to the two users with comments.

9. Oriya
Document: Comments received from David Corbett on L2/19-005 Proposal to encode ORIYA SIGN OVERLINE (feedback received May 7 2019)

> L2/19-005R "Proposal to encode ORIYA SIGN OVERLINE in the UCS" should explain where to put the proposed code point in a syllable. I assume it is either meant to immediately follow the vowel sign or to immediately follow the base consonant.
>
> Either way, Indic_Syllabic_Category=Vowel_Dependent is not appropriate for this character: above-base dependent vowels are encoded between pre- and post-base dependent vowels, so the overline would have to follow U+0B47 ORIYA VOWEL SIGN E but precede U+0B3E

ORIYA VOWEL SIGN AA. It should instead have InSC=Nukta or Syllable_Modifier, depending on where it is meant to go.

The proposal claims that "if the combining macron were to be used, it would not be supported in the general Indic rendering system implementation requirement. If the combining macron were used, script runs in Oriya would break". That is not true. U+0304 COMBINING MACRON is a common-script character so it does not break script runs. There is no Indic rendering system requirement that all marks be Indic-specific. For example, U+20F0 COMBINING ASTERISK ABOVE is used in Devanagari without any rendering system problems. Therefore that argument should be removed from the proposal.

Comments: We reviewed this comment. Indic_Syllabic_Category (InSC) is not yet designed to make a distinction between typical vowel signs and special vowel signs, such as ORIYA SIGN OVERLINE, so Vowel_Dependent is currently the correct choice for ORIYA SIGN OVERLINE. However, InSC values are subject to improvement of additional differentiation when deemed necessary. Also note that the InSC is not directly based on OpenType categories, which may override or otherwise depart from InSC values.

We recommend the author of L2/19-005 update page 2 of her proposal, providing character sequences for all the examples.

Recommendations: We recommend the UTC thank the author for his feedback and request Lorna Evans update L2/19-005 to include character sequences for all the examples on page 2.
(Note: The document has been revised and now includes the sequences, see L2/19-005R2.)

10. Toto
Document: L2/19-278 Introducing the Toto script -- Anderson/SEI

Comments: We reviewed this introductory proposal to the Toto script, which is used by a small group in India near Bhutan.
(Note: An earlier version of this proposal was reviewed by the Script Ad Hoc.)

The following summarizes the comments made during discussion:

- Provide information on how much is the script used today and in what contexts the community is planning to use it in the future.
- Correct the code points in the following character properties' entries:
  ```
  E644;TOTO LETTER BREATHY IU;Lo;0;L;E643 E64F;;;;N;;;;;
  E647;TOTO LETTER BREATHY E;Lo;0;L;E646 E64F;;;;N;;;;;
  E64C;TOTO LETTER BREATHY AE;Lo;0;L;E64B E64F;;;;N;;;;;
  ```
- Provide more information and clarification on diphthongs (cf. Wikipedia article on Toto, citing "The Phonological Study of Toto Language" in Language in India, 14:6 June 2014, which lists examples, such as /maibe/ 'flower'). How are diphthongs written?
- In the verb "to know", the rising tone is connected to the LETTER E (the second letter in the example below). Is the connection intentional or is it a feature of the particular font used here?

  𑵶𑶓𑶒𑶍

  Verb. know

- Verify the names of the consonants when pronounced (such as when reading out or reciting the alphabet), since character names in Unicode typically include a vowel after them, such as TOTO LETTER PE or TOTO LETTER PA. (Latin seems to be an exception, although the character names for Latin still suggest the actual pronunciation of the names of the letters, not pure consonant values.)
- The description of the Toto language in Wikipedia mentions /p^h/ and /t^h/, but no examples are given showing breathiness (or aspiration) after P or T. Does breathiness appear on the consonants or the following vowel? Provide examples.

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

SOUTHEAST ASIA, INDONESIA, AND OCEANIA
11. Balinese
Document: L2/19-259 Proposal to encode Balinese Archaic Jnya -- Yang and Perdana
(Note: An earlier version of this proposal was reviewed by the Script Ad Hoc.)

Comments: We reviewed this proposal for one archaic letter discovered when digitizing a Balinese manuscript for PanLex.

Modern Balinese represents JNYA as a stacked conjunct, with the sequence <BALINESE LETTER JA, BALINESE ADEG ADEG, BALINESE LETTER NYA>. An archaic form on JNYA is not, however, visually decomposable into separate JA and NYA characters. It was noted that the Javanese cognate is separately encoded, JAVANESE LETTER NYA MURDA.

Ample evidence is provided for this character, including evidence from a document containing the two forms of JNYA (figure 2). Discussion of alternative possibilities is also included, as is a discussion of the interaction with combining marks.

Recommendations: We recommend the UTC accept the character U+1B4C BALINESE LETTER ARCHAIC JNYA. We also recommend an Action Item be assigned to Ben Yang to work with the Editorial Committee to incorporate the appropriate changes to section 17.3 of the Core Spec.

12. Tagalog
Document: L2/19-258 The baybayin "ra", its origins and a plea for its formal recognition – Brennan
(Note: An earlier version of this proposal was reviewed by the Script Ad Hoc.)

Comments: We reviewed this proposal for the letter RA used in Tagalog (as known as Baybayin, the local name for the script). Whereas the RA character is encoded in two other Philippine scripts (Buhid and Hanunoo), no RA character was encoded in either Tagalog or Tagbanwa, although holes were left for it in the code charts.

The proposal provides ample evidence and justification for encoding TAGALOG LETTER RA in Tagalog. Indeed, the proposer notes that the spot left open for RA in Tagalog, U+170D, is already being used in modern fonts for RA and the character appears on Philippine currency. The proposal also provides evidence for an historic form of RA, sometimes called the "Zambales" form of RA. The Script Ad Hoc

recommended the historic form, which has a distinctive shape, be separately proposed, which has been added in the latest version of the proposal.

Two questions arose during discussion: Does Tagbanwa use a RA today? Also, are there other historic characters with radically different shapes, similar to ARCHAIC RA?

Recommendations: We recommend the UTC accept U+170D TAGALOG LETTER RA and also consider U+171F TAGALOG LETTER ARCHAIC RA, after reviewing the proposal.

Note: Offline the proposal author responded that according to his research, Tagbanwa appears to be chiefly an historic script today. Amongst those who are trying to revive the script, RA has not attracted much interest. Regarding other forms of Tagalog letters, he reports only TAGALOG LETTER YA has a slightly different shape than the code chart. Many in the community no longer draw it with a loop, since it could be confused with TAGALOG LETTER PA.

13. Western Cham
Documents: L2/19-217 Proposal to encode Western Cham – Hosken
Reference document: L2/16-198 Proposal to encode Western Cham in the SMP (WG2 N4734) -- Everson
(Note: An earlier version of L2/19-217 was reviewed by the Script Ad Hoc.)

Comments: We reviewed this proposal for Western Cham, which represents disunification from Eastern Cham, the script already encoded in Unicode as "Cham". This proposal was based on the earlier proposal L2/16-198. As stated in L2/16-198 and this proposal, the font style and glyphs generally differ between Western and Eastern Cham, so readers of one script are reportedly not able to read the other, prompting a request for separate encoding of Western Cham.

The following comments were made during discussion:

- Provide a chart showing the differences between Eastern and Western Cham. This can demonstrate the unintelligibility between the two scripts (cf. page 3 of L2/16-198).
- Confirm whether there are Western Cham font implementations using Eastern Cham code points. If so, it might be preferable to keep the same encoding model for the two scripts.
- Add glyphs after the code points in the Repertoire section (i.e., for FINAL G, FINAL NG, etc.). This makes it easier for reviewers to read the text.
- FINAL consonant letters appear to have a strong pattern of just extending an ordinary consonant letter's top-right hook longer downwards, which suggest the potential of encoding them as a sequence <letter, sign>, especially considering that this pattern might be productive. Instead of simply following the current Cham model in Unicode (for Eastern Cham), a conscious analysis of the encoding principles should be presented. Another consideration is that encoding such dependent structures as atomic characters can simplify text segmentation, which is especially tricky for such writing systems that do not use inter-word spaces.
- Because Eastern Cham encodes script-specific dandas, we recommend Western Cham dandas also be encoded.
- Rename LUNAR 5 DOTS OVER HEH to have a non-Arabic name.
- Consider re-naming WESTERN CHAM CONSONANT SIGN FINAL M to WESTERN CHAM SIGN ANUSVARA, and WESTERN CHAM CONSONANT SIGN FINAL H to WESTERN CHAM SIGN VISARGA.

However, it was noted that Eastern Cham employs the name FINAL M and FINAL H for the corresponding signs in Eastern Cham.

- Are there other Arabic communities using the five textual symbols (U+1E260..U+1E264)? If so, we would suggest moving the symbols to a new Arabic block where the characters do not have joining behavior (cf. Arabic Mathematical Alphabetic Symbols).
- Add the following confusables:
  - o U+1E267 WESTERN CHAM SIGN LUNAR 3 DOTS is confusable with U+0865 ARABIC LETTER JEEM WITH THREE DOTS ABOVE, which was recently accepted for Unicode 13.0
  - o U+1E268 WESTERN CHAM SIGN LUNAR 4 DOTS is confusable with U+0690 ARABIC LETTER DAL WITH FOUR DOTS ABOVE.
- Add examples and references provided by Debbie Anderson to supplement the evidence already included, with comments on how the text should be represented.

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

EAST ASIA
14. Hangul Jamo
Document: L2/19-230 Proposal to add the "6 letters" (Hangul Jamo) -- Obayashi Muneaki
Feedback:
L2/19-233 Comments on L2/19-230R (six hangul jamo letters) -- Jaemin Chung
L2/19-234 Oppose the L2/19230 document of Hangul Letters -- Sim CheonHyeong
L2/19-248 Comments on Mr. 小林 Muneaki's newly preparing document (hangul) -- Sim CheonHyeong
L2/19-251 Comments on L2/19-230R (six Hangul jamos) -- Kyongsok Kim
L2/19-253 My idea about the comments of 2 Drs. (hangul jamos) -- Obayashi Muneaki
L2/19-255 Comments on L2/19-253 (hangul jamos) -- Jaemin Chung
Note: L2/19-248 is a response to L2/19-253 (which was originally a tweet)

Comments: We reviewed this proposal for six Hangul Jamos, as well as various documents submitted in response to the proposal. The six characters were created as part of the New Korean Orthography in 1948, which was later abandoned.

There appears to be some academic interest in encoding the characters, as the author cites the characters' appearance in theses about the DRPK orthography. In order to properly evaluate the request, we recommend the author provide evidence of usage, citing the dates of usage, and provide scans of such documents with English translations.

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

15. Lisu Monosyllabic
Document: L2/19-208 Proposal to Encode the Lisu Monosyllabic Scripts (L2/19-208) – China NB
Note: This proposal was discussed at WG2 and comments based on that discussion have been sent to the proposers. The following is FYI for the UTC.

Comments: We reviewed this new proposal for 989 characters.

The following were raised during discussion:

- Provide background on the two documents appended to the proposal (i.e., starting pp. 76ff. and p. 91ff.), and their relation to one another.
- How are the characters ordered? Is there a native order or Romanized order? The order on page 76 appears to be left-to-right, but the document starting on p. 91 has the characters ordered vertically. Which is the traditional order?
- Because of the large repertoire, we recommend the proposal be reviewed carefully by experts.
- The glyphs in the "Lisu letter" column in the charts on pages 3-4 need to be embedded in the PDF to facilitate accurate review of the document.

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

16. Seal
Documents:
L2/19-228 Proposal to encode Small Seal Script -- TCA, China NB (Appendix Zip file of charts here)
L2/19-245 (=N5108) Report of Seal Script discussion in WG2 #68 -- suzuki toshiya
Related document:
N5089 Shuowen Seal Informal Meeting Report - Selena Wei et al.

Comments: We briefly reviewed the proposal for 11,093 characters and the two meeting reports on Seal.

The proposal focuses on a single source, whereas the approach in N5089 shows a multi-column chart (pp. 2-3) and extensive discussion of issues, such as how to handle duplicates (Appendix 1), etc.

The following points were raised during discussion:

- Encourage Richard Cook to submit a document to the UTC, describing what he believes should be done differently from L2/19-228 and next steps, including fruitful approaches for the upcoming meeting in Taiwan from 30 September – 1 October 2019 (such as an early posting of the agenda).
- Recommend the proposers use the script name "Seal", since the script covers more than just Small Seal. "Seal" includes variant styles.
- Suggest the proposers clarify the principles used in encoding Seal characters, using ca. 12 characters as examples. For example, provide clear cases of 2 or 3 characters that need to be distinguished (with rationale). Similarly, provide examples showing cases where characters were unified, and why they were unified.

Recommendations: We recommend the UTC review these documents and discuss them.

SYMBOLS, PUNCTUATION, MUSICAL AND NUMERICAL NOTATION
17. Jianzi Musical Notation
Document: L2/19-107 Preliminary proposal on encoding Jianzi Musical Notation and Jianzi Format Controls -- Culture and Art Publishing House / Eiso Chan

(Note: The following comments also appear in WG2 Expert Comments N5095)

Comments: The proposal includes a total of 357 characters (with 21 format controls). The notation uses components that are put together in a manner similar to ideographs. The proposal was discussed at the May 2019 IRG meeting, but the IRG decided Jianzi is out of scope, except for confusables (with CJK).

The notation is still being analyzed. A document focused on the structural analysis of the script and complete rationale behind the proposed encoding model is recommended.

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

18. Znamenny Musical Notation
Document: L2/19-053 Proposal to Encode Znamenny Musical Notation in Unicode (rev. April. 2019)-- Andreev and Simmons

Comments: This proposal has had several rounds of review, both in the Script Ad Hoc meetings and at the January 2019 and April-May 2019 Unicode Technical Committee meetings.

Two comments were made during discussion:

- In figure 21, it was noted that the original OpenType spec for applying marks to separate members of ligatures was incompletely specified. Hence, if two combining marks were to land alternatively on the second members of the ligature, or if the marks were divided between members of the ligature, the implementation may not work completely as expected. The way to get around this would be to not form a ligature in the third example, but allow the adjacent base glyphs to touch. This is an approach used by some Arabic fonts.
- It would be helpful to make the font available for testing, such as to Andrew Glass, MS, to see if there are any functional gaps.

Recommendations: We recommend the UTC approve 186 characters in a new block, Znamenny Musical Notation, from U+1CF00..U+1CFCF, and relay the comments above to the authors.

Note: Offline the proposal authors noted the font is available for testing on Github.

II. OTHER TOPICS
19. Error Report from Corbett on Identifier_Type=Obsolete characters
Document: Error Report from David Corbett (from L2/19-124 Public Review Issues)

Comments: The Error Report from David Corbett (from L2/19-124 Public Review Issues) was briefly discussed at the April-May 2019 meeting, but assigned to the Script Ad Hoc. The Error Report listed characters that Mr. Corbett deemed should have Identifier_Type=Obsolete.

In our opinion, this task falls outside our mandate, and should instead be coordinated with the ICANN effort for security profiles for IDNA. The topic properly belongs with a discussion on UTS #39.

Recommendations: We recommend the UTC take up this topic in the context of the maintenance of UTS #39 and in particular, identifier status and data type (obsolete, historic, etc.).

20. Space Character as Group Separator
Document: [L2/19-112](#) Proposal to define a space character as a group separator – Schneider

Comments: We reviewed this document, which discusses the non-breaking thin space. The proposed actions appear on page 8.

The following comments were made during discussion:

- The Unicode Standard includes several different space characters. It is acceptable to use THIN SPACE, and then use a higher-level protocol, a word joiner, or tailored line-breaking to prevent line-breaking. Alternatively, NNBSP can be used to force non-breaking behavior.
- CLDR inconsistencies should be directed to CLDR-TC.

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