

tTO: UTC

L2/14-268

FROM: Deborah Anderson, Ken Whistler, Rick McGowan, Roozbeh Pournader, Laurentiu Iancu, Andrew Glass, Peter Constable, and Michel Suignard

SUBJECT: Recommendations to UTC #141 October 2014 on Script Proposals

DATE: 24 October 2014

The recommendations below are based on documents available to the members of this group at the time they met, and do not include documents submitted later to the document registry.

SOUTH AND CENTRAL ASIA

Indic

1. Grantha

Documents:

[L2/14-020](#) Plain-text ligating virama representation for Grantha script –Ganesan

[L2/14-097](#) ZWJ Joiner for Chillu Consonants of Grantha Script – Ganesan

[L2/14-110](#) Comments on L2/14-097 re using ZWJ for Grantha "chillu" - Sharma

[L2/14-162](#) Control Characters (Joiners ZWNJ and ZWJ) in the Grantha Visible Virama and Chillu Consonants – Ganesan

[L2/14-164](#) Chillu examples – Ganesan

L2/14-XXX ZWJ for Grantha pre-pausal half-consonants (chillu) –Ganesan

Discussion: We reviewed the summary document by Naga Ganesan (L2/14-XXX), Shriramana's response to L2/14-097 (L2/14-110), and other documents.

While Shriramana maintains that the archaic consonant-virama ligatures (represented by Consonant + Virama + ZWJ) reflect presentation shapes of different orthographies (and not a semantic difference), we feel that if a user decided to make the visual distinction with ZWJ, they shouldn't be prevented from doing so, although they may run into problems, since there would be multiple means of showing the distinction.

Recommendations:

We recommend the UTC discuss this, and consider including text in the standard that states that ZWJ *can* be used, but its use may cause problems (with details) and hence should be employed with caution.

2. Tamil and Grantha

[L2/14-218](#) Unification of Tamil and Grantha numerals – Sharma

Discussion: We reviewed [L2/14-218](#). As noted by Sharma, the report from the meeting in India on Grantha, [L2/10-409](#), explicitly recorded the recommendation that "Tamil and Grantha share the same set of digits, numbers and fractions was accepted by all without debate. Digits 0-9 and numbers 10, 100 and 1000 are encoded in the Tamil block and can be used from the Tamil block. Experts affirmed that there is no need to separately encode Grantha numerals and the Tamil numerals (U+0BE6 to U+0BF2) should be used for Grantha." (Subsequent compromise code charts for Grantha did not include separate digits, except for the cantillation marks.)

Recommendations: We recommend the UTC review this document and discuss it.

3. Tamil

Documents:

[L2/14-210](#) Letter on Tamil Fraction Naming - Tamil Virtual Academy

[L2/14-212](#) Tamil names and annotations - Vasu Renganathan / INFITT

[L2/14-215](#) A Proposal as a Standardised Romanisation Scheme for Full Tamilwords Used Inside Code Pages as in Naming of Various Characters Etc & In CLDR – Logasundaram

[L2/14-216](#) Current status of Tamil symbols naming issue (W2 N4622) - Sharma

Discussion: We reviewed the various documents relating to Tamil names. As reported by Sharma [L2/14-216](#), a committee, commissioned by the Government of Tamil Nadu, met in July 2014 and came to an agreement on the transliteration, documented in [L2/14-210](#). The recommendations were then forwarded to the relevant department. However, the recommendations are still in bureaucratic limbo.

Recommendations: We recommend the UTC discuss the situation, and consider how to help progress the proposed name changes.

4. Malayalam

[L2/14-003](#) Proposal to encode Malayalam Anusvara Above - Sharma

[L2/14-029](#) Feedback on Malayalam Anusvara Above Proposal – Cibu

[L2/14-069](#) Evidence for considerable usage of the Malayalam anusvara above – Sharma

Discussion: We reviewed the documents relating to Malayalam Anusvara Above. This character appears in material that predates the time when Malayalam is considered a distinct script, which raises a number of questions: Is it meaningful to include this character in the modern script? If we encode Tulu and other characters ancestral to Grantha and Malayalam, will that present a problem? Would it be better to encode this character rather than have users introduce a hack into Malayalam script from some other script?

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

Central Asia

5. Nepaalalipi/Newar

Background docs:

[L2/14-220](#) Comparison between Newar and Nepaalalipi proposals – Anderson, SEI

[L2/14-221](#) Comparison between Ranjana proposals – Anderson, SEI

[L2/14-253](#) Recommendations to UTC from Script Meeting in Nepal – Anderson

([L2/14-258](#) Comments on the Recommendation for Nepalese Scripts – Pandey)

Discussion: We reviewed these documents.

A number of comments and questions were raised:

- The native names for Sinhala characters have created problems for implementers, so to speed implementation, adoption of the names used elsewhere (i.e., ANUSVARA, VISARGA, etc.) is highly recommended. The native names can be added in annotations.
- We recommend the option 2 for the independent vowels (encode vowel letters with combining formants, such as VISARGA), as this will also speed implementation. For input preferences,

recommendations can be made on the national level, though it is important to verify such input recommendations are agreed to by typical users.

- The order for the retroflex consonants in the “Recommendations” can be handled by collation tailorings
- On the spelling “NJ” vs. “NY”: to provide consistency in the standard, “NYA” is preferable
- On the spelling “V” vs. “W”: “W” is acceptable
- A script-specific character ABBREVIATION SIGN (without “CIRCLE”) seems warranted, as similar characters with the name “ABBREVIATION SIGN” are already encoded in other scripts (e.g., Kaithi, Mahajani, and Sharada)
- Before encoding an ABBREVIATION SIGN CROSS, examples in text are needed to see whether U+00D7 MULTIPLICATION SIGN might suffice
- The encoding of six “breathy consonants” will require discussion with the UTC

Recommendations: We recommend the UTC review the document, alongside a document from Whistler and the response document from Pandey ([L2/14-258](#)), which was not seen by the script group when it met.

6. *Ranjana*

[L2/09-192](#) Preliminary proposal for encoding the Rañjana script in the SMP (WG2 N3649)

[L2/14-221](#) Comparison between Ranjana Proposals - Anderson

[L2/13-243](#) Proposal to Encode Ranjana Script - Manandhar

[L2/14-253](#) Recommendations to UTC from Script Meeting in Nepal - Anderson

Discussion: We discussed these documents. Since decisions on the repertoire and encoding model for Ranjana depend upon those for “Nepaalalipi”, discussion on Ranjana was limited. It was noted that a future Ranjana proposal should also discuss the unification with Warty and Lanydza, and should provide details on any specific characters and behaviors of the script in Tibet and other locations outside Nepal.

Recommendations: We recommend the UTC review the document, but postpone discussion until after the “Nepaalalipi” encoding is resolved.

7. *Bhujinmola*

[L2/14-253](#) Recommendations to UTC from Script Meeting in Nepal

Discussion: We briefly discussed the section in the “Recommendations” on Bhujinmola. Bhujinmola has a characteristic wavy headline (see examples in “Roadmapping the Scripts of Nepal” [L2/09-325](#)). The question on whether Bhujinmola represents a stylistic variation of “Nepaalalipi” or should be separately encoded needs to be discussed in a separate document, with examples of how vowels and consonants join differently from “Nepaalalipi” and other rendering issues.

Recommendations: We recommend the UTC review the document, but wait for further research to support separately encoding Bhujinmola.

EUROPE

8. *Greek*

[L2/14-185](#) Proposal to change glyph for small GREEK LETTER YOT – Bobeck

([L2/14-255](#) Feedback on L2/14-185 Proposal to change glyph for small GREEK LETTER YOT – Anderson)

Discussion:

We discussed document [L2/14-185](#). We understand Deborah Anderson is submitting a new document on this topic ([L2/14-255](#) Feedback on L2/14-185 Proposal to change glyph for small GREEK LETTER YOT), which was not available during discussions.

Recommendations: We recommend the UTC discuss the document [L2/14-185](#) alongside the feedback document [L2/14-255](#).

AFRICA**9. *Adlam***

[L2/14-219](#) Proposal for encoding the Adlam script in the SMP – Everson

Discussion: We reviewed the documents. During the last UTC, discussion focused on two problematic issues: whether the script is joining and whether it is bicameral. There was also a question as to whether Adlam was one script or two (or more).

Recommendations: We recommend the UTC discuss the proposal and the outstanding issues.

AMERICAS**10. *Osage***

[L2/14-214](#) Final proposal to encode the Osage script (WG2 N4619) - Everson

Discussion: We discussed the Osage proposal, which WG2 recommended be put onto a ballot.

A number of questions and issues were raised:

- How should users encode existing text that was not originally bicameral?
- Nasal diphthongs appear in text as underlined characters (e.g., figure 10), but were later represented with an intrinsic dot (page 3), based on 2014 reforms to the orthography. Are the underlined characters part of an earlier orthography – and hence need to be supported – or are they a tentative orthography which doesn't need to be supported?
- The glyphs in the code chart have serifs, but the proposal shows no evidence of serif usage.
- The block introduction for the script should discuss input for the vowels, e.g., should a single key generate the base letter with combining dot, and a different key for acute?

Recommendations: We recommend the UTC discuss the points raised above, and request the glyphs for the code chart be changed to sans serif to match usage (such as the glyphs shown in figure 13).

EAST ASIA**11. *Tangut***

[L2/14-209](#) Tangut glyph corrections– West et al.

(Related: [L2/14-246](#) Ad hoc reports for Tangut and Khitan Large Script – Anderson)

Discussion: We discussed L2/14-209, which documented 61 glyph corrections, the reordering of 63 characters, and the addition of 1 character. We note that experts from China participated in the review, as did experts from other countries.

Recommendations: We recommend the UTC discuss the proposal, and decide what to do.

12. Tangut Radicals

[L2/14-228](#) Proposal to encode Tangut radicals in the UCS (WG2 N4636) – West
(Related: [L2/14-246](#) Ad hoc reports for Tangut and Khitan Large Script – Anderson)

Discussion: We discussed the proposal for Tangut radicals, which are in some respects more like components (about two-thirds are used as radicals, the other third appear as components on the right-hand side of the ideographs).

The Tangut Radical names list ([visible in the draft repertoire for PDAM 2.2](#), L2/14-271) includes annotations on the use of a given component in various characters. For example, for U+18900, the annotation reads: “used for 17000 .. 1702F”. Any change in the annotation would require hand-editing by the Editor, which could possibly introduce errors in the standard. Rather than include such information in the names list, in our view a better approach would be to have a Unicode Technical Note (or some other vehicle) containing such information, rather than the names list.

Recommendations: We recommend the UTC members discuss the proposal, and also consider making a ballot comment regarding the names list.

13. Khitan Large Script

[L2/14-234](#) Proposal on Encoding Khitan Large Script – China
[L2/14-233](#) Preliminary Review of Proposal on Encoding Khitan Large Script – West
[L2/14-246](#) Ad hoc reports for Tangut and Khitan Large Script – Anderson

Discussion: We reviewed these documents. As noted in [L2/14-233](#), the Khitan Large Script is largely undeciphered without any character list or recent dictionaries, vocabulary lists, or secondary linguistic materials, so the current proposal should be viewed as preliminary.

Also as mentioned in [L2/14-233](#), the script appears to have a significant percentage of characters (18%) that are either Han borrowings or identical in shape to already encoded CJK ideographs. A revised proposal should discuss the pros and cons of unifying those Khitan Large Script characters with CJK characters already encoded: what are the costs/benefits to unification? Because Khitan Large Script is an historical script, the security risk would not arise if Khitan Large Script used CJK characters, only if it encoded a large set of identical CJK characters.

Additionally, we suggest the proposal also create a “Uni-Khitan” database (or spreadsheet) to document sources.

Recommendations: We recommend the UTC members discuss these documents.

14. Nushu

[L2/14-236](#) Comments on Nushu in ISO/IEC 10646:2014 PDAM2 – Suzuki
[L2/14-244](#) Theory and Rules of Character Unification (in Nushu) – China
[L2/14-247](#) Nushu ad hoc report – Anderson

Discussion: We reviewed the documents. The WG2 Nushu ad hoc report ([L2/14-247](#)) relayed no substantive progress on the issues raised by Suzuki ([L2/14-236](#)), with the result that the script remained

in PDAM2. However, at WG2, experts were encouraged to review Suzuki's document and submit comments. Because the next WG2 meeting will take place in Japan, it is hoped that a face-to-face meeting between experts can resolve the remaining questions posed in [L2/14-236](#).

Recommendations: We recommend the UTC members review these documents. Because of the outstanding questions on the script, we recommend the UTC include ballot comments asking for the script to not progress to a DAM ballot.

15. *Small Seal Script*

[L2/14-242](#) Proposal to encode Small Seal Script – TCA and China

Discussion: We reviewed this proposal, which proposes 799 characters out of a projected 10,516. In our opinion, the proposal is still far from mature, and would benefit from coordinating work with experts in the U.S. and Japan in order to formalize mapping data, which is needed to evaluate a final proposal. The proposal should also provide demonstrated need for including the script in the international standard.

Recommendations: We recommend the UTC members review this proposal and consider sending the authors the comments above.

16. *Naxi Dongba*

[L2/14-241](#) Supplement on Proposal for Encoding Naxi Dongba Pictograph Script ([L2/11-178](#)) - China

[L2/14-245](#) Feedback on Naxi Dongba Supplement document - Anderson

Discussion: We reviewed the "Supplement" document, which answered questions posed at the June 2011 WG2 meeting in Helsinki, Finland (see Naxi Dongba Ad Hoc report, [L2/11-244](#)). Specifically, the authors in the "Supplement" confirmed that the encoding is for modern use, not traditional use of the characters, and that alphabetical ordering is preferred.

The "Feedback" document posed additional questions and made suggestions. During WG2 discussion, the Naxi Dongba proposal authors stated the script is both a logography and syllabary, and the variation shown in some glyphs is due to regional differences, but only one glyph per character is warranted in the encoding. They agreed to revise the proposal and provide information on the proposed characters, with glyphs, Romanized transcription, Chinese gloss (and English translation) and references.

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

17. *Shuishu*

[L2/14-243](#) Proposal for encoding Shuishu – China

Discussion: We reviewed this proposal, which is still at an early stage. In our view, it is not yet clear that Shuishu is an encodable writing system. In order to move forward, we recommend the authors prepare and publish a standard sign list for Shuishu, which can then be circulated for review by other scholars and gain scholarly support. The next version of the proposal should also provide a rationale for the digital representation of their sign list, answering the question why these shapes should be put into an international character encoding standard.

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

authors. The UTC may want to relay the suggestions to the authors above, regarding recommended next steps.

18. CJK Extension F

[L2/14-248](#) Proposal for CJK Unified Ideograph Extension F - IRG (WG2 [n4580SummaryForm.pdf](#);

Discussion: We discussed Extension F. (Extension F was formerly “F1”, since “F” was broken into two parts to speed review. “F2” is now Extension G.) As mentioned in the IRG meeting #42 report at the recent WG2 meeting (see WG2 N4581), the SAT project pulled 49 characters at the last moment. The 49 are documented in the supplementary files for [L2/14-248](#) as [14248-n4580SATWithdrawnCharacters.xls](#).

Recommendations: We recommend the UTC discuss Extension F. The UTC may wish to consider waiting until Extension G is complete before approving Extension F, so one large set of characters is published altogether. Since IRG decided to remove the 49 characters, the UTC should consider supporting this position in PDAM ballot comments.

SYMBOLS

19. Emoji

Modifiers and Portrait symbols

[L2/14-213](#) Skin tone modifier symbols – Unicode/Edberg

[L2/14-226](#) Proposal to encode Portrait symbols - Everson

[L2/14-227](#) Proposal of Tone Modifier Symbols for Emoji – Suzuki et al.

Discussion: We reviewed these documents.

Recommendations: We recommend that the UTC discuss these documents with other emoji issues.

20. [L2/14-229](#) “Afroji” Emoji Symbols Proposed for Encoding – White

Discussion: We reviewed this document.

Recommendations: We recommend that the UTC discuss this document with other emoji issues.

21. *Playing Cards*

[L2/14-223](#) Playing Card Variation Selectors – Davis

Discussion: We reviewed this document. We feel the subject requires a document with more detailed discussion, including identifying what the intended goal is.

Currently, Unicode has the playing card suits encoded in Miscellaneous Symbols, (U+2600..U+2667) and a separate block for Playing Cards (U+1F0A0..U+1F0FF). The playing card suit characters don’t all directly match up with the characters in the Playing Cards block, so there is no clarity about what is already encoded and their relations to one another, without even discussing multiple variations of presentation.

Recommendations: We recommend that the UTC discuss this document, taking into consideration the comments above.