TO: UTC   L2/14-129
FROM: Deborah Anderson, Ken Whistler, Rick McGowan, and Roozbeh Pournader
SUBJECT: Recommendations to UTC #139 May 2014 on Script Proposals
DATE: 2 May 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
1. L2/14-086 Proposal to Encode Nepaalalipi Script – Dev Dass Manandhar, et al.

Discussion: We reviewed this document, which differs from the earlier version of the Nepaalalipi proposal by the same authors (L2/12-349) in containing 37 fewer characters (including all the dependent vowels). The proposal mentions use of a Nepaalalipi-Non-Space-Joiner and a Nepaalalipi-Space-Joiner, though they are not in the set of requested characters. The proposed encoding model, which uses the Nepaalalipi-Space-Joiner to transform C + independent vowels into C + dependent vowel is radically different from encoding model used for most of the closely related North Indic scripts. (The proposal by Pandey, L2/12-003, follows the model used for most North Indic scripts.)

Recommendation: We recommend the UTC respond to the authors that the document is under review.

2. L2/14-091 Final Proposal to Encode the Bhaiksuki Script – Anshuman Pandey

Discussion: We reviewed this proposal, which builds on the earlier version (L2/14-036), seen at the last UTC meeting. In this final proposal, Pandey has modified certain vowel sign glyphs, and added a gap filler character (§3.14) and two digits (§3.15). (With the addition of the two digits, the "Nd" general category was retained.) Pandey also provided detailed information on the contextual forms of vowels signs (§3.10), along with more information on the digits (§3.16). The discussion about two-part vowels was removed.

Recommendation: We recommend the UTC review this proposal, in particular the new sections, and, after discussion, approve it.

3. L2/14-083 Proposal to include two Characters in the Bengali Bloc – Bangladesh ICT Division

Discussion: We reviewed this proposal for a script-specific danda and double danda characters. The proposal provides no evidence of contrastive use, so there is no apparent justification for encoding script-specific danda characters in Bengali.

Recommendation: We recommend the UTC request an annotation in the Bengali block names list for U+09E4 and U+09E5, identifying the name of the danda in Bangla as "darhi". (Currently, U+094E and U+094F have only cross-references to U+0964 DEVANAGARI DANDA and U+0965 DEVANAGARI DOUBLE DANDA.)

4. L2/14-097 ZWJ Joiner for Chillu Consonants of Grantha Script – Naga Ganesan

Discussion: We reviewed this document, which asked that text be added to the standard, spelling out

that the Grantha pre-pausal chillu be represented by the sequence <consonant, virama, ZWJ> in order to "remove ambiguity in plain-text representation of Chillus."

We note that Grantha doesn't follow the model of Malayalam regarding the contrastive usage of chillus (cf. GoI report L2/09-375 and L2/10-409, as well as several documents from S. Sharma). At UTC meeting #138, the UTC discussed S. Sharma's L2/14-002 "Finalizing the virama model for Grantha", and decided to incorporate information into the Grantha block introduction based on L2/14-002. Sharma recommended chillus in Grantha be represented in a manner similar to the way other vowelless consonants are represented in Telugu, Kannada, and Bhaiksuki – namely, in the font, and in a manner that differs from the sequence recommended by Ganesan in L2/14-097.

Recommendation: We recommend the UTC discuss this, and seek other expert opinions.

SOUTHEAST ASIAN
5. L2/14-090 Proposal to Deprecate and add 4 characters to the New Tai Lue block – Martin Hosken

Discussion: We reviewed this document, which puts forward three options on how to deal with New Tai Lue, a script whose Unicode encoding model – which prescribes logical order -- has not yet been widely implemented. At present, users tend to still use legacy fonts or a Unicode font in which the characters are stored in visual order, probably influenced by Thai. Because the script is still not widely implemented, the proposal author suggests this would be an opportune time to consider a change in the encoding model.

Three options are offered:
1. Keep the current model with logical order, i.e., pre-vowels are stored after the initial consonant
2. Deprecate four characters and introduce four new characters
3. Change the model to visual order, i.e., pre-vowels are stored before initial consonant. Martin Hosken also asks all the vowel signs (19B0..19C0) and the two tone marks (19C8-19C9) have their General Category changed from Mc to Lo and that New Tai Lue collation be modified to one similar to that for Thai.

Martin Hosken favors option 3. In our view, either option 1 or 3 would be viable. Option 2, in which characters would be deprecated and then added, would run the significant risk of making the situation for representation of New Tai Lue text even more complicated and confusing than it currently is.

Recommendation: Since the request involves a change of encoding model, we recommend the UTC review this proposal, and solicit feedback from members for whom such a change would have an impact on implementations and data. We recommend the UTC decide between options 1 or 3.

EUROPEAN
6. Glagolitic
Documents:
L2/14-087 **Proposal to Encode Additional Glagolitic Characters – Aleksandr Andreev, et al**
L2/14-103 **Expert Feedback on L2/14-087 Proposal to Encode Additional Glagolitic Characters – Cleminson and Birnbaum**

Discussion: We reviewed both documents, the proposal itself (L2/14-087) and a feedback document from two experts (L2/14-103).

The proposal has three main requests:

1. The addition of 38 combining Glagolitic superscript characters, described as analogous to the combining superscript letters in Cyrillic, to be included in a new Glagolitic Extended block.

2. A change in Scripts.txt for U+0487 COMBINING CYRILLIC POKRYTIE and U+A66F COMBINING CYRILLIC VZMET from "Cyrillic" to "Common" and an added annotation in the names list indicating these two characters are also used in Glagolitic.

3. The addition of a case pair for GLAGOLITIC LETTER BROAD ONU proposed in the two open spots in the Glagolitic block (at U+2C2F for the CAPITAL LETTER BROAD ONU and U+2C5F for the SMALL LETTER BROAD ONU).

In our view, the examples in the proposal do not make a sufficient case for considering the proposed superscript characters to be plain text.

The request to change Scripts.txt to "common" for the two overline characters, COMBINING CYRILLIC POKRYTIE and COMBINING CYRILLIC VZMET, seems reasonable, as does the request to modify the characters' annotations accordingly, as long as the characters are in use.

Regarding the proposed two new BROAD ONU characters, we agree with Cleminson and Birnbaum that more information on these characters seems warranted.

Recommendation: We recommend the UTC agree to the second request (i.e., make the changes to Scripts.txt and add the annotations for two characters) and discuss the other requests. (Note that if the UTC agrees to request #2, the two combining characters should be added to ScriptExtensions.txt.)

MIDDLE EASTERN
7. Arabic script: Quranic marks
Documents:
L2/14-095 **Proposal to encode Quranic marks used in Quran published in Pakistan – Lateef Sagar Shaikh**
L2/14-096 **Proposal to encode Quranic Alternate Dammatan used in Quran published in Pakistan – Lateef Sagar Shaikh**
L2/14-105 **Proposal to encode fourteen Pakistani Quranic marks – Roozbeh Pournader**

Discussion: We reviewed these three documents. The proposals by Lateef Sagar request 15 characters: 14 in L2/14-095 and 1 character in L2/14-096. The proposal by Roozbeh Pournader asks for 14 characters, to be located in the Arabic Extended-A block (from 08D5..08E2).

Recommendation: We recommend the UTC approve Roozbeh Pournader's proposal, after discussion of the differences between the proposals.

8. L2/14-104 **Supporting the Warsh orthography for Arabic script – Lorna Evans**

Discussion: We discussed this document, which outlined a problem in representing the Warsh orthography. The Warsh orthography, used widely in North and West Africa, shows a different patterning of dots on the skeleton of the letters, compared to other orthographies (such as Hafs and Al-Duri, as shown in the chart on page 1 of the document). While the other two orthographies can use a single code point to represent different forms of the letter feh, qaf, and noon, this is not possible in Warsh.

Document L2/14-104 asks for guidance on how to support the Warsh orthography, with three options:
1. Use different code points, depending upon where the character appears in a word (status quo)
2. Use variants (in a font) to represent the isolate and final dotless forms and medial dotted forms
3. Add three new characters to support the Warsh character requirements

In our opinion, the first option is the most advisable.

Recommendation: We recommend the UTC review this document and suggest the current practice, option 1, be maintained.

SYMBOLS
9. [L2/14-085](#) **Request Approval to add "Our New iDiversicons: Diverse Emoji Characters" – Katrina Parrott**

Discussion: We reviewed this document, which comes from Cub Club, a group that has a mobile application, iDiversicons, featuring over 900 diverse emoticons representing the "entire world of faces."

The document states that "Cub Club would like to work with the Unicode Consortium in an effort to update the standard with our images" and formally requests "our diverse characters…be included with the proposals to address [diversity]."

Recommendation: We recommend the UTC discuss this document and respond to the authors.

10. [L2/14-092](#) **Proposal to add a Gorilla Glyph – Kenneth Berman**

Discussion: We reviewed this document, which was requested by the President of GorillaTrades Inc.

Recommendation: We don't recommend the UTC accept this character, which has no established usage and no examples provided in plain text.