

DATE: 24 September 2016
DOC TYPE: Working Group Document
TITLE: Comments on Mongolian, Small Khitan, and other WG2 #65 documents
SOURCE: Deborah Anderson, Ken Whistler, Rick McGowan, Roozbeh Pournader, Andrew Glass, Laurentiu Iancu, and Lisa Moore
STATUS: Expert contribution
ACTION: For consideration by WG2
DISTRIBUTION: ISO/IEC JTC1/SC2/WG2

The following are comments by a group of experts on Mongolian, Small Khitan, and other WG2 #65 documents.

1. Mongolian

a. Document: N4753 WG2 #65 Mongolian Discussion Points – Eck and Orlog Ou Rileke (=L2/16-259)

Related doc: N4752 DS01 Mongolian Base Forms Positional Forms, & Variant Forms – Eck and Glass [=L2/16-258]

1.) UTC Request: NNBSP Deficiency – Proposal for new mechanism

The view of the majority of experts was that a new character would not necessarily solve the problem, since the proposed character would not likely be published in the standard until 2018 (Unicode 11) and would be implemented even later. In addition, the new character would require software to support both encoding models in perpetuity. A more productive approach would be work with vendors to fix current software and/or update fallback standards (such as CSS).

2.) UTC Request: U+1887 Deficiency – Change of Variation Selector Assignment [U+FE00]

The document states that the standard contains an error for the initial first variant of U+1887. The proposed solution is to add a new sequence for a new fifth of final form with the shape:

ᠮᠤ.

However, the current nameslist has the proposed shape listed as “second form (isolate)”:

1887 ➦ MONGOLIAN LETTER ALI GALI A
~ 1887 ➦ first form (initial)
~ 1887 ➦ first form (medial)
~ 1887 ➦ first form (final)
~ 1887 180B ➦ second form (isolate)
~ 1887 180B' ➦ second form (final)
~ 1887 180C } third form (final)
~ 1887 180D ➦ fourth form (final)

Can the authors verify the analysis?

If the proposed new sequence is deemed correct, then we agree a new sequence is warranted, and VS01 would be acceptable. If there is no opposition to the sequence, it could be added to Unicode 10.0 (to be published in June 2017) and ISO/IEC 10646. A proposal for the new sequence is invited. (The UTC will need to discuss how to handle the second form isolate sequence, since it is an error.)

3.) UTC Request: Unicode CDR 9.0 Corrections

In order to correct the error described for the U+182D final feminine form sequences, the following corrections should be made for Unicode 10.0:

- A change to [StandardizedVariants.txt](#)

The current variation sequences in are:

182D 180B; second form; initial medial # MONGOLIAN LETTER GA

182D 180B; feminine form; final # MONGOLIAN LETTER GA

182D 180C; third form; medial # MONGOLIAN LETTER GA

The appropriate change would be to the following (noted in **bold**):

182D 180B; second form; initial medial **final** # MONGOLIAN LETTER GA

182D 180C; third form; medial # MONGOLIAN LETTER GA

182D 180C; feminine third form; final # MONGOLIAN LETTER GA

- Font changes
 - The font will need to be remapped to get the 182D 180B second form final context mapped to the "z" shape.
 - The font will also need to add a mapping for 182D 180C feminine third form final context to the one currently shown for the 182D 180B feminine form final context.

4.) FYI: Nomenclature – Proposal for a Standard Canonical Nomenclature

- The workload required to produce code charts and nameslists as suggested would create an extremely heavy a burden for the Unicode and ISO/IEC 10646 editors. Instead, users are encouraged to create a [Unicode Technical Note](#) in the desired format. This UTN would be a post-processing action, but not one for code chart generation. (UTNs are relatively flexible documents, which can easily be updated and maintained by the authors. They are independent publications, which are not formally part of the Unicode Standard or Unicode specifications.)
- The use of gender in sequences (i.e., “feminine”) appears already in [StandardizedVariants.txt](#) (8x, all in Mongolian sequences). A strong justification should be provided to remove “feminine”, identifying the sequences, and providing discussion whether there would be any data collision if “feminine” were removed.

5.) FYI: MVS+A/E+Suffix – Problem & Solution

This item was noted by the group; no action by the character encoding committees is necessary.

6.) FYI: Difficulty in Displaying of Mongolian Web Pages – “Bookmarklet” Work-Around

This item should be noted by font developers, but involves no action by character encoding committees.

7.) FYI: Microsoft Word 2016 – Word-Count & Word-Jump

No action is required by character encoding committees on this item.

8.) FYI: DS01 Objectives ([N4752](#) = [L2/16-258](#))

This document was noted.

b. Document: WG2 [N4757](#) Modifications to Mongolian Encoding in UCS (=L2/16-261) – China

The following are comments on this document.

Page 1

“This proposal is for modifications to Mongolian encoding in UCS 2014.”

“UCS 2014” is the 4th edition of IEC/ISO 10646, and hence is not the latest version of ISO/IEC 10646,. The appropriate document to refer to is the 5th edition, currently under DIS ballot. An easy way to see what appears in the 5th edition is to view the Unicode 9.0 charts for [Mongolian](#) and the [Mongolian Supplement](#), which are identical in content to the 5th edition. The 5th edition is made up of “UCS 2014”, the two supplements ([ISO/IEC 10646:2014/Amd 1:2015](#) and [ISO/IEC 10646:2014/Amd 2:2016](#)), and other new character and script additions.

Part 1 Additions

1.) New characters

The submitters of this document should review the new characters already in 5th edition. (As noted above, see the Unicode 9.0 charts for [Mongolian](#) and the [Mongolian Supplement](#); the Unicode 8.0 charts had defects.) If characters are missing from the 5th edition (=Unicode 9.0 charts), then a proposal is invited.

2.) (Free) variation selectors

Sixteen Variation Selectors are available (see <http://www.unicode.org/charts/PDF/UFE00.pdf>).

3.) Birgas

Noted.

4.) Using variation selector characters within words

Please see the 5th edition (and Unicode 9.0 charts) for the changes in format. If a particular table format would be useful, users are encouraged to submit a [Unicode Technical Note](#) with the desired format.

5.) 3 levels of implementation

Levels of implementation are not part of character encoding, and hence should not be in ISO/IEC 10646. This is an implementation issue.

6.) Content needed to be added

For these items, put forward a proposal, containing the rationale for the addition (with evidence), and discuss how the proposed additions would affect existing data.

Part 2 Modifications

1.) Partial adjustments to UCA

- The request to delete 18A7 ALI GALI HALF YA from the nominal character set is out of scope for ISO/IEC 10646. The identification of which subsets are required for nominal Mongolian is an implementation issue, and not an encoding issue.
- The request to add rules for FVS in mandatory ligatures was unclear, and needs more detailed explanation.

2.) Direction of characters in code table and 3.) On punctuation

Noted.

4.) On encoded position

Such a table would be appropriate for a [Unicode Technical Note](#). (As noted above, UTNs are relatively flexible documents, which can easily be updated and maintained by the authors. They are independent publications, which are not formally part of the Unicode Standard or Unicode specifications.)

5.) Questions of groupings by scripts

A [Unicode Technical Note](#) could be created without tables for Todo, Sibe, or Manchu, though it would not be considered standardized (since the Notes are not part of a standard). Users are encouraged to submit more information on the problem, if such an approach is not adequate.

2. Khitan Small Script

Document: [L2/16-245](#) Final proposal to encode the Small Khitan Script in the SMP (WG2 N4738) - Everson et al.

Related documents:

[L2/16-243](#) Summary of Meeting on Khitan Scripts, 20 August 2016 (Yinchuan, China) - Ad Hoc Report #1 (WG2 N4736) - Anderson

[L2/16-244](#) Summary of Meeting on Khitan Scripts, 22 August 2016 (Yinchuan, China) - Ad Hoc Report #2 (WG2 N4737) - Anderson

The following were comments made on the script proposal:

2.1 Rendering

- The section needs more information. How are the “single cluster initial” and “double cluster initial” related to “stacker-1” and “stacker-2” (in [N4725r](#)) or the horizontal and vertical stacker (in the script ad hoc recommendations [L2/16-156](#))? Possible suggested names are: KHITAN SMALL SCRIPT VERTICAL JOINER and KHITAN SMALL SCRIPT HORIZONTAL JOINER (which would match the names for Egyptian, such as EGYPTIAN HIEROGLYPH HORIZONTAL JOINER).
- The format characters should go between the characters, and not be prefixed.

Code chart

- The two format characters should go into one of the Khitan script blocks and not in Ideographic Symbols and Punctuation, since they are Khitan-specific.
- The iteration character (U+18C60) was called out in the earlier document, “Towards an Encoding of Khitan Small Script” (N4725r=[L216-113r](#)) as being placed at the end of KSS lists, but now it is interfiled. Why was it interfiled?

Character properties

- The property assigned to all the KSS block characters currently is “Lo”, but the iteration character may have a different property (cf. Tangut Iteration Mark and Ideographic Iteration Mark, which are both Lm, whereas repetition marks treated as extenders may be Po or Lm).

3. Numbers

Document: [L2/16-256](#) **Proposal to encode additional circled numbers – West**

Because circled numbers belong to a potentially open set, a more appropriate way to handle the circled numbers, in our opinion, is to use markup or other mechanisms that tag it to a style.

The proposal provides examples, but the proposed characters are not known to be in a standard (unlike many of those currently encoded), and a strong justification for including the circled numbers in plain text not been made.

4. Game Symbols

Document: [L2/16-255](#) **Proposal to encode Xiangqi game symbols (WG2 N4748) – West**

In our view, dedicated characters for Xiangqi game symbols are warranted. However, the set should only include the obvious set: the black-colored ideographs and the negative circled white-coloured ideographs, as well as the alternate forms. We do not feel that the use of Variation Sequences is warranted.