WG2 #65 Mongolian Discussion Points
San Jose, CA
Sept. 26-30 2016

Greyson Translation Services
Greg Eck (by telephone) / Orlog Ou Rileke (in person)
greyson@postone.net

1.) Unicode CDR 9.0 Corrections
2.) MVS Deficiency – Proposal for new mechanism [2 new code-points]
3.) NNBSP Deficiency – Proposal for new mechanism [1 new code-point]
4.) New Mongolian FVS – Proposal for FVS4 [1 new code-point]
5.) Difficulty in Displaying of Mongolian Web Pages – Bookmarklet Work-Around
6.) Nomenclature – Proposal for a Standard Naming Structure
7.) Microsoft Word 2016 – Word-Count & Word-Jump
8.) DS01 Status – FYI
9.) Baiti Font enhancements – to show in person by Mr. Ou
    - Baluda / Third Baluda
    - MVS Deficiency Proof of Concept
10.)Todo GB xxxx-xxxx Comments/Typos (pass to Chinese Delegation)

Unicode CDR 9.0

There are two areas that are considered incorrect in the Unicode charts for the Mongolian block. We can argue from the Chinese Standard as the basis for the incorrectness.

U+182D – The Unicode charts from the beginning have placed the final feminine form at 180B (FVS1). The MGWBM, the TR170, the DS01, the Chinese Standard all place the final feminine form at 180C as below.

ZWJ+182D		*First Final Form* (NOTE: error in Chinese Standard - missing)
ZWJ+182D+FVS1		*Second Final Form* (needed to over-ride default context) (SIG+FVS1)
ZWJ+182D+FVS2		*Third Final Form* (context-driven) (ABSIG+FVS2)

U+1838 – the medial first variant form is missing

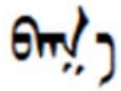
1838+ZWJ		*Initial Form*
ZWJ+1838+ZWJ		*First Medial Form*
ZWJ+1838+FVS1+ZWJ		*Second Medial Form* (context-driven)
ZWJ+1838		*First Final Form*
ZWJ+1838+FVS1		*Second Final Form*[&] (only implemented with MVS)

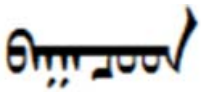
Call for Two New Characters to Handle MVS Deficiency – Orkhitz_A & Orkhitz_E

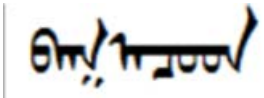
There is an advanced shaping case in regard to the MVS (U+180E) which is not being handled correctly by any of the existent Mongolian vertical fonts yet. This is the case where you have a STEM + MVS + A/E + SUFFIX. Proper shaping behavior in this case is like this … Given the STEM + MVS + A/E string, there is no problem. The word will form in most fonts correctly showing the left-ward sweeping orkhitz. By definition of the Mongolian script, no letter attaches to the orkhitz in final position. If the orkhitz is followed directly by an "attaching suffix", then the orkhitz A/E will transform back to its common A/E "shuud/single-tooth" medial form with no space in the shaping of the word. If the suffix following the orkhitz A/E is removed, then the space preceding the orkhitz A/E reappears.

Given the following word and suffix:

BAG-MVS-A – noun meaning "team" CHUD – plural suffix

Proper shaping of BAG-MVS-A is 

Proper shaping of BAG-MVS-A + CHUD(plural suffix) is 

Current shaping of BAG-MVS-A + CHUD – plural suffix is  which is incorrect (current BAITI font rendering).

The reason that this is incorrect is that the MVS (between the BAG and the ACHUD) …

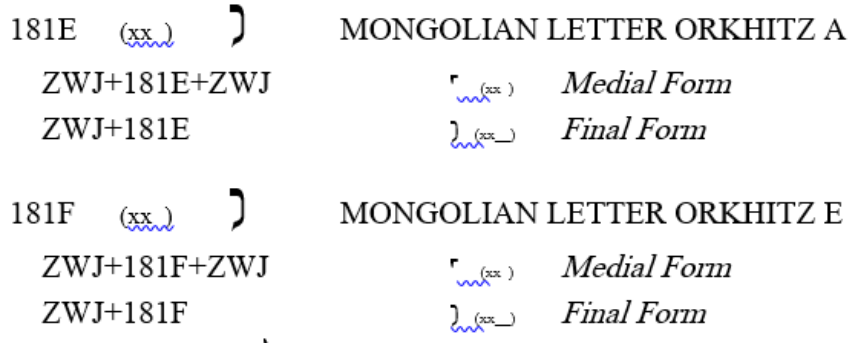
- has a space built into it – therefore it cannot NOT be displayed
- the early design of the MVS probably did not take this case into account
- OT rulings have no method to hide a character; even if the font "hid" the MVS character, the font would have to "unhide" the MVS once the characters following the Orkhitz A/E are removed

Therefore, I am calling for a new solution which includes deprecating the MVS in future fonts. The new solution might look like this …

1.) Add two new characters to our encoding – the orkhitz-A and the orkhitz-E. Instead of typing MVS+A or MVS+E to get the two orkhitzes shaped, the typist would type something like <shift><A> or <shift><E> to get the new characters. This would be similar to the handling of an upper-case character in English. In English, you type in <A> to get the lower-case letter "a". You type in <SHFT><A> to get the upper-case English letter "A".

2.) The rendering machines would not need to be changed as the font would do the display correctly through internal OT contextual rulings. Ongoing processing of the MVS would continue to allow for backwards-compatibility, however new fonts would not use the MVS anymore.
3.) OT rulings would …
    - display U+181E-Final_Form given final position as determined by the rendering device (typist depressing <SHFT><A>)
    - display U+181E-Medial_Form given medial position as determined by the rendering device (typist depressing <SHFT><A>)
    - display U+181F-Final_Form given final position as determined by the rendering device (typist depressing <SHFT><E>)
    - display U+181F-Medial_Form given medial position as determined by the rendering device (typist depressing <SHFT><E>)

The direct Unicode encoding impact would be the addition of two characters



The final forms would differ from the two current U+1820/U+1821 final variants in the addition of the leading space (since the MVS would no longer provide this). The medial forms would be identical to that of the U+1820/1821 medials.

Call for New NNBSP character in Mongolian Block

Even with the changes that were effected in the NNBSP re-specification (Unicode version 9.0), there are still deficiencies in proper functionality. These deficiencies include the following:

- Display problem if Fallback Font includes its own version of the NNBSP (U+202F). This problem occurs at the Application level.
- Certain applications such as PDF viewers will drop the NNBSP upon cut & paste as well as other operations

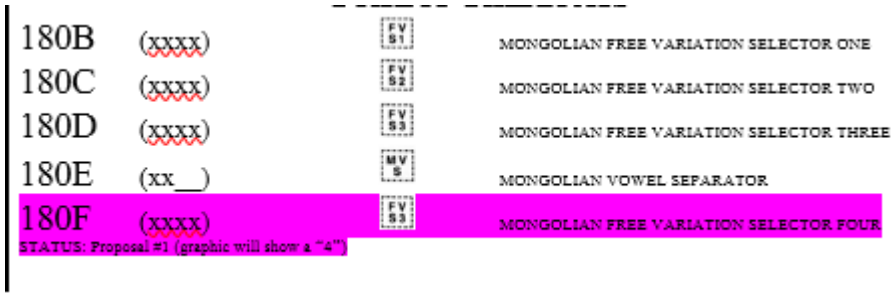
As much as we fix the NNBSP in Mongolian to display correctly in different situations, we continue to find more areas where it will fail. The most recent finding over the summer was that given an application that is displaying Mongolian text AND given a font selection that is not a proper vertical Mongolian font AND given NNBSP+Suffix forms in the text AND given a fall-back font that uses its own NNBSP, the display of the suffixal forms will fail. The reason is that the fall-back font definition of the NNBSP does not include the OT rulings to do proper and necessary substitutions based on the NNBSP context.

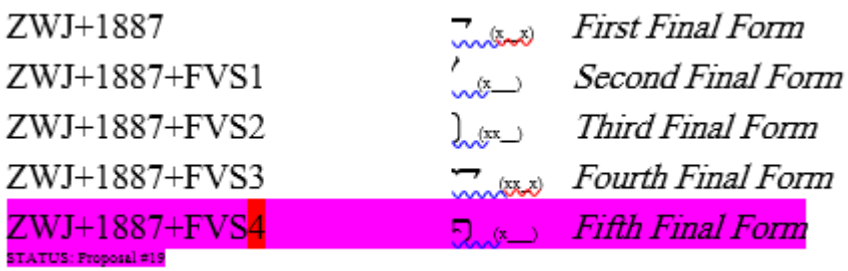
There is considerable frustration in the Mongolian user community with the ongoing problems we are having with this one control character. At the same time, we see much momentum in the Mongolian community to adopt the Unicode encoding if it will work correctly. This is significant as there are huge swaths of Mongolian text using ASCII legacy era vertical Mongolian fonts. We cannot miss this moment of opportunity to provide a simple solution to this one biggest trouble of our Mongolian encoding – the NNBSP. It is felt that once an NNBSP-replacement is encoded within the Mongolian block, both rendering engines as well as fonts will be able to exercise adequate control as to display suffixal forms correctly.

I am calling for a new control character “Mongolian Suffix Connector” placed at U+181A to replace the current functionality of the NNBSP (U+202F) in Mongolian contexts. Transitional fonts, of course, would have to implement both the NNBSP as well as the replacement control character for backwards compatibility. Keyboards could be changed to use the new control character instead of the old NNBSP mechanism. Even if it took a considerable period of time to cast the encoding, I think all parties dealing with Mongolian vertical fonts will embrace this move.

Call for New Free Variation Selector FVS4

Discussions with Professor Quejingzhabu have shown that the specification of the U+1887 Initial first variant is incorrect. The glyph is actually a final – the fifth of 5 final forms. Early in the specification period, a decision was made to incorrectly specify it as an initial in a work-around to including the form since there was no fourth variant. This should be corrected. Some call for usage of a generic Variation Selector. We call for a completely new variation selector, FVS4, to be included directly in the Mongolian block for several reasons. The main reason is to allow for specification of the U+1887-Final-fourth_variant. A second important reason is that there seems to be a need for a fourth Mongolian variation selector to fully specify other character variants. The suggested specification and justification follow:





Work-Around to Anomalous Web-Site Display of Traditional Mongolian

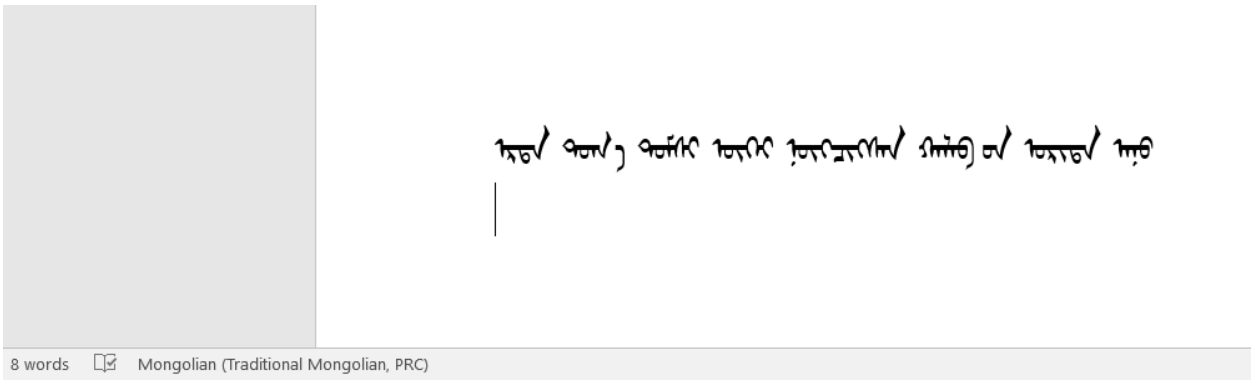
There have been several in the Mongolian user community who have complained of not being able to view good Mongolian on their pages. After looking at the pages, we found that they were indeed written in Unicode-compliant code-point. The deficiency was found to be at the font as called out by the particular browser. Depending on whether the page includes CSS specifications that call for an actual Mongolian font (rather than just a fall-back font), the page will either display well or display improperly. One of the main problems found is in the display of the NNBSP and associated suffixes.

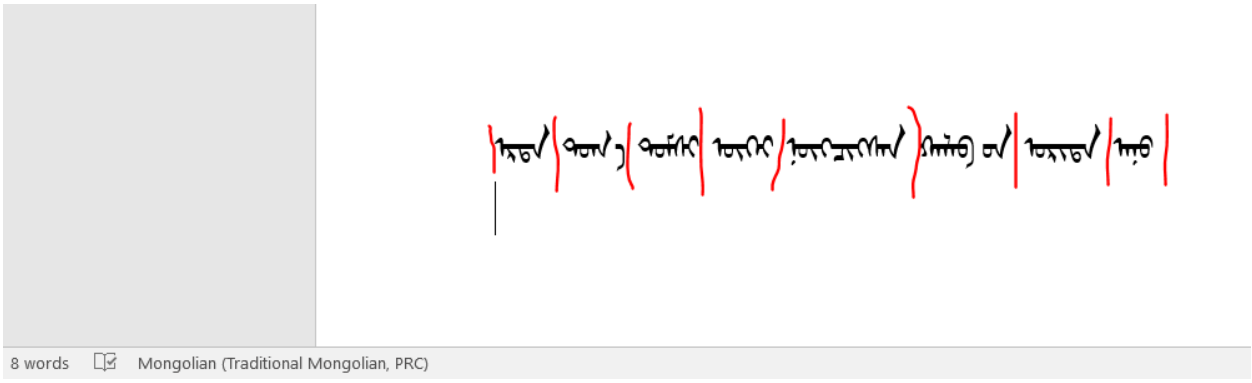
At a meeting on the Microsoft campus in Redmond this summer, the matter was discussed. A suggestion came up as to using a bookmarklet to control the actual font used to display the page. Microsoft employee, Aaron Bell, designed a bookmarklet, as attached that can be added to your favorites bar. Once executed, the display font is changed to Mongolian Baiti. Of course, the font could be Noto Sans Mongolian, MongolianWhite, or any other as desired. It is a work-around no doubt and still deficient at that. Only one font can be selected for a given page. But it is much better than having no control of the font on the given page. The bookmarklet can be found at the Greyson page (http://greyson.postone.net).

Microsoft Word 2016 – Word Count Feature Fixed

For some time, the word-count feature in Microsoft Word has been broken in the area of the NNBSP. Although the Unicode initiative to redefine some of the word-break features of the NNBSP (U+202F) will fix some of this functionality across all platforms/applications, there was a fix applied late summer 2016 which did fix the word-count feature under Microsoft Word 2016.

Given the string 





The word count is correct at 8 words. The second word includes an MVS before the orkhitz_A The seventh word includes an NNBSP before the genitive suffix UN. This is correct. Kudos to Zoey Fan and her Microsoft Word team on this fix.

Microsoft Word 2016 – Word Jump Fix in Process

The word-jump feature using <CTL>+<RIGHT/LEFT CLICK> is still in process as seen below ..

Cursor starts at beginning of the sentence

First <CTL><RIGHT_CLICK> - CORRECT – one word

Second <CTL><RIGHT_CLICK> - CORRECT – two words

Third <CTL><RIGHT_CLICK> - CORRECT – three words

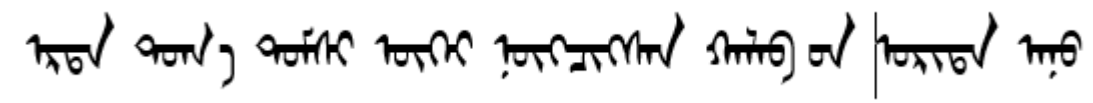
Fourth <CTL><RIGHT_CLICK> - CORRECT – four words

Fifth <CTL><RIGHT_CLICK> - CORRECT – five words

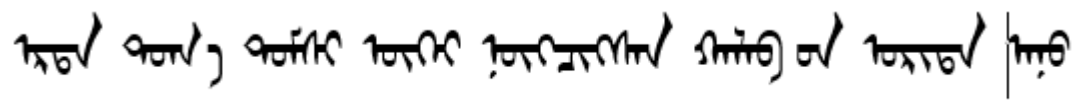
Sixth <CTL><RIGHT_CLICK> - WRONG – five words

Seventh <CTL><RIGHT_CLICK> - WRONG – five words

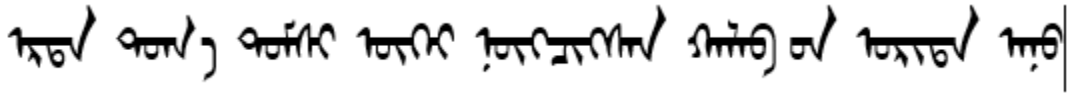
Eighth <CTL><RIGHT_CLICK> - WRONG – six words



Ninth <CTL><RIGHT_CLICK> - WRONG – seven words



Tenth <CTL><RIGHT_CLICK> - WRONG – eight words



The problem is that on the 6th and the 7th right-click, the NNBSP is being picked up as it still has work-break characteristics tied to it.

Nomenclature Standardization

There is a great need to standardize our notation in discussion regarding the Mongolian encoding. I suggest the following standards:

Index 1 – Unicode code-point

Index 2 – Position

- Isolate
- Initial
- Medial
- Final

Index 3 – Variant

- Default
- First variant (applies FVS1)
- Second variant (applies FVS2)
- Third variant (applies FVS3)
- Fourth variant (applies FVS4)

NOTE: Gender should not be specified at this level as it will only clutter the definition. Unambiguous and discrete definition can be made using just the three indices given. Further detail including gender, the particular script being applied to, etc. can be made in the footnotes.

Specification of a ligature can proceed in the following order:

- Initial + Medial
- Initial + Final
- Medial + Medial
- Medial + Final

Standardization of our nomenclature will help us understand each other's work faster and make the specification of the same to be more clear.

DS01 – for information only

TARGET AUDIENCE: Font developers

OBJECTIVE: Full specification of Mongolian/Todo/Sibe/Manchu/AliGali script base forms, positional forms, and variant forms. It should be noted that full specification of implementation logic is beyond the scope of this paper. It should be noted that the emphasis of this paper is stand-alone text rather than running text. However, all running text forms, aside from ligatures should be found here. This paper is not meant to be a replacement for the Chinese Standard nor the Ulaanbaatar Mongolian Standard. It is derived from these two standards. Suggestions such as in the yellow/violet highlighting are derived from external discussions such as the Unicode Forum.

The DS01 is in complete sync with the Unicode 9.0 charts. The DS01 is in complete sync (other than yellow/violet sections) with the current Chinese Standard. Comments showing where the DS01 is not in sync with these two documents are appreciated and should be sent to greyson@postone.net .