

Hohhot Mongolian Font Discussions

November 24-25, 2015

1.)NNBSP

- Explained that there was a lengthy discussion with Unicode Forum – considered creating new code-point specifically for Mongolian; found easier solution was to tweak the specification of U+202F such that it handles accurately Mongolian contexts while at the same time such specification does not break usage in languages other than Mongolian.
- Problem is that the NNBSP is not recognized as part of the word, therefore word-count, sort, search routines do not work correctly without a lot of work-arounds.
- Examples: A brown fox. (3 words); ᠠᠨᠨᠢᠪᠰᠤ (one word); ᠠᠨᠨᠢᠪᠰᠤ (2 words) ᠠᠨᠨᠢᠪᠰᠤ (3 words).
- Solution was to re-define characteristics of the NNBSP so that it is recognized as an integral part of the stem and the suffix.
- Timeline is February – June 2016 when apps will start to see changes as software developers implement changes in the new NNBSP spec.
- STATUS: Monitor

2.)ISOLATES

- Problem Statement: Unicode glyph representatives and Stand-alone glyph representatives are different in a variety of settings. Therefore, we need to have a specification for the stand-alone glyph set AND a specification for the Unicode representative glyph set. Overlap is fine. Professor Quejingzhabu is not for the idea of a separate FVS specification for each Unicode glyph but prefers the idea of specifying the 15 special cases of Unicode glyphs being referenced by the particular context they are found in (using ZWJ). A further problem as stated by Professor Quejingzhabu is that only vowels (plus the feminine consonants) have isolates. The issue was raised that a font developer must paint something on the screen for each letter in each position whether it is grammatical or not. The professor conceded that if the font developer really needs the isolate assignments, then they should use a programmatic and systematic approach to deciding which glyph should be the consonantal isolate. He suggested (as has already been suggested by at least Jirimutu) the following values in descending order of priority:
 - o Consider what the native script-writer would prefer as of prime import
 - o Given a masculine/feminine pair, assign the masculine first, feminine second
 - o Use the initial form if available
 - o Use the medial form if available
 - o Use the final form if there is no initial/medial

- Follow the Chinese Standard wherever possible (added by Greg)
- See Isolate Comparison Chart (Appendix) for conclusions drawn
- It was pointed out that a number of our isolate glyphs on the Font Comparator site (and therefore possibly in our fonts also) are lacking the extra ascender portion of the stem that sometimes differentiates the initial from a medial (including 1834, 1852, 1853, 1854, 1871, 1896)
- STATUS
 - Font developers must come to agreement on the 10 items (colored blue); chart shows determination following the principles above. As there is wide-spread variation in isolate specification, it is suggested that an agreed-upon set of principles such as those above decide the choice of default consonantal isolate, choice of potential second isolate, etc.
 - Propose to the UTC the addition of 15 FVS assignments to isolates (grey cells) which will allow ALL Unicode base form glyphs to be displayed either as the default isolate or as UnicodePoint+FVSx. The purpose of this is to allow for an easy method of determining how to display a given Unicode point. In most cases, the stand-alone isolate glyph is also the Unicode presentation form. However, in 15 code-points, this is not the case. U+1824 is a case in point. Without an FVS assignment, you would specify this isolate glyph as U+1824_SecondInitial. You would specify the U+1826 Unicode presentation glyph as U+1826_ThirdInitial. Without the FVS assignment, you will need to consider in some cases both the glyph's position as well as its variant. This is considered to be so troublesome as to justify the assignment of a specific FVS to allow display of Unicode presentation forms for 1824, 1826, 184A, 184B, 184F, 1859, 185D, 185E, 1861, 1862, 1863, 1873, 1874, 1875, 1876.
 - Inform the UTC that the 1887_SecondIsolate glyph is actually a final as confirmed by Professor Quejingzhabu. It should be U+1887_FifthForm demanding a new FVS4 specification or the use of a VS selector. This is also a mistake in the Chinese Standard. Discussion needed as to whether a fourth FVS should be proposed.
 - Inform the UTC that our NP proposed specification set does not follow the Chinese Standard in at least one case – that of the U+182D undotted feminine isolate. The Chinese Standard uses a ZWJ in the specification of this glyph which is highly unique and implies certain implementation decisions which might not be general to all fonts. It is felt that the standard approach using an FVS as in all other cases will be easier to implement and easier for the user to remember. Further, it is felt that the use of a control character such as the ZWJ for one case where the FVS is used elsewhere is a bit dangerous as extra testing will be necessary.
 - Forum discussion needed on 1868, 1869 leading to resolution as to whether there are really two isolates
 - After discussions with Professor Quejingzhabu, I am withdrawing my proposal to re-map FVS Mis-Matches in point 3 below. Reasoning is given

below. Therefore, the new isolates as noted on the attached Chart will not be proposed. This follows a decision to not re-map FVS Mis-Matches. Such isolates are important however and need to be newly specified as finals. I count six such occurrences – U+1822_SecondIsolate (per current NP Proposal), U+1824_ThirdIsolate (per current NP Proposal), U+1824_FourthIsolate (per current NP Proposal), U+1826_FourthIsolate (per current NP Proposal), U+1826_FifthIsolate (per current NP Proposal), U+1873_ThirdIsolate (per current NP Proposal).

3.) FVS Mis-Match

1820 Medial+FVS2 -> Initial+FVS1

(Note: This variant is used only in the post-NNBSP context.)

1828 Medial+FVS2 -> Final+FVS1

(Note: This variant is used only in the pre-MVS context.)

182C Medial+FVS2 -> Final+FVS1

(Note: The Mongolian grammar does not allow for a final Q. This variant is used only in the pre-MVS context.)

182C Medial+FVS3 -> Final

(Note: The Mongolian grammar does not allow for a final Q. This variant is used only in the pre-MVS context.)

182D Medial+FVS2 -> Final+FVS3

(Note: This variant is used only in the pre-MVS context.)

1835 Medial+FVS1 -> Final+FVS1

(Note: This variant is used only in the pre-MVS context.)

1836 Medial+FVS2 -> Final+FVS1

(Note: Mongolian grammar does not allow for a final Y. This variant is used only in the pre-MVS context.)

- STATUS: I am withdrawing my idea of re-mapping these mis-matched FVS assignments for the following reasons:

- First and foremost reason is that our matching the Chinese Standard to the best of our abilities will be helpful in the years to come. The Chinese data set will be the largest in the world as they are the primary users of the Mongolian Traditional vertical script. There is new information coming out in the 2015 Chinese Standard that will be best facilitated if any publications that we produce match this standard as closely as possible.
- The Mis-Match FVS proposal that I put forth is based on a particular view. It is the font developers' view. The current view is that of the grammarian. Neither is right or wrong. Let's match the current and make sure that the differences between glyph implementation and FVS assignment are well documented.
- There may be problems of FVS space usage especially at the U+1828_Medial are and the U+182D_Medial area. I am not sure how to handle this. If these two slots toggle, then the space allocation problem may be alleviated.

Important !!!

- Continuation of the current FVS specification here will be helpful in maintaining compatibility and consistency with the current library of data “out there”.
- We will need to be sure and document the mis-matches for the benefit of future font developers. On my part, I will be using my DS01 document to secure such information. All of my DSxx docs will be available at <http://greyson.postone.net> under the help files option.

4.) Special Spellings Cases (' = FVS1, "=FVS2)

The following are Professor Quejingzhabu’s view on proper spelling on several difficult words (last item of each line is a graphic image):

- Medial I NAYMA 
- Medial N AN'AR 
(Baiti is rendering this with <U+1828><U+180B> for N)
- Medial G COG'A'GULA 
(Baiti is rendering this with <U+182D><U+180C> for the first G)
- Final G SIG/SIG'/ABSHIG" 
(Baiti is rendering this with no FVS on the first SIG, FVS1 following the second SIG, and FVS2 following the ABSHIG)
- Medial Y SAI IQAN/SAYIQAN/SAY'IQAN 
(Baiti is rendering this with <U+1836><U+180B> for the SAYIQAN)
- Final UE SUE' 
(Baiti is rendering this with <U+1826><U+180B>)
- Medial UE GUESHI 

There is another possible spelling is NAIMA, we all taught by teachers as NAIMA. It is better to support both in the font !

This word spelling is not correct !

We support this encoding <U+1836><U+180B> is for SAYIQAN, not for SAY'IQAN

5.) Ligatures (blue highlighting denotes Prof. Quejingzhabu’s answer)

Mongolian

- 1.) The 1889/1892/1893 set has 8 ligatures (at 1826) in the Baiti font. However, the Chinese Standard seems to specify only 4 (the four with the connected FVS1). Should the Mongolian Script use all 8 or only the 4? In other words, should the 1889/1892/1893 take the dots also. YES
- 2.) Version 9.0 also shows 1853, 1858, 185B, 185C. How are these Todo characters used in the Mongolian Script? Still can be used

Todo

- 1.) The Chinese Standard specifies 184D + 1847. NO, this is incorrect. There is no 184D + 1847
- 2.) The Chinese Standard specifies 184E + 1847. NO, this is incorrect. There is no 184E + 1847
- 3.) The proper encoding is 1889+1820, 1889+1844, 1889+1845, 1889+1846, and 1889+1849. Is this correct? YES 11/25
- 4.) The proper encoding is 1892+1820, 1892+1844, 1892+1845, 1892+1846, and 1892+1849. Is this correct? YES 11/25
- 5.) The proper encoding is 1893+1820, 1893+1844, 1893+1845, 1893+1846, and 1893+1849. Is this correct? YES 11/25

Thanks to him
accepting mistake
correction for the GB !

Manchu

- 1.) How to form the 183A+1887? There is no ligature here.

Mongolian Ligatures

182A + 1820, 1821, 1822, 1823, 1824, 1825(8), 1826(8), 1827, 1888
182B + 1820, 1821, 1822, 1823, 1824, 1825(8), 1826(8), 1827, 1888
182C + 1821, 1822, 1825(16), 1826(16), 1827
182D + 1821, 1822, 1825(16), 1826(16), 1827
1839 + 1820, 1821, 1822, 1823, 1824, 1825(8), 1826(8), 1827, 1888
183A + 1820, 1821, 1822, 1823, 1824, 1825(8), 1826(8), 1827, 1888
183B + 1820, 1821, 1822, 1823, 1824, 1825(8), 1826(8), 1827, 1888
1889 + 1820, 1823, 1826, 1827, 1888
1892 + 1820, 1823, 1826, 1827, 1888
1893 + 1820, 1823, 1826, 1827, 1888

Todo Ligatures

184B + 1820, 1844, 1845, 1846, 1847, 1848, 1849
184C + 1820, 1844, 1845, 1846, 1847, 1848, 1849
184D + 1844, 1845, 1848, 1849
184E + 1844, 1845, 1848, 1849
1857 + 1820, 1846, 1847
1858 + 1820, 1846, 1847
1889 + 1820, 1844, 1845, 1846, 1847
1892 + 1820, 1844, 1845, 1846, 1847
1893 + 1820, 1844, 1845, 1846, 1847

Sibe Ligatures

182A + 1820, 1823, 185D, 185E, 1860, 1861
183A + 1820, 1823, 1861
1863 + 185D, 185E, 1860
1864 + 185D, 185E, 1860
1865 + 185D, 185E, 1860
1866 + 1820, 1823, 185D, 185E, 1860, 1861
186C + 1820, 1823, 1861

186D + 1820, 1823, 1861

Manchu Ligatures

182A + 1820, 1823, 185D, 1860, 1861, 1873, 1888

183A + 1820, 1823, 185D, 1860, **1887**, 1888

1864 + 185D, 1860, 1873

1865 + 185D, 1860, 1873

1866 + 1820, 1823, 185D, 1860, 1861, 1873, 1888

186C + 1820, 1823, 185D, 1860, 1888

186D + 1820, 1823

1874 + 185D, 1860, 1873

1892 + 1820, 1823, 185D, 1860, 1888

18A8 + 1820, 1823, 185D, 1860, 1888

6.) Special Letters

- 180A Nirugu – final hook position added as default – this is included in the Quejingzhabu specification 9.0
- 1885 Baluda – We informed Professor Quejingzhabu that the forum is proposing this be redefined as a “mark” rather than a letter
- 1886 Triple Baluda – same as above

Appendix – Isolates

Mongolian Block Isolate Set						
<i>(see legend at end of chart)</i>						
	Unicode Glyph	Isolate Glyph Default	Isolate #2 FVS1	Isolate #3 FVS2	Isolate #4 FVS3	Isolate #5 ?FVS4/VS01?
1820	ᠠ	ᠠ	ᠠ			
1821	ᠡ	ᠡ				
1822	ᠢ	ᠢ	Accusative Form			
1823	ᠣ	ᠣ				
1824	ᠤ	ᠤ	ᠤ	Genitive Form	Genitive Form (with N-dot)	
1825	ᠥ	ᠥ				
1826	ᠦ	ᠦ	ᠦ	ᠦ	Genitive Form	Genitive Form (with N-dot)
1827	ᠨ	ᠨ				
1828	ᠯ	ᠯ				
1829	ᠪ	ᠪ				
182A	ᠭ	ᠭ				

182B	𐎡	𐎡				
182C	𐎢	𐎢	𐎣	𐎣		
182D	𐎤	𐎤	𐎥	𐎥	Note: Chinese Std. (Isol #3 Q+ZWJ)	Greg suggests that we reject the use of the ZWJ here and use standard method of FVS assignment
182E	𐎦	𐎦				
182F	𐎧	𐎧				
1830	𐎩	𐎩				
1831	𐎪	𐎪				
1832	𐎬	𐎬				
1833	𐎭	𐎭				
1834	𐎮	𐎮				
1835	𐎯	𐎯				
1836	𐎱	𐎱				
1837	𐎲	𐎲				
1838	𐎴	𐎴				
1839	𐎶	𐎶				

I agree this !

183A	ੴ	ੴ					
183B	ਕ	ਕ					
183C	ਮ	ਮ					
183D	ਸ	ਸ					
183E	ਹ	ਹ					
183F	ੳ	ੳ					
1840	ੲ	ੲ					
1841	ਯ	ਯ					
1842	ਰ	ਰ					
	TODO						
1843	ਕਿ	ਕਿ					
1844	ਕੇ	ਕੇ					
1845	ਕੁ	ਕੁ					
1846	ਕੌ	ਕੌ					
1847	ਕੂ	ਕੂ					

1848	ቄ	ቄ					
1849	ቅ	ቅ					
184A	ቆ	ቆ	ቆ	ቆ	Principle #1/#4 Used in choice Of default isolate		
184B	ቇ	ቇ	ቇ	ቇ	Principle #1/#3 Used in choice Of default isolate		
184C	ቈ	ቈ					
184D	቉	቉	቉	቉			
184E	ቊ	ቊ	ቊ	ቊ			
184F	ቋ	ቋ	ቋ	ቋ	Principle #1/#3 Used in choice Of default isolate		
1850	ቌ	ቌ					
1851	ቍ	ቍ					
1852	቎	቎					
1853	቏	቏					
1854	ቐ	ቐ					
1855	ቑ	ቑ					
1856	ቒ	ቒ					

1857	ੴ	ੴ					
1858	ੴ	ੴ					
1859	ੴ	ੴ	ੴ	ੴ	Principle #1/#3 Used in choice Of default isolate		
185A	ਊ	ਊ					
185C	ਘ	ਘ					
185C	ਙ	ਙ					
	SIBE						
185D	ਞ	ਞ	ਞ				
185E	ਝ	ਝ	ਝ				
185F	ੜ	ੜ					
1860	ਛ	ਛ					
1861	ਞ	ਞ	ਞ				
1862	ਝ	ਝ	ਝ				
1863	ੜ	ੜ	ੜ	ੜ			
1864	ੴ	ੴ	ੴ				

1865	ᲚᲟ	ᲚᲟ	ᲚᲟ				
1866	ᲚᲟ	ᲚᲟ					
1867	ᲚᲟ	ᲚᲟ					
1868	ᲚᲟ	ᲚᲟ	Note: Need to settle on second isol				
1869	ᲚᲟ	ᲚᲟ	Note: Need to settle on second isol				
186A	ᲚᲟ	ᲚᲟ					
186B	ᲚᲟ	ᲚᲟ					
186C	ᲚᲟ	ᲚᲟ					
186D	ᲚᲟ	ᲚᲟ					
186E	ᲚᲟ	ᲚᲟ					
186F	ᲚᲟ	ᲚᲟ					
1870	ᲚᲟ	ᲚᲟ					
1871	ᲚᲟ	ᲚᲟ					
1872	ᲚᲟ	ᲚᲟ					

	MANCHU						
1873	ᡞ	ᡞ	ᡞ	Accusative Form			
1874	ᡞ	ᡞ	ᡞ	ᡞ		Principle #1/#2 Used in choice Of default isolate	
1875	ᡞ	ᡞ	ᡞ			Principle #1/#3 Used in choice Of default isolate	
1876	ᡞ	ᡞ	ᡞ			Principle #1/#3 Used in choice Of default isolate	
1877	ᡞ	ᡞ					
	ALI GALI						
1880	ᡞ	ᡞ					
1881	ᡞ	ᡞ					
1882	ᡞ	ᡞ					
1883	ᡞ	ᡞ					
1884	ᡞ	ᡞ					
1885	ᡞ	ᡞ					
1886	ᡞ	ᡞ					
1887	ᡞ	ᡞ	ᡞ	This is a final - mistake in Chinese Standard documentation.			

This is Ok.

1888	Ɔ	Ɔ					
1889	Ɔ	Ɔ		Note: Should have a tail			
188A	Ɔ.	Ɔ.					
188B	Ɔ	Ɔ					
188C	Ɔ	Ɔ					
188D	B	B					
188E	Ɔ	Ɔ					
188F	Ɔ	Ɔ					
1890	Ɔ	Ɔ					
1891	Ɔ	Ɔ					
1892	Ɔ	Ɔ					
1893	Ɔ	Ɔ					
1894	Ɔ	Ɔ					
1895	Ɔ	Ɔ					
1896	Ɔ	Ɔ					

1897	ㄷ	ㄷ					
1898	ㅌ	ㅌ					
1899	ㄹ	ㄹ					
189A	ㄹ [◦]	ㄹ [◦]					
189B	ㅍ	ㅍ					
189C	ㅍ [◦]	ㅍ [◦]					
189D	ㅑ	ㅑ					
189E	ㅑ [◦]	ㅑ [◦]					
189F	ㅓ	ㅓ					
18A0	ㅓ [◦]	ㅓ [◦]					
18A1	ㅕ	ㅕ					
18A2	ㅕ [◦]	ㅕ [◦]					
18A3	ㅗ	ㅗ					
18A4	ㅗ [◦]	ㅗ [◦]					
18A5	ㅛ	ㅛ					

18A6	◄	◄					
18A7	↵	↵					
18A8	◌◦	◌◦					
18A9	◌◦	◌◦					
18A0	◌◦	◌◦					
	LEGEND						
	Blue: 10 isolates differing from current NP Proposal						
	Grey: 15 Unicode presentation glyphs needing FVS specification						
	Yellow: NP-proposed isolates to be changed to final position thus matching grammarian view						

I am attaching my notes from meetings with Professor Quejingzhabu in Hohot last week. I have had a few days now to process my thoughts from the talks and present in the attached file both the main points that we discussed as well as conclusions drawn. Comments are welcome.

I am suggesting the following action points:

ISOLATES

- 1.) Come to agreement on the principles used to select the isolates for non-assigned consonants. This set of principles will allow us to drive the actual determination of such isolates without bias on any one party's personal views.
- 2.) Given that we come to agreement on #1 above, then we propose the 10 consonantal isolates as listed above (blue highlighted items).
- 3.) I suggest that we propose the 15 Unicode FVS assignments as shown in the attached chart (grey highlighted items).
- 4.) I suggest that we propose the 6 new variants for FVS assignments (yellow highlighted items) changing them all to be final variants.

FVS MisMatch Situation

- 1.) Change the NP such that it does not reflect the FVS Mis-Match change as discussed earlier.

Special Spelling Cases, Ligature Discussion, and Special Letters section are FYI only.

Greg