Title	Notes on Mongolian Baiti
Author:	Andrew West
Date:	2006-01-31
Font:	Mongolian Baiti version 0.93
Rendering Engine:	Uniscribe version 1.606.5112.0
Operating System:	Windows XP

References

Report 170. Myatav Erdenechimeg, Richard Moore and Yumbayar Namsrai, *Traditional Mongolian Script in the ISO/IEC 10646 and Unicode Standards* (UNU/IIST Report No. 170, August 1999).

MGWBM. Choijinzhab 确精扎布, *Mengguwen Bianma* 蒙古文编码 (University of Inner Mongolia Press, 2000).

1. Narrow No Break Space

1.1 Pre-NNBSP Letter Form

The letter before U+202F NARROW NO-BREAK SPACE (NNBSP) should always be in the final form (**MGWBM** p.104). However, with Mongolian Baiti the letter preceding NNBSP is always in the medial form, as shown in the following example:

GURUN-I <1864 1860 1875 1860 1828 202F 1873>

Actual Appearance : 

Expected Appearance : 

1.2 Post-NNBSP Letter Form

The form of the letter taken after the NNBSP varies according to the particular suffix involved, but may be the initial form, normal medial form, a variant medial form or the final form (see **MGWBM** pp.104-105). However, with Mongolian Baiti the letter form after NNBSP is always the normal medial form or the final form. Those suffixes which have an incorrect initial letter form with Mongolian Baiti are shown in Table 1-1.

Table 1-1. Post-NNBSP Letter Forms

Suffix	Character Sequence	Actual Appearance	Expected Appearance	Notes
-ača	202F 1820 1834 1820			A takes the 3rd medial form
-eče	202F 1821 1834 1821			E takes the initial form
-nar	202F 1828 1820 1837			NA takes the initial form
-ner	202F 1828 1821 1837			NA takes the initial form
-tur	202F 1832 1824 1837			TA takes the initial form
-tür	202F 1832 1826 1837			TA takes the initial form
-tu	202F 1832 1824			TA takes the initial form
-tü	202F 1832 1826			TA takes the initial form
-tai	202F 1832 1820 1822			TA takes the initial form
-tei	202F 1832 1821 1822			TA takes the initial form
-taɣan	202F 1832 1820 182D 1820 1828			TA takes the initial form
-tegen	202F 1832 1821 182D 1821 1828			TA takes the initial form
-yin	202F 1836 1822 1828			YA takes the 2nd initial form
-igi	202F 1845 184E 1845			I takes the initial form

Suffix	Character Sequence	Actual Appearance	Expected Appearance	Notes
-ētse	202F 1844 1843 1854 1844			E takes the initial form
-de	202F 1869 185D			DA takes the initial form
-deri	202F 1869 185D 1875 1873			DA takes the initial form
-ni	202F 1828 1873			NA takes the initial form

2. Mongolian Vowel Separator

2.1 Pre-MVS Letter Form

The letter before U+180E MONGOLIAN VOWEL SEPARATOR (MVS) should take the appropriate final or variant medial letter form, but when a word ending in MVS + A/E is not immediately followed by a space or another Mongolian letter or punctuation mark, then the rendering system does not select the correct form for the letter preceding the MVS. Thus, a word spelled with MVS will be rendered incorrectly if it occurs at the end of a line or is immediately followed by anything other than a Mongolian letter or Mongolian punctuation mark. This is illustrated by the example below, where the rendering system does not select the expected form of U+182D MONGOLIAN LETTER GA when the word is at the end of line, but does when the word is followed by a space before the end of line:

GA + MVS + A at end of line	GA + MVS + A followed by a space
182D 180E 1820	182D 180E 1820 0020
	

My expectation would be that when typing a word with an MVS, the letter preceding the MVS would take the correct pre-MVS form as soon as the MVS is typed, as shown below for the word BAGA.

Letters Typed	Code Sequence	Actual Appearance	Expected Appearance
BA	182A		
BA + A	182A 1820		
BA + A + GA	182A 1820 182D		
BA + A + GA + MVS	182A 1820 182D 180E		
BA + A + GA + MVS + A	182A 1820 182D 180E 1820		
BA + A + GA + MVS + A + ,	182A 1820 182D 180E 1820 1802		

2.2 Pre-MVS Form of letter JA

The rendering system normally selects the wrong form of U+1835 MONGOLIAN LETTER JA before an MVS. The rendering system only selects the correct 2nd medial form when the letter JA starts a run, and incorrectly selects the normal final form of the letter elsewhere (but note that the glyph is also wrong -- see 8.1):

BA + U + I + NNBSP + JA + MVS + A + Space		JA + MVS + A + Space
182A 1824 1822 202F 1835 180E 1820 0020		1835 180E 1820 0020
Actual Appearance	Expected Appearance	Appearance
		

3. Incorrect Shaping Behaviour

3.1 Mongolian Letter GA

U+182D MONGOLIAN LETTER GA may take either a masculine or a feminine form finally or medially before a consonant. Final GA and medial GA before a consonant in a feminine word are incorrectly rendered with the masculine letter form when the vowel preceding the GA is the letter I (a neuter vowel) and the first consonant-vowel cluster preceding the letter I is a ligature with a feminine or neuter vowel. Some examples are given in Table 3-1.

Table 3-1. Incorrect Gender Forms for Mongolian GA

Word	Character Sequence	Actual Appearance	Expected Appearance
BILIG	182A 1822 182F 1822 182D		
EGESHIG	1821 182D 1821 1831 1822 182D		
BOEJIGCI	182A 1825 1835 1822 182D 1834 1822		

Note that this problem only seems to occur in words with ligatures, so that the following examples, with no ligature preceding the -IG are rendered correctly with the feminine form of the letter GA.

CECEGLIG <1834 1821 1834 1821 182D 182F 1822 182D> 

CIGLEN <1834 1822 182D 182F 1821 1828> 

CIRIG <1834 1822 1837 1822 182D> 

3.2 Manchu Letter KA

U+1874 MONGOLIAN LETTER MANCHU KA may take either a masculine or a feminine form finally or medially before a consonant. The wrong gender selection rules appear to have been applied for final KA and medial KA before a consonant. The feminine form of the letter KA should only be selected by default in a syllable-final position if it follows the syllables KU <1874 1860>, GU <1864 1860>, HU <1865 1860> or HUU <1865 1861>, or if it follows U+185D MONGOLIAN LETTER SIBE E other than the syllable TE <1868 185D>. In all other cases the masculine form should be selected. However, with Mongolian Baiti the masculine form is selected in all cases after U+1820 MONGOLIAN LETTER A, U+1823 MONGOLIAN LETTER O and U+1861 MONGOLIAN LETTER SIBE U; and the feminine form is selected in all cases after the letters U+185D MONGOLIAN LETTER SIBE E, U+1873 MONGOLIAN LETTER MANCHU I and U+1860 MONGOLIAN LETTER SIBE UE.

Table 3-2. Incorrect Gender Forms for Manchu KA

Word	Character Sequence	Actual Appearance	Expected Appearance
IKTAMBI	1873 18741868 1820 182E 182A 1873		
CIKSIN	1834 1873 1874 1830 1873 1828		
UKCAMBI	1860 1874 1834 1820 182E 182A 1873		
FUKJIN	1876 1860 1874 1835 1873 1828		
TEKSILEMBI	1868 185D 1874 1830 1873 182F 185D 182E 182A 1873		

In addition to the above problem, when a medial KA occurs in either a masculine or femine context and is immediately preceded or followed by a consonant-vowel ligature, then the undotted masculine form is incorrectly selected, as shown in the following examples:

BEKDUN <182A 185D 1874 1869 1860 1828>

Actual Appearance : 

Expected Appearance : 

BOKTO <182A 1823 1874 1868 1823>

Actual Appearance : 

Expected Appearance : 

3.3 Sibe Letter TA

In a medial position before a consonant U+1868 MONGOLIAN LETTER SIBE TA should always take the third medial form. With Mongolian Baiti, the first medial form is incorrectly selected when Sibe TA precedes a consonant-vowel ligature, as shown in the following example:

BITHE <182A 1873 1868 1865 185D>

Actual Appearance : 

Expected Appearance : 

---

4. Isolate Forms

In principal, the glyph form of a letter rendered in isolation should be:

- the special isolate form for vowels (1820..1827, 1844..1849, 185D..185E, 1860..1861, 1873 and 1887..1888)
- the medial form for letters which do not occur initially (1829, 1843, 184A, 185F and 1862)
- the feminine isolate form for consonants with gender-specific forms (182C..182D, 184D..184E, 1863..1865, 1874 and 189A)
- the initial form (but with no stem extension) for all other letters (1828, 182A..182B, 182E..1842, 184B..184C, 184F..185C, 1866..1872, 1875..1877, 1887..1899, 189B..18A8).

However, in order to ensure that each character in the Mongolian block has a unique glyph form, the representative glyphs shown in the Unicode character charts do not always show the isolated form of the character (e.g. U+1825 and U+1826 have identical isolate forms, so the code chart glyph for U+1825 is the

isolate form, whereas the code chart glyph for U+1826 is the initial form). The ramifications of this are that a font's CMAP should map each code point in the Mongolian block to its Unicode code chart glyph form, but the rendering system should select the appropriate isolated form when a letter is rendered in isolation. This ensures that user expectations are met when typing in Mongolian text, whilst also allowing an application that needs to display the code chart forms of letters (e.g. a character map application) to do so by by-passing the rendering system (e.g. Uniscribe) and extracting the correct glyph directly from the CMAP table.

With Uniscribe and Mongolian Baiti there is inconsistency with regards to which glyph form is selected for a letter in isolation, in some cases the code chart glyph form is selected (e.g. 182C..182D), whereas for other letters the initial glyph form is selected (e.g. 1863..1865). All the letters in the Mongolian block that have code chart glyphs that differ from the expected isolated form are listed in Table 4-1.

Table 4-1. Isolated Glyph Forms

Character	Character Name	Code Chart Glyph	Expected Isolated Glyph	Actual Isolated Glyph
1824	MONGOLIAN LETTER U	ᠤ	ᡠ	ᡠ
1826	MONGOLIAN LETTER UE	ᠦ	ᡡ	ᡡ
182C	MONGOLIAN LETTER QA	ᠬ	ᢀ	ᠬ
182D	MONGOLIAN LETTER GA	ᠭ	ᢁ	ᠭ
1833	MONGOLIAN LETTER DA	ᠳ	ᡩ	ᠳ
184A	MONGOLIAN LETTER TODO ANG	ᡃ	ᡂ	ᡃ
184B	MONGOLIAN LETTER TODO BA	ᡄ	ᡅ	ᡄ
184D	MONGOLIAN LETTER TODO QA	ᡆ	ᡇ	ᡆ
184E	MONGOLIAN LETTER TODO GA	ᡈ	ᡉ	ᡈ
184F	MONGOLIAN LETTER TODO MA	ᡊ	ᡋ	ᡊ
1859	MONGOLIAN LETTER TODO HAA	ᡙ	ᡚ	ᡙ
185D	MONGOLIAN LETTER SIBE E	ᡝ	ᡜ	ᡜ
185E	MONGOLIAN LETTER SIBE I	ᡞ	ᡟ	ᡟ
1861	MONGOLIAN LETTER SIBE U	ᡡ	ᡠ	ᡠ
1862	MONGOLIAN LETTER SIBE ANG	ᡢ	ᡣ	ᡢ

Character	Character Name	Code Chart Glyph	Expected Isolated Glyph	Actual Isolated Glyph
1863	MONGOLIAN LETTER SIBE KA			
1864	MONGOLIAN LETTER SIBE GA			
1865	MONGOLIAN LETTER SIBE HA			
186A	MONGOLIAN LETTER SIBE JA			
1872	MONGOLIAN LETTER SIBE ZHA			
1873	MONGOLIAN LETTER MANCHU I			
1874	MONGOLIAN LETTER MANCHU KA			
1875	MONGOLIAN LETTER MANCHU RA			
1876	MONGOLIAN LETTER MANCHU FA			

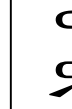
Note that for U+182C MONGOLIAN LETTER QA and U+182D MONGOLIAN LETTER GA, the initial form produced with ZWJ (i.e. `<182C 200D>` and `<182D 200D>`) is the feminine isolate form in Mongolian Baiti, when according to **MGWBM** it should be the mssculine initial form. That is to say, the first initial and first isolate forms are swapped with respect to the positional definitions in **MGWBM**.

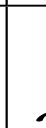
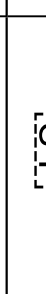
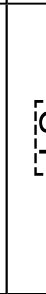
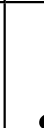
5. Ligatures

5.1 Required Ligatures

Table 5-1 lists all the required ligatures that are incorrect or missing according to **Report 170** and **MGWBM**.

Table 5-1. Incorrect or Missing Ligatures

Character Sequence	Raw Glyphs	Actual Appearance				Expected Appearance				Reference
		Isolate	Initial	Medial	Final	Isolate	Initial	Medial	Final	
`182A 1861`										MGWBM #8
`182B 1844`						Should be ligatured				MGWBM #161

Character Sequence	Raw Glyphs	Actual Appearance				Expected Appearance				Reference
		Isolate	Initial	Medial	Final	Isolate	Initial	Medial	Final	
182B 1845						Should be ligatured				MGWBM #162, 163
182B 1847						Should be ligatured				MGWBM #164, 165
182C 180B 1825										MGWBM #35
182C 180B 1826										MGWBM #35
182C 1827										MGWBM #39, 40 See **Note 1**
183A 1844						Should be ligatured				MGWBM #176
183A 1845						Should be ligatured				MGWBM #177, 178
183A 1847						Should be ligatured				MGWBM #179, 180
183B 1847						Should be ligatured				MGWBM #181, 182
1892 1826										MGWBM #157, 158

Character Sequence	Raw Glyphs	Actual Appearance				Expected Appearance				Reference
		Isolate	Initial	Medial	Final	Isolate	Initial	Medial	Final	
1892 1873										MGWBM #151, 152
1893 1826										MGWBM #172, 173
18A8 1873										MGWBM #189, 190 See **Note 2**

Note 1

According to **Report 170** and **MGWBM** the dotted form should be selected by default, without the need to apply an FVS. With Mongolian Baiti the undotted form is selected by default, and FVS-1 is required to select the dotted form.

Note 2

Acccording to **Report 170** and **MGWBM** the vowel in this ligature is U+1873 MONGOLIAN LETTER MANCHU I, but with Mongolian Baiti the ligature can be made with U+1888 MONGOLIAN LETTER ALI GALI I, but not with the ordinary Manchu I.

5.2 Optional Ligatures

Mongolian Baiti provides some optional ligatures, as shown in Table 5-2:

Table 5-2. Optional Ligatures

Ligature	Character Sequence	Default Appearance	Ligature Inhibited
BL	182A 182F		
Ali Gali KL	1889 182F		

Whilst there is nothing wrong with these ligatures, they are not required, but may be used or not used according to the whims of the individual user. As such, I would recommend that the ligatures are not selected by default, but should only be selected if the user explicitly requests ligation by inserting ZWJ between the letters. For example <182A 182F> would select the unligatured BL, and <182A 200D 182F> would select the ligatured BL (at present the behaviour is exactly opposite, so that ZWJ can be used to inhibit the default ligature behaviour).

6. Variation Sequences

6.1 Variant Isolate Forms

There are a number of variant isolate forms defined in Mongolian Baiti that are not defined in **MGWBM** and that use variation sequences that are not defined by Unicode. All of these variant isolate forms are identical to another positional form of the letter, and can be selected using the appropriate combination of ZWJ and FVS. I do not believe any of these are true isolate forms which require special variation sequences other than the already defined sequences for when they occur in a non-isolate position.

Table 6-1. Undefined Isolate Variants

Code Point	Character Name	Glyph	Isolate Selection	Non-Isolate Selection	Notes
1820	MONGOLIAN LETTER A		1820 180C	200D 1820 180B	Same as second final form
1821	MONGOLIAN LETTER E		1821 180B	200D 1821 180B	Same as second final form
1822	MONGOLIAN LETTER I		1822 180B	200D 1822	Same as final form
1824	MONGOLIAN LETTER U		1824 180B	200D 1824	Same as final form
1826	MONGOLIAN LETTER UE		1826 180C	200D 1826	Same as first final form
182D	MONGOLIAN LETTER GA		182D 180B	200D 182D 180D 200D	Same as feminine medial form
1835	MONGOLIAN LETTER JA		1835 180B	200D 1835 180B 200D	Same as second medial form

---

6.2 Variant Initial and Final Forms

There are three variant initial or final forms of letters that that are not defined in **MGWBM** and which use undefined variation sequences. These are identical to the medial forms of the same letter that are selected after NNBSP or before NVS (see Tables 6-2 amd 6-3), and I do not think that they are true initial or final forms. Any usage in initial or final position (e.g. when discussing a suffix in isolation) can be dealt with using the existing, defined variation sequences and ZWJ where appropriate (e.g. the suffix ACA that occurs after NNBSP can be represented in isolation as <200D 1820 180C 1834 1820>, without requiring a special initial variant).

Table 6-2. Undefined Initial Variants

Code Point	Character Name	Glyph	Initial Variant (undefined)	Medial Variant (defined)	Notes
1820	MONGOLIAN LETTER A		1820 180B 200D	200D 1820 180B 200D	Initial variant is the same as the second medial form (used after NNBSP)

Code Point	Character Name	Glyph	Initial Variant (undefined)	Medial Variant (defined)	Notes
1826	MONGOLIAN LETTER UE	ᠦ	1826 180B 200D	200D 1826 200D	Initial variant is the same as the first medial form (used after NNBSP)

Table 6-3. Undefined Final Variants

Code Point	Character Name	Glyph	Final Variant (undefined)	Medial Variant (defined)	Notes
1828	MONGOLIANLETTER NA		200D 1828 180C	200D 1835 180C 200D	Final variant is the same as the third medial form (used before MVS)

6.3. Variant Forms of Mongolian Letter LHA

U+1840 MONGOLIAN LETTER LHA has undefined initial and medial variation sequences:

LHA + FVS-1 (initial) 1840 180B 200D 

LHA + FVS-1 (medial) 200D 1840 180B 200D 

Not only are these variants not defined in **MGWBM** or by Unicode, but they are not even variant forms of the Mongolian letter LHA, but are in fact forms of an as yet unencoded Manchu Ali Gali letter. This Manchu letter LHA is a completely different letter from Mongolian LHA (Mongolian letter LHA is in origin a ligature of LHA plus HAA, whereas Manchu LHA is the letter LHA with a circle diacritic), and should be encoded separately. I have already submitted a proposal to encode the Manchu Ali Gali letter LHA for consideration at the February meeting of the UTC. Until such time as Manchu Letter LHA is encoded, I suggest that it is removed from Mongolian Baiti, or, at the very least, the undefined FVS sequences are removed.

7. Ali Gali I

The use of U+1888 MONGOLIAN LETTER ALI GALI I in Manchu is very problematic. In Mongolian there is a clear distinction between the use of ordinary Mongolian I (U+1822) and the use of Ali Gali I. The latter is used to represent Sanskrit/Tibetan I, and the final positional form has a distinct, rounded glyph shape, whereas the former is used to represent Mongolian I and Sanskrit/Tibetan E. In Manchu, there is no such distinction in either usage or glyph form, and all occurences of I in both Manchu and Ali Gali Manchu could and should be represented by U+1873 MONGOLIAN LETTER MANCHU I. However, in Mongolian Baiti it is necessary to use either Manchu I or Ali Gali I, in an inconsistent manner, in order to get the correct form of the letter I for some Manchu Ali Gali sequences, as shown in Table 7-1, where the correct appearance is shown in green and the incorrect appearance is shown in red. Some sequences are only correctly rendered using Ali Gali I, some sequences are only correctly rendered using Manchu I, and other sequences are rendered correctly using either Manchu I or Ali Gali I (with the help of FVS or ZWJ in places).

Table 7-1. Manchu Syllables with Ali Gali I

Syllable	Consonant	With Manchu I		With Ali Gali I	
		Character Sequence	Appearance	Character Sequence	Appearance
ZI	U+186F MONGOLIAN LETTER SIBE ZA	186F 1873		186F 1888;	
JHI	U+189D MONGOLIAN LETTER MANCHU ALI GALI JHA	189D 1873		189D 1888;	
BHI	U+18A8 MONGOLIAN LETTER MANCHU ALI GALI BHA	18A8 1873		18A8 1888;	
KI	U+186C MONGOLIAN LETTER SIBE GAA	186C 1873		186C 1888	
KHI	U+183A MONGOLIAN LETTER KA	183A 1873		183A 1888	
GI	U+1864 MONGOLIAN LETTER SIBE GA	1864 200D 1873		1864 1888 180C	
GHI	U+189A MONGOLIAN LETTER MANCHU ALI GALI GHA	189A 1873		189A 1888 180C	

8. Glyph Issues

8.1 Final Form of Letter JA

The glyph for the final form of U+1835 MONGOLIAN LETTER JA is identical to the glyph for the final form of U+1834 MONGOLIAN LETTER CHA:

This is wrong, as the final form of JA should look like the medial form of JA, but with a tail.

8.2 Final Form of Sibe Letter ZA

The final glyph form of U+186F MONGOLIAN LETTER SIBE ZA has a tail:

The final form of the letter should be the same as the medial form, with no tail. The tail in this glyph is the form of the letter I that occurs after ZA, and so the glyph in fact represents the syllable ZI.

9. Multiple Stacked FVSs

Unexpected behaviour is seen when multiple FVSs are attached to a single character. When two FVSs (any combination of FVS-1, FVS-2, FVS-3) are attached to a base letter, the character is unexpectedly rendered twice, as shown by the examples in Table 8-1. I would expect that any FVS immediately following another FVS would be ignored by the rendering system.

Table 9-1. Unexpected behaviour with multiple FVSs

Letters	Character Sequence	Actual Appearance	Expected Appearance
A + FVS-1 + FVS-2	1820 180C 180D		
BA + FVS-2 + FVS-3	182A 180C 180D		
QA + FVS-1 + FVS-1	182C 180B 180B