Comments on the Mongolian block

Weizhe Zheng (wzheng@math.ac.cn)

October 30, 2016

This is a revised version of L2/16-292. More details are added pursuant to feedback from Greg Eck. [**Added on Nov. 1:** Proposed changes that coincide with L2/16-309 are shown in <mark>yellow</mark>.]

Contents

Acknowledgment. These comments and proposals are based on research of many people, and in particular that of Yihua Wang.

## 1 Comments on the Unicode Standard 9.0, Section 13.5

Pronunciations of "velar" letters. On page 532, it is stated that "For Mongolian and Todo, the masculine and feminine forms of these letters have different pronunciations." The restriction to Mongolian and Todo in this statement is unnecessary. In Manchu and Sibe, the masculine uvular letters ꡃ ꡃ ꡃ and the feminine velar letters ꡃ ꡃ ꡃ are unified in Unicode, despite being treated as different letters in both traditional syllabaries and modern orthography[1]. To avoid confusion, this unification should perhaps be explained more explicitly in the Standard, as it cannot be inferred from the code chart whether the feminine velar letters ꡃ ꡃ ꡃ are unified with the masculine velar letters ꡃ ꡃ ꡃ or with the masculine uvular letters ꡃ ꡃ ꡃ.

Syllable-final "velar" letters in Manchu and Sibe. On page 532–533, it is stated that the rendering system should select the correct gender form of these letters based on "the gender of the preceding vowel". This statement is inaccurate. In Sibe, the only "velar" letter appearing in syllable-final position is U+1863 MONGOLIAN LETTER SIBE KA, and it always takes the masculine form in such positions. [**Added on Nov. 1:** The Todo case is similar to Sibe.] In Manchu, the gender of U+1874 MONGOLIAN LETTER MANCHU KA in syllable-final positions is determined by the following rules:

1. masculine if preceded by $a/i/o/u$; feminine if preceded by $e/\bar{u}$. (u=U+1860 ꡧ, $\bar{u}$=U+1861 ꡧ)

2. (overriding rule 1) masculine if preceded by te; feminine if preceded by $ku/gu/hu$.

By rule 1, the gender of syllable-final U+1874 often differs from the gender of the preceding vowel. By rule 2, the gender of syllable-final U+1874 depends not only on the preceding vowel, but also on the letter preceding the vowel in some cases.

---

[1] See the official Sibe orthography 《ꡃ ꡃ ꡃ ꡃ ꡃ ꡃ ꡃ ꡃ》.

Baluda. On page 534, U+1885 ᢅ ᴍᴏɴɢᴏʟɪᴀɴ ʟᴇᴛᴛᴇʀ ᴀʟɪ ɢᴀʟɪ ʙᴀʟᴜᴅᴀ and U+0F85 ࿅ ᴛɪʙᴇᴛᴀɴ ᴍᴀʀᴋ ᴘᴀʟᴜᴛᴀ are associated with Sanskrit *avagraha*. This association does not seem to be accurate. Both U+1885 and U+0F85 are *pluta* signs indicating extra long vowels and are related to Sanskrit numeral three (U+0969 ३ ᴅᴇᴠᴀɴᴀɢᴀʀɪ ᴅɪɢɪᴛ ᴛʜʀᴇᴇ).

2 Proposed changes to glyphs and glyph presentation on the Unicode 9.0 code chart

code points and position	Unicode 9.0	desired	comments
1820 180B second form (medial)	ᚭ		The glyph of the second medial form on the Unicode 9.0 code chart is a duplicate of the third medial form ᚭ. This is a typo in Unicode 9.0. Other sources (Unicode 8.0 list of standardized variants, GB standards, 《蒙古文编码》, DS01) all give the desired ᴨ as the second medial form.
186F 180B second form (initial)			The short stroke (circled) in the current glyphs should be removed. The second forms of 186F are only used when followed by ᚭ or ᚭ. Yet the short stroke prevents the glyph from joining smoothly with the following letter. The short stroke is not present in Unicode 8.0, GB/T 26226-2010, or 《蒙古文编码》. Compare ᚭ (current) with ᚭ (desired).
186F 180B second form (medial)			
185E 180C third form (medial)			The current glyph of the third medial form is a duplicate of the first medial form. This is not desired, because the third medial form is not used to override context-driven forms. Rather it is itself a context-driven form used only when preceded by the letter ᴢᴀ (ᚭ or ᚭ). Yet the current glyph does not join correctly with the preceding ᴢᴀ. The desired glyph is identical to the fourth medial form of U+1873, whose curly shape allows it to join correctly with the preceding ᴢᴀ. Both 《蒙古文编码》 and GB/T 26226-2010 give the desired glyph and identify it with the fourth medial form of U+1873. Compare ᚭ (current) with ᚭ (desired).
185D first form (isolate)			The first (and only) isolate forms of these Manchu and Sibe vowel letters should be shown on the Unicode code chart since they differ from the code point glyphs. Isolate forms of Mongolian vowel letters are shown on the Unicode 9.0 code chart whenever they differ from the code point glyphs (see U+1824 and U+1826 on the code chart). Manchu and Sibe vowel letters should be treated in the same way.
185E first form (isolate)	not shown		
1873 first form (isolate)			

(continued)

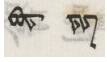| code points and position	Unicode 9.0	desired	comments
1869 first form (final)	⟨glyph⟩	not shown or <mark>marked as unattested</mark>	These glyphs added in the Unicode 9.0 code chart are not attested in the indicated positions. They are not included in 《蒙古文编码》 or GB standards. They are appropriate for inclusion in documents such as DS01 where they are clearly marked as unattested. Including them in the Unicode code chart without marking them as unattested causes confusion, and makes the code chart unnecessarily unstable, because some attestation in the position may later be discovered and the glyph will have to be changed. In the case of U+188A ALI GALI NGA, there are at least three attested final forms: ⟨glyph⟩, ⟨glyph⟩ 《大藏全咒》, ⟨glyph⟩ (Mongolian Kanjur). It is likely that one of them will eventually be defined as the first final form of U+188A.
1876 first form (final)	⟨glyph⟩		
1887 first form (initial)	⟨glyph⟩		
1887 first form (medial)	⟨glyph⟩		
188A first form (final)	⟨glyph⟩		
1887 first form (isolate)	not shown	⟨glyph⟩	The desired glyphs are that of Mongolian and Todo Ali Gali A. The current glyph ⟨glyph⟩ is the final form of Manchu Ali Gali AH, but the two letters cannot be unified. See the next heading.
1887 code point glyph	⟨glyph⟩		
1887 first form (final)	⟨glyph⟩	⟨glyph⟩	
1887 180D fourth form (final)	⟨glyph⟩	⟨glyph⟩	see below

The fourth final form of U+1887 MONGOLIAN LETTER ALI GALI A is given the glyph ⟨glyph⟩ in Unicode 8.0, and ⟨glyph⟩ in Unicode 9.0. Neither glyph should be encoded as a final form of U+1887:

- In Mongolian, these glyphs are precursors in an older style of the backward tail ⟨glyph⟩ of U+1820 A, U+1821 E, and U+1828 NA, and should be encoded at these code points. Several variants of these glyphs have indeed been proposed as final forms of U+1820 and U+1828 in N4757 (L2/16-261).

- In Manchu Ali Gali, ⟨glyph⟩ transcribes Sanskrit long vowel ā and the current glyph is a decorative variant thereof. Either glyph is in fact a sequence of two letters: ⟨glyph⟩ (medial form of U+1820) followed by ⟨glyph⟩ (final form of MANCHU ALI GALI AH).

Mongolian/Todo Ali Gali A and Manchu Ali Gali AH. [**Added on Nov. 1:** This heading proposes no changes in itself, but provides explanations for the changes proposed for 1887 above and for the character 18AE proposed in the next section.]

The contextual forms of the two letters are detailed in the following table.

character	⊤⊐ 18AE Manchu Ali Gali ᴀʜ	ꞁ 1887 Mongolian/Todo Ali Gali ᴀ	
isolate form	⊤⊐	ꓛ	
initial form	⊤	none	
medial form	none	none	
final form	⊐	ꞁ first form default	
		⁄ second form context-driven: 18A6 1887 ✓	
		ꞁ third form context-driven: 188A 1887 ꞁ)/ꞁ)	
		⊐ꞁ fourth form post-MVS, limited usage	

In Mongolian/Todo, the shape ⊐ does not represent ᴀ, but rather is part of the final forms of several letters. In the example 𑐀𑐀 *bud dha* from 《大藏全咒》, the shape ⊐ at the end of the first syllable is part of the final form of U+1891 ᴍᴏɴɢᴏʟɪᴀɴ ʟᴇᴛᴛᴇʀ ᴀʟɪ ɢᴀʟɪ ᴅᴀ[2], representing the consonant *d* with no following vowel.

As can be seen from the table above, Mongolian/Todo Ali Gali ᴀ and Manchu Ali Gali ᴀʜ do not share any common glyph. They also have different linguistic values: Mongolian/Todo Ali Gali ᴀ is a vowel, while Manchu Ali Gali ᴀʜ transcribes U+0F60 ཀ, Tibetan consonant letter -ᴀ. For these reasons, the two letters cannot be unified and should be encoded as different characters.

The following table from 《同文韻統》 gives examples of Manchu Ali Gali ᴀʜ (circled) corresponding to another Mongolian/Todo Ali Gali letter, U+1897 ᴍᴏɴɢᴏʟɪᴀɴ ʟᴇᴛᴛᴇʀ ᴀʟɪ ɢᴀʟɪ ᴀʜ.

3 Preliminary proposal to encode four letters in Todo and Manchu Ali Gali
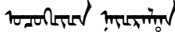
[**Added on Nov. 1:** This section should be superseded by a separate document.]

The proposed characters and glyphs are the following:

- 18AB ꞁ ᴍᴏɴɢᴏʟɪᴀɴ ʟᴇᴛᴛᴇʀ ᴛᴏᴅᴏ ᴀʟɪ ɢᴀʟɪ ʀᴇᴠᴇʀsᴇᴅ ɪ

  ~ 18AB ꞁ (final)

- 18AC ꓘ ᴍᴏɴɢᴏʟɪᴀɴ ʟᴇᴛᴛᴇʀ ᴍᴀɴᴄʜᴜ ᴀʟɪ ɢᴀʟɪ ᴠᴏᴄᴀʟɪᴄ ʀ
  ~ 18AC ꓘ (initial)
  ~ 18AC ꓘ (medial)

- 18AD ᴛ́ ᴍᴏɴɢᴏʟɪᴀɴ ʟᴇᴛᴛᴇʀ ᴍᴀɴᴄʜᴜ ᴀʟɪ ɢᴀʟɪ ɢᴀ

  ~ 18AD ᴔ first form (initial)

  ~ 18AD ᴛ́ first form (medial)

  ~ 18AD 180B ᴦ́ second form (medial)

- 18AE ⊤⊐ ᴍᴏɴɢᴏʟɪᴀɴ ʟᴇᴛᴛᴇʀ ᴍᴀɴᴄʜᴜ ᴀʟɪ ɢᴀʟɪ ᴀʜ
  ~ 18AE ⊤⊐ (isolate)

---

[2]In Baiti, the final form would look like ⊓.

~ 18AE ⊤ (initial)
~ 18AE ⌐ (final)

U+18AB Todo Ali Gali Reversed I. This letter transcribes U+0F80 ྀ ᴛɪʙᴇᴛᴀɴ ᴠᴏᴡᴇʟ ꜱɪɢɴ ʀᴇᴠᴇʀꜱᴇᴅ ɪ and is used in the transcription of the Sanskrit vowels ṛ and ḷ. The glyph is derived by reflection from ᠥ, the final form of U+1845 ᴍᴏɴɢᴏʟɪᴀɴ ʟᴇᴛᴛᴇʀ ᴛᴏᴅᴏ ɪ. The two letters have different sound values. Examples can be found in the following sample of Zaya Bandida's ᠪᠣᠯᠣᠷᠤᠨ ᠳᠤᠷᠠᠰᠬᠠᠯ ᠪᠢᠴᠢᠭᠡᠰᠦ.
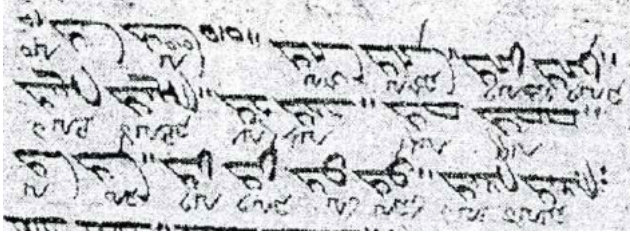

On a side note, the above sample also shows that U+1843 ᠃ ᴍᴏɴɢᴏʟɪᴀɴ ʟᴇᴛᴛᴇʀ ᴛᴏᴅᴏ ʟᴏɴɢ ᴠᴏᴡᴇʟ ꜱɪɢɴ is not a letter, but a diacritic mark.

U+18AC Manchu Ali Gali Vocalic R. This letter is used in the transcription of the Sanskrit vowel ṛ. It differs from U+1875 ᴍᴏɴɢᴏʟɪᴀɴ ʟᴇᴛᴛᴇʀ ᴍᴀɴᴄʜᴜ ʀᴀ in shape and sound value, as shown in the following examples from 《同文韻統》:

- 18AC 1873 ᠊᠊᠊ 𑖀 ṛ

- 1875 1873 ᠊᠊᠊ 𑖀 ri

Manchu Ali Gali vocalic R also appears in the medial position, as shown in the circled items of the following table from 《同文韻統》.

Manchu Ali Gali Vocalic L. The Manchu transcription of the Sanskrit vowel ḷ uses a letter that differs from Manchu ʟᴀ in shape in the initial position, as shown by the following examples from 《同文韻統》:

- ???? 1873 ᠊᠊᠊ ḷ

- 182F 1873 ᠊᠊᠊ li

However, in the medial position, there seems to be no difference with U+182F. Since attestations of ḷ in the initial position are rare, more discussion may be needed on the encoding of this letter.

U+18AD Manchu Ali Gali GA. This letter is derived from the masculine form of U+1864 ᴍᴏɴɢᴏ-ʟɪᴀɴ ʟᴇᴛᴛᴇʀ ꜱɪʙᴇ ɢᴀ. In contrast to U+1864, U+18AD takes the masculine form even when followed by the vowels i and e. When followed by i and e, the two letters differ in shape and sound value, as shown by the following examples from 《同文韻統》 (see the table above for context):

18AD	᠊᠊᠊ gi	᠊᠊᠊ ge	
1864	᠊᠊᠊ ki	᠊᠊᠊ ke	

The sequence 18AD 185D *ge* forms a ligature in the isolate, initial, and medial positions. In the final position, 18AD 185D *ge* is attested in two different shapes, making a second medial form of 18AD necessary. Examples of this sequence are given below.

- 18AD 185D (isolate) ge 《同文韻統》

- 18AD 185D (initial) gei 《同文韻統》

- 18AD 185D (medial) 'gel 《五體清文鑑》

- 18AD 185D (final) ngge 《大藏全咒》

- 18AD 180B 185D (final) dge 《五體清文鑑》

U+18AE Manchu Ali Gali AH. See the last part of Section 2.

4 Proposal to define six variation sequences

Many variants in the Mongolian block were proposed in N4752R2 (L2-16/258, DS01) and N4757 (L2-16/261). We propose five additional variants and provide more information on one previously proposed variant.

Very common variants

In Manchu and Sibe words consisting of a consonant (or a sequence of consonants in Manchu Ali Gali) followed by U+1860 ᡠ MONGOLIAN LETTER SIBE UE, the vowel takes one of the following forms, depending on the preceding consonant.

proposed form	glyph	comments
1860 180C third form (final)	ᠥ	context-driven: used in monosyllabic words when preceded by a consonant other than 1868/1869/18A0/18A1.
1860 180D fourth form (final)	ᠥ	context-driven: used in monosyllabic words when preceded by 1868/1869/18A0/18A1. Examples: .

The third final form is included in GB/T 26226-2010 and DS01. It is included here because of its high frequency and to provide more information.

Both forms are very commonly used in native words and in transcriptions of other languages. Without either of the two forms, the most basic syllabary (*a e i uju*) cannot be rendered, as shown in the following sample of 《清漢對音字式》.
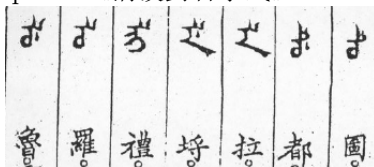
Other variants

proposed form	glyph	comments	examples
1859 180B second form (medial)	ᡙ	context-driven: used in transcriptions of aspirated voiced consonants	see the Todo sample in Section 3
185D 180C third form (medial)	▪	context-driven: 189B 185D (initial) ᡳ 18AE 185D (initial) ᡨ	《同文韻統》 ᠠ ngei ᠠ ngen
185D 180C third form (final)	)	context-driven: 189B 185D (isolate) ᡰ 18AE 185D (isolate) ᠵ 18AD 185D (final) ᠨ	《同文韻統》 ᠠ nge 《大藏全咒》 ᠠ ngge
1865 180B second form (medial)	ᡥ	The default medial/final form of $h\bar{u}$ is 1865 1861 ᠨᠣ/ᠨᠣ. In a few Manchu words, $h\bar{u}$ takes the exceptional form 1865 180B 1861 ᡥᠣ/ᡥᠣ.	《五體清文鑑》 ᠨᠠᡥᠣᠨᠠ ᠨᡥᠣ