

Preliminary comments on L2/16-309

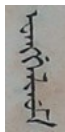
Weizhe Zheng
November 2, 2016

Many of the proposed changes and additions are indeed very much needed. There are however a few problematic glyphs/forms.

1.

1807 (__x) 𐎆 MONGOLIAN SIBE SYLLABLE BOUNDARY MARKER
~ 1807 𐎆 (medial)

As mentioned on page 535 of the Unicode Standard and in DS01, 1807 is used not only in Sibe



but also in Manchu. Example: *ewanggeli-um* from the Manchu Bible. A more accurate glyph would be 𐎆.

2. Some proposed forms are in fact combinations of two characters.

2.1

~ 1873 180D 𐎆 (__x) 4th form (final)

This proposed final form of Manchu I is better analyzed as the sequence 1807 Syllable Boundary



Marker followed by 1873 Manchu I. Example: *kui-i*.

2.2

~ 1864 180B 𐎆 (__x) 2nd form (initial)

This proposed form of Sibe GA transcribes Sanskrit and Tibetan *ge*, and is in fact a ligature of two letters: the consonant Ali Gali GA followed by the vowel 185D Sibe E. This transcription is explained in more details in Section 3 of L2/16-292, with examples of its usage in all positions. As explained there, another reason for not encoding it with 1864 is that when followed by 185D, 1864 already transcribes Sanskrit and Tibetan *k*.

2.3

189A (__x) 𐎆 MONGOLIAN LETTER MANCHU ALI GALI GHA
~ 189A 𐎆 1st form (__x) (initial)
~ 189A 180B 𐎆 2nd form (__x) (initial)
~ 189A 𐎆 (__x) (medial)

This proposed form transcribes Sanskrit *ghe*, and is in fact a ligature of two letters: the consonant 189A Ali Gali GHA followed by the vowel 185D Sibe E.

2.4

18A6 (xx_) 𑖠 MONGOLIAN LETTER ALI GALI HALF U

- ~ 18A6 𑖠 (____) (initial)
- ~ 18A6 𑖠 (____) (medial)
- ~ 18A6 𑖠 (xx_) 1st form (final)
- ~ 18A6 180B 𑖠 (x_) 2nd form (final)

The proposed first final form transcribes Tibetan *wa*, and is perhaps better analyzed as a ligature of two letters: 18A6 (medial) followed by 1887/1820 A. In fact, this is a slight variation of

𑖠, another attested transcription of *wa*, which is encoded as 18A6 1887. It would be strange to encode such glyph variations of the same syllable using different numbers of base characters.

The proposed second final form of 18A6 is a duplicate of the second final form of 1887 Ali Gali A.

- ~ 1887 180B 𑖠 (xx_) 2nd form (final)
 - context driven

2.5

18A7 (x_) 𑖡 MONGOLIAN LETTER ALI GALI HALF YA

- ~ 18A7 𑖡 (____) (initial)
- ~ 18A7 𑖡 (x_) (medial)
- ~ 18A7 𑖡 (x_) (final)

The proposed final form transcribes Tibetan *ya*, and is perhaps better analyzed as a ligature of two letters: 18A7 followed by 1887/1820 A.

3. Some forms are very unusual, and it would be helpful if some attestations could be provided.

Examples:

~ 1867 𑖢 (____x) (final)

~ 1874 180C 𑖣 (____x) 3rd form, feminine (isolate)

~ 1874 180C 𑖣 (____x) 3rd form, feminine (initial)

1887 (xx_x) 𑖤 MONGOLIAN LETTER ALI GALI A

~ 189B 𑖤 (____x) (final)

For more comments on 1887, see Section 2 of L2/16-292.

1888 (x_x) 𑖥 MONGOLIAN LETTER ALI GALI I

~ 1888 𑖥 (x_x) 1st form (final)

1888 Ali Gali I is marked as used in Manchu. However, in Manchu Ali Gali, I is written with 1873

Manchu I. It would be helpful if some attestations of 1888 in Manchu could be provided.

4. Inconsistent indication of context-driven forms

Some context-driven forms are not marked as such. Examples:

~ 1860 180C ^ᠴ (_ _xx) 3rd form (final)

~ 1863 180B ^ᠳ (_ _x) 2nd form (medial)

For the context of the third final form of 1860, see Section 4 of L2/16-292.

Some forms are only found in context-driven ligatures, and they should be treated consistently. Compare the feminine initial forms of 1863, 1864, 1865 (not marked as context-driven) with

~ 1874 180B ^ᠴ (_ _x) 2nd form, feminine (initial)

• context driven

Preliminary comments on L2/16-309 (continued)

Weizhe Zheng

November 10, 2016

4. Inconsistent indication of context-driven forms (continued)

Some forms are incorrectly marked as context-driven. Examples:

~ 1843 ¹ (x_) (medial)

- context driven

~ 1843 ¹ (x_) (final)

- context driven

~ 1873 180B ¹ (x_) 2nd form (medial)

- context driven

5. Mongolian and Todo old forms

Many proposed Todo isolate/initial forms presumably represent forms in an older writing style.

Examples:

~ 1844 180B ¹ (x_) 2nd form (initial)

~ 1845 180B ¹ (x_) 2nd form (isolate)

~ 1845 180B ¹ (x_) 2nd form (initial)

~ 1846 180B ¹ (x_) 2nd form (isolate)

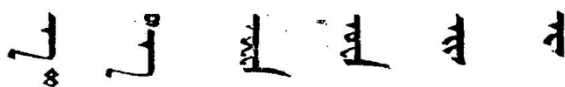
~ 1846 180B ¹ (x_) 2nd form (initial)

~ 1820 180C ¹ (x_) 3rd form (isolate)

~ 1840 180B ¹ (x_) 2nd form (initial)

There are two approaches to handle the older writing style: one can either use different fonts for different writing styles, or one can define variation sequences for forms in the older style. The proposal assigns variation sequences only to selected Todo forms, but this older style is used in both Mongolian and Todo. If the second approach is desired, one should assign variation sequences to all forms in the older style.

Mongolian examples:



Todo examples (not assigned variation sequences): 1820 isolate, 1828 initial



Another problem is that the proposed glyphs for the vowel letters do not seem to be accurate, as can be inferred from the above examples. It should be ᠠᠨ ᠠᠨ ᠠᠨ etc.

6. Some Todo Ali Gali forms are assigned to wrong code points.

1892 ᠠᠨ MONGOLIAN LETTER ALI GALI PA

- ~ 1892 ᠠᠨ 1st form (initial)
- ~ 1892 180B ᠠᠨ 2nd form (initial)
- ~ 1892 180C ᠠᠨ 3rd form (initial)
- ~ 1892 ᠠᠨ 1st form (medial)
- ~ 1892 180B ᠠᠨ 2nd form (medial)
- ~ 1892 180C ᠠᠨ 3rd form (medial)

While the proposed second forms of 1892 indeed transcribe pa (Tibetan པ), the proposed third forms transcribe pha (Tibetan ཕ) and should be encoded as variant forms of 184C.

188B ᠴᠠ MONGOLIAN LETTER ALI GALI CA

- ~ 188B ᠴᠠ 1st form (initial)
- ~ 188B 180B ᠴᠠ 2nd form (initial)
- ~ 188B ᠴᠠ 1st form (medial)
- ~ 188B 180B ᠴᠠ 2nd form (medial)
- ~ 188B ᠴᠠ (final)

In Mongolian Ali Gali the first forms transcribe Tibetan tsa ཅ (corresponding to Sanskrit ca, hence the name of the letter). Attestations in Todo Ali Gali are less common. It seems however that in Todo Ali Gali the first forms transcribe Tibetan ཅ cha (rare), while the proposed second forms transcribe Tibetan ཅ ca (rare) or ཅ tsha. Since the first and proposed second forms have different sound values in Todo Ali Gali, they should not be encoded at the same code point. The proposed second forms of 188B (at least when representing Tibetan ཅ tsha) should perhaps be unified with 183C, which has the same function and very similar glyphs.

7. Todo usage of 183A

The code points 183A and 1858 have the same shape (when attested):

183A (xxxx) ᠎ MONGOLIAN LETTER KA

→ 0445 x cyrillic small letter ha

~ 183A ᠎ (xxxx) (initial)

~ 183A ᠎ (xxxx) (medial)

~ 183A ᠎ (xx_) (final)

1858 (xx_) ᠎ MONGOLIAN LETTER TODO GAA

~ 1858 ᠎ (xx_) (initial)

~ 1858 ᠎ (xx_) (medial)

~ 1858 ᠎ (___) (final)

In Mongolian, the letter represents two different sound values:

(1) In Mongolian Ali Gali, it represents the voiced consonant /g/ (Tibetan ᠭ).

(2) In modern Mongolian, it represents the aspirated consonant /k^h/.

(1) is encoded as 1858 and (2) as 183A.

In Todo, however, it seems that the letter consistently represents the unaspirated consonant /k/ (romanized as k or g according to the scheme). Since there is no difference in shape or sound value, it is desirable to use only one code point (1858) for this letter in Todo and to change the usage of 183A to (x_xx).