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. 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  madrid.

2. Some proposed forms are in fact combinations of two characters.
   2.1
   \[ 1873 \text{ Syllable Boundary Marker} \]
   \[ 1873 180D \] 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
   \[ \text{Sibe GA transcribes Sanskrit and Tibetan } ge \text{, and is in fact a ligature of two letters: the consonant Ali Gali GA followed by the vowel 185D Sibe E.} \]
   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
   \[ 189A \text{ Manchu Ali Gali GHA} \]
   \[ 189A 180D \] 2nd form (initial)
   \[ 189A (\_\_\_) \] (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 \(\text{ 중심} \) MONGOLIAN LETTER ALI GALI HALF U
~ 18A6 \(\text{ 중심} \) (initial)
~ 18A6 \(\text{ 중心} \) (medial)
~ 18A6 \(\text{ 중심} \) 1st form (final)
~ 18A6 180B \(\text{ 중심} \) 2nd form (final)

The proposed first final form transcribes Tibetan wa, and is a perhaps better analyzed as a ligature of two letters: 18A6 (medial) followed by 1887/1820 A. In fact, this is a slight variation of \(\text{ wa} \), 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 \(\text{ 중심} \) 2nd form (final)
• context driven

2.5

18A7 \(\text{ 중심} \) MONGOLIAN LETTER ALI GALI HALF YA
~ 18A7 \(\text{ 중심} \) (initial)
~ 18A7 \(\text{ 중心} \) (medial)
~ 18A7 \(\text{ 중심} \) (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 \(\text{ 여} \) (final)

~ 1874 180C \(\text{ 여} \) 3rd form, feminine (isolate)

~ 1874 180C \(\text{ 여} \) 3rd form, feminine (initial)

1887 \(\text{ 여} \) MONGOLIAN LETTER ALI GALI A
~ 189B \(\text{ 여} \) (final)

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

1888 \(\text{ 여} \) MONGOLIAN LETTER ALI GALI I
~ 1888 \(\text{ 여} \) 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:

\[
\sim 1860 \ 180C \ (\_xx) \ 3rd \ form \ (final)
\]

\[
\sim 1863 \ 180B \ (\_x) \ 2nd \ form \ (medial)
\]

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

\[
\sim 1874 \ 180B \ (\_x) \ 2nd \ form, \ feminine \ (initial)
\]

• 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:

\[
\begin{align*}
\sim 1843 & (x\ldots) \text{ (medial)} \\
& \cdot \text{ context driven} \\
\sim 1843 & (x\ldots) \text{ (final)} \\
& \cdot \text{ context driven} \\
\sim 1873 & 180B \, x, \text{ 2nd form (medial)} \\
& \cdot \text{ context driven}
\end{align*}
\]

5. Mongolian and Todo old forms

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

\[
\begin{align*}
\sim 1844 & 180B \, \, (x\ldots) \text{ 2nd form (initial)} \\
\sim 1845 & 180B \, \, (x\ldots) \text{ 2nd form (isolate)} \\
\sim 1845 & 180B \, \, (x\ldots) \text{ 2nd form (isolate)} \\
\sim 1846 & 180B \, \, (x\ldots) \text{ 2nd form (isolate)} \\
\sim 1846 & 180B \, \, (x\ldots) \text{ 2nd form (isolate)} \\
\sim 1820 & 180C \, \, (x\ldots) \text{ 3rd form (isolate)} \\
\sim 1840 & 180B \, \, (x\ldots) \text{ 2nd form (initial)}
\end{align*}
\]

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 \( \text{ البحر} \) \( \text{ البحر} \) etc.

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

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

In Mongolian Ali Gali the first forms transcribe Tibetan ཛ (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 བ ཙ བ (rare), while the proposed second forms transcribe Tibetan ཝ ཝ ཝ (rare) or བ བ བ. 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 ཝ ཝ) 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):
In Mongolian, the letter represents two different sound values:

1. In Mongolian Ali Gali, it represents the voiced consonant /g/ (Tibetan དག). (1)
2. In modern Mongolian, it represents the aspirated consonant /kʰ/. (2)

(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).