

Preliminary analysis of L2/19-221 (WG2 N5073), China national body's comments on Mongolian

To: Unicode Technical Committee
From: 梁海 Liang Hai <lianghai@gmail.com>
Date: 25 July 2019

This is a preliminary analysis of China national body's comments on the Mongolian section of ISO/IEC 10646:2019 (6th edition) CD.2, which is in sync with the current Mongolian code chart and names list in the Unicode Standard 12.1:

- L2/19-221 (WG2 N5073), *Comments on Mongolian in UCS6CD2*

Other reviewers should note that in the analyzed document, what look like names list excerpts are not excerpts from the current names list in the Unicode Standard, but largely reflect a proposed look for the names list, and are interwoven with comments and proposals.

1 Proposal

1. Improve the representative glyphs of U+1834 MONGOLIAN LETTER CHA and related characters, per *Section 3.1*.
2. Experts working on the Mongolian UTN should investigate the requests triaged in *Section 3.3* and *3.4*.

2 Background

The current revision (Unicode 11.0–12.1) of variant information available in the Mongolian block's names list follows the original *Users' Convention*, with variants of both characters and standardized variation sequences listed. See the following documents for background of the editorial revision happened for Unicode 11.0:

- L2/17-368 (WG2 N4900), *Considerations on Mongolian Code Chart*
- L2/17-332 (WG2 N4884), *Positional Mismatches in Mongolian Encoding*

In Unicode 13.0, the Mongolian names list will no longer contain the variant information. See the following:

- L2/19-247 (WG2 N5054), *Recommendations from WG 2 meeting 68: Recommendation M68.16 (Mongolian)*
- L2/19-229 (WG2 N5103), *SC2/WG2 Mongolian Ad Hoc Report*
- Proposed Draft UTR #54 (L2/19-211, WG2 N5091), *Unicode Mongolian 12.1 Baseline*

The variant information is being improved and transferred into a Unicode Technical Note (UTN). The UTN is being drafted to allow clearer cooperation on the Mongolian

issues, and to eventually provide comprehensive specification for Mongolian encoding and shaping. The initial draft only covers the contemporary Hudum writing system:

- L2/19-130 (WG2 5104, MWG/3-N8R), *Towards a well-formed Mongolian specification that allows interoperable implementations*

3 Preliminary analysis and triage

The requests for filling *unattested cursive joining positions* (shown as square-enclosed question marks in the current names list) with a character’s *representative glyph* (referred to as “canonical character” in the analyzed document) are intentionally excluded in this analysis. Unattested positions should be handled with a specific fallback mechanism in the shaping process, which is to be further discussed and will be specified in the UTN.

Also excluded in this analysis are the requests that *would make no difference* in a future variant set where *positional mismatches* have been cleaned up (see L2/17-332) because they either originated from the issue of positional mismatches (a mismatched version of the requested variant already exists)—

- U+1828 NA, dotted final form, page 2–3
- U+182D GA, feminine initial form, page 3
- U+1836 YA, final form, page 4–5
- U+1844 TODO E, final form, page 6

—or would further contribute to the issue of positional mismatches (e.g., the request regarding U+182E MA on page 3). Note that experts already generally agree to correct positional mismatches:

- L2/18-139 (WG2 5061, MWG/3-N18R2), *Mongolian Working Group Meeting 3 (MWG3) Report: MR6. Positional mismatching*

3.1 Eligible for a future revision of the chart font

The following request affects a representative glyph:

- U+1834 CHA, properly drawn initial form, page 25

The current representative glyph of U+1834 CHA (which is its initial form) does not accurately reflect the conventional letterform. However the issue does not appear to be urgent, as it is not significant in the midst of the current chart font’s generally problematic design quality.

When a future revision of the chart font happens, this requested improvement should be included. The UTN should reflect this letterform convention as well.

Also, there are graphically related characters that should be considered along:

- *Need to improve too:* U+1871 SIBE CHA, U+1878 CHA WITH TWO DOTS.
- *Okay:* U+183C TSA, U+183D ZA, U+185C TODO DZA, U+188B ALI GALI CA.

- *Suspicious*: U+1853 TODO JA, U+1896 ALI GALI ZA, ...

3.2 Already included in the draft UTN

This is only an FYI that the following requested variant is a well-known oversight of the *Users' Convention*, and is already included in Table 4 of the draft UTN:

- U+1824 U, tailless final form, page 2

3.3 Likely eligible for a future revision of the UTN

The following requested variants are noted now, and will likely be included in a future revision of the UTN once the involved writing systems are covered, although investigation will be needed:

- U+1860 SIBE UE, tailless final forms, page 6-7
- U+189E MANCHU ALI GALI TTA, final form, page 8 (Note this matches the proposal of L2/17-115R, *Glyph corrections for Ali Gali final forms in the Mongolian block.*)
- U+18A6 ALI GALI HALF U, final form, page 10
- U+18A7 ALI GALI HALF YA, final form, page 10

3.4 Investigation is needed

Investigation is needed for determining actions for the following requests:

- U+1864 SIBE GA, initial form with alternative positioning of the dot, page 7
- U+1866 SIBE PA, initial and medial forms with an opening, page 7-8
- U+1887 ALI GALI A, removing a stylistic variant, page 27
- U+189A MANCHU ALI GALI GHA, initial form with alternative positioning of the dot, page 8
- U+18A0 MANCHU ALI GALI TA, initial and medial forms with protruding stroke, page 9
- U+18A1 MANCHU ALI GALI DHA, initial and medial forms with protruding stroke, page 9
- U+18A4 MANCHU ALI GALI ZHA, alternative initial and medial forms, page 9-10

Particularly careful investigation is needed for the following requests, as they would further increase confusability because of the controversial analysis of diphthongs:

- U+1838 WA, medial form for a diphthong offglide role (confusable with U+1824 U and U+1826 UE, etc.), page 5
- U+1855 TODO YA, final form (confusable with U+1845 TODO I), page 6