

Resolving Cross-Writing-System Ambiguity in Mongolian Script

Kushim Jiang
<kushim_jiang@outlook.com>

2026-07-26

Currently, the Mongolian fonts used on major platforms (*Noto Sans Mongolian*, *Mongolian Baiti*) all cover the entire Mongolian block. As noted in the *Core Specification*, the Mongolian block is a multi-writing-system environment including Hudum, Todo, Sibe, Manchu and so on. Because multiple different writing systems share the same characters, this gives rise to writing-system-related issues, which fall primarily into the following three categories.

Different letterforms. A typical example is the letter *m* in Sibe and Manchu, which was mentioned in L2/17-009. Although the proposal itself was not adopted, the facts it describes are indeed correct. This proposal focuses on its final form, and we have:

U+182E	isolated	initial	medial	final
Hudum (MNG)				
Sibe (SIB)				
Manchu (MCH)				

As the proposal states, the final *m* in Hudum is designed with a small tail, while the final *m* in Sibe and Manchu is designed with a large tail. There are many other similar design differences, but this is one example of the shared character.

Our current implementation handles this via localized shaping per writing system; we now have four relevant OTL language system tags: MNG for Traditional Mongolian (Hudum), TOD for Todo, SIB for Sibe, and MCH for Manchu (RBU for Buryat is not considered for now). In this example, shaping rules declared under the SIB and MCH language system tags can replace the final *m* in the small tail with the final *m* in the large tail.

Different FVS allocations. The unification occurs not only when all variants of a character are identical in usage, but also when the variant sets of characters overlap. For example, Hudum, Sibe, and Manchu all use U+182A MONGOLIAN LETTER BA; the difference is that Hudum uses the final form with FVS1, while Sibe and Manchu do not. According to the China national standard test set, for Sibe and Manchu, <U+182A, FVS1> is an invalid string, and FVS1 should be displayed as an error form. For Hudum, however, <U+182A, FVS1> is a valid string and should be displayed in its correct actual form.

U+182E	isolated	initial	medial	final	final FVS1
Hudum (MNG)					
Sibe (SIB)					
Manchu (MCH)					

Our current implementation also handles this with localized shaping per writing system. In the example above, shaping rules declared under the SIB and MCH language system tags replace the FVS1 final form with the default final form and an FVS1 error form. Since China has incorporated Ali Gali into its corresponding national standard, the validity of strings is determined based on national standards. In other words, a string that is valid for Hudum Ali Gali is also valid for Hudum, because they are declared in the same national standard.

Different default forms. For example, U+188A MONGOLIAN LETTER ALI GALI NGA is used for both Hudum Ali Gali and Todo Ali Gali, where Todo Ali Gali uses the default initial form ᠨᠠ and default medial form ᠨᠡ, while Hudum Ali Gali also uses its default final form ᠨᠢ, as mentioned in [L2/17-115R](#) #5. For final *-ng*, Todo Ali Gali will use the final form of U+184A MONGOLIAN LETTER TODO ANG, which is ᠨᠭ. In this situation, it might be more natural for users to see the following result:

U+188A	isolated	initial	medial	final
Hudum Ali Gali (MNG)	ᠨᠠ	ᠨᠠ	ᠨᠡ	ᠨᠢ
Todo Ali Gali (TOD)	ᠨᠠ	ᠨᠠ	ᠨᠡ	ᠨᠡ

In other words, when U+188A appears at the end of a word, users of Hudum Ali Gali should see the actual final form ᠨᠢ, while users of Todo Ali Gali should see the virtual form (or the fabricated variant) derived from its medial form ᠨᠡ.

For the example above, we could also solve the problem using localized shaping per writing system. However, we encounter the following issue: we cannot distinguish between Manchu and Manchu Ali Gali, because we do not have language tags for Hudum Ali Gali, Manchu Ali Gali, and Todo Ali Gali (considering that they are all used to transcribe Sanskrit and Tibetan, perhaps we should use SAN and TIB, but that would not resolve the conflicts between these Ali Gali, so it is not feasible). For example, something similar happens with U+186C MONGOLIAN LETTER SIBE GAA, where we have:

U+188A	isolated	initial	medial	final
Sibe (SIB)	ᠨᠠ	ᠨᠠ	ᠨᠡ	ᠨᠢ
Manchu (MCH)	ᠨᠠ	ᠨᠠ	ᠨᠡ	ᠨᠢ
Manchu Ali Gali (MCH)	ᠨᠠ	ᠨᠠ	ᠨᠡ	ᠨᠢ

This situation is also mentioned in [L2/17-115R](#), #2. In other words, for users of Sibe and Manchu, the actual form ᠨᠢ found in Manchu Ali Gali seems puzzling; they should perceive it as a virtual form which reuses the medial form ᠨᠡ. Here are a few more similar examples:

U+1807	isolated	initial	medial	final
Sibe (SIB)	ᠨᠠ	ᠨᠠ	ᠨᠡ	ᠨᠢ
Manchu (MCH)	ᠨᠠ	ᠨᠠ	ᠨᠡ	ᠨᠢ
Manchu Ali Gali (MCH)	ᠨᠠ	ᠨᠠ	ᠨᠡ	ᠨᠢ

U+1877	isolated	initial	medial	final
Manchu (MCH)	ᠨᠠ	ᠨᠠ	ᠨᠡ	ᠨᠢ
Manchu Ali Gali (MCH)	ᠨᠠ	ᠨᠠ	ᠨᠡ	ᠨᠢ

To this end, Option 1 is to establish separate language tags for Hudum Ali Gali, Manchu Ali Gali, and Todo Ali Gali, which would require an adjustment to the BCP 47 tags. Option 2 is to allow the actual cursive variant of another writing system to be exposed, given that there is no consensus among users regarding undefined cursive variants. Option 3 is to include the actual form used by Ali Gali in a feature that is disabled by default; however, we also note that this fragile API contract is difficult to use.

Please consider the guidance regarding how we should properly address the three issues mentioned above in a font that covers all writing systems within the entire Mongolian block.

(End of Document)