

# Consonant-Vowel Ligatures for Hudum, Todo, Sibe, Manchu, Ali Gali Scripts

Doc Type: Working Group Document

Title: Consonant-Vowel Ligatures for Hudum, Todo, Sibe, Manchu, Ali Gali Scripts

Author: Greg Eck (greyson@postone.net)

Date: March 23, 2018

## Justification

There has never been an exhaustive set of test data to use to determine whether a Mongolian font is correctly shaping syllable patterns. This paper attempts to provide a set of character sequences showing all possible Consonant-Vowel (CV) combinations or syllables – grammatical or ungrammatical – the attempt is to show how a given font is shaping a given CV sequence.

## Requirements for fonts using this data

The control characters used in this set of test data are the ZWJ (U+200D) and the FVS set (U+180B, 180C, 180D). The MVS (U+180E) and the NNBS (U+202F) are not used in the test data as they are not within the scope of this paper.

## Layout method of the charts

Each script is divided into two parts – the native section first and the transliteration section second. Each of these two sections is further divided into four sections dealing with the positional make-up of the CV sequence. These are the initial\_consonant plus medial\_vowel, initial\_consonant plus final\_vowel, medial\_consonant plus medial\_vowel and medial\_consonant plus final\_vowel.

## What this paper does not concern itself with

The main intent of this paper has been to provide test data for font developers. But, there are several things that might be confusing to the reader. The breadth of current Mongolian fonts is wide encompassing 5 scripts. The breadth of research dealing with the five scripts as a whole is low as compared to older scripts such as Devanagari or Hebrew. Therefore there are potential code-points which may be seen in the future to be duplicates. There are other issues such as the ordering of the code-points following traditional orthographical order rather than a phonetic grouping.

In an attempt to provide a more intuitive and quicker analysis of a given font's shaping behavior, the author has placed the code-point sequences in a manner which starts with the traditional orthographical order, but then is modified by moving some of the code-points around to group more naturally with code-points with similar phonetic/sound qualities. Sometimes the code-point name does not match the Uniscribe-ascribed name to the code-point. This is not intended to be a suggested name change. It happened as a result of trying to understand how the CV syllable structure works within each script, how diacritics are applied and where there might be overlap between the scripts. Apologies are extended in this light.

## Media used

There are necessarily two formats for this paper in two different files. The presentation format of this paper will be (DS02.pdf). The actual textual file (DS02.docx or DS02.odt) will be just as valuable however as the intent will not be to illustrate the correct shaping behavior of a given CV pattern. The textual file will provide the data stream for the font developer to compare to the expected shaping behavior as seen in the presentation .pdf file.





## Questions / Observations on the Hudum Section

- 1.) Is it correct to say that Ali Gali will never use 1824/1825? There is only one velar (1826) and only one uvular (1823)? Therefore the sequences suggesting use of 1824/1825 do not need to ligature as that is only an extra task to handle the case of incorrect spelling?
- 2.) Why does 1897 not ligature? (same question in Todo)
- 3.) Is the 1833-D initial sequence correct? Compare Sibe 1868/1869.
- 4.) Does the Hudum script actually use the AG 1887 final?
- 5.) Grey highlighting indicates that the particular glyph needs further work. It may be a candidate for ligaturing. Or it may be that the diacritics need refinement. Or it may be that the C+FVS / V+FVS sequence is not used in running text – it is only used for the display of individual glyphs for purposes of instruction/presentation.





## Questions / Observations on the Todo Section

- 1.) Does the Todo script actually use the AG 1887 final?
- 2.) Yellow highlighting is used to denote syllables which demonstrate ambiguity. Consideration might be given as to how to disambiguate such sequences.



## Questions / Observations on the Sibe Section

- 1.) There is no transliteration section for the Sibe script.

Init-Medi

N<sub>(1828)</sub> ᠨ ᠨ ᠨ ᠨ ᠨ ᠨ ᠨ  
 Ng<sub>(1829)</sub> ᠨᠭ ᠨᠭ ᠨᠭ ᠨᠭ ᠨᠭ ᠨᠭ  
 B<sub>(182A)</sub> ᠪ ᠪ ᠪ ᠪ ᠪ ᠪ  
 P<sub>(1866)</sub> ᠪ ᠪ ᠪ ᠪ ᠪ ᠪ  
 G<sub>(1864+I)</sub> ᠨ ᠨ ᠨ ᠨ ᠨ ᠨ  
 H<sub>(1865+I)</sub> ᠨ ᠨ ᠨ ᠨ ᠨ ᠨ  
 K<sub>(1874+2I)</sub> ᠨ ᠨ ᠨ ᠨ ᠨ ᠨ  
 Gh<sub>(186C)</sub> ᠨ ᠨ ᠨ ᠨ ᠨ ᠨ  
 Hh<sub>(186D)</sub> ᠨ ᠨ ᠨ ᠨ ᠨ ᠨ  
 Kh<sub>(183A)</sub> ᠨ ᠨ ᠨ ᠨ ᠨ ᠨ  
 M<sub>(182E)</sub> ᠮ ᠮ ᠮ ᠮ ᠮ ᠮ  
 L<sub>(182F)</sub> ᠯ ᠯ ᠯ ᠯ ᠯ ᠯ  
 S<sub>(1830)</sub> ᠰ ᠰ ᠰ ᠰ ᠰ ᠰ  
 Sh<sub>(1867)</sub> ᠰ ᠰ ᠰ ᠰ ᠰ ᠰ  
 T<sub>(1868+I)</sub> ᠲ ᠲ ᠲ ᠲ ᠲ ᠲ  
 D<sub>(1869+I)</sub> ᠲ ᠲ ᠲ ᠲ ᠲ ᠲ  
 C<sub>(1834)</sub> ᠴ ᠴ ᠴ ᠴ ᠴ ᠴ  
 J<sub>(1835)</sub> ᠵ ᠵ ᠵ ᠵ ᠵ ᠵ  
 Y<sub>(1836)</sub> ᠶ ᠶ ᠶ ᠶ ᠶ ᠶ  
 R<sub>(1875)</sub> ᠷ ᠷ ᠷ ᠷ ᠷ ᠷ  
 W<sub>(1838)</sub> ᠸ ᠸ ᠸ ᠸ ᠸ ᠸ

Init-Fina

N<sub>(1828)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 Ng<sub>(1829)</sub> ᠨᠭᠢ ᠨᠭᠢ ᠨᠭᠢ ᠨᠭᠢ ᠨᠭᠢ ᠨᠭᠢ  
 B<sub>(182A)</sub> ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ  
 P<sub>(1866)</sub> ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ  
 G<sub>(1864+I)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 H<sub>(1865+I)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 K<sub>(1874+2I)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 Gh<sub>(186C)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 Hh<sub>(186D)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 Kh<sub>(183A)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 M<sub>(182E)</sub> ᠮᠢ ᠮᠢ ᠮᠢ ᠮᠢ ᠮᠢ ᠮᠢ  
 L<sub>(182F)</sub> ᠯᠢ ᠯᠢ ᠯᠢ ᠯᠢ ᠯᠢ ᠯᠢ  
 S<sub>(1830)</sub> ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ  
 Sh<sub>(1867)</sub> ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ  
 T<sub>(1868+I)</sub> ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ  
 D<sub>(1869+I)</sub> ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ  
 C<sub>(1834)</sub> ᠴᠢ ᠴᠢ ᠴᠢ ᠴᠢ ᠴᠢ ᠴᠢ  
 J<sub>(1835)</sub> ᠵᠢ ᠵᠢ ᠵᠢ ᠵᠢ ᠵᠢ ᠵᠢ  
 Y<sub>(1836)</sub> ᠶᠢ ᠶᠢ ᠶᠢ ᠶᠢ ᠶᠢ ᠶᠢ  
 R<sub>(1875)</sub> ᠷᠢ ᠷᠢ ᠷᠢ ᠷᠢ ᠷᠢ ᠷᠢ  
 W<sub>(1838)</sub> ᠸᠢ ᠸᠢ ᠸᠢ ᠸᠢ ᠸᠢ ᠸᠢ

Medi-Medi

N<sub>(1828+M)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 Ng<sub>(1829)</sub> ᠨᠭᠢ ᠨᠭᠢ ᠨᠭᠢ ᠨᠭᠢ ᠨᠭᠢ ᠨᠭᠢ  
 B<sub>(182A)</sub> ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ  
 P<sub>(1866)</sub> ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ  
 G<sub>(1864+M)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 H<sub>(1865+M)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 K<sub>(1874+3M)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 Gh<sub>(186C)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 Hh<sub>(186D)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 Kh<sub>(183A)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 M<sub>(182E)</sub> ᠮᠢ ᠮᠢ ᠮᠢ ᠮᠢ ᠮᠢ ᠮᠢ  
 L<sub>(182F)</sub> ᠯᠢ ᠯᠢ ᠯᠢ ᠯᠢ ᠯᠢ ᠯᠢ  
 S<sub>(1830)</sub> ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ  
 Sh<sub>(1867)</sub> ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ  
 T<sub>(1868+2M)</sub> ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ  
 D<sub>(1869+M)</sub> ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ  
 C<sub>(1834)</sub> ᠴᠢ ᠴᠢ ᠴᠢ ᠴᠢ ᠴᠢ ᠴᠢ  
 J<sub>(1835)</sub> ᠵᠢ ᠵᠢ ᠵᠢ ᠵᠢ ᠵᠢ ᠵᠢ  
 Y<sub>(1836)</sub> ᠶᠢ ᠶᠢ ᠶᠢ ᠶᠢ ᠶᠢ ᠶᠢ  
 R<sub>(1875)</sub> ᠷᠢ ᠷᠢ ᠷᠢ ᠷᠢ ᠷᠢ ᠷᠢ  
 W<sub>(1838)</sub> ᠸᠢ ᠸᠢ ᠸᠢ ᠸᠢ ᠸᠢ ᠸᠢ

Medi-Fina

N<sub>(1828+M)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 Ng<sub>(1829)</sub> ᠨᠭᠢ ᠨᠭᠢ ᠨᠭᠢ ᠨᠭᠢ ᠨᠭᠢ ᠨᠭᠢ  
 B<sub>(182A)</sub> ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ  
 P<sub>(1866)</sub> ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ ᠪᠢ  
 G<sub>(1864+M)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 H<sub>(1865+M)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 K<sub>(1874+3M)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 Gh<sub>(186C)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 Hh<sub>(186D)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 Kh<sub>(183A)</sub> ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ ᠨᠢ  
 M<sub>(182E)</sub> ᠮᠢ ᠮᠢ ᠮᠢ ᠮᠢ ᠮᠢ ᠮᠢ  
 L<sub>(182F)</sub> ᠯᠢ ᠯᠢ ᠯᠢ ᠯᠢ ᠯᠢ ᠯᠢ  
 S<sub>(1830)</sub> ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ  
 Sh<sub>(1867)</sub> ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ ᠰᠢ  
 T<sub>(1868+2M)</sub> ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ  
 D<sub>(1869+M)</sub> ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ ᠲᠢ  
 C<sub>(1834)</sub> ᠴᠢ ᠴᠢ ᠴᠢ ᠴᠢ ᠴᠢ ᠴᠢ  
 J<sub>(1835)</sub> ᠵᠢ ᠵᠢ ᠵᠢ ᠵᠢ ᠵᠢ ᠵᠢ  
 Y<sub>(1836)</sub> ᠶᠢ ᠶᠢ ᠶᠢ ᠶᠢ ᠶᠢ ᠶᠢ  
 R<sub>(1875)</sub> ᠷᠢ ᠷᠢ ᠷᠢ ᠷᠢ ᠷᠢ ᠷᠢ  
 W<sub>(1838)</sub> ᠸᠢ ᠸᠢ ᠸᠢ ᠸᠢ ᠸᠢ ᠸᠢ

