MWG/2-N4R2

L2/18-102

An improved graphetic model for the Mongolian encoding
改进的蒙古文编码字形模型

Liang Hai (梁海)
lianghai@gmail.com

3 April 2018
2018 年 4 月 3 日

1 Introduction

1.1 Related earlier documents

- L2/17-328: *Script Ad Hoc Group Recommendations on Mongolian Text Model*, 9 August 2017.
- L2/17-335: *A graphetic approach for the Mongolian encoding model (draft)*, 31 August 2017.
- L2/17-334: *On migration issues of the graphetic model*, 18 September 2017.
- L2/17-347: *Mongolian Ad Hoc Report (Hohhot, Inner Mongolia)*, 29 September 2017.

1.2 Acknowledgements

The author received feedback from Shen Yilei (沈逸磊) and the Script Ad Hoc group. The prototype Mongolian font is based on Bolorsoft's open-source font MongolianScript.

1.3 Scope

This document discusses[1] a crucial subset of the Mongolian script's usage, the so-called *Hudum* (᠋᠋᠋᠋᠋ or худам) writing system of the modern Mongolian language. Various other writing systems and historical stages (Hudum Ali Gali, historical Hudum, Todo, Todo Ali Gali, Manchu, Manchu Ali Gali, historical Manchu, Sibe, etc) are excluded.

[1]For a precise comparison with the current encoding in the Unicode Standard, note this document covers the equivalence of letters U+1820..U+1842 and format control characters U+180A..U+180E, in terms of representation capability.

1.4 Mongolian phonetic letters

A text encoding model generally encodes letters as characters, taking advantage of the established letter concepts that generally have a predictable (although sometimes complex) mapping to written forms.

However, users of the Hudum writing system conventionally identify letters on a more phonetic level, where the relationship between letters and written forms is not only complex, but also largely ambiguous and often unpredictable.

See *Table 1* for one of the ways to list these *phonetic letters*. Loanword letters are in brackets. Note *ŋ* is often not listed in the alphabet because it's not a valid syllable onset in the typical (C)V(C) pattern[2] of orthographical syllables, while *ë* is often identified as a variant of *e* instead of an independent letter. Also, *ž* and *č* are usually not listed independently, but are bound in *ži* and *či*, two of the special syllables for transcribing Chinese.

Table 1: Phonetic letters and transliterations

Vowels:	*a* *e* *i* *o* *u* *ö* *ü* [*ë*]
Consonants:	*n* (*ŋ*) *b* *p* *x* *g* *m* *l* *s* *ś* *t* *d* *ć* *j* *y* *r* *w* [*f*] [*k*] [*c*] [*z*] [*h*] [*ř*] [*ł*] [*ž*] [*č*]

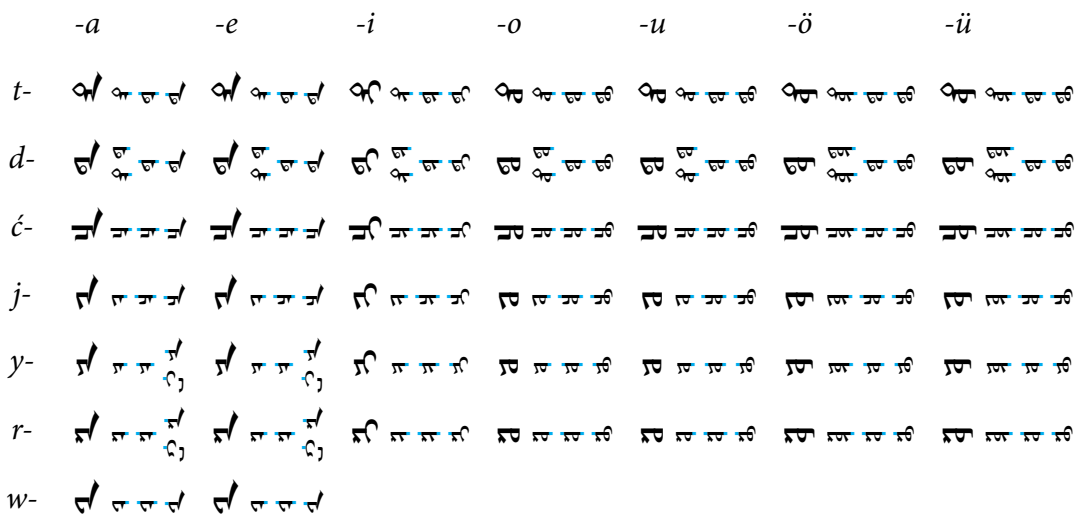
Note the concept of letters is a late invention for Hudum. When learning and using Hudum, users still largely rely on a syllable-based traditional analysis, *the Twelve Syllabaries*. From the point of view of phonetic letters, the Twelve Syllabaries exhibit some typical written forms (but not all written forms) of enumerated (C)V(C) combinations of phonetic letters, when these combinations are written as a whole word or at the beginning, middle, or end of a word. See *Table 2* for one of the ways to arrange *the First Syllabary* (the (C)V syllables) of the twelve.

Table 2: The First Syllabary

	-a	*-e*	*-i*	*-o*	*-u*	*-ö*	*-ü*
ø-							
n-							
b-							
p-							
x-							
g-							
m-							
l-							
s-							
ś-							

[2] An optional onset consonant, a required vowel, and an optional coda consonant.

Table 2: (continued)



From the First Syllabary it's already observable that encoding the phonetic letters as characters (the *phonetic* approach) leads to has serious inherent defects:

1. A large number of combinations are confusable in written forms, and such ambiguity means low reliability. For example, text inputting and editing are highly prone to visually invisible errors. Notable confusable groups include *a/e*, *o/u/ö/ü* (in particular, *o* is always written the same as *u*, and *ö* the same as *ü*), *x/g*, *t/d*, etc.

2. Even though complex contextual rules can decide most written forms, additional information is required when an expected form is grammatical or lexical thus is not predictable with only a context of phonetic letters. This uncertainty is underrepresented in the Twelve Syllabaries (where the uncertainty is only observable in *a/e* and *d*), but is common in actual words. Thus a phonetic encoding requires inserting arbitrary format control characters to select writing forms, leading to low usability for users.

1.5 Graphetic approach

In light of the unsuitability of the phonetic approach, a group of experts have been exploring a *graphetic*[3] approach that encodes cursive joining graphemes instead of phonetic letters, still utilizing the Unicode–OpenType cursive joining model, but avoiding the aforementioned ambiguity and uncertainty inherent in the phonetic approach. After analyzing written forms of phonetic letters according to their alternation between different cursive joining positions, revealed cursive joining graphemes are further decomposed and unified to avoid highly confusable characters. Such a relatively simple encoding model provides solid support for higher-level text processes.

Taking the graphetic approach, a specific encoding model has been designed for discussing and testing, and is introduced in this document.

[3]The rather special term *graphetic* is used to represent a concept opposed to *phonetic*. It means "relating to graphemes or graphs". Just like *phonetic* is often preferred in text encoding discussions as it's a broader concept than *phonemic* (strictly relating to phonemes), the broader wording *graphetic* is here chosen over *graphemic* (strictly relating to graphemes).

2 Character repertoire

The model requires a repertoire of 30 characters. See *Table 3* for a provisional list of representative glyphs and character names. Note representative glyphs are not relevant to actual shaping. Slashes (/) instead of the word "OR" are used in provisional names for better readability. Transliterations of phonetically definite characters are simply borrowed from their corresponding phonetic letters (note these transliterations are all in lowercase, and colored in blue to distinguish from phonetic transliterations), while phonetically ambiguous characters have distinct transliterations in uppercase letters or other symbols.

Table 3: Character repertoire

No.	Repr.	Name	Translit.
1.	᠊	MONGOLIAN CHARACTER NIRUGU	-
2.	ᠠ	MONGOLIAN CHARACTER ALEPH/A/E/NA	·
3.	ᠡ	MONGOLIAN CHARACTER A/E	*A*
4.	ᠢ	MONGOLIAN CHARACTER I/JA/YA	*I*
5.	ᠣ	MONGOLIAN CHARACTER O/U/OE/UE	*U*
6.	ᠤ	MONGOLIAN CHARACTER O/U/OE/UE/WA	*Ú*
7.	ᠥ	MONGOLIAN CHARACTER OE/UE	*Ü*
8.	ᠨ	MONGOLIAN CHARACTER NA	*n*
9.	ᠪ	MONGOLIAN CHARACTER BA	*b*
10.	ᠫ	MONGOLIAN CHARACTER PA	*p*
11.	ᠬ	MONGOLIAN CHARACTER XA/GA	*X*
12.	ᠭ	MONGOLIAN CHARACTER XE/GE	*G*
13.	ᠬ᠋	MONGOLIAN CHARACTER GA	*g*
14.	ᠮ	MONGOLIAN CHARACTER MA	*m*
15.	ᠯ	MONGOLIAN CHARACTER LA	*l*
16.	ᠰ	MONGOLIAN CHARACTER SA	*s*
17.	ᠱ	MONGOLIAN CHARACTER SHA	*ś*
18.	ᠲ	MONGOLIAN CHARACTER TA/DA	*T*
19.	ᠳ	MONGOLIAN CHARACTER DA/TA	*D*
20.	ᠴ	MONGOLIAN CHARACTER CHA	*ć*
21.	ᠵ	MONGOLIAN CHARACTER JA	*j*
22.	ᠶ	MONGOLIAN CHARACTER YA	*y*
23.	ᠷ	MONGOLIAN CHARACTER RA	*r*
24.	ᠸ	MONGOLIAN CHARACTER WA/EE	*W*
25.	ᠹ	MONGOLIAN CHARACTER FA	*f*

Table 3: (continued)

No.	Repr.	Name	Translit.
26.	ᠺ	MONGOLIAN CHARACTER KA	*k*
27.	ᠴ	MONGOLIAN CHARACTER CA	*c*
28.	ᠵ	MONGOLIAN CHARACTER ZA	*z*
29.	ᠹ	MONGOLIAN CHARACTER HA/ZHI	*H*
30.	ᠶ	MONGOLIAN CHARACTER RHA	*ř*

3 Text representation

The model requires two stages of contextual shaping to select the orthographically correct form for every character: first the *cursive joining* stage (*Section 3.1*), then the *round consonant* stage (*Section 3.2*).

Note certain written forms of phonetic letters are encoded as character sequences:

1. The *aleph*-ed (prepended with an extra cap or tooth) forms of vowel letters (*a*: ᠠ᠋, ᠡ᠋, ᠤ᠋; *i*: ᠢ᠋, ᠣ᠋, ᠥ᠋; etc) have their aleph encoded as a character: ᠠ ALEPH/A/E/NA.
2. The *long tooth*-ed forms of *ö/ü* (ᠣᠷ᠋, ᠥᠷ᠋, ᠤᠷ᠋) have their long tooth encoded as a character: ᠷ I/JA/YA.
3. *ŋ* (ᠩ᠋, ᠨᠭ᠋) is encoded as a sequence: < ᠠ ALEPH/A/E/NA, ᠺ XE/GE >.
4. The syllable coda form of *d* (ᠳ᠋, ᠳᠠ᠋) is encoded as a sequence: < ᠤ O/U/OE/UE, ᠠ ALEPH/A/E/NA >.
5. The word-initial form of *h* (ᠬᠠ᠋) is encoded as a sequence: < ᠠ ALEPH/A/E/NA, ᠹ HA/ZHI >.
6. *ɬ* (ᠯᠠ᠋, ᠯᠡ᠋) is encoded as a sequence: < ᠯ LA, ᠹ HA/ZHI >.
7. *č* (ᠴᠴ᠋) is encoded as a sequence: < ᠤ O/U/OE/UE, ᠤ O/U/OE/UE >.

3.1 Cursive joining stage

According to joining types of adjacent characters, the appropriate positional form is selected for every character. Note cursive joining positional forms are irrelevant to the grammatical definition of "word".

See *Table 4*, *Table 5*, and *Table 6* for positional forms of characters of the three joining types.[4] Joined ends of positional forms are emphasized with an extra stroke in blue (᠆). Positions with parentheses are unattested. A couple of recommended fallbacks are in the parentheses, but futher discussion is necessary for the unattested isolate positions, which are the most problematic and are all left empty for now.

[4]The standard terminology and notation in the Unicode Standard for joining types and positional forms are avoided here, because they are confusing for vertically written scripts. For the background, see "Cursive Joining" in *Section 14.4, Phags-pa, The Unicode Standard, Version 10.0*.

Table 4: Dual-joining characters

Character	Translit.	Isolate	Initial	Medial	Final
MONGOLIAN CHARACTER NIRUGU	-	᠆	᠆	᠆	᠆
MONGOLIAN CHARACTER ALEPH/A/E/NA	·	ᠠ	ᠠ	᠊	᠉
MONGOLIAN CHARACTER I/JA/YA	*I*	ᠢ	ᠢ	᠊	᠌
MONGOLIAN CHARACTER O/U/OE/UE	*U*	( )	ᠤ	᠊	᠋
MONGOLIAN CHARACTER NA	*n*	( )	ᠨ	᠊	᠈
MONGOLIAN CHARACTER BA	*b*	( )	ᠪ	᠊	᠋
MONGOLIAN CHARACTER PA	*p*	( )	ᠫ	᠊	᠋
MONGOLIAN CHARACTER XA/GA	*x*	( )	ᠬ	᠊	᠍
MONGOLIAN CHARACTER XE/GE	*G*	( )	ᠭ	᠊	᠎
MONGOLIAN CHARACTER GA	*g*	( )	ᠭ᠋	᠊	᠈
MONGOLIAN CHARACTER MA	*m*	( )	ᠮ	᠊	᠊
MONGOLIAN CHARACTER LA	*l*	( )	ᠯ	᠊	᠊
MONGOLIAN CHARACTER SA	*s*	( )	ᠰ	᠊	᠊
MONGOLIAN CHARACTER SHA	*ś*	( )	ᠱ	᠊	᠊
MONGOLIAN CHARACTER TA/DA	*T*	( )	ᠲ	᠊	᠋
MONGOLIAN CHARACTER DA/TA	*D*	( )	ᠳ	᠊	᠋
MONGOLIAN CHARACTER CHA	*ć*	( )	ᠴ	᠊	᠋
MONGOLIAN CHARACTER JA	*j*	( )	(ᠵ)	᠊	᠋
MONGOLIAN CHARACTER YA	*y*	( )	ᠶ	᠊	(ᠶ)
MONGOLIAN CHARACTER RA	*r*	( )	ᠷ	᠊	᠌
MONGOLIAN CHARACTER WA/EE	*w*	( )	ᠸ	᠊	᠋
MONGOLIAN CHARACTER FA	*f*	( )	ᠹ	᠊	᠋
MONGOLIAN CHARACTER KA	*k*	( )	ᠺ	᠊	᠏
MONGOLIAN CHARACTER CA	*c*	( )	ᠼ	᠊	᠋
MONGOLIAN CHARACTER ZA	*z*	( )	ᠽ	᠊	᠋
MONGOLIAN CHARACTER HA/ZHI	*H*	( )	ᠾ	᠊	᠋
MONGOLIAN CHARACTER RHA	*ř*	( )	ᠿ	᠊	(ᠿ)

Table 5: Top-joining characters

Character	Translit.	Isolate	Final
MONGOLIAN CHARACTER O/U/OE/UE/WA	*Ú*	ᠦ	ᡂ
MONGOLIAN CHARACTER OE/UE	*Ü*	( )	ᡃ

Table 6: Non-joining characters

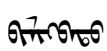
Character	Translit.	Isolate
MONGOLIAN CHARACTER A/E	*A*	᠈

See *Table 7* for examples of text representation that only depend on this stage of contextual shaping. Both phonetic (including notable alternatives) and graphetic transliterations are provided.

Table 7: Examples

Phonetic	Output	Graphetic	Character sequence
a/en	᠕	··	< ᠠ ALEPH/A/E/NA , ᠠ ALEPH/A/E/NA >
e/a	᠔	·	< ᠠ ALEPH/A/E/NA >
a/e	᠈	*A*	< ᠈ A/E >
i/ei	᠖	·*I*	< ᠠ ALEPH/A/E/NA , ᠢ I/JA/YA >
i	᠌	*I*	< ᠢ I/JA/YA >
o/u	ᠣ	·*U*	< ᠠ ALEPH/A/E/NA , ᠤ O/U/OE/UE >
ö/ü	ᠥ	·*Ü*	< ᠠ ALEPH/A/E/NA , ᠦ OE/UE >
ü/eü	ᠧ	·*Ú*	< ᠠ ALEPH/A/E/NA , ᡂ O/U/OE/UE/WA >
u/ü	ᡂ	*Ú*	< ᡂ O/U/OE/UE/WA >
ordo/urtu	ᠣᠷᠳᠤ	·*UrDÚ*	< ᠠ ALEPH/A/E/NA , ᠤ O/U/OE/UE , ᠷ RA , ᠳ DA/TA , ᡂ O/U/OE/UE/WA >
ada/ende	ᠠᠨᠳᠠ	··*D*·	< ᠠ ALEPH/A/E/NA , ᠠ ALEPH/A/E/NA , ᠳ DA/TA , ᠠ ALEPH/A/E/NA >
xana	ᠬᠠᠨᠠ	*X*·*n*·	< ᠬ XA/GA , ᠠ ALEPH/A/E/NA , ᠨ NA , ᠠ ALEPH/A/E/NA >
xana	ᠬᠠᠨ᠈	*X*·*nA*	< ᠬ XA/GA , ᠠ ALEPH/A/E/NA , ᠨ NA , ᠈ A/E >
ćerig	ᠴᠡᠷᠢᠭ	*ć*·*rIG*	< ᠴ CHA , ᠠ ALEPH/A/E/NA , ᠷ RA , ᠢ I/JA/YA , ᠭ XE/GE >
jarlig	ᠵᠠᠷᠯᠢᠬ	*I*·*rlIX*	< ᠵ I/JA/YA , ᠠ ALEPH/A/E/NA , ᠷ RA , ᠯ LA , ᠢ I/JA/YA , ᠬ XA/GA >
altan odo	ᠠᠯᠲᠠᠨ ᠣᠳᠤ	··*lD*··␣·*UDÚ*	< ᠠ ALEPH/A/E/NA , ᠠ ALEPH/A/E/NA , ᠯ LA , ᠲ DA/TA , ᠠ ALEPH/A/E/NA , ᠠ ALEPH/A/E/NA , ␣ SPACE , ᠠ ALEPH/A/E/NA , ᠤ O/U/OE/UE , ᠳ DA/TA , ᡂ O/U/OE/UE/WA >
altanodo	ᠠᠯᠲᠠᠨᠣᠳᠤ	··*lD*···*UDÚ*	< ᠠ ALEPH/A/E/NA , ᠠ ALEPH/A/E/NA , ᠯ LA , ᠲ DA/TA , ᠠ ALEPH/A/E/NA , ᠠ ALEPH/A/E/NA , ᠠ ALEPH/A/E/NA , ᠤ O/U/OE/UE , ᠳ DA/TA , ᡂ O/U/OE/UE/WA >
bilig batu	ᠪᠢᠯᠢᠭ ᠪᠠᠲᠤ	*b*·*lIG*␣*b*·*DÚ*	< ᠪ BA , ᠢ I/JA/YA , ᠯ LA , ᠢ I/JA/YA , ᠭ XE/GE , ␣ SPACE , ᠪ BA , ᠠ ALEPH/A/E/NA , ᠲ DA/TA , ᡂ O/U/OE/UE/WA >

Table 7: (continued)

Phonetic	Output	Graphetic	Character sequence
biligbatu		*bIlIGb·DÚ*	< BA , I/JA/YA , LA , I/JA/YA , XE/GE , BA , ALEPH/A/E/NA , DA/TA , O/U/OE/UE/WA >
ed		*·D*	< ALEPH/A/E/NA , DA/TA >
gxir/xxir		*GGIr*	< XE/GE , XE/GE , I/JA/YA , RA >
radio/radiu		*r·DIU*	< RA , ALEPH/A/E/NA , DA/TA , I/JA/YA , O/U/OE/UE >
tŋri/tegri		*T·GrI*	< TA/DA , ALEPH/A/E/NA , XE/GE , RA , I/JA/YA >
naima		*n·Im·*	< NA , ALEPH/A/E/NA , I/JA/YA , MA , ALEPH/A/E/NA >
sayixan		*s·yIX··*	< SA , ALEPH/A/E/NA , YA , I/JA/YA , XA/GA , ALEPH/A/E/NA , ALEPH/A/E/NA >
sain/sayin		*s·II·*	< SA , ALEPH/A/E/NA , I/JA/YA , I/JA/YA , ALEPH/A/E/NA >
uu/üü		*UÚ*	< O/U/OE/UE , O/U/OE/UE/WA >

For controlling cursive joining, although the general characters U+200C ZERO WIDTH NON-JOINER and U+200D ZERO WIDTH JOINER can be used, in day-to-day usage it is recommended to produce joined forms in isolation using ▪ MONGOLIAN CHARACTER NIRUGU, which as a visible character is easier for users to manipulate. See *Table 8* for examples.

Table 8: Examples

Phonetic	Output	Graphetic	Character sequence
ga		*g·-*	< GA , ALEPH/A/E/NA , ▪ NIRUGU >
ga		*-g·-*	< ▪ NIRUGU , GA , ALEPH/A/E/NA , ▪ NIRUGU >
ga		*-gA*	< ▪ NIRUGU , GA , A/E >
ba		*b·-*	< BA , ALEPH/A/E/NA , ▪ NIRUGU >
na		*-nA*	< ▪ NIRUGU , NA , A/E >

3.2 Round consonant stage

Certain combinations of positional forms that involve the five so-called round consonants (BA, PA, XE/GE, FA, KA) require special treatment. The rules for selecting the required variants forms are formally defined below. See *Table 9* for explanation of the notation.

Rule 1 Cr + Va → Cr + Va′

Rule 2 Cr + Vu → Cr′ + Vu

Table 9: Notation

Notation	Positional forms	Note
Cr	𐴀 𐴁 𐴂 𐴃 𐴄 𐴅 𐴆 𐴇 𐴈	Initial and medial forms of round consonants: BA, PA, XE/GE, FA, KA.
Cr′	𐴉 𐴊 𐴋 𐴌 𐴍 𐴎 𐴏 𐴐 𐴑 𐴒	Corresponding variant forms of Cr.
Va	𐴓 𐴔	Final forms of certain vowels: ALEPH/A/E/NA, I/JA/YA.
Va′	𐴕 𐴖	Corresponding variant forms of Va.
Vu	𐴗 𐴘 𐴙	Medial and final forms of certain vowels: O/U/OE/UE, OE/UE.

See *Table 10* for examples of affected text representation cases. A column of intermediate glyphs (the output positional forms from the last stage) is provided for comparison.

Table 10: Examples

Phonetic	Interm.	→	Output	Graphetic	Rule	Character sequence
ba/be	*𐴀𐴓	→	𐴁	*b·*	1	< 𐴀 BA , 𐴓 ALEPH/A/E/NA >
bi	*𐴀𐴔	→	𐴂	*bI*	1	< 𐴀 BA , 𐴔 I/JA/YA >
bo/bu/bö/bü	*𐴀𐴗	→	𐴃	*bU*	2	< 𐴀 BA , 𐴗 O/U/OE/UE >
bö/bü	*𐴀𐴙	→	𐴄	*bÜ*	2	< 𐴀 BA , 𐴙 OE/UE >

Only the shape changes relevant to encoding are specified here. In actual fonts, depending on the specific style, glyphs (including the glyphs discussed above) might require more typographical adjustments that do not affect encoding. In particular, Vu glyphs usually need a change of connection point in order to properly join to Cr′ glyphs, although the resulted glyph changes can be subtle.

Other optional variant forms, such as 𐴆 → 𐴇 *bl*, are also commonly seen in fonts.

4 Adjustments to the character repertoire

The character repertoire presented in *Table 3* is a result of maximizing the "low ambiguity" advantage of the graphetic approach, therefore most confusable graphemes are unified or decomposed. Considering the benefits of recording more information of phonetic letters,[5] it might be desirable to moderately disunify or recompose some graphemes while tolerating more ambiguity. Some candidates are shown in *Table 11*.

Table 11: Candidates for disunification

From	Disunify/recompose	Note
𐴓 ALEPH/A/E/NA	ALEPH	Only initial and medial forms are attestted.
𐴔 I/JA/YA	LONG TOOTH OF OE/UE	Change joining type of 𐴙 OE/UE to dual-joining to have a medial form that has the disunified as a part.

[5]The key benefit is easier syllabification, which is helpful for collation, line-breaking, typographical preferences, etc.

Table 11: (continued)

From	Disunify/recompose	Note
ᠵ WA/EE	EE	Only medial and final forms are attestted.
ᠠ ALEPH/A/E/NA	CODA NA	Only medial and final forms are attestted.
< ᡅ O/U/OE/UE, ᠠ ALEPH/A/E/NA >	CODA DA	Only medial and final forms are attestted.
ᡆ I/JA/YA	INITIAL JA	The disunified becomes initial ᡇ JA.
< ᠠ ALEPH/A/E/NA, ᡂ HA/ZHI >	INITIAL HA	The recomposed becomes initial ᡂ HA/ZHI, pushing the existing intial form to be disunified.
ᡂ HA/ZHI	ZHI	Only initial form is attestted.
< ᡅ O/U/OE/UE, ᡅ O/U/OE/UE >	CHI	Only initial form is attestted.

Note although the medial forms ᡬ XA/GA and ᡭ GA are theoretically confusable with < ᠠ ALEPH/A/E/NA, ᠠ ALEPH/A/E/NA > and < ᡮ NA, ᡮ NA >, respectively, it is not beneficial to decompose them, because they are well distinguished by users (especially since the related intial and final forms are distinct) and decomposing only the medial forms would introduce problematic medial-less dual-joining characters.

5 Text processes

Text encodings play a fundamental role in the life cycle of digitalized text: users → *input* (keyboards, OCR, speech-to-text, etc) → Unicode encoding → *output* (display, collation, text-to-speech, etc) → users. The encoding model's implications on various text processes are discussed here.

5.1 Output: collation

When idealistic phonetic collation is expected, for example in dictionaries, phonetic information should be carried as metadata.

Otherwise, in day-to-day use cases where text must be collated automatically without metadata, a folded collation (certain homographic written forms are folded into a single collation element) is a reliable solution. For an introduction, see Shen Yilei's document MWG/2-N5 *G2P Sorting: An Automated Natural Sorting Method for Graphetically Encoded Mongolian*.

5.2 Input: keyboards and input methods

The preference of inputting phonetically should be fulfilled in the keyboard and input methods for average users. See *Section 7, Input* in L2/17-334 *On migration issues of the graphetic model* for a full discussion.

In order to prevent confusing fallbacks on undefined cursive joining positions (for example, if an undefined isolate BA were shown with its intial form ᢛ, it would be confused with ᢛ, an isolate O/U/OE/UE/WA), unattested positions must either have distinct and reasonable fallback forms or be explicitly marked as invalid.

Table 4 shows some recommended fallback forms in parentheses, but unattested isolate positions are especially problematic and need discussion. To explicitly mark invalid positions, visual aids similar to Indic scripts' dotted circles (◌) can be considered, for example, arrows pointing to the direction where a character is required. See *Table 12* for how arrows can improve user experience.

Table 12: Input experience

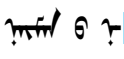
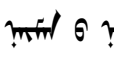
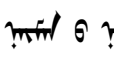
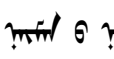
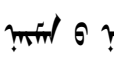
Keystroke	Output	Graphetic	Note
N	𐬬→|	*n*	𐬬 NA doesn't have a valid initial form.
A	𐬬𐬀|	*n*·	
I	𐬬𐬁|	*n*·*I*	
M	𐬬𐬀𐬨|	*n*·*Im*	
A	𐬬𐬀𐬨𐬀|	*n*·*Im*·	
Shift-N	𐬬𐬀𐬨𐬀𐬬|	*n*·*Im*··	
[Space]	𐬬𐬀𐬨𐬀𐬬 |	*n*·*Im*··␣	
Shift-U	𐬬𐬀𐬨𐬀𐬬 𐬵|	*n*·*Im*··␣*Ú*	
[Space]	𐬬𐬀𐬨𐬀𐬬 𐬵 |	*n*·*Im*··␣*Ú*␣	
N	𐬬𐬀𐬨𐬀𐬬 𐬵 𐬬→|	*n*·*Im*··␣*Ú*␣*n*	𐬬 NA doesn't have a valid initial form.
I	𐬬𐬀𐬨𐬀𐬬 𐬵 𐬬𐬁|	*n*·*Im*··␣*Ú*␣*nI*	
Shift-G	𐬬𐬀𐬨𐬀𐬬 𐬵 𐬬𐬇|	*n*·*Im*··␣*Ú*␣*nIG*	
E	𐬬𐬀𐬨𐬀𐬬 𐬵 𐬬𐬁𐬇|	*n*·*Im*··␣*Ú*␣*nIG*·	

To avoid artificial visual aids, thus providing a more natural user experience, the output can be always appended with a 𐬺 NIRUGU until an "end of word" signal (either a special keystroke or a character that is not bottom-joining) is detected. See *Table 13*.

Table 13: Alternative input experience

Keystroke	Output	Graphetic	Note
N	𐬬|	*n*-	
A	𐬬|	*n*·-	
I	𐬬|	*n*·*I*-	
M	𐬬|	*n*·*Im*-	
A	𐬬|	*n*·*Im*·-	
Shift-N	𐬬|	*n*·*Im*··-	
[End of Word]	𐬬𐬀𐬨𐬀𐬬 |	*n*·*Im*··␣	𐬺 NIRUGU replaced by a ␣ SPACE.
Shift-U	𐬬𐬀𐬨𐬀𐬬 𐬵 |	*n*·*Im*··␣*Ú*␣	[End of Word] implied by characters that are not bottom-joining.

Table 13: (continued)

Keystroke	Output	Graphetic	Note
N		*n·Im··␣Ú␣n-*	
I		*n·Im··␣Ú␣nI-*	
Shift-G		*n·Im··␣Ú␣nIG-*	
E		*n·Im··␣Ú␣nIG·-*	
[End of Word]		*n·Im··␣Ú␣nIG·␣*	· NIRUGU replaced by a ␣ SPACE.