

Progress report on
**UTN #57: Encoding and Shaping of
the Mongolian Script**

Kushim Jiang

Overview

- New standards for Mongolian encoding
- Fonts that support the new standards
- UTN #57 that automatically describe the new standards

Part I

**New standards for
Mongolian encoding**

New standards

- [GB/T 25914—2023](#): Hudum and Hudum Ali Gali
 - *Information technology — Traditional Mongolian nominal characters, presentation characters and use rules of controlling characters*
- GB/T 36649—xxxx: Todo and Todo Ali Gali
- GB/T 36641—xxxx: Sibe
- GB/T 36645—xxxx: Manchu and Manchu Ali Gali

Before GB/T 25914—2023

- [GB/T 25914—2010](#)
 - A very vague description of a small part of the shaping rules
 - $-(n)C \rightarrow \text{nun}$,
 $-(n)V \rightarrow \text{nun with a dot}$

1828	↪	蒙古文字母 NA MONGOLIAN LETTER NA
0005	↪	*na 词中第一形式 na first medial form
0014	↪	na 词中第二形式 na second medial form

^d 表示传统蒙古文元音前不带点的 ↪ NA。

^e 表示传统蒙古文辅音前的带点的 ↪ NA。

(d) FVS3 is used for NA without a dot when preceding a Hudum vowel.

(e) FVS1 is used for NA with a dot when preceding a Hudum consonant.

Before GB/T 25914—2023

- GB/T 25914—2010
- *Hudum shaping rules for nominal characters to presentation characters (version 9)*
 - [BabelStone](#)
 - *User agreements for Mongolian encoding system implementation*
 - [L2/17-251](#)

9. ◌→ 1828; 蒙 – 键盘位置: n 键 (下档)

① ◌ (◌→) ... ⇒ ◌→ 1828

② ◌ (◌→) ^[FV]_[S1]180B... ⇒ ◌→ 0097

③ ◌◌◌ (◌→) 元V音 ⇒ ◌→ 0014

④ ◌◌◌ (◌→) ^[FV]_[S1]180B & 元V音 ⇒ ◌→ 0005

⑤ ◌◌◌ (◌→) 辅C音 ⇒ ◌→ 0005

⑥ ◌◌◌ (◌→) ^[FV]_[S1]180B & 辅C音 ⇒ ◌→ 0014

Before GB/T 25914—2023

- GB/T 25914—2010
- *Hudum shaping rules for nominal characters to presentation characters (version 9)*
- [GB/T 32912—2016](#)
 - *Information technology — Specification of word form of traditional Mongolian words — Basic set*

ᠠᠨᠢᠯᠵᠠᠭᠤᠷ	aniljagur	M	ᠠᠨᠢᠯᠵᠠᠭᠤᠷ	A
ᠠᠨᠢᠰᠤ	anisu	M	ᠠᠨᠢᠰᠤ	N
ᠠᠨᠢᠰᠤᠠ	anix_a	M	ᠠᠨᠢᠰᠤᠠ	N
ᠠᠨᠢᠶᠤᠠ	aniy_a	M	ᠠᠨᠢᠶᠤᠠ	N
ᠠᠨᠢᠷ I	anir	M	ᠠᠨᠢᠷ	N
ᠠᠨᠢᠷ II	anir	M	ᠠᠨᠢᠷ	D
ᠠᠨᠢᠷᠯᠠᠰᠤ	anirlaxu	M	ᠠᠨᠢᠷᠯᠠᠰᠤ	V
ᠠᠨᠢᠷᠰᠢᠬᠤ	anirsixu	M	ᠠᠨᠢᠷᠰᠢᠬᠤ	V
ᠠᠨᠢᠷᠰᠢᠯ	anirsil	M	ᠠᠨᠢᠷᠰᠢᠯ	N
ᠠᠨᠢᠷᠲᠤ	anirtu	M	ᠠᠨᠢᠷᠲᠤ	A
ᠠᠨᠢᠷᠲᠤᠰᠤ	anirtuxu	M	ᠠᠨᠢᠷᠲᠤᠰᠤ	V
ᠠᠨᠤᠬᠠᠨ	anuxan	M	ᠠᠨᠤᠬᠠᠨ	A
ᠠᠨᠤᠷᠠᠳ	anurad	M	ᠠᠨᠤᠷᠠᠳ	N
ᠠᠨᠰᠠᠮᠪᠢ	ansambli	M	ᠠᠨᠰᠠᠮᠪᠢ	N
ᠠᠨᠲᠢᠨᠠ	antin_a	M	ᠠᠨᠲᠢᠨᠠ	N
ᠠᠨᠳᠠ I	anda	M	ᠠᠨᠳᠠ	N

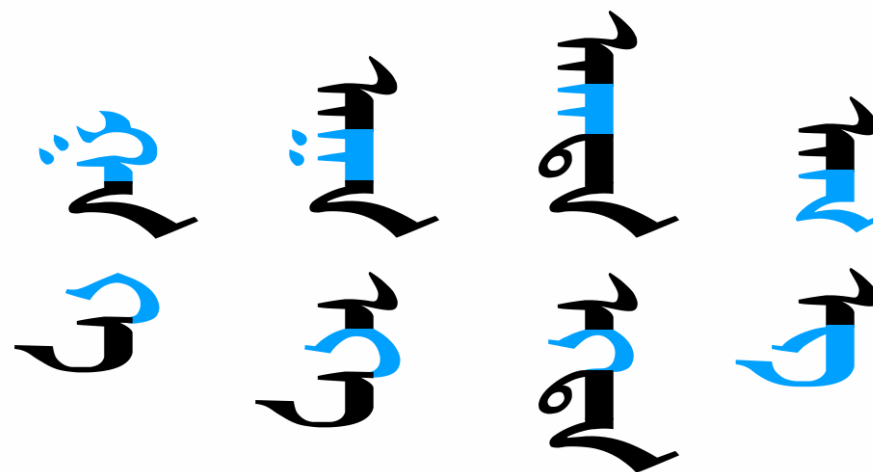
GB/T 25914—2023

- [GB/T 25914—2023](#)
 - Include the contents of the *Shaping rules*, so it is no longer vague, but rather accurate and formalized

1	0005	ʀ	2	①	[NIRUGU]+[U+ 1828 <i>n</i>]+[NIRUGU]
				②	[ExtMCS]+[U+ 1828 <i>n</i>]+[Cons]+[ExtMCS WB]
				③	[ExtMCS]+[U+ 1828 <i>n</i>]+[FVS2]+[ExtMCS]
2	0014	ʀ	1	①	[ExtMCS]+[U+ 1828 <i>n</i>]+[Vow]+[ExtMCS WB]
				②	[ExtMCS]+[U+ 1828 <i>n</i>]+[FVS1]+[ExtMCS]

GB/T 25914—2023

- [GB/T 25914—2023](#)
 - Include the contents of the *Shaping rules*, so it is no longer vague, but rather accurate and formalized
 - FVS is associated with specific variant
 - Specify the procedure for calculating the masculine or feminine context of QA and GA



ga a.ga ag.da ag
ge e.ge eg.de eg

GB/T 25914—2023

Table: masculine context for medial QA and GA

1	0006	ྐ	①	[ExtMCS]+[U+182C]+[NIRUGU]{0, ∞ }+[U+1820 U+1823 U+1824]+[ExtMCS WB]
2	001C	ྐ	①	[ExtMCS]+[U+182D]+[NIRUGU]{0, ∞ }+[U+1820 U+1823 U+1824]+[ExtMCS WB]
3	0006	ྐ	①	[MCS]+[NIRUGU FVS]{0, ∞ }+[1820 1823 1824]+[NIRUGU FVS]{0, ∞ }+[182C 182D]+[NIRUGU]{0, ∞ }+[Cons]+[ExtMCS WB]
4			②	[MCS]+[NIRUGU FVS]{0, ∞ }+[1820 1823 1824]+[NIRUGU FVS]{0, ∞ }+[182C 182D]+[NIRUGU]{1, ∞ }+WB
5			③	[MCS]+[NIRUGU FVS]{0, ∞ }+[1820 1823 1824]+[1822 Cons NIRUGU FVS]{0, ∞ }+[1822]+[NIRUGU FVS]{0, ∞ }+[182C 182D]+[NIRUGU]{0, ∞ }+[Cons]+[ExtMCS WB]
6			④	[MCS]+[NIRUGU FVS]{0, ∞ }+[1820 1823 1824]+[1822 Cons NIRUGU FVS]{0, ∞ }+[1822]+[NIRUGU FVS]{0, ∞ }+[182C 182D]+[NIRUGU]{1, ∞ }+WB
7			⑤	[1822 Cons NIRUGU FVS]{0, ∞ }+[1822]+[NIRUGU FVS]{0, ∞ }+[182C 182D]+[NIRUGU]{0, ∞ }+[Cons]+[1822 Cons NIRUGU FVS]{0, ∞ }+[1820 1823 1824]+[ExtMCS WB]

GB/T 25914—2023

- [GB/T 25914—2023](#)
 - Include the contents of the *Shaping rules*, so it is no longer vague, but rather accurate and formalized
 - FVS is associated with specific variant
 - Specify the procedure for calculating the masculine or feminine context of QA and GA
 - Using this standard, the font implementers can start to work, but then they can't start to work
 - Due to the precise shaping rules for characters to glyphs, it is possible to evaluate whether a font meets the requirements or not
 - The standard shaping rules have no connection to the OpenType Layout (OTL), and the shaping phases cannot be inferred directly from the rules

Part II

Fonts that support the new standards

Fonts

- [*Noto Sans Mongolian*](#) and *Mongolian Baiti*
- Manual maintenance of glyph set and OTL based on [Liang's approach](#) and expanded to other writing systems
- Checking for new shaping rules introduced by the new standard
- NNBSB has been merged to MVS since [Unicode 16.0](#)

<i>Shaping phase</i>		<i>Shaping step</i>
IA. General		• Basic character-to-glyph mapping
IIA. Cursive script		• Initiation of cursive positions
III. Hudum-specific <i>Reduction of phonetic letters to written units</i>	<i>Phonetic:</i>	1. Syllabic 2. Gender-specific 3. MVS-involving 4. NNBSB-involving
	<i>Graphemic:</i>	5. Devsger i 6. Post-bowed
	<i>Uncaptured:</i>	7. FVS-selected
IIB. Cursive script (continued) <i>Sub-written unit variations</i>		• Variation involving bowed written units • Cleanup of format controls • Optional treatments
IB. General (continued) <i>Typography</i>		• Vertical forms of punctuation marks • Optional treatments

Fonts

- *Noto Sans Mongolian*
and Mongolian Baiti
 - Manual maintenance of glyph set and OTL
 - Separate maintenance of fonts for the seven writing systems, and finally merged into a universal writing system
 - Test the OTL by [EAC testing website](#)
 - 3545 words

	No.	form	fvs	Testing Words			
				ruleid	Expected		DraftEAC-Test.otf
ISOL	1	𐌲	3	MAD10	1	𐌲	𐌲
					2	𐌲	𐌲
				MAD11	1	𐌲	𐌲
				MAD12	1	𐌲	𐌲
					2	𐌲𐌲	𐌲𐌲
	2	𐌴	1	MAD20	1	𐌴	𐌴
				MAD21	1	𐌴	𐌴
					2	𐌴𐌴	𐌴𐌴
	3	𐌶	2	MAD30	1	𐌶	𐌶
				MAD31	1	𐌶𐌶	𐌶𐌶
					2	𐌶𐌶	𐌶𐌶
				MAD32	1	𐌶	𐌶
					2	𐌶𐌶	𐌶𐌶

Part III

**UTN #57 that automatically
describe the new standards**

UTN #57

- Data
 - Transliteration for phonetic letters and written units
 - Variants and locale data for each letter arranged by joining position
 - Contextual conditions and categories for each writing systems
 - Required particles and required ligatures

Files

main

Go to file

> .vscode

✓ data

aliases.ts

export.ts

ligatures.ts

locales.ts

misc.ts

particles.ts

variants.ts

writtenUnits.ts

> docs

> lib

> tests

> web

.gitattributes

mongolian-utn / data / variants.ts

Code

Blame

8934 lines (8931 loc) · 168 KB

```
87      },
88      "MONGOLIAN LETTER A": {
89        isol: {
90          "1": {
91            written: ["A"],
92            locales: {
93              MNG: {
94                gb: "00B3 a second isolated form",
95                eac: "MAD2",
96              },
97              MNGX: {},
98              TOD: {
99                gb: "1821 a second isolated form",
100                eac: "TAD2",
101              },
102              TODX: {
103                lvs: true,
104              },
105            },
106          },
107          "2": {
108            written: ["Aa"],
109            locales: {
110              MNG: {
111                conditions: ["chachlag", "particle"],
112                gb: "00B4 a third isolated form",
113                eac: "MAD3",
114              },
115            },
116          },
117        },
118      },
119    },
120  },
121}
```


UTN #57

- Data
- Web-based UTN
 - unicode.org/notes/tn57
 - mongolian-utn.pages.dev (editor's draft)
 - Based on Draft UTN by Liang ([L2/19-368](#)) and expanded to seven writing systems
 - Generate written units, variants and conditions from data files

docs

Introduction

Architecture

Hudum

Todo

Sibe

Manchu

Hudum Ali Gali

Todo Ali Gali

Manchu Ali Gali

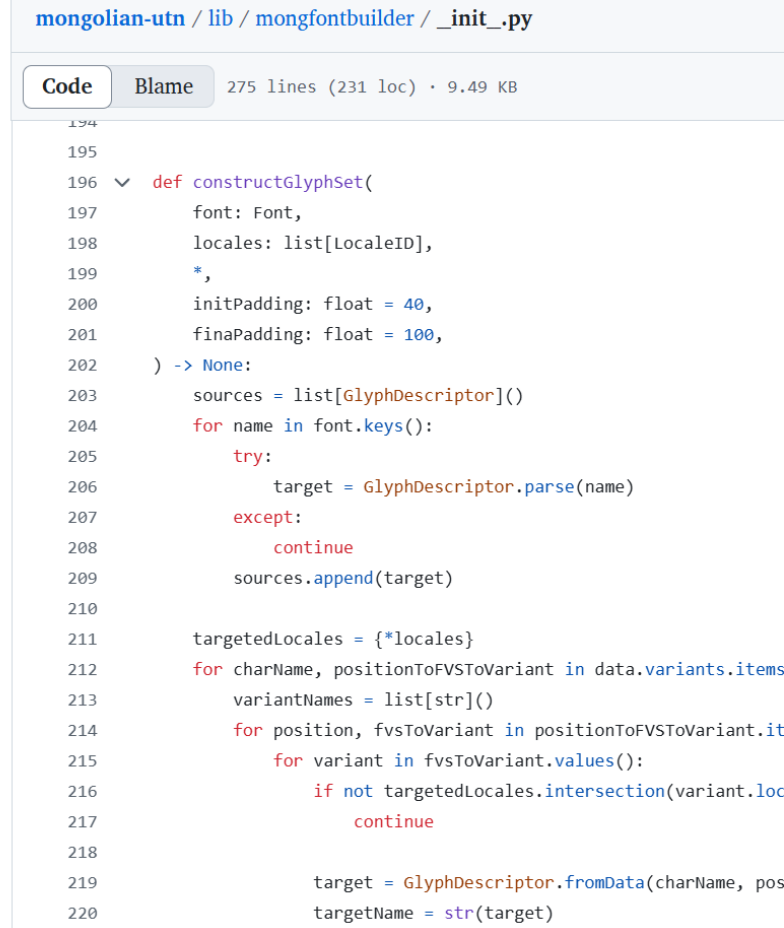
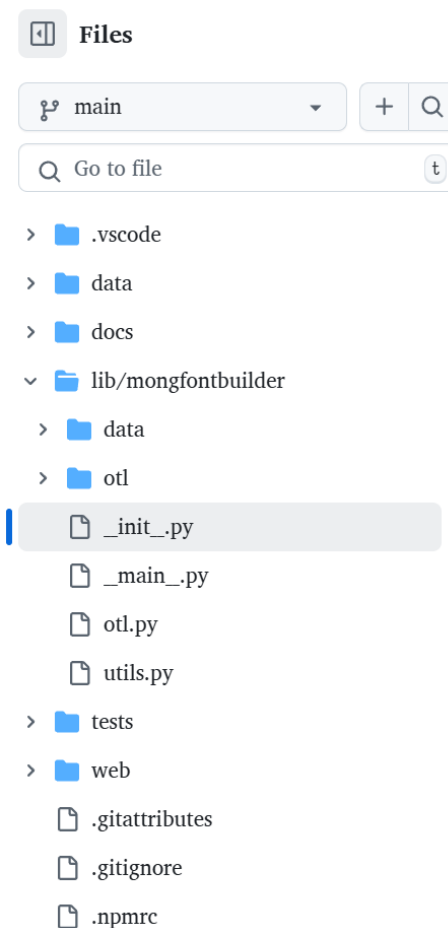
Comparison

Mongolian-specific phase

Shaping step	Letters	Condition	Mapping type
1. Chachlag	<i>a, e</i>	if follows an MVS:	chachlag
	<i>a, e</i>	if follows an MVS and precedes an FVS:	default
2. Syllabic	<i>o, u, ö, ü</i>	if follows an <i>initial consonant</i> :	marked
	<i>o, u, ö, ü</i>	if precedes an FVS or follows an FVS:	default
	<i>ö, ü</i>	if final form precedes an FVS2 or FVS4 that precedes <i>h</i> or <i>g</i> :	marked
	<i>ö, ü</i>	if medial form follows a <i>consonant</i> cluster that starts with an initial <i>consonant</i> :	marked
	<i>d</i>	if precedes a final <i>vowel</i> that does not precedes an FVS:	marked
	<i>n, j, w</i>	if precedes an MVS that precedes an isolated <i>a</i> or an isolated <i>e</i> :	chachlag_onset
	<i>h, g</i>	if precedes an MVS that precedes an isolated <i>a</i> :	chachlag_onset
	<i>g</i>	if precedes an MVS that precedes an isolated <i>e</i> :	chachlag_devsger
	<i>n, t, d</i>	if precedes a <i>vowel</i> :	onset
		else if follows a <i>vowel</i> :	devsger
	<i>h, g</i>	if precedes a <i>masculine vowel</i> :	masculine_onset
		else if precedes a <i>feminine vowel</i> or a <i>neuter vowel</i> :	feminine
		else if follows a <i>masculine vowel</i> :	masculine_devsger

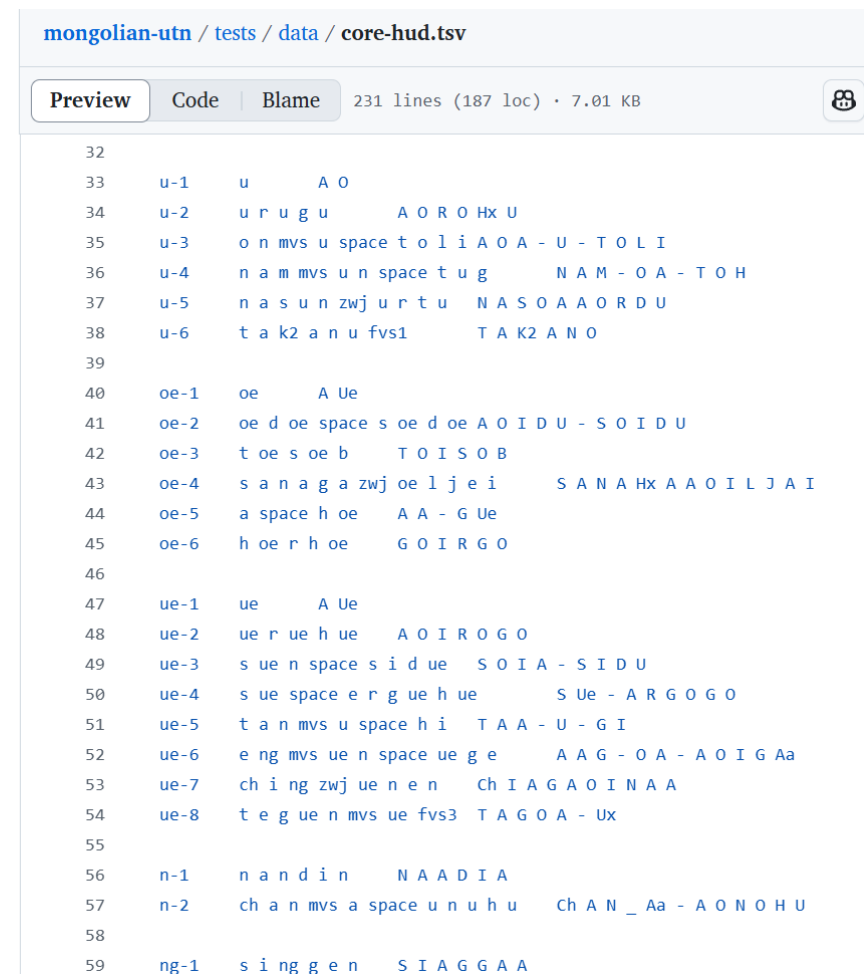
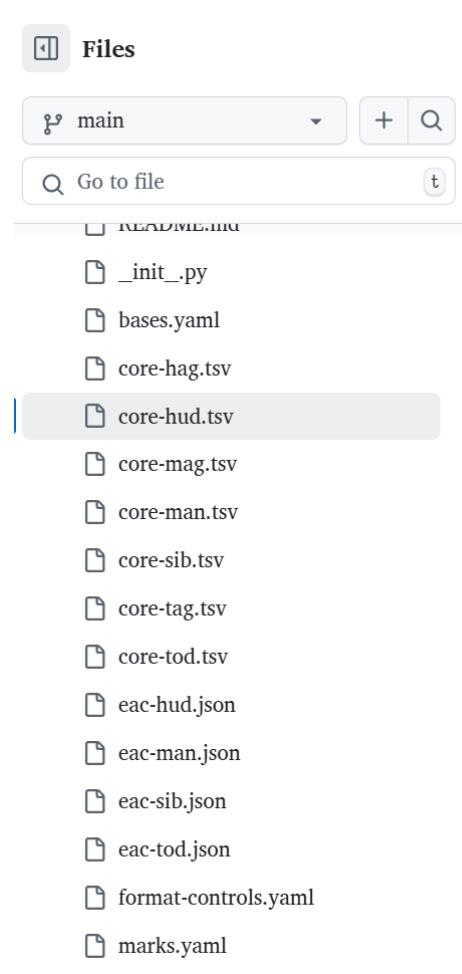
UTN #57

- Data
- UTN websites
- mongfontbuilder
 - Construct complete glyph set for given set of writing systems from data files
 - Generate OTL given set of writing systems from data files
 - [tptq.feacomposer](https://github.com/mongoliansoftware/tptq-feacomposer)



UTN #57

- Data
- UTN websites
- mongfontbuilder
- Testing
 - [Crowbar](#)
 - Letter transliteration > character > shaping engine (font) > glyph name > written unit name > written unit transliteration



Progress report on
**UTN #57: Encoding and Shaping of
the Mongolian Script**

Kushim Jiang

Thank you for listening.