[Draft]

Mongolian text representation and shaping:
Preliminary review of the EAC standard

To: Unicode Technical Committee
From: Liang Hai / 梁海 <lianghai@gmail.com>
Date: 15 May 2020

1   Introduction

This is an preliminary review of the industry standard, 信息技术 传统蒙古文名义字符 到变形显现字符的转换补充规则 / *Information technology—The Transferring Rules of Traditional Mongolian Nominal Form to Variant Form*, published by 内蒙古自治区民族 事务委员会 (Ethnic Affairs Committee of the Inner Mongolia Autonomous Region, commonly referred to as *EAC*).

(A literal translation of the standard's title would be: "information technology— supplemented rules for converting traditional Mongolian script nominal characters to transformation–presentation characters".)

This standard document is commonly referred to as *the EAC standard*, and its latest published version has been archived at the author's public repo *Mongolian resources*, as the EAC website has been unreliable (follow the link for more explanation).

2   The variant set

Table 1 shows the variant set of the EAC standard, with a normalized format to match the draft UTN for easier comparison. Notes on the format:

1. The first two columns show the phonetic letters (in the draft UTN's transliteration scheme) and their encoded characters. Phonetic letters involving automatic Hudum shaping (i.e., automatic contextual shaping beyond cursive joining) are highlighted in yellow. Only the non-highlighted phonetic letters are comparable to Arabic letters.

2. Variant glyphs of each phonetic letter are grouped into four columns according to their cursive joining positions (isolated, initial, medial, and final). The joined ends of variants are marked with a blue nirugu (▪). The alignment of variants between positions is informative, mostly according to the variants' constituent written units.

3. Four more columns, parallel to the four columns of variants, show each variant's constituent written units and conditions. The cursive joining positions are not marked for most variants as variants are already grouped into each position's

column. However, variants mismatched to positions are grayed out (along with their glyphs) and their joined ends are explicitly marked with a dash (-).

4. Thus each variant can be unambiguously referred to according to their phonetic letter, cursive joining position, and constituent written units (as well as the mismatched position, if any). For example, 🐎 *a*.init.ᴀᴀ and 🐴 *ŋ*.init.ᴀɢ.medi (an initial variant of *ŋ*, but actually is a medial).

5. On each cursive joining position, there is one and only one default positional variant, which is not highlighted.

6. Automatically triggered variants (with the automatic Hudum shaping, which includes the MVS and NNBSP mechanisms, among others) are marked with a question mark (?) highlighted in <mark>yellow</mark>.

7. The FVS assignment is marked with a digit 1–4. The manual variants that are neither default or automatic (thus rely on the FVS mechanism) are highlighted in <mark>red</mark>. A default positional variant's FVS digit is parenthesized.

Table 1. Variant set of the EAC standard

Phonetic letter & character		Written forms: default positional forms, additional automatic variants, & manual variants				Note			
a	1820 A					AA (3)	AA (2)	AA 1	
						A 1	A ?1	A	A (2)
						Á ?2			Á ?1
e	1821 E						AA 1		
						A (2)	A	A	A (2)
						Á ?1			Á ?1
i	1822 I					AI (3)	AI (2)	AI 1	
						I ?1	I ?1	I (3)	I
								II ?2	
						I̦ 2			
o	1823 O					AO	AU	AU 1	
								U	U (2)
								O ?1	
u	1824 U					AO (3)	AU (2)	AU 1	
						U ?1	U ?1	U	U (2)
								O ?1	
						U̦ 2			

ö	1825 OE	〓	〓	〓		AÜ	AUI	AUI 2		
		〓				AU 1				
				〓	〓			U (3)	U (3)	
				〓	〓			UI ?1	Ü ?1	
					〓				O ?2	
ü	1826 UE	〓	〓	〓		AÜ	AUI (2)	AUI 2		
		〓				AU 1				.isol.AÜ—*lack of FVS*
		〓	〓	〓	〓	U ?2	U ?1	U (3)	U (3)	
				〓	〓			UI ?1	Ü ?1	
					〓				O ?2	
		〓				Ų 3				
é	1827 EE	〓	〓			AW	AW			
				〓	〓			W	W	
n	1828 NA	〓	〓	〓	〓	N-	N	N ?1	N ?1	
		〓	〓	〓	〓	A- 1	A 1	A (2)	A (2)	
ŋ	1829 ANG	〓	〓	〓	〓	-AG-	-AG-	AG	AG	
b	182A BA	〓	〓	〓	〓	B-	B	B	B	
p	182B PA	〓	〓	〓	〓	P-	P	P	P	
b	182C QA	〓	〓	〓	〓	H-	H (3)	H (3)	H	
		〓	〓	〓	〓	Ğ- 1	Ğ 1	Ğ 1	Ğ 1	
		〓	〓	〓		G- 2	G ?2	G ?2		
		〓	〓	〓		Ğ- 4	Ğ 4	Ğ 4		
g	182D GA	〓	〓	〓	〓	H- 1	H 1	H (3)	H (1)	
		〓	〓	〓	〓	Ğ-	Ğ (3)	Ğ ?1	Ğ ?3	
		〓	〓	〓	〓	G- 2	G ?2	G ?2	G ?2	
		〓	〓	〓		Ğ- 4	Ğ 4	Ğ 4		
m	182E MA	〓	〓	〓	〓	M-	M	M	M	
l	182F LA	〓	〓	〓	〓	L-	L	L	L	
s	1830 SA	〓	〓	〓	〓	S-	S	S	S	

ś	1831 SHA					Ś-	Ś (2)	Ś (2)	Ś	
						S- 1	S ?1	S ?1		
t	1832 TA					T-	T	T 1	T	
								D		
d	1833 DA					T- 1	T			
						D-	D ?1	D ?1	D 1	.init.т—*lack of FVS*
								Đ (2)	Đ	
ć	1834 CHA					Ć-	Ć	Ć	Ć	
j	1835 JA							J	J	.isol.ı—*lack of FVS*
						I ?	I		I ?	.isol.ı.init—*lack of FVS*
						I-				.fina.ȷ—*lack of FVS*
										.fina.ı—*lack of FVS*
y	1836 YA					Y-	Y (2)	Y (3)		
						I- 1	I ?1	I ?1	I	
								II 2		
r	1837 RA					R-	R	R	R	
w	1838 WA					W-	W	W	W	
								U 1	U ?1	.fina.w—*lack of FVS*
f	1839 FA					F-	F	F	F	
ḱ	183A KA					Ḱ-	Ḱ	Ḱ	Ḱ	
k	183B KHA					K-	K	K	K	
c	183C TSA					C-	C	C	C	
z	183D ZA					Z-	Z	Z	Z	
ħ	183E HAA					H-	Ħ	Ħ	Ħ	
ř	183F ZRA					Ř-	Ř	Ř	Ř	
ł	1840 LHA					LH-	LĦ	LĦ	-LH-	
ž	1841 ZHI					Ž-	Ž	Ž-	Ž-	
č	1842 CHI					Č-	Č	Č-	Č-	

4

2.1 Irregularity on the FVS assignment

Theoretically, on each cursive joining position of a phonetic letter, as long as there are more than one automatic variants (including the default variant), every automatic variant needs to be assigned with an FVS, so the FVS mechanism can be complete. However, certain variants in the EAC standard's set lack an FVS assignment (noted in the table).

2.2 Disagreements with the draft UTN

The following are the high-level, intentional differences:

1. The EAC standard is conservative, thus it keeps *ḱ* U+183A ᠺ ᴋᴀ and does not utilize U+1807 SSBM to address the compound word issue.

2. Therefore the aleph-ed vowel variants in compound words, such as *a*.init.ᴀᴀ, only exist in the EAC standard.

3. All cursive joining positions are normatively filled in the EAC standard, while many of them appear to be positional mismatches and are merely duplications of certain fallback positions. The UTN prefers to specify the fallback with a separate mechanism and potentially different forms.

4. There are default positional forms in the EAC standard. While in the draft UTN, all automatic variants on a position of a phonetic letter are considered parallel to each other.

5. The EAC standard includes .fina.Á variants of *a* and *e*, because its analysis of the bowed consonants' shaping behavior.

6. *ö*.isol.ᴀᴜ, parallel to the commonly understood *ü*.isol.ᴀᴜ, is included in the EAC standard.

7. Variants of *y* and *w* that are used as offglides in diphthongs are included as manual variants in the EAC standard.

8. *j*.fina.ɪ exists in the EAC standard as a pre-*chachlag* variant.

9. Pre-modern dotted variants of *i*, *u*, and *ü* only exist in the EAC standard. The UTN should include these. See also [L2/20-057](#), *Proposal to add* ᴍᴏɴɢᴏʟɪᴀɴ ꜰʀᴇᴇ ᴠᴀʀɪᴀᴛɪᴏɴ sᴇʟᴇᴄᴛᴏʀ ꜰᴏᴜʀ.

10. Pre-modern variants of *h* and *g* also only exist in the EAC standard. However, all the additional variants besides the ǧ ones are questionable as they are the result of a highly phonetic analysis.

11. Pre-modern *a*-alike initial form of *e* is also included in the EAC standard.

3 Recommendation for further analyses

In order to investigate a possible convergence between the EAC standard and the eventual Mongolian UTN, the following tasks are recommended as the next steps:

1. Independent implementations and thorough testing of the EAC standard. Note that machine readable data is available at the author's public repo *Mongolian resources*, under the directories [/archived-eac-publications/data](#) and [/archived-eac-publications/properties](#); a font production workflow is being built under [/sandbox-font](#).

2. The EAC standard needs to be radically transformed into a streamlined format so it is comprehensible and reviewable. The section 2 of this document is an attempt at transforming the part about the variant set, in order to gain a wider reviewer audience. A lossless transformation is not practical as the EAC standard's text is not coherent, thus the focus of the transformation should be trying to cover the overall strategies and provide machine readable data that can drive implementation attempts.

3. Investigate the pre-modern variants of *e, i, u, ü, h,* and *g.*

<center>⋆ EOF ⋆</center>