

# Proposal to encode two Mongolian letters

**Badral Sanlig**  
badral@bolorsoft.com  
**Jamiyansuren Togoobat**  
onoltnmn@yahoo.com  
**Munkh-Uchral Enkhtur**  
munkhuchral@bolorsoft.com  
Bolorsoft LLC, Mongolia

## 1 Introduction

This is a proposal to encode two additional mongolian letters that are most actively used for writing texts in traditional Mongolian writing system. These letters are at the present partially implemented as variant forms of correspondingly QA, GA. The first letter is Mongolian KE, which is known as feminine form of QA and second letter is Mongolian GE, which is known as feminine form of GA.

Pos.	KE	QA	GE	GA
initial	ᠳ	ᠳ	ᠳ	ᠳ
medial	ᠳ	ᠳ	ᠳ	ᠳ
final	- ᠰ	ᠳ	ᠳ	ᠳ

Table 1: Forms of KE(QA) and GE(GA).

In current encoding scheme, only final and medial form of GE are encoded and all other forms of GE, KE (such as initial GE, KE, medial KE) can be illustrated only through open type font algorithms. On top of that, those currently encoded forms of GE are only as variant of GA (medial form of GE is second variant by FVS1, whereas final form of GE is fourth variant by FVS3) implemented. QA, KE, GA, GE are most frequently used characters in Mongol script, as most of the heading words are started by these letters, all nominal forms of verb are built by these letters and all long vowels are illustrated by these letters. To back up our argument, we have done the frequency analysis of Mongol script letters in our lexical database, which contains 41808 non-inflected distinct words (lemma), result of our analysis are shown in Table 2.

No.	Letter	Name	Code point	Occurrence
1	Ѐ	MONGOLIAN LETTER A	U 1820	44245
2	Ӄ	MONGOLIAN LETTER I	U 1822	29090
3	ӄ + Ӆ	MONGOLIAN LETTER GA	U 182D	11105+16507
4	ӄ	MONGOLIAN LETTER E	U 1821	23269
5	ӄ	MONGOLIAN LETTER U	U 1824	19942
6	ӄ	MONGOLIAN LETTER LA	U 182F	19865
7	ӄ	MONGOLIAN LETTER RA	U 1837	18383
8	ӄ + Ӆ	MONGOLIAN LETTER QA	U 182C	7069+6243
9	ӄ	MONGOLIAN LETTER DA	U 1833	11596
10	ӄ	MONGOLIAN LETTER UE	U 1826	10988

Table 2: Top 10 frequently used letters in Mongol script

Due to its frequent occurrences, almost all font vendors try to implement KE and GE through open type font algorithms, which would then automatically represent all correct forms. But to display KE and GE properly in all possible practices (as there are lots of exceptionally written words, such as neutral Mongolian words, special words, foreign and loan words) results in unlimited number of rules, which in lots of time breaks one another and failure to display the correct form. To recognize female GA (GE) in all loan words like Aghsin, Oyu or in foreign (or imported) words like Angli are impossible without checking whole words. Almost all modern scientific terms are directly transliterated or transcribed in Mongolian. The reason is, scientific terms are very hard to translate or intentionally not translated to keep precision of the term. How should we write foreign human or location names that starts one of the GE, KE, QA, GA letters? The current only possible solution is to release new version of font, which would then include all open type font rules for newly recognized patterns or even words. Also, to recognize pattern correctly and writing contextual rules are quite complex [L2/14-031]. It's the main reason why all Mongolian fonts are so heterogeneous and incomplete. Such inconsistency of fonts leads to the instability of encoding in turn. Let us assume, we want to write  "Bicig" meaning

of "Script" in English as representative sample of Mongolian neutral words. This word will be typed as 'bicig'. There are many fonts that have implemented rules to recognize 'bicig' and display the correct final form of GE (). But of course, there are many fonts that haven't implemented the rules. In such cases, the user will have to type as 'bicig' + FVS1/FVS2 to display the correct final form

of GE, otherwise without any free variation selector will be incorrect word **ᠭ**  
displayed. This is very common scenario and highly destabilizes the encoding.

Input sequence	Mongolian script (v4)	Mongolian Baiti	Noto sans mongolian	Mongolian white	Menksoft	Code2000	From Unicode specification
ᠭ 182A ᠳ 1822 ᠳ 1834 ᠳ 1822 ᠳ 182D	ᠮ	ᠮ	ᠮ	ᠮ	ᠮ	ᠮ	182C ᠶ MONGOLIAN LETTER QA 182C ᠶ first form (initial) = 182C ᠶ first form (isolate) 182C ᠶ first form (medial) 182C 18B8 ᠶ second form (initial) 182C 18B8 ᠶ second form (isolate) 182C 18B8 ᠶ second form (medial) 182C 18B8 ᠶ third form (initial) 182C 18B8 ᠶ fourth form (initial) 182D ᠶ MONGOLIAN LETTER GA 182D ᠶ first form (initial) = 182D ᠶ first form (isolate) 182D ᠶ first form (medial) 182D 18B8 ᠶ first form (final) 182D 18B8 ᠶ second form (initial) 182D 18B8 ᠶ second form (isolate) 182D 18B8 ᠶ second form (medial) 182D 18B8 ᠶ third form (final) 182D 18B8 ᠶ fourth form (medial)
ᠶ 182A ᠳ 1822 ᠳ 1834 ᠳ 1822 ᠳ 182D FVS1 180B	ᠮ	ᠮ	ᠮ	ᠮ	ᠮ	ᠮ	
ᠶ 182A ᠳ 1822 ᠳ 1834 ᠳ 1822 ᠳ 182D FVS2 180C	ᠮ	ᠮ	ᠮ	ᠮ	ᠮ	ᠮ	

Figure 1: Font mismatches

From above table, we can easily conclude that users would ignore to type free variation selectors, if they use "Noto Sans Mongolian" or "Menksoft" fonts. "Mongolian Script" font users have to type FVS1 after GA but "Mongolian Baiti" fonts users must type FVS2 to get proper results. The correct input sequence is displayed in middle row highlighted in table 1 as in clearly specified in Unicode standard, that we use FVS1 for feminine GA. The "Noto Sans Mongolian" font represents all input sequences correctly. However, it's not recommended to type users anything without consideration. The reason, why some font developers implement such rules to shape automatically, is in every text processing software, which shows the preview of "Mongol Bicig" in font select menu, "ᠮᠶᠱᠲᠳ" is represented incorrectly as "ᠮᠶᠱᠲᠳ". The scenario is illustrated in figure 2.

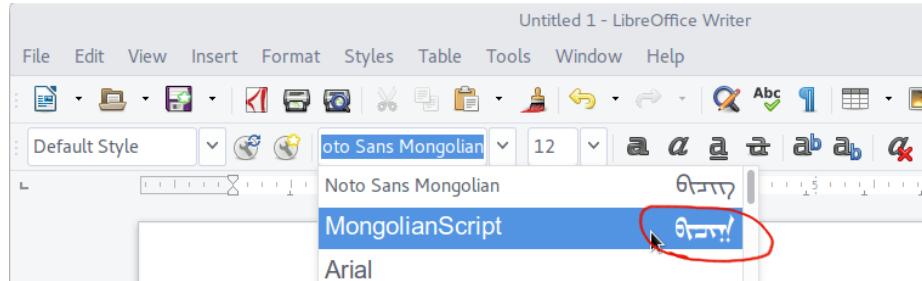


Figure 2: Preview in font selection menu

Mongolian users feel very unconfident while writing GA, QA in neutral and feminine words, as the representation of this character is totally font dependent and really unpredictable how it will displayed. These problems are clearly solved by introducing KE, GE separately from QA and GA.

It can be confusing to understand GE, KE are letters or characters. They are letter, as both represent a sound in speech and both are taught as individual letters in almost all Mongolian alphabet tables. The Mongolian letter KE represents [k], whereas the Mongolian letter QA represents the [q]. Mongolian letter GE represents [g], whereas GA represents [ ]. The details on form of these letters are designed in [Mongolian aesthetics]. We attached some alphabet tables as example in section "Figures". The recommended alphabet table in Mongolia is the table defined in "The School Textbook for Mongolian Script", which is printed in 1986 by Dr. Prof. Choi. Luvsanjav.

Here is the translation of a few snippets of text from the book.

"The Mongolian script has thirty-two letters: seven vowels and twenty-five consonants. Twenty consonants are used for Mongolian speech sounds while five consonants such as f, z, c, k and h for foreign language sounds. The consonants are: (See blue mark in figure 10)

Thus, seven vowels and twenty consonants in total twenty-seven letters are sufficient to represent Mongolian sounds\* (See red mark 2 in figure 10). \*A consideration of GA and KE as QA only and GA and GE as GA only, will make the number of consonants eighteen. A consideration to remove NGA and LHA due to compound letters, will make the number of consonants sixteen. A consideration to add five letters for foreign words, the number of consonants will add up to twenty-one. Then with the seven vowels, it will be twenty-eight letters in the Mongolian alphabetical order. Twelve syllable closing consonants:

2. GA (ГА). Medial form is 但不限制 and final form is 但不限制 when it closes syllable. For example: тартар-а, 但不限制 (тар) (See red mark 3 in figure 10)

3. GE (ГЭ). Medial form is 但不限制 and final form is 但不限制 when it closes syllable. (See red mark 4 in figure 10)"

In 1995, Choi. Luvsanjav, Ts. Shagdarsuren, O. Chilkhaasuren also proposed<sup>1</sup> to encode KE, GE separately in the Mongolian transliteration and transcription table. (See figure 13) However, the table did not provide any explanation why KE and GE should be considered as separate letters. We met Prof. Dr. Ts. Shagdarsuren to ask about his argument in 10 September 2018. He developed the following arguments.

1. KE and GE letters are graphically different from QA and GA letters. They must be considered systematically in contrast with QA and GA.
2. KE and GE must not be encoded with same code. GE can close syllables but KE never closes syllables. According to Mongolian phonetic law, there must be only KE when a feminine syllable is closed by d or s consonant. For example, when one writes "beledkebe" (in English prepared) the user must not write "G" instead of "K".

---

<sup>1</sup><https://www.unicode.org/L2/topical/mongolian/mnism19931214.pdf>

In our study on the Mongolian transliteration, we have identified the four letters are considered always as individual letters. For the registration of all bibliographic items, almost all universities and libraries use the Vladimirtsov-Mostaert system (V-M), which is the only transliteration system for the Mongolian vertical script which is used commonly throughout Mongolian studies worldwide<sup>2</sup>. These letters are also considered as individual letters in Vladimirtsov-Mostaert system (V-M)<sup>3</sup>.

## 2 Why is an additional encoding necessary?

Let us trace the KE and GE in the history of Mongolian Linguistics. In the first Mongolian grammar "Jirüken-ü tolta" written by Sasakiya Pandita (1243), in his Mongol script alphabet, "QA", "KE", "GA", "GE" are described with independent letters [2]. Cosgi Odser also follow Sasakiya Pandita on the KE and GE [3]. Later Mongolian linguists explain that Sasakiya Pandita counted eighteen consonant letters including QA as Q, KE as K, GA as γ and GE as G as four independent letters [4] [5] [7].

The unification of KE and GE, QA and GA is one of the most critical defects of Mongolian Unicode architecture, which has to be fixed urgently. The letter QA/GA occurs only in masculine words and KE/GE in feminine or neutral words but the concept of word gender is grammatical term. Grammatically, QA and KE, GA and GE could be same, but orthographically they are totally different.

The following benefits arise by the additional encoding of these letters.

- The font implementation will be simplified significantly.

At the moment, immaculate and well functioning Mongolian font does not (also cannot) exist at all, due to number of open type font rules needs to be written to solve the problem. If the syllable structure of Mongol script is followed strictly and there exist not a single, or with very few well known exception, then it would have been totally fine to solve it through open type font rules. Unfortunately there exist too many exceptional cases that are not included in the syllable structure, which then leads to unlimited number of font rules. By separating KE, GE from QA, GA and encoding them additionally, the number of rules will be significantly reduced. Currently, "MongolianScript" font from Bolorsoft contains more than hundred contextual rules for automatic determination of GE, KE, which is still far from completion. The "Mongolian Baiti" font includes definitively several hundred rules just for this reason.

- Due to required entangled contextual rules and deep embedded linguistic knowledge, currently only limited numbers of font developers are involved

---

<sup>2</sup><https://collab.its.virginia.edu/wiki/tibetan-script/Transliteration%20Schemes%20for%20Mongolian%20Vertical%20Script.html>

<sup>3</sup><http://www.loc.gov/catdir/cpso/romanization/mongolia.pdf>

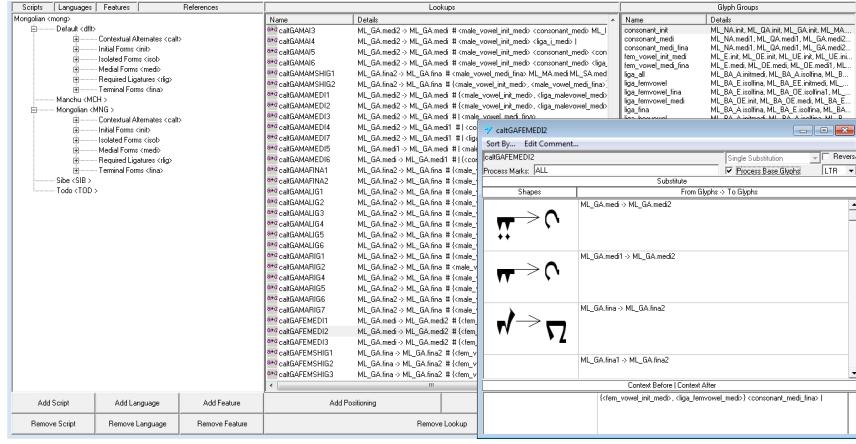


Figure 3: From contextual rules of MongolianScript font in VOLT

with Mongol script font development. But by separating and additionally encoding those 2 letters, complexity will be significantly reduced. Thus, it will make Mongol script attractive for wider circle of font developers and shall result in more fonts.

- Usage of free variation selectors will be decreased. Currently, all three variation selectors are used only for QA, GA, NA letters. The fourth medial variant of NA is not used in practice. Therefore, by separating and additionally encoding those 2 letters, FVS3 will be eliminated.
- The non-interoperability of different implementations and fonts will be dramatically reduced. Any issues, caused in it by an unstable stream of patches with unstable rationales will be removed.
- It will be easier for end users to memorize written Mongolian words by typing these letters explicitly.

### 3 Can one grapheme be used for both letters?

No, absolutely not. They are different letters. The usage, transliteration, transcription are different. Even forms are slightly different. See [Mongolian aesthetics]. The differences are clearly recognizable, if two characters shall be written after each other. For example,



If these characters are merged into one grapheme, then it will break current encoding model.

## 4 Can Ali Gali KA (U+1889) used at least as KE?

The KE letter in Ali Kali system is devoted to describe Sanskrit "KHA" and Tibetan "KA". This means ALI GALI KA is used to indicate masculine words. We can find it in the following examples.

- Mo. gandarvis (BT 161b.10) < skr. Gandhārva
- Mo. kandig (Burq. 19a.7) < skr. Kanthaka
- Mo. galab (Burq.3a.20, 62a) < skr. Kalba
- Mo. gšan (Burq. 26a) < Skr. Ksana
- Mo. agaru (Burq.46b) < Skr. Aparu<sup>4</sup>

In the eighth part titled *Ambiguous names* of his *Mongolian grammar Mongγol üsiüg-ün γool yosun-i todadqayči toli*, Delegjungnai writes that to write *nüken* one should use *ü* with *long teeth* in the form to start a word, and to write **lüge** *ü* without a *long teeth* (silbi ügei) in the middle form. However, their gender division can be identified in the context with *partnering letters* (neren-ü doturaki nökor-iyer medekü)." *Nüken* is a word which means a whole, **lüge** is a case suffix. In the case of **lüge**, *partnering letters* in the word is GE. This GE in **lüge** helps to identify the belly indicates u but ü. Thus, GE is used as a key consonant to identify that **lüge** is a feminine word [8].

If current model had KE, GE separately encoded from QA, GA, it is unlikely that Ali Gali KA would have been ever defined. It is more likely to assume that KE would be used as Ali Gali KA. We would not believe that works in reversed way. If it is strictly necessary to reuse Ali Gali KA as Mongolian KE then this character has to be completely redefined and renamed. It is absolutely necessary to freeze all characters in Mongolian sub block and maybe allow usage in derived sub blocks like Todo, Manchu, Shibe, Buriat, Ali Gali but not vice versa. Not a character, which is taken from derived sub block (Todo, Manchu etc.), should be used in basic block (Mongolian).

## 5 Unicode Properties

KE and GE both are consonant letters, which indicate feminine Mongolian word. We suggest encoding the proposed character KE at U+1879 and GE at U+187A, which are the first reserved spaces after the basic Mongolian letters. The suggested first character name is "MONGOLIAN LETTER KE", which describes speech of k and suggested second character name is "MONGOLIAN LETTER GE", which describes speech of g. The proposed letters have following positional forms.

---

<sup>4</sup>BT stands for Bodhicaryāvatar-yin tayilburi. Burq. means Twelve deeds of Buddha (Burqan-u arban qoyra jokiyanggui uiles) [6].

<b>1879</b>	<b>MONGOLIAN LETTER KE</b>	ᠺ
	Initial form	ᠺ
	Medial form	ᠺ
	Final form	ᠺ
<b>187A</b>	<b>MONGOLIAN LETTER GE</b>	ᠳ
	Initial form	ᠳ
	Medial form	ᠳ
	Final form	ᠳ

There exist no final form of KE, as it belongs to non-closing consonants (ülü debiskerlekü), therefore it has to be followed by a vowel. Nevertheless, a final form of a character is needed to display at the writing moment in computer system. Thus, we can define the final form as a medial like form, which builds user's intuitiveness to type something after that. This practice could be applied generally for every final positional form, which is not defined in Unicode standard (shown in standard as █).

Figure 4 shows the visual differences of these two letters.

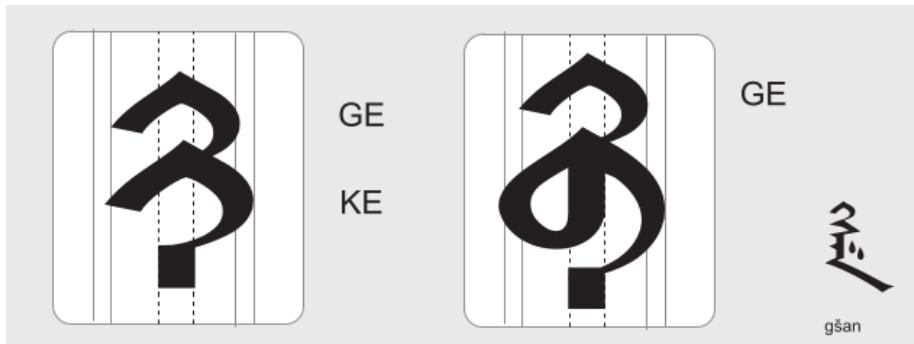


Figure 4: Differences between two letters

## 5.1 UCD Properties

1879;MONGOLIAN LETTER KE;Lo;0;L;;;;;N;;;;;
--

Line break: AL

Joining type: D (Dual Joining)

Script: Mongolian

187A;MONGOLIAN LETTER GE;Lo;0;L;;;;;N;;;;;
--

Line break: AL

Joining type: D (Dual Joining)

Script: Mongolian

## 6 Note on collation

The alphabetical order of MONGOLIAN LETTER KE is equal to the order of MONGOLIAN LETTER QA and the order of MONGOLIAN LETTER GE is exactly same as the order of MONGOLIAN LETTER GA. Therefore, the collation algorithms should handle KE identical as QA and GE identical as GA. It is analog to German collation for umlaut letters, such as "ä" is equal to the "a", "ö" is equal to the "o", "ü" is equal to the "u" and "ß" is same as "ss". Such equations are well known common practice for collation algorithms.

Sequentially, MONGOLIAN LETTER KE is located after MONGOLIAN LETTER QA, MONGOLIAN LETTER GE is placed just after MONGOLIAN LETTER GA. See figure 10.

1.  -A(1820)	2.  -E(1821)	3.  -I(1822)	4.  -O(1823)
5.  -U(1824)	6.  -Ö(1825)	7.  -Ü(1826)	8.  -EE(1827)
9.  -NA(1828)	10.  -ANG(1829)	11.  -BA(182A)	12.  -PA(182B)
13.  -QA(182C)	14.  -KE(1879)	15.  -GA(182D)	16.  -GE(187A)
17.  -MA(182E)	18.  -LA(182F)	19.  -SA(1830)	20.  -SHA(1831)
21.  -TA(1832)	22.  -DA(1833)	23.  -CHA(1834)	24.  -JA(1835)

Continued on next page

25. <b>ئ</b> -YA(1836)	26. <b>ئ</b> -RA(1837)	27. <b>ئ</b> -WA(1838)	28. <b>ئ</b> -FA(1839)
29. <b>ئ</b> -KA(183A)	30. <b>ئ</b> -KHA(183B)	31. <b>ئ</b> -TSA(183C)	32. <b>ئ</b> -ZA(183D)
33. <b>ئ</b> -HAA(183E)	34. <b>ئ</b> -ZHA(183F)	35. <b>ئ</b> -LHA(1840)	36. <b>ئ</b> -ZHI(1841)
37. <b>ئ</b> -CHI(1842)			

Table 3: Mongolian alphabet in sequential order

The following table shows the alphabetical sorting of some Mongolian words, which contains QA, GA, KE, GE letters.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ئوشتىم	ئۇغۇزلىرىن	ئەشىل	ئەسىسەن	ئەتىم	ئەشقا	ئەشىقى	ئەپا	ئەسەم	ئەشىل	ئەپەم	ئەشىقەنۇنىڭ	ئەتىكىن	ئەپەن	ئەپەزىز
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
ئەھىزىز	ئەپەيەستەنەن	ئەشىسىنەن	ئەپەن	ئەپەيەنەن	ئەپەيەنەن	ئەپەن	ئەپەن	ئەپەن	ئەپەن	ئەپەن	ئەپەن	ئەپەن	ئەپەن	ئەپەن

Table 4: Example of alphabetical sorting

## 7 Figures



Figure 5: From the Jirüken-ü tolta-yin tayilburi [1]

Figure 6: Alphabet table on [15]

*Schrift und Aussprache.*

c) Consonanten.

			d	q"	q" a" d" q"
			t	q"	q" a" d"
			r	q"	q"
			l	q"	q"
Am Anfang	{	b	g	g"	g" l"
In der Mitte	{	ch	gh	gh"	gh" l"
Am Ende	{	k	g	g"	g" l"
Am Anfang	{	s, ds	ts	ss	sch
In der Mitte	{	ts	ts	ts	w
Am Ende	{	ss	ts	ts	w

§. 4. Es ist bei den Mongolen nicht gebräuchlich, die Consonanten als abgesonderte Schriftzeichen zu behandeln: sie werden immer mit einem der Vocale verbunden, und als einfache Sylben dargestellt und ausgesprochen, wie aus folgender Tabelle zu ersehen:

a. E.	i. d. M.	i. A.	a. E.	i. d. M.	i. A.
q" a" l"	q" a" l"	q" a" l"	cha	qa"	qa" l"
ke	qa"	qa" l"	ke	qa"	qa" l"
ki	qa"	qa" l"	ki	qa"	qa" l"

Figure 7: Alphabet table on [9]

### Letters

### III. Script

58. The old script is written or printed vertically from left to right. Most letters have three different forms according to their position in the word: an initial one used at the beginning of words, a medial one used in the middle of words, and a final one used at the end of words. The letters will be discussed in the Mongolian alphabetic order.

59. The letters are given in the following table.

Letters of the Mongolian Alphabet

Number	Transcription	Characters		
		Initial	Medial	Final
1	a	ᠠ	ᠤ	ᠥ
2	e	ᠡ	ᠤ	ᠥ
3	i	ᡢ	ᠤ	ᠥ
4	o u	ᠽ	ᠶ	ᠹ
5	ə ɨ	ᠱ	ᠳ	ᠴ
6	n	ᠶ	ᠳ	ᠴ
7	ng	ᠩ	ᠩ	ᠩ
8	g	ᠭ	ᠩ	ᠩ
9	γ	ᠳ	ᠳ	ᠳ
10	b	ᠪ	ᠶ	ᠶ
11	p	ᠶ	ᠶ	ᠶ
12	s	ᠶ	ᠶ	ᠶ
13	đ	ᠶ	ᠶ	ᠶ
14	t d	ᠶ	ᠶ	ᠶ
15	l	ᠯ	ᠯ	ᠯ
16	m	ᠮ	ᠮ	ᠮ
17	č	ᠴ	ᠴ	ᠴ
18	j	ᠵ	ᠵ	ᠵ
19	y	ᠶ	ᠶ	ᠶ
20	k g	ᠶ	ᠶ	ᠶ
21	k	ᠶ	ᠶ	ᠶ
22	r n	ᠶ	ᠶ	ᠶ
23	v	ᠶ	ᠶ	ᠶ
24	h	ᠶ	ᠶ	ᠶ

**Монгол бичигийн цагаан толгой**

No.	МОНГОЛ ҮСЭГ			ДУУДАЛГЫН (tc.) БОЛОН ҮСГИЙН (tl.) ГАЛИГ			
	ЭХИН	ДУНД	АДАГ	(tc.)	(tl.) ЭХИН	ДУНД	АДАГ
01	ᠡ	ᠳ	ᠱ	a   e	"	'	/   (-) / \
02	ᠡ	ᠳ	ᠱ	i	'i	i	k
03	ᠳ	ᠳ	ᠱ	o   u	'w	w	b
04	ᠳ	ᠳ	ᠱ	ö   ü	'wi	w	b
05	-	-	-	n	n	n   '	n-   \
06	ᠩ	ᠩ	ᠩ	ng	-	'k	'i /
07	ᠴ	ᠴ	-	q	q	"	-
08	ᠴ	ᠴ	ᠴ	γ	γ	γ   "	γ   " /
09	ᠵ	ᠵ	ᠵ	b	b	b	b   w /
10	ᠵ	ᠵ	ᠵ	p	p	p	p
11	ᠰ	ᠰ	ᠰ	s	s	s	s   '
12	ᠰ	ᠰ	-	š	š	š	-
13	ᠶ	ᠶ	ᠶ	d	t	d   w'	d   w \
14	ᠶ	ᠶ	-	t	t	d	-
15	ᠯ	ᠯ	ᠯ	l	l	i	i
16	ᠮ	ᠮ	ᠮ	m	m	m	m
17	ᠴ	ᠴ	-	č	č	č	-
18	ᠵ	ᠵ	-	j	i	j	-
19	ᠶ	ᠶ	ᠶ	y	i   y	i   y	k
20	ᠳ	ᠳ	-	k	k	k	-
21	ᠳ	ᠳ	ᠳ	g	k	k	i /
22	ᠷ	ᠷ	ᠷ	r	ր	r	ր
23	ᠸ	ᠸ	-	v (w)	v	v	-
24	ᠳ	ᠳ	-	h	'h	h	-

Figure 9: Alphabet table on [11]

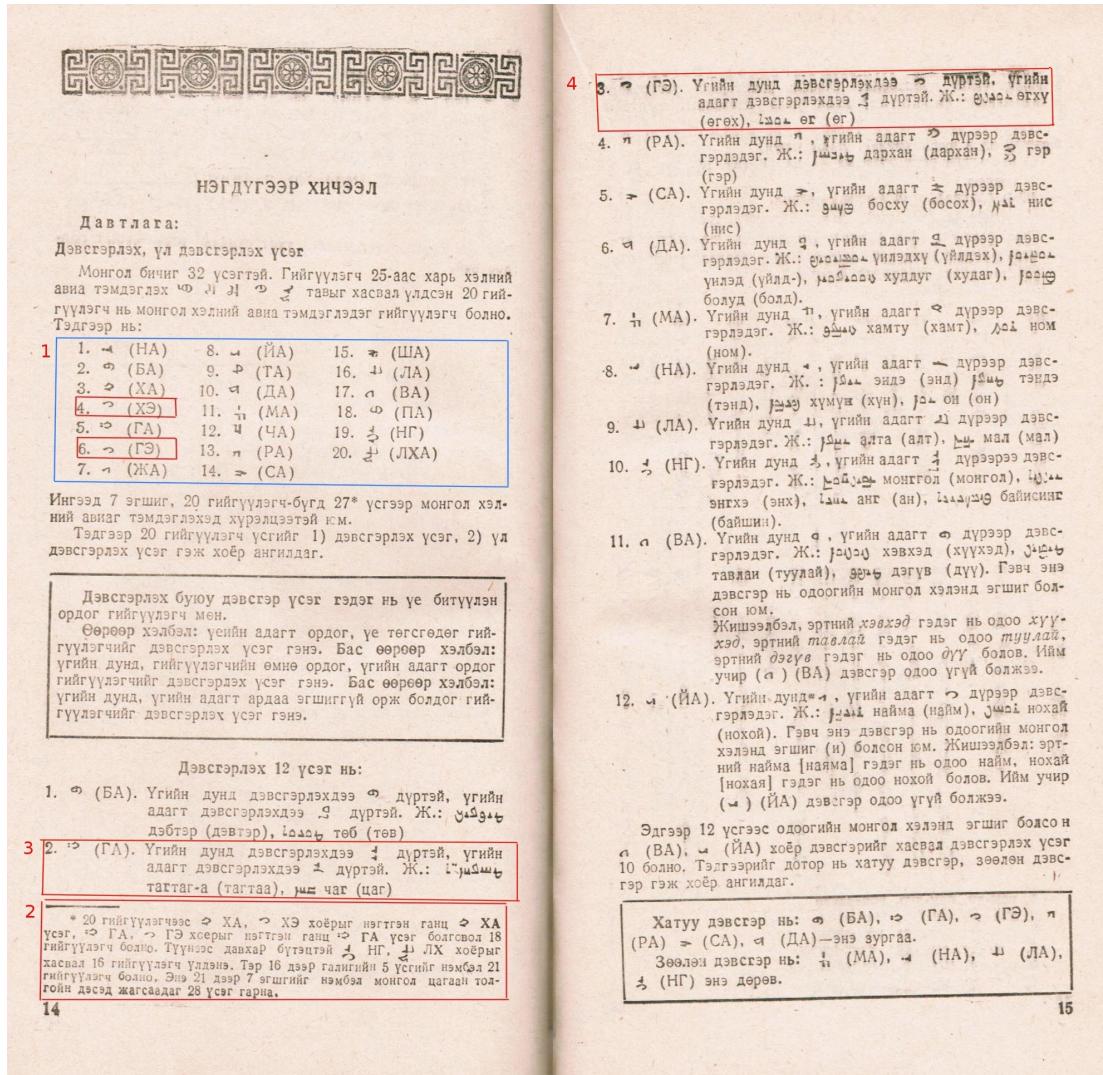


Figure 10: Alphabet table on school book [12]

### Гийгүүлэгчийг тэмдэглэх нь

	н	б	x(a)	x(э)	г(а)	(Э)г г(э)	(а)г	ж	й	т	д	м	ч
Эхэнд	ӟ	ӟ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ
дунд	ӝ	ӟ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ
сүүлд	ӝ	ӟ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ	ӝ
жишээ	нээм нээм нээм нээм												

Figure 11: Alphabet table on [13]

## 2. THE MONGOLIAN ALPHABETS

### 2.1 The Classical Alphabet

#### 2.1.1 General Table of the Classical Alphabet

We know that this attempt to compare Mongolian sounds with English sounds doesn't necessarily follow linguistic rules. It is intended as a way to get started for beginners. We assume that Mongolian pronunciation and conversation won't be undertaken as individual study, but with a Mongolian teacher, so we will do without an exact phonetic study.

	Print	Cursive	Transcription	Sound in a Mongolian word	Transcription	Sound in an English word
1.	𠂇	ӟ	a	амагалт	amagalta	gather
2.	𠂈	ӟ	e	енгэг	engge	bed
3.	𠂉	ӟ	i	илли	illi	easy
4.	𠂊	ӟ	o	огуул	oguu	bought
5.	𠂋	ӟ	u	унгуул	unguu	similar to 'bone'
6.	𠂌	ӟ	ö	өмүдүү	omudu	similar to 'girl/pur'
7.	𠂍	ӟ	ü	үндүүлүү	ündüü	soup
8.	𠂎	ӟ	n	ном (нэмэртэй)	nom nemertay	name sing
9.	𠂆	ӟ	b	баруу	baruu	bird
10a	ӟ	ӟ	q	զաբ	qabar	loch (Scottish)
10b	ӟ	ӟ	k	кеlekүү	keleku	non-existent in Engl.
11a	ӟ	ӟ	y	yal	yal	get
11b	ӟ	ӟ	g	гөрөл	gorel	get
12.	𠂔	ӟ	l	јама	jama	lateral fricative "l" (non-existent in Engl.)
13.	𠂖	ӟ	m	моду	modu	my
14.	𠂗	ӟ	s	самур	samur	song
15.	𠂘	ӟ	š	шату	šatu	shot
16.	ӟ	ӟ	t	төвүүлүү	tovusu	tall
17.	ӟ	ӟ	d	дэгебүри	degeburi	day
18.	𠂑	ӟ	č	чингис	chingis	church
19.	𠂐	ӟ	j	жүржүү	jurju	joke
20.	𠂑	ӟ	y	ябуюу	yabuu	yacht
21.	𠂑	ӟ	r	яртуу	jartuu	rolled "r" (non-existent in Engl.)
22.	𠂑	ӟ	v	явар	yavar	yase

The following letters don't belong to the original Classical Mongolian alphabet, but were introduced because of the use of loan words.

23.	ڣ	ڣ	p	پارك	park
24.	ڻ	ڻ	f	ڻانڪ	franci
25.	ڱ	ڱ	k	ڳوڻ	king
26.	ڦ	ڦ	ê	ڳڦڻ	german
27.	ڻ	ڻ	c	ڻانڙ	its
28.	ڻ	ڻ	z	ڻانڙ	zandan
29.	ڦ	ڦ	h	ڳڦڻ	house
30.	ڦ	ڦ	ڙ	ڙونال	joke

#### Explanations:

to 2) and 26): some dialects pronounce the "c" closed, some open. In Inner Mongolia the closed "c" is written with letter 2), the open "c" for foreign words with number 26. In Khalkh, letter 2) is pronounced open, similar to 'egg'.  
 to 5): this open "u" (back language) is difficult to pronounce; 'bone' is only similar.  
 to 6): the closed "o" (front language) is pronounced quite differently in each dialect; the British 'girl' follows the western Mongolian dialect; American 'pur' reflects Khalkh.  
 to 8): for the two different pronunciations of "n" see [3rd pronunciation rule \(p.23\)](#).  
 to 10): this sound is absent in English, but the Scottish word 'loch' is pronounced like the back language "q"; the front language "k" is even more difficult to pronounce for English speakers (like the German "mich").  
 to 12): this Mongolian "l" is **lateral fricative**, not like the English "l" (practice well!).  
 to 18): this letter also has for **different pronunciations** depending on the dialects. In Inner Mongolia this sound is "ch"; in Khalkh it could also be pronounced as "ts".  
 to 19): see 18), in Inner Mongolia "ch" in Khalkh it could also be "ds".  
 to 21): the Mongolian "y" must be rolled.

#### 2.1.2 Polyphotetic Alphabet

As we mentioned on page 3, in a polyphotetic alphabet, the same letter can symbolize different sounds. In the Classical Mongolian alphabet, this is true of many letters. In the past, scholars attempted time and again to change the Mongolian alphabet to a monophotetic one:

Pag-pa Script	14th century
Tod Script	16th century (Oirat)
Soyombo Script	17th century (Khalkh)
Vagindra Script	19th century (Buriat)
Cyrillic Script	20th century (Khalkh)

These attempts were largely handicapped by the fact that they only reflected the pronunciation of one dialect. It is very difficult to reflect justly all Mongolian dialects with one monophotetic alphabet. For this reason the Classical Mongolian alphabet has the advantage that it gives all the Mongolian dialects one common script.

Figure 12: Alphabet table on [14]

4. TRANSLITERATION AND TRANSCRIPTION TABLES

ANNEX A.1

Transcription and transliteration table for the basic characters of Mongolian script:

N	Characters			Transcription	Transliteration		
	initial	medial	final		initial	medial	final
1	ᠡ	ᠾ	ᠱ	a/e	"	'	'/(')+
2	ᠡ	ᠾ	ᠱ	i	'i	i	i
3	ᠾ	ᠾ	ᠾ	o/u	'w	w	w
4	ᠾ	ᠾ	ᠾ	ö/ü	'wi	w	w
5	ᠶ	ᠶ	ᠶ	n	n	n	' / (' ) +
6	-	ᠩ	ᠩ	ng/ñ	-	nk	'ŋ
7	ᠪ	ᠪ	ᠪ	b	b	b	b w'
8	ᠬ	ᠬ	-	q	q	"	-
9	ᠴ	ᠴ	-	k	g	g	-
10	ᠶ	ᠶ	ᠶ	y	y	y	" -
11	ᠶ	ᠶ	ᠶ	g	g	g	i'
12	ᠶ	ᠶ	-	j	i	i	-
13	ᠶ	ᠶ	ᠶ	y	i y	y	i
14	ᠶ	ᠶ	-	t	t	d	-
15	ᠶ	ᠶ	ᠶ	d	t	d	w'/(w')+
16	ᠮ	ᠮ	ᠮ	m	m	m	m
17	ᠴ	ᠴ	-	č	č	č	-
18	ᠷ	ᠷ	ᠷ	r	r	r	r
19	ᠶ	ᠶ	-	s	s	s	" -
20	ᠶ	ᠶ	-	š	š	š	-
21	ᠯ	ᠯ	ᠯ	l	l	l	l
22	ᠶ	ᠶ	-	p	p	p	-
23	ᠶ	ᠶ	-	v	v	v	-
24	ᠶ	ᠶ	-	h	h	h	-

+ In the cases of the absence of the printing possibility

- 7 -

Figure 13: Annex A1 of ISO\_CD 14522

ANNEX A.2

Transcription and transliteration table of the character of Mongolian script for the foreign words:

N	Characters			Trans- cription	Transliteration		
	initial	medial	final		initial	medial	final
25	đ	đ	đ	é	'e	e	e
26	đ	đ	đ	f	f	f	f
27	đ	đ	đ	k	k	k	k
28	đ	đ	đ	z	z	z	z
29	đ	đ	đ	c	c	c	c
30	đ	đ	đ	z	z	z	z

A.3 Punctuation:

Punctuational signs in  
the text of Mongolian script

Transcribing  
Signs

Significances

- |    |   |    |               |
|----|---|----|---------------|
| 1. | egčim or biry-a,<br>the sign at the<br>beginning of books<br>or chapters, at the<br>beginning of the<br>recto of each pages | .  | (,)           |
| 2. | čeg "the point"   | .  | (,)           |
| 3. | dabqur čeg "double<br>point"  | :  | (,), (;), (.) |
| 4. | dorbeljin čeg 'the<br>square point' -at the<br>end of the chapters or<br>books  | :: | (.)           |

## 8 Acknowledgments

I express my gratitude to Odgerel Boldbaatar for her careful proofreading of the paper. I am also thankful to Erdenechimeg Myatav for providing detailed information on Technical Report 170 and valuable comments on the history of the Mongolian encoding model.

## References

- [1] Rabjamba Danjindagba. *Jirüken-ü tolta-yin tayilburi Ogtaruyin maani*, XVIII Century
- [2] Serjee, Zh. *Early Mongolian Linguistic texts*, Treasures of Mongolian culture and Tibeto-Mongolian Buddhism, Vol. 4 Budapest, 2005: 28-29; (*Jirüken-ü tolta-yin tayilburi oγtarγui-yin mani kemekü sudur orosibai, Mongγol kele-sinjilel-ün durasqaltu bičig-üd*), Köke qota, 1983:3
- [3] Serjee, Zh. *Jirüken-ü tolta-yin tayilburi oγtarγui-yin mani kemekü sudur orosibai, Mongγol kele-sinjilel-ün durasqaltu bičig-üd*, Köke qota, 1983:7-8
- [4] Tumurtogoo, D. *About wa and we in "Jiruken-u tolta" written by Sasakiya Pandita Gunggajalcan*, International conference of Mongolists, II, Vol. 2; 1973:163-165
- [5] Tumurtogoo, D. *Issues on theory and history of Mongolian linguistics*, International conference of Mongolists, II, Vol. 2; Ulaanbaatar, 2002: 475-479
- [6] Tumurtogoo, D. *Features of writing forms of Sanskrit words in the Mongol script primary sources written in XIII-XIV*, International conference of Mongolists, II, Vol. 2; Ulaanbaatar, 2002:342-351
- [7] Zayabaatar, D. *Clarification on concepts of Gunggajalcan's 44 letters and Cosgi Odser's 123 letters in Traditional linguistic sources*, Research journal of Scool of Mongolian Language and Culture, National University of Mongolia, Vol. XXXVII Ulaanbaatar, 2013: 39-43
- [8] Delegjungnai, D. *Mongγol üsüg-ün γool yosun-i todadqayči toil*, Primary sources of Mongolian linguistic works Köke qota, 1959: 115
- [9] Schmidt I.J. *Grammatik der mongolischen Sprache*. Number Bd. 1-2. Kaiserliche Akademie der Wissenschaften, 1831. Available at <https://books.google.be/books?id=zrpLAAAACAAJ>
- [10] N.N. Poppe. *Grammar of Written Mongolian*. Far Eastern and Russian Institute publications on Asia. Harrassowitz, 1954. LCCN: 56036254

- [11] Ts. Shagdarsürg. *Study of Mongolian Scripts (Graphic Study or Grammatology)*. Enlarged 2nd edition. Bibliotheca Mongolica: Monograph 1, 2001. Ulaanbaatar: Center for Mongol Studies, National University of Mongolia.
- [12] *School Textbook for Mongolian Script*. Ulaanbaatar, 1986.
- [13] Nadmin Ya. *Монголын бичигийн орфографийн толгойн толгот* (Orthographic dictionary for Mongolian script). Ulaanbaatar, 1990
- [14] Rita K., Tserenpil D. *Mongolian Grammar*. Hong Kong, 1996 ISBN: 9628457055
- [15] Choimaa Sh. *Монголын бичиг* (Mongol Bichig). Ulaanbaatar, 1989
- [TR170, P. 10-11] Erdenechimeg, M., Moore, R., Namsrai, Yu. *UNU/IIST Technical Report No. 170 - Traditional Mongolian Script in the ISO/Unicode Standards*, Macau, 1999
- [Mongolian aesthetics] *The Aesthetics of Mongolian Script in 800 Years*, Jamyangsureng Togoobat Available at <https://www.unicode.org/~lisa/mongoliandocs/mwg2-16MongolianAesthetics11.0beta-Badral.pdf> (21.03.2018, Unicode)
- [L2/14-031] Aaron Bell, Greg Eck, Andrew Glass *Proposal to update the code charts for Mongolian*, <https://unicode.org/L2/L2014/14031-mong-chart-upd.pdf>