

Proposal to encode two Mongolian letters

Badral Sanlig

badral@bolorsoft.com

Jamiyansuren Togoobat

onoltnn@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	᠎	᠎	᠎	᠎
medial	᠎	᠎	᠎	᠎
final	- ᠎	᠎	᠎	᠎


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	ᠠ	MONGOLIAN LETTER A	U 1820	44245
2	ᠢ	MONGOLIAN LETTER I	U 1822	29090
3	ᠭ + ᠠ	MONGOLIAN LETTER GA	U 182D	11105+16507
4	ᠡ	MONGOLIAN LETTER E	U 1821	23269
5	ᠤ	MONGOLIAN LETTER U	U 1824	19942
6	ᠯ	MONGOLIAN LETTER LA	U 182F	19865
7	ᠷ	MONGOLIAN LETTER RA	U 1837	18383
8	ᠬ + ᠠ	MONGOLIAN LETTER QA	U 182C	7069+6243
9	ᠳ	MONGOLIAN LETTER DA	U 1833	11596
10	ᠰ	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  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
ᠠᠨᠠᠭᠤᠨᠠᠨᠠᠭᠤᠨ	ᠠᠨᠠᠭᠤᠨᠠᠨᠠᠭᠤᠨ	ᠠᠨᠠᠭᠤᠨᠠᠨᠠᠭᠤᠨ	ᠠᠨᠠᠭᠤᠨᠠᠨᠠᠭᠤᠨ	ᠠᠨᠠᠭᠤᠨᠠᠨᠠᠭᠤᠨ	ᠠᠨᠠᠭᠤᠨᠠᠨᠠᠭᠤᠨ	ᠠᠨᠠᠭᠤᠨᠠᠨᠠᠭᠤᠨ
182A						
1822						
1834						
1822						
182D						
182A						
1822						
1834						
1822						
182D						
FVS1 180B						
182A						
1822						
1834						
1822						
182D						
FVS2 180C						

From Unicode specification

182C ᠠ MONGOLIAN LETTER GA
 ~ 0445 x cyrillic small letter ha
 ~ 182C ᠠ first form (isolate)
 ~ 182C ᠠ first form (initial)
 ~ 182C ᠠ first form (medial)
 ~ 182C ᠠ first form (final)
 ~ 182C 180B ᠠ second form (isolate)
 ~ 182C 180B ᠠ second form (initial)
 ~ 182C 180B ᠠ second form (medial)
 ~ 182C 180C ᠠ third form (medial)
 ~ 182C 180C ᠠ fourth form (medial)

182D ᠡ MONGOLIAN LETTER GHA
 ~ 0433 x cyrillic small letter ghe
 ~ 182D ᠡ first form (isolate)
 ~ 182D ᠡ first form (initial)
 ~ 182D ᠡ first form (medial)
 ~ 182D ᠡ first form (final)
 ~ 182D 180B ᠡ second form (initial)
 ~ 182D 180B ᠡ second form (medial)
 ~ 182D 180C ᠡ second form (final)
 ~ 182D 180C ᠡ second form (medial)
 ~ 182D 180C ᠡ third form (medial)
 ~ 182D 180C ᠡ fourth form (medial)

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, "ᠠᠨᠠᠭᠤᠨᠠᠨᠠᠭᠤᠨ" is represented incorrectly as "ᠠᠨᠠᠭᠤᠨᠠᠨᠠᠭᠤᠨ". The scenario is illustrated in figure 2.

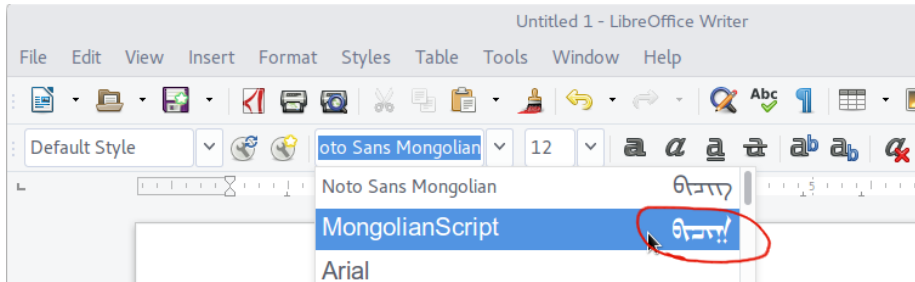


Figure 2: Preview in font selection menu





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 [γ]. 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 (ΓΑ). Medial form is  and final form is  when it closes syllable. For example:  (tartar-a),  (uar) (See red mark 3 in figure 10)

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

In 1995, Choi. Luvsanjav, Ts. Shagdarsuren, O. Chilkhaasuren also proposed¹ 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".

¹<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 worldwide². These letters are also considered as individual letters in Vladimirtsov-Mostaert system (V-M)³.

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 Saskiya Pandita (1243), in his Mongol script alphabet, "QA", "KE", "GA", "GE" are described with independent letters [2]. Cosgi Odser also follow Saskiya Pandita on the KE and GE [3]. Later Mongolian linguists explain that Saskiya 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

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

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

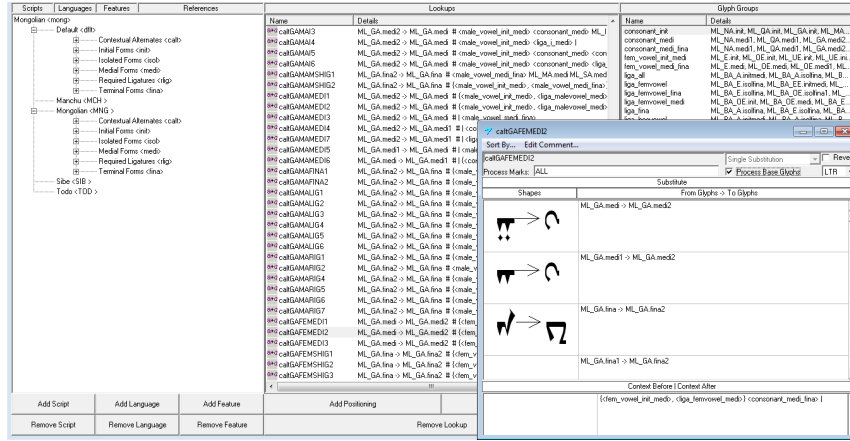



Figure 3: From contextual rules of MongolianScript font in VOLT

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. Aparu⁴









In the eighth part titled *Ambiguous names* of his *Mongolian grammar Mong-γol üsüg-ün γool yosun-i todadqarči toli*, Delejungnai 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.

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

1879	MONGOLIAN LETTER KE	
	Initial form	
	Medial form	
	Final form	
187A	MONGOLIAN LETTER GE	
	Initial form	
	Medial form	
	Final form	

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 𐠫).

Figure 4 shows the visual differences of these two letters.

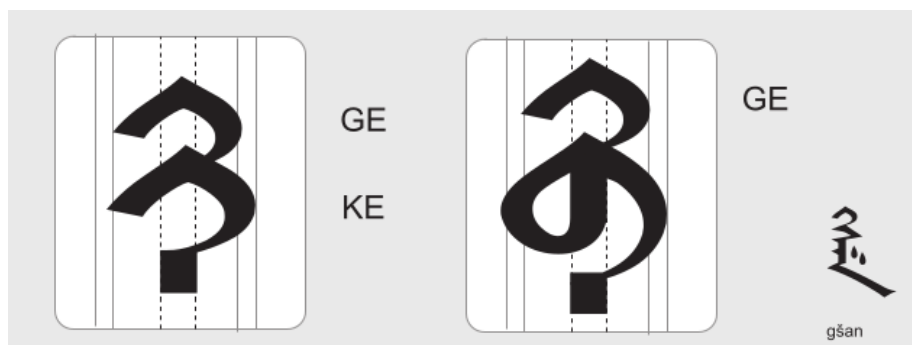


Figure 4: Differences between two letters

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


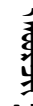
























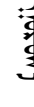
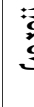


1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
														
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
														

Table 4: Example of alphabetical sorting

7 Figures

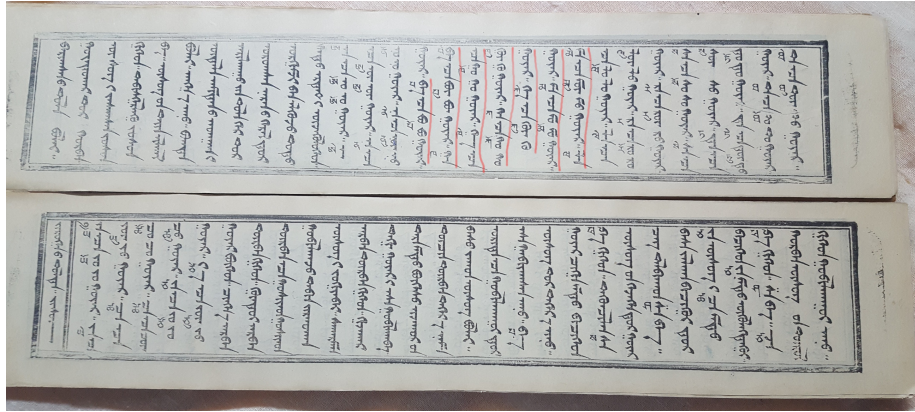


Figure 5: From the Jirüken-ü tolta-yin tayilburi [1]

МОНГОЛ БИЧГИЙН ЦАГААН ТОЛГОЙ

ᠮᠣᠩᠭᠣᠯᠪᠢᠴᠢᠭᠢᠨᠴᠠᠭᠠᠨᠲᠣᠯᠭᠣᠢ

Кирилл бичиг	Латин бичиг	ЧГИЙН					
		эхэн		гүнэ		агаа	
		бич- мөр	дөр- мөр	бич- мөр	дөр- мөр	бич- мөр	дөр- мөр
а	a	ᠠ	ᠡ	ᠢ	ᠣ	ᠤ	ᠤ
э	e	ᠡ	ᠢ	ᠣ	ᠤ	ᠤ	ᠤ
и	i	ᠢ	ᠣ	ᠣ	ᠣ	ᠣ	ᠣ
у	u	ᠣ	ᠣ	ᠣ	ᠣ	ᠣ	ᠣ
о	o	ᠣ	ᠣ	ᠣ	ᠣ	ᠣ	ᠣ
х	x	ᠬ	ᠬ	ᠬ	ᠬ	ᠬ	ᠬ
г	g	ᠭ	ᠭ	ᠭ	ᠭ	ᠭ	ᠭ
д	d	ᠳ	ᠳ	ᠳ	ᠳ	ᠳ	ᠳ
н	n	ᠨ	ᠨ	ᠨ	ᠨ	ᠨ	ᠨ
п	p	ᠯ	ᠯ	ᠯ	ᠯ	ᠯ	ᠯ
м	m	ᠮ	ᠮ	ᠮ	ᠮ	ᠮ	ᠮ
хэр	xe	ᠬᠡᠷ	ᠬᠡᠷ	ᠬᠡᠷ	ᠬᠡᠷ	ᠬᠡᠷ	ᠬᠡᠷ
хэм	xe	ᠬᠡᠮ	ᠬᠡᠮ	ᠬᠡᠮ	ᠬᠡᠮ	ᠬᠡᠮ	ᠬᠡᠮ
гэр	ge	ᠭᠡᠷ	ᠭᠡᠷ	ᠭᠡᠷ	ᠭᠡᠷ	ᠭᠡᠷ	ᠭᠡᠷ
гэм	ge	ᠭᠡᠮ	ᠭᠡᠮ	ᠭᠡᠮ	ᠭᠡᠮ	ᠭᠡᠮ	ᠭᠡᠮ

Кирилл бичиг	Латин бичиг	ЧГИЙН					
		эхэн		гүнэ		агаа	
		бич- мөр	дөр- мөр	бич- мөр	дөр- мөр	бич- мөр	дөр- мөр
ё	ё	ᠢ	ᠣ	ᠣ	ᠣ	ᠣ	ᠣ
нг	ng	ᠨᠭ	ᠨᠭ	ᠨᠭ	ᠨᠭ	ᠨᠭ	ᠨᠭ
т	t	ᠲ	ᠲ	ᠲ	ᠲ	ᠲ	ᠲ
г	d	ᠭ	ᠭ	ᠭ	ᠭ	ᠭ	ᠭ
р	r	ᠷ	ᠷ	ᠷ	ᠷ	ᠷ	ᠷ
ж	ʒ	ᠵ	ᠵ	ᠵ	ᠵ	ᠵ	ᠵ
й	y	ᠶ	ᠶ	ᠶ	ᠶ	ᠶ	ᠶ
с	s	ᠰ	ᠰ	ᠰ	ᠰ	ᠰ	ᠰ
ш	ʃ	ᠱ	ᠱ	ᠱ	ᠱ	ᠱ	ᠱ
ц	č	ᠴ	ᠴ	ᠴ	ᠴ	ᠴ	ᠴ
в	w	ᠪ	ᠪ	ᠪ	ᠪ	ᠪ	ᠪ

Чигсэн зурлагын нэр
 1 тугим 1-ийд, өргөс, аян
 2 гээс 1 шилд, урт ийд
 3 нум (суя) орчим,
 1 нуруу цайлаг

Дэвсгэр цсэг
 Чигийн гүнэ, агаат аргаа
 зөвшөөсг, чс битүүчж
 өдөөг гийдүүлээг чигийн
 дэвсгэр цсэг гэжэ.
 ᠠ ᠡ ᠢ ᠣ ᠤ ᠶ ᠷ ᠰ ᠱ ᠴ ᠵ
 ᠨ ᠨ ᠶ ᠨ ᠨ ᠶ ᠨ ᠨ ᠶ ᠨ
 ᠮ ᠮ ᠶ ᠮ ᠮ ᠶ ᠮ ᠮ ᠶ ᠮ

Хатгуу дэвсгэр гэрэг
 1 ᠠ ᠡ ᠢ ᠣ ᠤ ᠶ ᠷ ᠰ ᠱ ᠴ ᠵ
 ᠨ ᠨ ᠶ ᠨ ᠨ ᠶ ᠨ ᠨ ᠶ ᠨ
 ᠮ ᠮ ᠶ ᠮ ᠮ ᠶ ᠮ ᠮ ᠶ ᠮ

Зөвлөн дэвсгэр гэнэ.
 (1᠖), (1᠘) хоёр дэвсгэр
 цсэг нь ороогийн
 мангол хэлэнг згшиг
 болжээ. Хатгуу дэвсгэрийн
 «болгорсод» гэж
 тогтоогдун!

Хатгуу, зөвлөн дэвсгэрийн сайн мэдэх нь зөв бичихэд
 чухал холбоотой.

Figure 6: Alphabet table on [15]

Schrift und Aussprache.

c) Consonanten.

	n	b	ch	gh	k	g	m	l	r	t	d
Am Anfang	ᠨ	ᠪ	ᠴ	ᠭ	ᠬ	ᠭ	ᠮ	ᠯ	ᠷ	ᠲ	ᠳ
In der Mitte	ᠨᠠ	ᠪᠠ	ᠴᠠ	ᠭᠠ	ᠬᠠ	ᠭᠠ	ᠮᠠ	ᠯᠠ	ᠷᠠ	ᠲᠠ	ᠳᠠ
Am Ende	ᠨᠢ	ᠪᠢ			ᠬᠢ		ᠮᠢ	ᠯᠢ	ᠷᠢ	ᠲᠢ	ᠳᠢ
	(ᠰ)	s, ds	ts	ss	sch	w					
Am Anfang	ᠰ	ᠰ	ᠲᠰ	ᠰᠰ	ᠰᠢ	ᠰᠢ					
In der Mitte	ᠰᠠ	ᠰᠠ	ᠲᠰᠠ	ᠰᠰᠠ	ᠰᠢᠠ	ᠰᠢᠠ					
Am Ende					ᠰᠢᠢ						

§. 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.	a. A.		a. E.	i. d. M.	a. A.		a. E.	i. d. M.	a. A.
na	ᠨᠠ	ᠨᠢ	ᠨᠤ	ba	ᠪᠠ	ᠪᠢ	ᠪᠤ	cha	ᠴᠠ	ᠴᠢ	ᠴᠤ
ne	ᠨᠡ	ᠨᠢ	ᠨᠤ	be	ᠪᠡ	ᠪᠢ	ᠪᠤ	ke	ᠬᠡ	ᠬᠢ	ᠬᠤ
ni	ᠨᠢ	ᠨᠢ	ᠨᠤ	bi	ᠪᠢ	ᠪᠢ	ᠪᠤ	ki	ᠬᠢ	ᠬᠢ	ᠬᠤ

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	ᠡ	ᠢ ᠨ
2	e	ᠡ	ᠢ	ᠣ ᠨ
3	i	ᠢ	ᠣ	ᠤ
4	o u	ᠣ	ᠣ	ᠥ
5	ö ü	ᠣ	ᠣ ᠡ	ᠥ
6	n	ᠨ	ᠨ	ᠨ
7	ng		ᠨ	ᠨ
8	q	ᠠ	ᠠ	ᠠ
9	γ	ᠠ	ᠠ ᠠ	ᠠ ᠠ
10	b	ᠪ	ᠪ	ᠪ
11	p	ᠪ	ᠪ	
12	s	ᠰ	ᠰ	ᠰ ᠠ
13	š	ᠰ	ᠰ	ᠰ
14	t d	ᠲ	ᠲ ᠲ	ᠲ
15	l	ᠯ	ᠯ	ᠯ
16	m	ᠮ	ᠮ	ᠮ
17	č	ᠴ	ᠴ	
18	ʃ	ᠴ	ᠴ	
19	y	ᠶ	ᠶ	ᠶ
20	k q	ᠬ	ᠬ	ᠬ
21	k	ᠬ	ᠬ	ᠬ
22	r	ᠷ	ᠷ	ᠷ
23	v	ᠸ	ᠸ	
24	h	ᠬ	ᠬ	

Figure 8: Alphabet table on [10]

Монгол бичигийн цагаан толгой

No.	монгол үсэг			дуудалгын (tc.) болон үсгийн (tl.) галиг			
	эхин	дунд	адаг	(tc.)	(tl.) эхин	дунд	адаг
01	ᠠ	ᠡ	ᠢ	a e	'	'	/ (-) / \
02	ᠡ	ᠢ	ᠣ	i	'i	i	k
03	ᠢ	ᠣ	ᠤ	o u	'w	w	b
04	ᠣ	ᠣ	ᠤ	ö ü	'wi	w	b
05	ᠣ	ᠣ	ᠣ	n	n	n '	n- \
06		ᠣ	ᠣ	ng	-	'k	'i /
07	ᠣ	ᠣ	-	q	q	'	-
08	ᠣ	ᠣ	ᠣ	γ	γ	γ ''	γ ''
09	ᠣ	ᠣ	ᠣ	b	b	b	b w /
10	ᠣ	ᠣ	ᠣ	p	p	p	p
11	ᠣ	ᠣ	ᠣ	s	s	s	s '
12	ᠣ	ᠣ	-	š	š	š	-
13	ᠣ	ᠣ	ᠣ	d	t	d w'	d w \
14	ᠣ	ᠣ	-	t	t	d	-
15	ᠣ	ᠣ	ᠣ	l	l	l	l
16	ᠣ	ᠣ	ᠣ	m	m	m	m
17	ᠣ	ᠣ	-	č	č	č	-
18	ᠣ	ᠣ	-	j	i	j	-
19	ᠣ	ᠣ	ᠣ	y	i y	i y	k
20	ᠣ	ᠣ	-	k	k	k	-
21	ᠣ	ᠣ	ᠣ	g	k	k	i /
22	ᠣ	ᠣ	ᠣ	r	r	r	r
23	ᠣ	ᠣ	-	v (w)	v	v	-
24	ᠣ	ᠣ	-	h	'h	h	-

Figure 9: Alphabet table on [11]

Гийгүүлэгчийг тэмдэглэх нь													
	н	б	х(а)	х(э)	г(а)	(э)г г(э)	(а)г	ж	й	т	д	м	ч
Эхэнд	ᠨ	ᠪ	ᠬ	ᠬ	ᠭ	ᠭ		ᠵ	ᠶ	ᠲ	ᠳ	ᠮ	ᠴ
дунд	ᠨ, ᠨ	ᠪ	ᠬ	ᠬ	ᠭ	ᠭ	ᠬ	ᠵ	ᠶ	ᠲ	ᠳ, ᠳ	ᠮ	ᠴ
сүүлд	ᠨ	ᠪ	ᠬ		ᠭ	ᠭ	ᠬ	ᠵ	ᠶ	ᠲ	ᠳ, ᠳ	ᠮ	
жишээ	ᠨᠠᠮᠤ, ᠨᠠᠮᠤ, ᠨᠠᠮᠤ	ᠪᠠᠮᠤ, ᠪᠠᠮᠤ, ᠪᠠᠮᠤ	ᠬᠠᠮᠤ, ᠬᠠᠮᠤ, ᠬᠠᠮᠤ	ᠬᠠᠮᠤ, ᠬᠠᠮᠤ, ᠬᠠᠮᠤ	ᠭᠠᠮᠤ, ᠭᠠᠮᠤ, ᠭᠠᠮᠤ	ᠭᠠᠮᠤ, ᠭᠠᠮᠤ, ᠭᠠᠮᠤ	ᠬᠠᠮᠤ, ᠬᠠᠮᠤ, ᠬᠠᠮᠤ	ᠵᠠᠮᠤ, ᠵᠠᠮᠤ, ᠵᠠᠮᠤ	ᠶᠠᠮᠤ, ᠶᠠᠮᠤ, ᠶᠠᠮᠤ	ᠲᠠᠮᠤ, ᠲᠠᠮᠤ, ᠲᠠᠮᠤ	ᠳᠠᠮᠤ, ᠳᠠᠮᠤ, ᠳᠠᠮᠤ	ᠮᠠᠮᠤ, ᠮᠠᠮᠤ, ᠮᠠᠮᠤ	ᠴᠠᠮᠤ, ᠴᠠᠮᠤ, ᠴᠠᠮᠤ

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.	ᠠ	ᠠ	a	ᠠᠮᠠᠷᠠᠭᠠ	amarᠠᠭᠠ	father
2.	ᠡ	ᠡ	e	ᠡᠮᠡᠭᠡ	emeᠭe	bed
3.	ᠢ	ᠢ	i	ᠢᠵᠢ	ijil	easy
4.	ᠣ	ᠣ	o	ᠣᠷᠠᠭᠤ	orugu	bought
5.	ᠤ	ᠤ	u	ᠤᠮᠤᠭᠤ	umugu	similar to 'bone'
6.	ᠥ	ᠥ	ö	ᠥᠮᠢᠳᠤ	emidu	similar to 'girl'/'put'
7.	ᠦ	ᠦ	ü	ᠦᠨᠳᠢᠰᠢ	fndijsi	soup
8.	ᠨ	ᠨ	n	ᠨᠠᠮᠤᠰᠠᠭᠤ	nom	name
9.	ᠪ	ᠪ	b	ᠪᠠᠷᠠᠭᠤ	baragu	bird
10a	ᠬ	ᠬ	q	ᠬᠠᠪᠠᠷ	qabar	loch (Scottish)
10b	ᠬ	ᠬ	k	ᠬᠡᠯᠡᠭᠦ	keleku	non-existent in Engl.
11a	ᠭ	ᠭ	x	ᠭᠠᠯ	gal	get
11b	ᠭ	ᠭ	g	ᠭᠡᠷᠡᠯ	gerel	get
12.	ᠯ	ᠯ	l	ᠯᠠᠮᠤ	lama	lateral fricative "l" (non-existent in Engl.)
13.	ᠮ	ᠮ	m	ᠮᠣᠳᠤ	modu	my
14.	ᠰ	ᠰ	s	ᠰᠠᠮᠤᠷ	samur	song
15.	ᠱ	ᠱ	š	ᠱᠠᠲᠤ	šatu	shot
16.	ᠲ	ᠲ	t	ᠲᠣᠶᠤᠰᠤ	toyusu	tall
17.	ᠳ	ᠳ	d	ᠳᠡᠭᠡᠪᠦᠷᠢ	dgegburi	day
18.	ᠴ	ᠴ	č	ᠴᠢᠩᠭᠢᠰᠢ	činggis	church
19.	ᠵ	ᠵ	j	ᠵᠣᠭᠤ	jurj	joke
20.	ᠶ	ᠶ	y	ᠶᠠᠪᠤᠭᠤ	yabuqu	yacht
21.	ᠷ	ᠷ	r	ᠷᠠᠭᠤ	ragu	rolled "r" (non-existent in Engl.)
22.	ᠸ	ᠸ	v	ᠸᠠᠶᠠᠷ	xavar	xase

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

23.	ᠫ	ᠫ	p	ᠫᠠᠷᠠ	park	park
24.	ᠭ	ᠭ	f	ᠭᠠᠨᠠᠭᠤ	franci	fine
25.	ᠬ	ᠬ	k	ᠬᠢᠯᠠ	kilo	king
26.	ᠭ	ᠭ	e	ᠭᠡᠷᠠᠨ	german	egg
27.	ᠮ	ᠮ	c	ᠮᠠᠨᠵᠠᠨ	ganzan	its
28.	ᠮ	ᠮ	z	ᠮᠠᠨᠵᠠᠨ	gandan	sceds
29.	ᠮ	ᠮ	h	ᠮᠠᠨᠵᠠᠨ	himalai	house
30.	ᠮ	ᠮ	ž	ᠮᠠᠨᠵᠠᠨ	žornal	joke

Explanations:

- to 2) and 26): some dialects pronounce the "e" closed, some open. In Inner Mongolia the closed "e" is written with letter 2), the open "e" 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): this **closed "ö"** (front language) is pronounced quite differently in each dialect; the British 'girl' follows the western Mongolian dialect; American 'put' reflects Khalkh pronunciation.
- 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 "nich").
- to 12): this Mongolian "l" is **lateral fricative**, not like the English "l" (practice well!)
- to 18): this letter allows 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 "j", in Khalkh it could also be "ds".
- to 21): the Mongolian "r" must be rolled.

2.1.2 Polyphonic Alphabet

As we mentioned on page 3, in a polyphonic 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 monophonic 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 monophonic 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	ᠠ	ᠡ	ᠢ	a/e	"	"	'/(')+
2	ᠡ	ᠢ	ᠣ	i	'i	i	i
3	ᠢ	ᠣ	ᠤ	o/u	'w	w	w
4	ᠣ	ᠣ	ᠤ	ô/û	'wi	w	w
5	ᠣ	ᠣ	ᠣ	n	n	n	'/(')+
6	ᠣ	ᠣ	ᠣ	ng/ñ	-	nk	'i'
7	ᠣ	ᠣ	ᠣ	b	b	b	b w'
8	ᠣ	ᠣ	ᠣ	q	q	"	-
9	ᠣ	ᠣ	ᠣ	k	g	g	-
10	ᠣ	ᠣ	ᠣ	y	y	y	y "
11	ᠣ	ᠣ	ᠣ	g	g	g	i'
12	ᠣ	ᠣ	ᠣ	j	i	j	-
13	ᠣ	ᠣ	ᠣ	y	i y	y	i
14	ᠣ	ᠣ	ᠣ	t	t	d	-
15	ᠣ	ᠣ	ᠣ	d	t	d w'	d w'/(w)+
16	ᠣ	ᠣ	ᠣ	m	m	m	m
17	ᠣ	ᠣ	ᠣ	č	č	č	-
18	ᠣ	ᠣ	ᠣ	r	r	r	r
19	ᠣ	ᠣ	ᠣ	s	s	s	s 's
20	ᠣ	ᠣ	ᠣ	š	š	š	-
21	ᠣ	ᠣ	ᠣ	l	l	l	l
22	ᠣ	ᠣ	ᠣ	p	p	p	-
23	ᠣ	ᠣ	ᠣ	v	v	v	-
24	ᠣ	ᠣ	ᠣ	h	'h	h	-

+ In the cases of the absence of the printing possibility

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			Transcription	Transliteration		
	initial	medial	final		initial	medial	final
25	ᠠ	ᠠ	ᠠ	é	'é	é	e
26	ᠡ	ᠡ	ᠡ	f	f	f	f
27	ᠢ	ᠢ	ᠢ	k	k	k	k
28	ᠣ	ᠣ	ᠣ	z	z	z	z
29	ᠤ	ᠤ	ᠤ	c	c	c	c
30	ᠥ	ᠥ	ᠥ	z	z	z	z

A.3 Punctuation:

	Punctuational signs in the text of Mongolian script	Transcribing Signs	Significances
1.	egčim or biry-a, the sign at the beginning of books or chapters, at the beginning of the recto of each pages		
2.	čeg "the point"	.	(.)
3.	dabqur čeg "double point"	:	(.), (:), (.)
4.	dorbeljin čeg "the square point" -at the end of the chapters or 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 Saskiya 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 todadqarčai 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=zrpLAAAcAAJ>
- [10] N.N. Poppe. *Grammar of Written Mongolian*. Far Eastern and Russian Institute publications on Asia. Harrassowitz, 1954. LCCN: 56036254

[11] Ts. Shagdarsürüng. *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. *ᠣᠪᠤᠨᠠᠳᠢᠨ ᠶ᠋ᠠ. ᠮᠣᠩᠭᠣᠯᠢ ᠰᠣᠷᠦᠯᠢᠯᠤᠯᠤᠰ (Orthographic dictionary for Mongolian script)*. Ulaanbaatar, 1990

[14] Rita K., Tserenpil D. *Mongolian Grammar*. Hong Kong, 1996 ISBN: 9628457055

[15] Choimaa Sh. *ᠴᠣᠶᠢᠮᠠᠸ ᠰᠡ. ᠮᠣᠩᠭᠣᠯᠢ ᠪᠢᠴᠢᠭ (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-Badralf.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>

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest *Roadmaps*.

A. Administrative

1. Title:	Proposal to encode two Mongolian letters		
2. Requester's name:	Badral Sanlig, Munkh-Uchral Enkhtur, Jamiyansuren Togoobat		
3. Requester type (Member body/Liaison/Individual contribution):	Individual contribution		
4. Submission date:	2018-09-10		
5. Requester's reference (if applicable):			
6. Choose one of the following:			
This is a complete proposal:			Yes
(or) More information will be provided later:			

B. Technical – General

1. Choose one of the following:			
a. This proposal is for a new script (set of characters):			No
Proposed name of script:			
b. The proposal is for addition of character(s) to an existing block:			Yes
Name of the existing block:			Mongolian
2. Number of characters in proposal:			2
3. Proposed category (select one from below - see section 2.2 of P&P document):			
A-Contemporary	B.1-Specialized (small collection)	X	B.2-Specialized (large collection)
C-Major extinct	D-Attested extinct		E-Minor extinct
F-Archaic Hieroglyphic or Ideographic			G-Obscure or questionable usage symbols
4. Is a repertoire including character names provided?			Yes
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?			Yes
b. Are the character shapes attached in a legible form suitable for review?			Yes
5. Fonts related:			
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	Badral Sanlig		
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	MongolianScript, SIL licensed, open font		
6. References:			
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?			Yes
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?			Yes
7. Special encoding issues:			
Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?			Yes
See the "Note on collation" section in the proposal			

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database (<http://www.unicode.org/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	No
If YES explain	
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?	Yes
If YES, with whom? W3, MWG/2 group members	
If YES, available relevant documents:	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	No
Reference:	
4. The context of use for the proposed characters (type of use; common or rare)	Common
Reference:	
5. Are the proposed characters in current use by the user community?	No
If YES, where? Reference:	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	Yes
If YES, is a rationale provided?	
If YES, reference: See above proposal	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	Yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character?	Yes
If YES, is a rationale for its inclusion provided?	
If YES, reference: See above proposal	
11. Does the proposal include use of combining characters and/or use of composite sequences?	No
If YES, is a rationale for such use provided?	
If YES, reference:	
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	
If YES, reference:	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	Yes
If YES, describe in detail (include attachment if necessary)	
See above proposal	
13. Does the proposal contain any Ideographic compatibility characters?	No
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	