

SOME POSSIBLE IMPROVEMENT IDEAS TO THE CURRENT MONGOLIAN ENCODING

ANAND Orkhonbaatar

Contents

1. Unifying the identical letter which represent more than one sound
2. Eliminating historical and stylistic variations from the letters
3. Possible ways to write the suffixes without the usage of NNBS

1. Unifying identical letters

Unifying the identical letter which represent more than one sound

- The letters o and u (ᠣ and ᠤ)
- The letters ö and ü (ᠥ and ᠦ)
- The letters d and t (ᠳ and ᠲ)
- Separating the k and g from x and γ (ᠬ from ᠭ ;
ᠬ from ᠭ) then coding them as a single letter in one code point

Why we should unify them

- This letters pronunciations differ by the dialects
- When we adopted the script from the old Uyghur script we used many letters to represent more than one sound in Mongolian due to the limited numbers of the letters
- By not unifying them, we will have various database
- There are already some letters which encoded by single code point

Different pronunciation by different dialects.

- ᠪᠠᠭᠤᠨ (boy)
 - köwüün in western dialects;
 - höwgüün in khalkha dialect;
 - hübüün in buryat dialect;
- ᠪᠠᠭᠤᠨ (bridge)
 - hüürge in buryat dialect;
 - güür in khalkha dialect;
- ᠨᠠᠭᠤᠨ (part, percent)
 - howi in Inner Mongolian dialects
 - huwi in most parts of Mongolia
- ᠨᠠᠭᠤᠨ (lay down)
 - hewteh in some dialects
 - gewteh in some dialects
 - kewteh in some dialects
- ᠨᠠᠭᠤᠨ (stinky)
 - ömhii in some dialects
 - ümhii in some dialects
- ᠨᠠᠭᠤᠨ (pull out)
 - datah in some dialects
 - tatah in some dialects
- ᠨᠠᠭᠤᠨ (by the way)
 - dashiramd in some dialects
 - tashiramd in some dialects
- ᠨᠠᠭᠤᠨ (living adj.)
 - amid(u) in the pronunciation but the construction is **ami+tu**

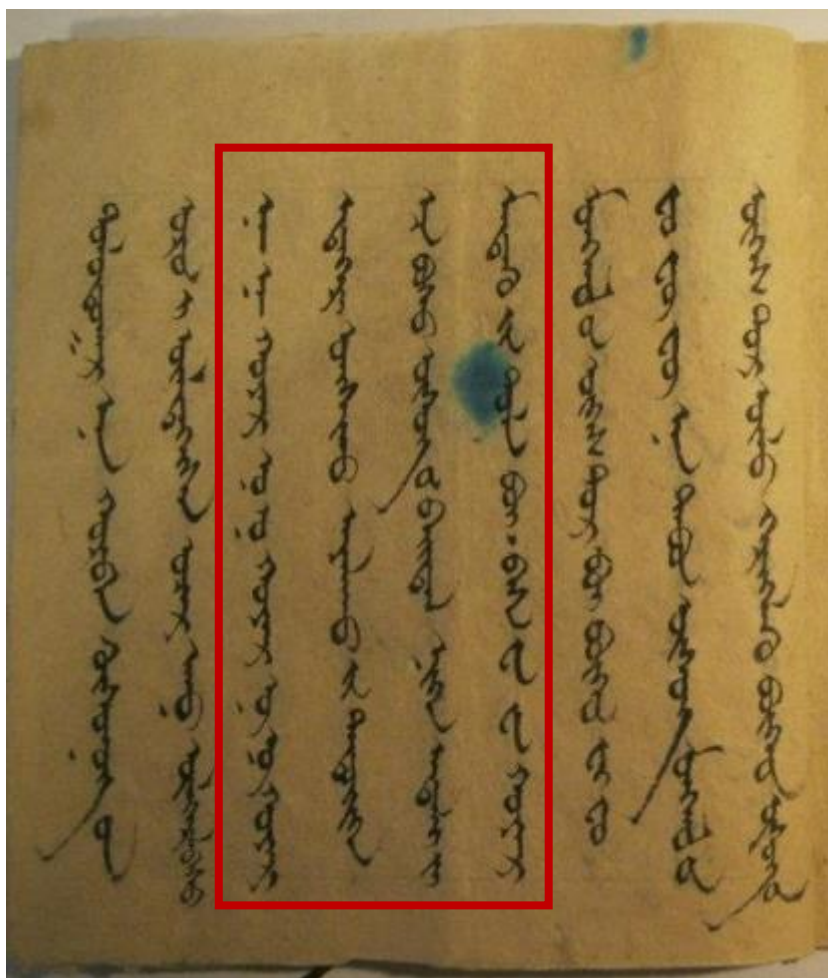
There are already many letters coded as a single letter, but pronounced by more than one sounds which are:

- **Რ** is pronounced either as **ts** and **tʃ**
- **Ს** is pronounced either as **dz** and **dʒ**
- **Ტ** is pronounced either as **b** and **w**

(mostly w with the exception of initial position)

- **Უ** is pronounced as **ɣ**, but in before consonant it's pronounced as **g**
- **Ფ** is pronounced as **ʃ** when it's followed by **Ქ** (i)

Documents



		१	२	३	४	५	६	७
१	१	२	३	४	५	६	७	८
	२	३	४	५	६	७	८	९
	३	४	५	६	७	८	९	१०

പ്ര	പ്രകൃതിയെ പ്പറുത്തുവെക്കുക - 4
പ്രകൃതി	പ്രകൃതിയെ പ്പറുത്തുവെക്കുക - 4
പ്രകൃതി	പ്രകൃതിയെ പ്പറുത്തുവെക്കുക - 4

Uyghur to Mongolian

NUN	ن	ن	ن	ن
SAMEKH	س	س	س	س
PE	پ	پ	پ	پ
SADHE	ه	ه	ه	ه
RESH	ر	ر	ر	ر
SHIN	ش	ش	ش	ش
TAW	ت	ت	ت	ت
LESH	ل	ل	ل	ل

	Dual-joining letters			
	iso	fin	med	init
ALEPH	ا	ا	ا	ا
BETH	ب	ب	ب	ب
GIMEL-HETH	ג	ג	ג	ג
WAW	ו	ו	ו	ו
ZAYIN	ז	ז	ז	ז
YODH	י	י	י	י
KAPH	כ	כ	כ	כ
LAMEDH	ל	ל	ל	ל
MEM	מ	מ	מ	מ

Uyghur to Mongolian

[illegible]

2. Eliminating historic and stylistic variations

Historic and Stylistic variations

- Historic forms that are no longer in use. For example:
 - ſ (s)
 - ʃ (e)
 - ʒ (i)
 - ʔ (o)
 - ƙ (k) and so on...
- Stylistic forms that could be shaped by font designs
 - - ʔ etc.

3. Possible ways to write
separated suffixes without NNBS

Using FVSx to write suffixes

- NNBS and SP would be the choice of the person who is using.
- In many documents, there are lines starting with suffixes.

Using ZWJ to write suffixes

- ZWJ is used to show the joining variation, just what exactly Mongolian separated suffixes nature is

Proposal which already made on the usage of FVSx

- <https://www.unicode.org/L2/L2018/18099-mwg2-1-mong-proposal.pdf>

Using ZWJ to write suffixes

q+o+t+a+n(2nd v.)
+SP/NNBSP+ZWJ+a(2nd v.)

ᠨᠣᠲᠠᠨ ᠠ

n+o+m+SP/NNBSP+ZWJ+u+n

ᠨᠣᠮᠤ ᠤ

n+a+r+a+SP/NNBSP+ZWJ+y(2nd v.)+i

ᠨᠠᠷᠠᠠ ᠠ

APPENDIX IV MONGOLIAN SUFFIXES AS CONNECTED BY NNBSP

VOCATIVE CASE

NNBSP+1820

᠋

Vocative Case

᠋᠋

NNBSP+1821

᠋

GENITIVE CASE

NNBSP+1836+1822+1828

᠋᠋

Genitive Case

᠋᠋᠋᠋᠋᠋

NNBSP+1824+1828

᠋᠋

NNBSP+1826+1828

᠋᠋

NNBSP+1824

᠋

NNBSP+1826

᠋

ACCUSATIVE CASE

NNBSP+1822

᠋

Accusative Case

᠋᠋

NNBSP+1836+1822

᠋᠋