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 NNBSP
1. Unifying identical letters
Unifying the identical letter which represent more than one sound

- The letters o and u (ᠶ and ᠥ)
- The letters ö and ü (ᠥ and ᠥ)
- The letters d and t (ᠤ and ᠤ)
- Separating the k and g from x and ɣ (ᠬ from ᠧ; ᠪ from ᠧ) 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.

- **ᠬᠦᠪᠡᠭᠦᠨ** (boy)
  - köwüün in western dialects;
  - höwgüün in khalkha dialect;
  - hübüün in buryat dialect;

- **ᠬᠡᠭᠦᠷᠭᠡ** (bridge)
  - hüürge in buryat dialect;
  - güür in khalkha dialect;

- **ᠬᠦᠪᠢ** (part, percent)
  - howi in Inner Mongolian dialects
  - huwi in most parts of Mongolia

- **ᠢᠯᠥᠭᠤᠷ** (lay down)
  - hewteh in some dialects
  - gewteh in some dialects
  - kewteh in some dialects

- **ᠭᠦᠭᠦᠷᠭᠡ** (stinky)
  - ömhii in some dialects
  - ümhii in some dialects

- **ᠤᠮᠥᠬᠡᠢ** (pull out)
  - datah in some dialects
  - tatah in some dialects

- **ᠲᠠᠲᠤᠬᠤ** (by the way)
  - dashiramd in some dialects
  - tashiramd in some dialects

- **ᠠᠮᠢᠳᠤ** (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:

• ṣ is pronounced either as ts and tʃ
• ḋ is pronounced either as dz and dʒ
• ḍ is pronounced either as b and w
  (mostly w with the exception of initial position)
• ḡ is pronounced as ɣ, but in before consonant it’s pronounced as g
• ḩ is pronounced as ʃ when it’s followed by ṣ (i)
Documents
Uyghur to Mongolian

<table>
<thead>
<tr>
<th>Letter</th>
<th>Uyghur</th>
<th>Mongolian</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUN</td>
<td>𝑁</td>
<td>🅱️</td>
</tr>
<tr>
<td>SAMEKH</td>
<td>🅱️</td>
<td>🅱️</td>
</tr>
<tr>
<td>PE</td>
<td>🅱️</td>
<td>🅱️</td>
</tr>
<tr>
<td>SADHE</td>
<td>🅱️</td>
<td>🅱️</td>
</tr>
<tr>
<td>RESH</td>
<td>🅱️</td>
<td>🅱️</td>
</tr>
<tr>
<td>SHIN</td>
<td>🅱️</td>
<td>🅱️</td>
</tr>
<tr>
<td>TAW</td>
<td>🅱️</td>
<td>🅱️</td>
</tr>
<tr>
<td>LESH</td>
<td>🅱️</td>
<td>🅱️</td>
</tr>
</tbody>
</table>

Dual-joining letters

<table>
<thead>
<tr>
<th>Letter</th>
<th>iso</th>
<th>fin</th>
<th>med</th>
<th>init</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALEPH</td>
<td>🆑</td>
<td>🆑</td>
<td></td>
<td>🆑</td>
</tr>
<tr>
<td>BETH</td>
<td>🆑</td>
<td>🆑</td>
<td>🆑</td>
<td>🆑</td>
</tr>
<tr>
<td>GIMEL-HETH</td>
<td>🆑</td>
<td>🆑</td>
<td></td>
<td>🆑</td>
</tr>
<tr>
<td>WAW</td>
<td>🆑</td>
<td>🆑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZAYIN</td>
<td></td>
<td></td>
<td>🆑</td>
<td></td>
</tr>
<tr>
<td>YODH</td>
<td></td>
<td></td>
<td>🆑</td>
<td></td>
</tr>
<tr>
<td>KAPH</td>
<td>🆑</td>
<td>🆑</td>
<td></td>
<td>🆑</td>
</tr>
<tr>
<td>LAMEDH</td>
<td>🆑</td>
<td>🆑</td>
<td>🆑</td>
<td>🆑</td>
</tr>
<tr>
<td>MEM</td>
<td>🆑</td>
<td></td>
<td>🆑</td>
<td>🆑</td>
</tr>
</tbody>
</table>
Uyghur to Mongolian
2. Eliminating historic and stylistic variations
Historic and Stylistic variations

• Historic forms that are no longer in use. For example:
  • ɂ (s)
  • Ǝ (e)
  • Ǝ (i)
  • Ǝ (o)
  • Ǝ (k) and so on...

• Stylistic forms that could be shaped by font designs
  • - Ǝ etc.
3. Possible ways to write separated suffixes without NNBSP
Using FVSx to write suffixes

• NNBSP 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

\[ q+o+t+a+n(2^{\text{nd}} \text{ v.}) \]
\[ +\text{SP/NNBSP}+\text{ZWJ}+a(2^{\text{nd}} \text{ v.}) \]

ᠬᠤᠲᠠᠨ‍ᠠ᠋

\[ n+o+m+\text{SP/NNBSP}+\text{ZWJ}+u+n \]

ᠨᠤᠮ ᠤᠨ

\[ n+a+r+a+\text{SP/NNBSP}+\text{ZWJ}+y(2^{\text{nd}} \text{ v.})+i \]

ᠨᠠᠷᠠᠶ᠋ᠢ

### APPENDIX IV
MONGOLIAN SUFFIXES AS CONNECTED BY NNBSP

<table>
<thead>
<tr>
<th>Vocative Case</th>
<th>Genitive Case</th>
<th>Accusative Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNBSP+1820</td>
<td>NNBSP+1836+1822+1828</td>
<td>NNBSP+1822</td>
</tr>
<tr>
<td>NNBSP+1821</td>
<td>NNBSP+1824+1828</td>
<td>NNBSP+1836+1822</td>
</tr>
<tr>
<td>NNBSP+1826</td>
<td>NNBSP+1826</td>
<td></td>
</tr>
<tr>
<td>NNBSP+1824</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NNBSP+1824</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>