

[Draft]

Mongolian text representation and shaping: Potential changes that may improve the Mongolian encoding model

To: Unicode Technical Committee
 From: Liang Hai / 梁海 <lianghai@gmail.com>
 Date: 24 September 2020

1 Merging the particle shaping responsibility of U+202F NNBSP into U+180E MVS

The opportunity of overloading MVS with NNBSP's responsibility is exhibited in the draft UTN's appendix *Generally recognized particles*. See rows Á (*·a/·e*) and A (*a/e*) in the appendix's chart.

2 Changes to basic letter representation

The following six potential changes are sorted from less disruptive to more.

1. Unify U+183A *KA* into U+183B *KHA*.
2. Decompose the following six A-prefixed variants (used in compound words), then unify the written unit A into U+1807 *SSBM*.
 - Variants *a.medi.AA*, *i.medi.AI*, *o/u.medi.AU*, and *ö/ü.medi.AUI* of U+1820 *A*, U+1821 *E*, U+1822 *I*, U+1823 *O*, U+1824 *U*, U+1825 *OE*, and U+1826 *UE*.
 - Then change the General_Category value Po (Other_Punctuation) of U+1807 *SSBM* to Lm (Modifier_Letter) or Lo (Other_Letter).

The two above have been already applied in the draft UTN.

3. Decompose U+1840 *LHA* into <182F *LA*, 183E *HAA*>.
4. Decompose U+1829 *ANG* into <1828 *NA*, 182D *GA*>.
 - U+182D *GA* is subjective to adjustment 6.
5. Unify U+1823 *O* into U+1824 *U*, and U+1825 *OE* into U+1826 *UE*.
6. Unify the following five G variants (often considered feminine) into U+1889 *ALI GALI KA*.
 - Variants *b.init/medi.G* and *g.init/medi/fina.G* of U+182C *QA* and U+182D *GA*.
7. ...

3 Additional variants

...

4 Other changes to property values

1. Change the General_Category value Po (Other_Punctuation) of U+180A *NIRUGU* to Lm (Modifier_Letter).
 - To be the same as U+0640 ARABIC TATWEEL.

* EOF *