Proposal to add
MONGOLIAN FREE VARIATION SELECTOR FOUR

To: Unicode Technical Committee
From: 梁海 Liang Hai <lianghai@gmail.com>
Date: 23 January 2020

1 Proposal

Encode MONGOLIAN FREE VARIATION SELECTOR FOUR at U+180F, with a representative glyph and character properties all consistent with those of the existing three Mongolian free variation selectors (U+180B..U+180D).

2 Rationale

This proposed character, MONGOLIAN FREE VARIATION SELECTOR FOUR (FVS4), is necessary because:

1. According to the existing Mongolian encoding model's principle of "phonetic letters are encoded as characters", certain written forms that only appear in pre-contemporary orthographies need to be represented as additional variants of phonetic letters (encoded characters).

    - For example: ᠊ *i*.ɪ.isol (i.e., isolated yodh form of U+1822) with a dot, ᠊ *u/ü*.ᴜ.isol (isolated waw form of U+1824/U+1826) with a dot, and ᠊ ᠊ *h/g*.ɢ.init/medi (i.e., initial/medial kaph form of U+182C/U+182D) with two dots.

    - For an introduction of the encoding model's principles and the notation scheme the author uses to describe variants, see <u>L2/19-368</u>, *Draft technical note: Text representation and shaping specification of the Mongolian script*.

2. Although it is controversial whether some of the earliest Mongolian texts should be considered and thus encoded as the Mongolian script or the Old Uyghur script (see <u>L2/20-003</u>, *Revised proposal to encode Old Uyghur*), the aforementioned variants apparently belong to the Mongolian script.

3. Encoding the aforementioned variants as characters separate from *i/u/ü/h/g* would create special cases for text representation, and thus would further complicate the analysis required for determining the representation.

4. Encoding combining mark for the dots is inconsistent with how similar written forms are encoded in the existing encoding model, and would significantly worsen the issue of confusables.

5. It is architecturally dangerous to use a general variation selector (VS, e.g., U+FE00 VARIATION SELECTOR-1) as it may introduce unexpected complications.

- The general variation selectors are intended for different purposes (in particular, largely optional) in Unicode text representation.

- Also, they are expected by the OpenType specification and de facto implementations to take effect at a different stage of shaping (VSes at cmap vs. FVSes at GSUB).

- Although using general VSes seem to be doable according to at least the Founder Type / 方正字库's preliminary tests for the additional variants proposed in L2/16-309 **(WG2 N4880)**, *Proposed additions for Mongolian in 5th edition of UCS.*

6. FVS4 would only be needed by pre-contemporary, historical texts, and therefore the standardization and implementation process (mostly, upgrading shaping engines to recognize this new character) ahead is relatively acceptable.

∗ EOF ∗