WG2 N5002

L2/18-213

To: WG2 and Unicode Technical Committee
Source: Deborah Anderson, SEI, UC Berkeley; Peter Constable, Microsoft
Title: Khitan Small Script Ad Hoc Report (London)
Date: 20 June 2018

Attendees included: Andrew West, Michael Everson, Wu Yingzhe, Deborah Anderson, Andrew West, Toshiya Suzuki, Wei Lin-Mei, Chen Zhuang, Peter Constable, Karljurgen Feuerherm, Kyongsok Kim, Yang Wangsung, and Richard Ishida.

There was a long discussion of the different models for Khitan Small Script:

> (1) the one proposed by Andrew West (et al.) in WG2 N4943,
> (2) the "Egyptian Hieroglyphic" model with a horizontal and vertical joiner,
> (3) a model with a "glue" character between each character in a cluster (with [2] and [3] both summarized in N4977), and
> (4) a new suggestion by Peter Constable and Andrew Glass which relied on a Tibetan-type model.

Option (2) was rejected as that model is designed to provide layout options (e.g., 3 elements on a row) that are not relevant for Khitan.

Options (3) and (4) would both provide an explicit representation in text that clustering/gluing behavior should occur, but in different ways. (3) would use a single format control between every two graphic characters within a cluster. For (4), each ideographic character in the repertoire would get a paired subjoined character. The subjoining characters would have character properties that would have gluing behavior, without the need to encode sequences for stacks with a gluing format control character between each pair. Criticisms of (4) were that binary searches for a particular Khitan character would return incomplete results if separate non-subjoined and subjoined characters were separately encoded.

One of the main concerns with model (1) expressed by Unicode experts was the editing behavior, and that the user community should be satisfied with it.

Andrew West demonstrated his model (as described in N4943), with three prototype fonts that illustrate three layout modes that are used: horizontal LTR linear (the preferred default), vertical with clustering, and horizontal LTR with vertical-stack clusters. (These are the three modes are used with Mongolian.)

Attendees agreed to the following points:

- The repertoire for the CD will remain as in the PDAM 2.2, with the exception of the two points below.
- Since the ITERATION MARK functions in cluster formation like other KSS characters, it will be retained in the KSS block, but its name will be changed to be algorithmic.
- The model agreed to is the one described in N4943, but with a script-specific format character, U+16FE4 KHITAN SMALL SCRIPT FILLER, instead of a COMBINING GRAPHEME JOINER. The FILLER character's general category properties and linebreak properties would be the same as for U+034F COMBINING GRAPHEME JOINER.

- The expected use of the FILLER is after the first element in a cluster. The proposal author will submit a document providing guidance on what happens when the KHITAN SMALL SCRIPT FILLER appears in a position that is *not* between first and second graphic characters in a cluster.
- Any new Khitan Small Script characters that are discovered should be documented and submitted to WG2 for possible inclusion. The current block has 42 open slots.
- Unicode character properties related to vertical layout (UAX #50) are essential to determine how KSS fonts should be implemented.