Comments on

COMMENTS ON L2/17-036 (MONGOLIAN SUFFIX CONNECTOR) by Liang Hai
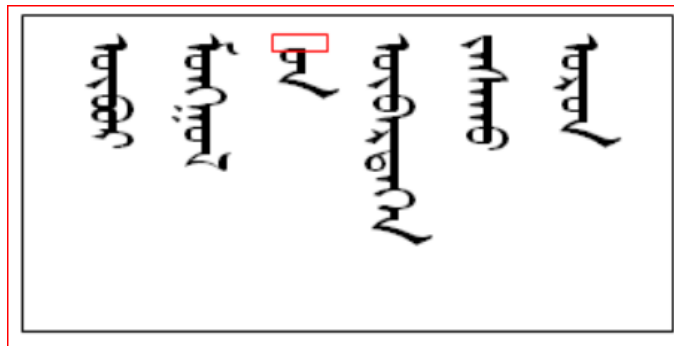
by Greg Eck

April 26, 2017

I would like to make a few comments on Liang Hai's paper L2/17-052.

1.) General controversies on the NNBSP issue

1.2 I would tone down the statement to read "*The space before suffixes is <u>sometimes</u> okay for line breaking*". It is very rare that a Mongolian suffixed word can be broken. In fact the only occasion that I know of is on a sign-board or a name-card. This is the case where you have very limited space and you want to space out the word such as …

I would invite others to give examples of other cases where it is considered permissible to use a line-break within a suffixed word.

1. I would add the following, "*But maybe a default procedure of no line-breaks should be adopted so that implementation methods are clear and simple.*" I suggest further that such line-breaking might be done in upper-level layout rather included in lower-level of encoding.

2. I would again restate this item to downplay the commonness of such a case, "*In printed materials there is one case where we see a suffix at the beginning of a word*" instead of "*it's common to see*". Again, **I would invite discussion leading to examples showing all of the cases where line-breaking is considered to be acceptable.**

3. This is really a matter of opinion. The late Dr. Dandy-Yadam Tserenpil, lists exactly one hundred suffixes, no more no less, in his book Mongolian Grammar. Of course, we know that as languages evolve, the suffix set evolves with it and this number will change. The point is however that in many parties' mind, myself included, the suffix set is very fixed and determined. The definition of a suffix is very clear. I attach my document DS00 – Mongolian Suffixes to layout this understanding.

4. .

5. Yes, agreed. This is possible with the ZWJ.
6. No, I disagree. The out-working of such an idea at this low level will be too difficult, especially at this stage of stabilizing the Mongolian Encoding. Let there be a default saying that the default is to NEVER break the suffixed word. The anomalous 1% or less cases where such is acceptable can be handled in upper-level layout.

STANDARD MONGOLIAN MSC MODEL

March 4, 2017   *(greyson@postone.net)*

STEM +++ MSC +++ SUFFIX

SuffixSet =

{ V | CV | VC |

CVC | VCV |

CVCV | VCVC |

CVCVC | VCVCV |

VCVCVC | CVCVCVC }

N

$I + X$

$I + M + X$

$I + M_1 ... M_n + X$

$I_{ligature}$

$I + X_{ligature}$

$I + M + X_{ligature}$

$I + M_1 ... M_n + X_{ligature}$

$+ \text{MSC} + \text{SuffixSet} \ [\ + \text{MSC} + \text{SuffixSet}\ ]^*$

NOTES
MSC = Mongolian Suffix Connector (traditionally implemented by U+202F)
N = a digit or solitary Mongolian character, I = Initial Mongolian character, M = Medial Mongolian character, X = Mongolian character (Medial or Final)
MS Universal Shaping Engine (USE) applies <fina> feature to X
MS Universal Shaping Engine (USE) applies <isol> feature to SuffixSet$_{first\_letter\_of\_one}$, applies <init> to SuffixSet$_{first\_letter\_of\_more\_than\_one}$
Harfbuzz – Same as MS – USE
Apple AAT – Same as MS – USE