**Proposal for New Character Encoding**

**Block: Mongolian (U+1800 – 18AF)**

**Name: Mongolian Suffix Connector (MSC)**

**General Category: Cf**

**Decomposition: None**

**Shaping Behavior: The main function of the MSC is to act as a connecting character which has no form, but connects the preceding character to the following character with a bit of space between. The MSC has no word boundary characteristics – the word will not break before or after it. Given a word count, the MSC is considered as another simple character – in other words there is not a word before and a word after it – it is an integral part of the word made up by the characters before and after. Given a word jump, the MSC is again considered as an integral part of the characters fore and aft. The MSC will never appear at the beginning of the line nor at the end of the line. It will never appear at the beginning of a word nor at the end of the word. The character preceding and the character following the MSC generally change their shape. In this sense, it is better considered to be a format character that contains some space rather than a space character.**

**Appendix**

Answering the questions below will provide basic information to allow the Unicode Technical Committee members to determine a character's properties. Provide a description of each character's use, with examples if possible.

A. Some scripts have *case*, and if so, it will be necessary to know:

* Is it uppercase, lowercase, or uncased? If uppercase or lowercase, what are the *case mappings*? (These mappings refer to a property that identifies the other element of a case pair, for example the uppercase mapping of "m" is "M".) Uppercase and titlecase characters must have lowercase mappings. Uncased
* Is it a titlecase digraph?, E.g. the Unicode character U+01F2 LATIN CAPITAL LETTER D WITH SMALL LETTER Z (which looks like "Dz") No
* Does it have complex or non-standard case mapping behavior? (e.g., Turkish dotless i) No

B. Is the character an ordinary letter of an alphabet or syllabary (non-CJK ideograph)? No Or is it a stand-alone symbol? (For CJK ideographs, see the special section above.) No

C. Is the character a white-space character, or does it cause visible separation between other characters? The character does cause visible separation between characters before and after (about 1/3 of a regular space) however it should not be considered a space character as it should never break the word. It should be considered a format character rather than a space character. The character should not be considered a white\_space character and there should be no decomposition down to a space character.

D. Does the character have a numeric value? If so, is it a decimal digit, or is it a "digit" of some other non-decimal numbering system? No

* If the character is a true decimal digit (i.e., it forms decimal radix numbers like European numbers), then the General\_Category value is Nd and all three numeric fields should have a numeric value filled in (for example, for CHAKMA DIGIT NINE, the General\_Category is Nd and 9 is inserted in the three numeric fields: 1113F;CHAKMA DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;) N/A
* If the character is any other kind of number, even if it has a numeric value from 1 through 9, then the General\_Category value is No (or Nl), and only the third numeric field should be filled in (for example, AEGEAN NUMBER NINE is not a decimal radix number, so it is No and 9 appears only in the third numeric field: 1010F;AEGEAN NUMBER NINE;No;0;L;;;;9;N;;;;;). N/A

E. Is it a "base letter" or does it combine with letters or symbols? No, it binds tightly the character before and the character after. Both the characters before and the characters after may change form in the presence of this character.

F. If it is a combining character: No

* How does it combine? Above? Below? After? Are there particularly strong restrictions on how it is displayed, such as being centered or to the left/right of a base character? N/A
* Does it bind very tightly to letters, such as some vowel signs do? N/A
* Is it completely non-spacing, or does it combine but also have spacing characteristics? N/A

G. If this is a punctuation character: No

* Is it terminal punctuation (i.e., ending a clause or a sentence)? No
* Is it paired with anything else? For example, "(" is paired with ")",  "[" is paired with "]". No
* Does it separate words? If so, does it occur exclusively before or after words? Never
* Does it occur within words? Yes
* Does it occur within (as opposed to at the end of) sentences? Only between the characters of a word
* Can it appear at the end of a line? Beginning of a line? Never. Never
* Does it come between letters and cause them to *not* be breakable at the end of a line? It will come between letters but it will never break the sequence of letters before nor aft. In this sense, the character does make the sequence unbreakable at the end of a line.