L2/18-099

The Proposal for deprecation of MSC/NNBSP Mongolian Suffix Form Controlling Behavior

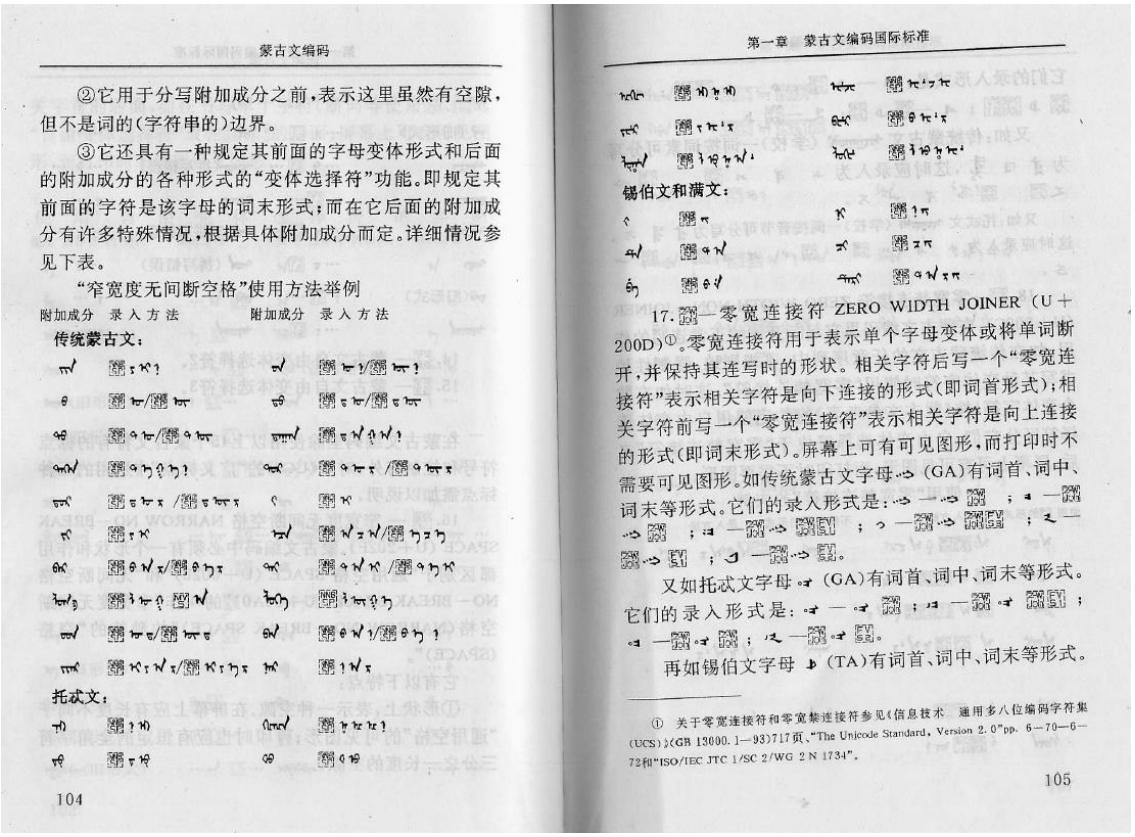
Jirimutu – jrmt@almas.co.jp
Siqin – siqin@almas.co.jp
Bao Haishan – baohaishan@delehi.com
Burigudu – burgood@vip.sina.com
Menghejiya – csmhjy@126.com

1. Background

As we know, the introduction of Mongolian Suffix Connector in Unicode Mongolian Encoding proposal is coming after 1998. As Professor Quejingzhabu had declared on Hohhot ad hoc meeting, he had proposed new code point for Mongolian Suffix Connecter (Maybe it was named as Mongolian Space in his proposal in early time), but WG2 group alternatively selected NNBSP as the Mongolian Suffix Connector in the standard.

In the publication work of Professor Quejingzhabu in 2002, the "Meng Gu Wen Bian Ma (The Mongolian Encoding)" book, we can find his introduction of the NNBSP Mongolian Suffix Connecter model usage example as Picture 1-1.

We had also found exactly same list in TR170 document on 11th page written by Myatav Erdenechimeg, Richard Moore and Yumbayar Namsrai. Please find the document from Internet (http://www.unicode.org/~asmus/mongolian/MD001-unu-tr170.html).



蒙古文编码

②它用于分写附加成分之前，表示这里虽然有空隙，但不是词的(字符串的)边界。

③它还具有一种规定其前面的字母变体形式和后面的附加成分的各种形式的“变体选择符”功能。即规定其前面的字符是该字母的词末形式；而在它后面的附加成分有许多特殊情况，根据具体附加成分而定。详细情况参见下表。

“窄宽度无间断空格”使用方法举例

附加成分	录入方法	附加成分	录入方法
传统蒙古文：			
托忒文：			

104

第一章 蒙古文编码国际标准

锡伯文和满文：			

17. —零宽连接符 ZERO WIDTH JOINER (U+200D)①。零宽连接符用于表示单个字母变体或将单词断开，并保持其连写时的形状。相关字符后写一个“零宽连接符”表示相关字符是向下连接的形式(即词首形式)；相关字符前写一个“零宽连接符”表示相关字符是向上连接的形式(即词末形式)。屏幕上可有可见图形，而打印时不需要可见图形。如传统蒙古文字母 (GA)有词首、词中、词末等形式。它们的录入形式是： ； ； ； ； 。

又如托忒文字母 (GA)有词首、词中、词末等形式。它们的录入形式是： ； ； ； 。

再如锡伯文字母 (TA)有词首、词中、词末等形式。

① 关于零宽连接符和零宽禁连接符参见《信息技术 通用多八位编码字符集(UCS)》(GB 13000.1—93)717页、“The Unicode Standard, Version 2.0”pp. 6—70—6—72和“ISO/IEC JTC 1/SC 2/WG 2 N 1734”。

105

Picture 1-1

We had also found similar list of NNBSP Mongolian Suffix Connector usage in GB/T 26226-2010 as Picture 1-2. In this list, there were several additional suffixes to “Meng Gu Wen Bian Ma(The Mongolian Encoding)”.

o) [FVS3]——蒙古文自由变体选择符 3。

在蒙古文编码里除使用以上蒙古文特有的标点符号和控制字符外，还使用多个通用标点，见附录 E。另有从 GB 13000—2010 的“广义标点”中引用的 3 种标点，说明如下：

a) [NNBSP]——窄无间断空格 NARROW NO-BREAK SPACE (U+202F)。其形状和作用区别于“通用空格 SPACE (U+0020)”和“无间断空格 NO-BREAK SPACE (U+00A0)”，具有以下特点：

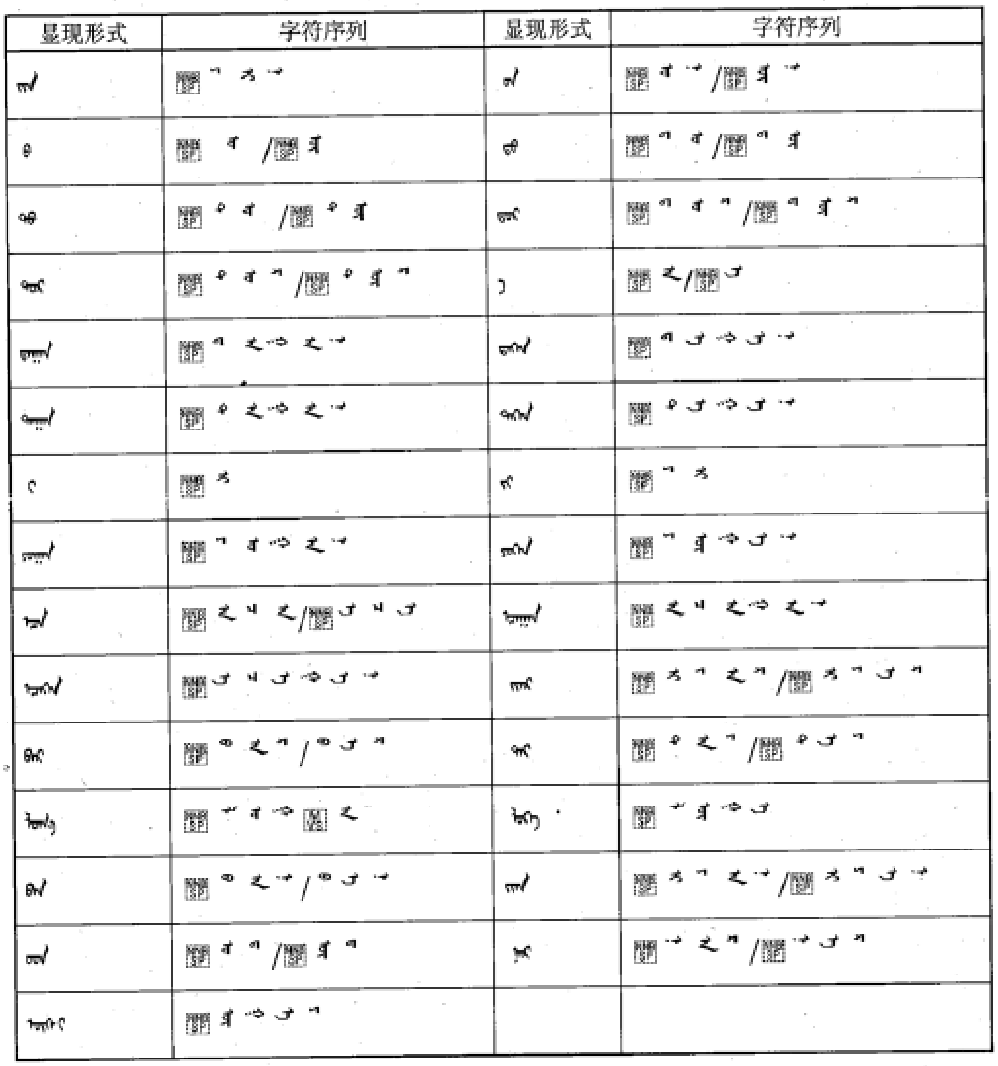
1) 其高度是恒定的，为全角字符的三分之一；

2) 用于分写附加成分之前，但不是词的边界；

3) 规定其前后字母的变体形式。

传统蒙古文窄无间断空格的使用方法见表 10。

表 10 传统蒙古文窄无间断空格的使用方法



Picture 1-2

After 2007, it is possible to use OpenType technology to implement the Mongolian Unicode Encoding in font/system level. Microsoft cooperation provided Unicode Mongolian font on Windows Vista system firstly. After several years' efforts, many local vendors have also been providing many kinds of Mongolian fonts now. Although all of these fonts have slightly differences between each other on implementation details and we have been all struggling on the Mongolian Unicode Encoding stabilization efforts for several years as you know.

In the Mongolian Unicode encoding, we find one significant issue exists on MSC/NNBSP usage now. The solution of the MSC/NNBSP for Mongolian suffix Connecting usage will depends on the Font Rendering Engine. But when we use NNBSP for Mongolian Suffix Connecting (actually, it is controlling the following Mongolian Suffix as well), we are facing at Mongolian Suffixes all break on current existing font rendering engine, such as Harfbuzz and Uniscribe. On the Microsoft Uniscribe font rendering engine, although it is rendering on the most recent version in the regular case, but on the some special cases, we found that there are no proper ways to get the all Mongolian Suffix correctly formed. We had also failed on Microsoft Office 2013 and 2016 for one or two years even on the regular case as well. The most significant impact is that the Mongolian Suffix all fails on Mobile Phone system like Android and iOS as well as PC Operating system like Mac OS X and Linux now.

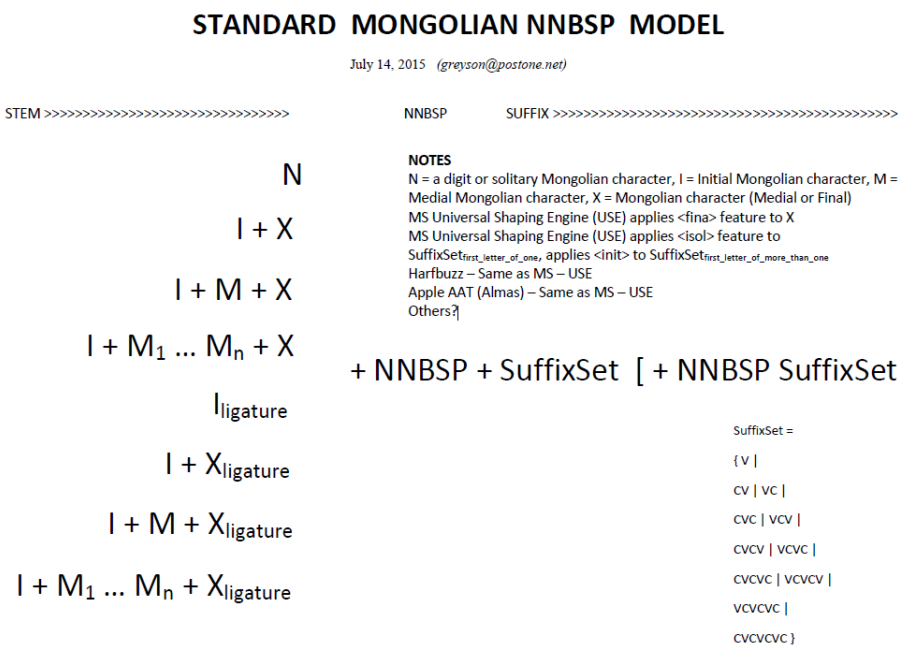
2. Introduction of current solution

We did find the NNBSP Mongolian Suffix connecting model problem several years ago and have been discussing to improve the NNBSP Mongolian Suffix connector usage in Mongolian Unicode Encoding for several years.

We had discussed NNBSP Mongolian Suffix connector model and it's improving solution in W3C forum for a long time during November of 2014 to December of 2015. The Model of the MSC/NNBSP Mongolian suffix model is defined as Picture 2-1 in the forum discussion.

The Mongolian Suffix connecting problem of current solution is actually on using NNBSP as a Mongolian Suffix Connector. It is because that the NNBSP properties could not match all of Mongolian Suffix Connecting and Controlling requirements, For this reason, we had proposed one more code point U+180F as the MSC (Mongolian Suffix Connector) in 2016. But the proposal has not been approved by UTC and WG2 yet.

We have listed all of the Mongolian Suffixes in the MSC model solution as Appendix I.



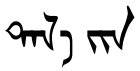
Picture 2-1

3. Existing issue on the MSC/NNBSP model solution

1) The issue occurring when using NNBSP as the MSC(Mongolian Suffix Connector)

The following is three examples of Mongolian Suffix.

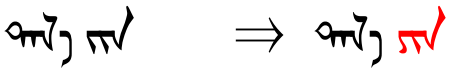
〈U+1832〉〈U+1820〉〈U+182F〉〈U+180E〉〈U+1820〉〈U+202F〉〈U+1836〉〈U+1822〉〈U+1828〉

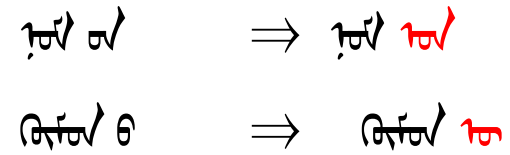


〈U+1828〉〈U+1823〉〈U+182E〉〈U+202F〉〈U+1824〉〈U+1828〉

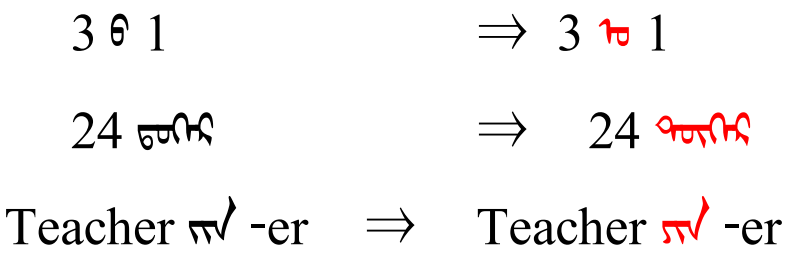
〈U+182C〉〈U+1826〉〈U+182E〉〈U+1826〉〈U+1828〉〈U+202F〉〈U+1826〉

On the current font rendering engine like Harfbuzz, the font will be displayed with the wrong presentation form like bellow. The Mobile phone system Android, iOS and Mac OS X, Linux system all experiencing this issue now, Only some version of Microsoft Uniscribe correctly rendering the Mongolian Suffixes like listed above





2) The issue occurring when the Mongolian Suffix is led by digit, punctuation or other language characters in the NNBSP as MSC. We cannot get correct Mongolian Suffix presentation form on this case on most of the current existing system. For example：



3) The issue existing when we copy and paste Mongolian text from PDF document, the NNBSP will be lost in the pasting text.

4) As a solution for these issues, we had proposed one more new Unicode code point on U+180F. But the proposal still have not been accepted or approved by neither UTC meeting nor WG2 meeting till now. We have to wait for a long time to solve this issue in the Unicode Mongolian Encoding standard.

5) The issue exists on the MSC model itself. Even if we define new MSC code point on U+180F, we will face at following issues continuously as well.

   A) The Mongolian Suffix rule is restricted in the predefined Suffix list. It is difficult for us to increase or extend the Suffixes list in the future. But the Mongolian Suffix list is still under arguing that which is needed to be included or excluded in Mongolian Suffix list The Mongolian Suffix list maybe will change according to the grammar definition or linguistic progress.

   B) From the encoding perspective, we find that the MSC (or NNBSP) Mongolian Suffix Form Control Model is not a proper solution. According to our several years' experience, we can say that the MSC Mongolian Suffix Form Control Model solution is an unstable, inadequate and problematic solution. We need to deprecate the Suffix Form Controlling behavior of MSC (or NNBSP) and just use it as a connector. The **Mongolian Suffix Form followed after MSC/NNBSP** should be explicitly defined by FVSs.

   C) In current solution, we cannot display or print most of Mongolian Suffix

without MSC. But in the real world, it is necessary to display or print Mongolian Suffix individually.

D) According to some experts declared on the Hohhot Mongolian Ad Hoc meeting, the publishing houses and news agencies are selecting the Mongolian Suffix Non-Breakable Line Mode and Breakable Line Mode on their preference. If the Mongolian Suffix Breakable mode is common in the world, we do not need to define the Mongolian Suffix Model so restricted and inflexible way. As a Mongolian script user, I know it is true we are using both Mongolian Suffix Non-Breakable and Breakable written style in daily life.

6) This issue is the one biggest issue which had been blocking us to use Unicode Mongolian on mobile phone system for several years. In Inner Mongolia, we have started to use following work around solution for this issue and have been utilizing Unicode Mongolian on mobile phone widely now.

4. Best Practices of work around solution

Since the summer of 2017, we have started using one work around solution for current MSC Mongolian Suffix Connecting model which is only use NNBSP as a Narrow-Non-Break-Space like other language in our mobile phone applications and Mac OS X system applications as well as on the Linux system. After half year utilization, we feel it is the most proper solution for Mongolian Suffix.

The work around solution is like below.

1) Use NNBSP as the Mongolian Suffix Connector as usual. The NNBSP still act like our current MSC/NNBSP Mongolian Suffix model. Even if it is getting error on some Suffix, we will just ignore the error on this stage in our work around solution for remaining the compatibility with the previous solution.
2) Additionally, we defined the irregular written form of some Mongolian Character presentation form in Mongolian Suffix as a new form (or use certain form if exists) for specified character and use explicit FVS1-3 to shape the form in the Suffix.
3) Modified all of our fonts to support this logic. (Actually Almas font include this logic from the beginning, The Ling Jinbao's encoding rule for Inner Mongolia Autonomous Region Ethnic Affairs Committee's project included this rule as well).
4) We have modified all of our IME (Input Method Editor) tools on different system and provided new work around solution to end users for half year.

According to the end users' feedback, this solution is acceptable anyway.

5) Currently, we are widely using this work around solution to get over the MSC/NNBSP Mongolian Suffix Controlling/Connecting big issue and utilizing Mongolian Unicode in our daily life now.
6) We should learn intelligence from this kind of best practice to improve the Mongolian Suffix Encoding solution to break through the big issue which is restricting Mongolian Unicode fail on most of the operating system as well as the other popular application.

5. The Proposal for Mongolian Suffixes

According to the best practice mentioned above, we would like to prepare following proposal for Mongolian Suffix Encoding Model to discuss on the UTC Mongolian ad hoc meeting.

1) Change current MSC/NNBSP Mongolian Suffix Encoding model.
2) Continue to use NNBSP as a Narrow-Non-Break-Space in Mongolian. The property and behavior of NNBSP will be exactly same with in other language.
3) NNBSP can be used for Mongolian as a non-breakable Mongolian Suffix Connector.
4) The Mongolian Suffix will be encoded explicitly using FVS1-3 to select the displaying character form.
5) The Mongolian Suffix displaying form will not depend on the leading NNBSP, but it will depend on normal word font rendering rule. If use explicit FVS1-3, the display form of the character will be specified to certain defined form.
6) The Mongolian Suffix can be used as Line Breakable Writing Mode. In the Mongolian Suffix Line Breakable Writing Mode, we will use normal space (not NNBSP) for leading the Mongolian Suffix. Even without NNBSP, the Mongolian Suffix will select correct form in any case.
7) In another word, this proposal can be called as the proposal to depreciate Mongolian Suffix Controlling behavior of MSC/NNBSP.

Please refer the Appendix II to check the Detailed Mongolian Suffix encoding list. Here we have only listed currently existing Mongolian Suffixes. According to the encoding rule, we can assume any word as a Mongolian Suffix and write it in the line with Non-Breakable Mode.

6. Evaluation and Conclusion

Advantage of the proposal

1) It can be utilize soon after approval.
2) No need to wait another 2-or 3 years for new code point approval.
3) It is Mongolian language internal improving solution, no impact to other language,
4) No need to make any changes on font rendering engine.
5) It will correctly work on current popular PC and Mobile operating system
6) It has similarity with previous solution and no need to educate end user again.
7) Limited impact for both Vendor and End user level.
8) The Mongolian Suffix encoding is based on normal rule.
9) The Mongolian Suffix encoding rule is simple and easy to share, communicate and standardize.
10) The solution is flexible, extendable, and stable.

Disadvantage of the proposal

1) Vendor need to change font logic for Mongolian Suffix Connecting model. Anyway vendors need to do this for improving MSC model.
2) Vendor need to modify IME tools to support new suffix code.
   Anyway vendors need to modify their IME if we defined new MSC code point.
3) The existing text and corpus will get incorrect form. Need to provide auto Mongolian Suffix auto correction tools to simplify the correction.
   If we define new MSC code point, the existing text and corpus correction is also necessary. Also need to auto exchange NNBSP in the text with MSC.

Appendix I
The Encoding List of Mongolian Suffixes
Under Current MSC/NNBSP model

Vocative Case

᠋ ᠋

VOCATIVE CASE

NNBSP+1820 ᠋

NNBSP+1821 ᠋

Genetive Case

ᠤ ᠤ ᠤ ᠠ ᠠ

GENITIVE CASE

NNBSP+1836+1822+1828 ᠤ

NNBSP+1824+1828 ᠤ

NNBSP+1826+1828 ᠤ

NNBSP+1824 ᠠ

NNBSP+1826 ᠠ

Accusative Case

ᠢ ᠨ

ACCUSATIVE CASE

NNBSP+1822 ᠢ

NNBSP+1836+1822 ᠨ

Dative-Locative Case

ᠳᠤ ᠳᠦ ᠲᠤ ᠲᠦ ᠳᠤᠷ ᠳᠦᠷ ᠲᠤᠷ ᠲᠦᠷ ᠳᠠᠬᠢ ᠳᠡᠬᠢ ᠠ ᠡ

DATIVE-LOCATIVE CASE

NNBSP+1833+1824	ᠳᠤ
NNBSP+1833+1826	ᠳᠦ
NNBSP+1832+1824	ᠲᠤ
NNBSP+1832+1826	ᠲᠦ
NNBSP+1833+1824+1837	ᠳᠤᠷ
NNBSP+1833+1826+1837	ᠳᠦᠷ
NNBSP+1832+1824+1837	ᠲᠤᠷ
NNBSP+1832+1826+1837	ᠲᠦᠷ
NNBSP+1833+1820+182C+1822	ᠳᠠᠬᠢ
NNBSP+1833+1821+182C+1822	ᠳᠡᠬᠢ
NNBSP+1820	ᠠ
NNBSP+1821	ᠡ

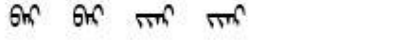
Ablative Case

ᠠᠴᠠ ᠡᠴᠡ

ABLATIVE CASE

NNBSP+1820+1834+1820	ᠠᠴᠠ
NNBSP+1821+1834+1821	ᠡᠴᠡ

Instrumental Case



INSTRUMENTAL CASE

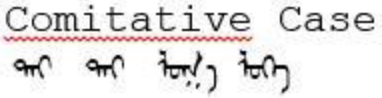
NNBSP+182A+1820+1837

NNBSP+182A+1821+1837

NNBSP+1822+1836+1820+1837 

NNBSP+1822+1836+1821+1837 

Comitative Case



COMITATIVE CASE

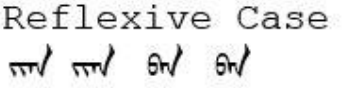
NNBSP+1832+1820+1822

NNBSP+1832+1821+1822

NNBSP+182F+1824+182D+180E+1820 

NNBSP+182F+1826+182D+1821 

Reflexive Case



REFLEXIVE CASE

NNBSP+1822+1836+1820+1828 

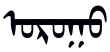
NNBSP+1822+1836+1821+1828 

NNBSP+182A+1820+1828 

NNBSP+182A+1821+1828 

Directive

DIRECTIVE CASE (may or may not use NNBSP)

NNBSP+1824+1837+1824+182D+1824 

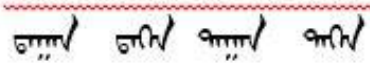
Reflexive+Accusative

REFLEXIVE + ACCUSATIVE CASE

NNBSP+1836+1824+182D+1820+1828 

NNBSP+1836+1826+182D+1821+1828 

Reflexive+Dative-Locative



REFLEXIVE+DATIVE-LOCATIVE CASE

NNBSP+1833+1820+182D+1820+1828 

NNBSP+1833+1821+182D+1821+1828 

NNBSP+1832+1820+182D+1820+1828 

NNBSP+1832+1821+182D+1821+1828 

Reflexive+Ablative

REFLEXIVE+ABLATIVE CASE

NNBSP+1820+1834+1820+182D+1820+1828 

NNBSP+1821+1834+1821+182D+1821+1828 

Reflexive+Comitative

REFLEXIVE+COMITATIVE CASE

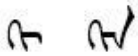
NNBSP+1832+1820+1836+1822+182D+1820+1828 

NNBSP+1832+1821+1836+1822+182D+1821+1828 

NNBSP+1832+1820+1822+182D+1820+1828 

NNBSP+1832+1821+1822+182D+1821+1828 

CASE-BOUND POSSESSION

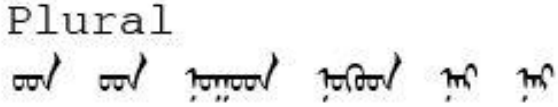


CASE-BOUND POSSESSIVE CASE

NNBSP+182C+1822 

NNBSP+182C+1822+1828 

Plural



PLURAL (first form may connect directly also)

NNBSP+1824+1833 

NNBSP+1826+1833 

NNBSP+1828+1824+182D+1824+1833 

NNBSP+1828+1826+182D+1826+1833 

NNBSP+1828+1820+1837 

NNBSP+1828+1821+1837 

Negation

NEGATION (may or may not use NNBSP)

NNBSP+1826+182D+1821+1822

Ordinal

ORDINAL

NNBSP+1833+1824+182D+1820+1837

NNBSP+1833+1826+182D+1821+1837

NNBSP+1833+1820+182C+1822

NNBSP+1833+1821+182C+1822

Regular

REGULAR ACTION (may or may not use NNBSP)

NNBSP+1833+1820+182D

NNBSP+1833+1821+182D

Appendix II
The Encoding List of Mongolian Suffixes
Using New Proposed Solution

Vocative Case

᠋ ᠋

VOCATIVE CASE

1820+FVS2 ᠋

1821+FVS1 ᠋

Genetive Case

᠌ ᠌ ᠌ ᠊ ᠊

GENITIVE CASE

1836+FVS1+1822+1828 ᠌

1824+FVS1+1828 ᠌

1826+FVS1+1828 ᠌

1824+FVS1 ᠊

1826+FVS2 ᠊

Accusative Case

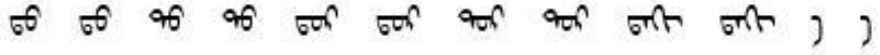
ᠴ ᠵ

ACCUSATIVE CASE

1822+FVS1 ᠴ

1836+FVS1+1822 ᠵ

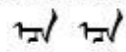
Dative-Locative Case



DATIVE-LOCATIVE CASE

1833+FVS1+1824	
1833+FVS1+1826	
1832+1824	
1832+1826	
1833+FVS1+1824+1837	
1833+FVS1+1826+FVS3+1837	
1832+1824+1837	
1832+1826+FVS3+1837	
1833+FVS1+1820+182C+1822	
1833+1821+182C+1822	
1820+FVS1	
1821+FVS1	

Ablative Case



ABLATIVE CASE

1820+FVS1+1834+1820	
1821+1834+1821	

Instrumental Case

INSTRUMENTAL CASE

182A+1820+1837 

182A+1821+1837 

1822+FVS1+1836+FVS1+1820+1837 

1822+FVS1+1836+FVS1+1821+1837 

Comitative Case

COMITATIVE CASE

1832+1820+1822 

1832+1821+1822 

182F+1824+182D+180E+1820 

182F +1826+FVS3+182D+1821 

Reflexive Case

REFLEXIVE CASE

1822+FVS1+1836+FVS1+1820+1828 

1822+FVS1+1836+FVS1+1821+1828 

182A+1820+1828 

182A+1821+1828 

Directive

DIRECTIVE CASE (may or may not use NNBSP)

1824+1837+1824+182D+1824 

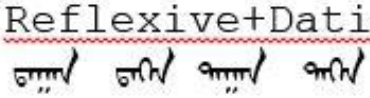
Reflexive+Accusative

REFLEXIVE + ACCUSATIVE CASE

1836+1824+182D+1820+1828 

1836+1826+FVS3+182D+1821+1828 

Reflexive+Dative-Locative



REFLEXIVE+DATIVE-LOCATIVE CASE

1833+FVS1+1820+182D+1820+1828 

1833+FVS1+1821+182D+1821+1828 

1832+1820+182D+1820+1828 

1832+1821+182D+1821+1828 

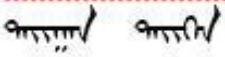
Reflexive+Ablative

REFLEXIVE+ABLATIVE CASE

1820+FVS1+1834+1820+182D+1820+1828 

1821+1834+1821+182D+1821+1828 

Reflexive+Comitative



REFLEXIVE+COMITATIVE CASE

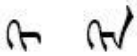
1832+1820+1836+FVS1+1822+182D+1820+1828 

1832+1821+1836+FVS1+1822+182D+1821+1828 

1832+1820+1822+182D+1820+1828 

1832+1821+1822+182D+1821+1828 

CASE-BOUND POSSESSION

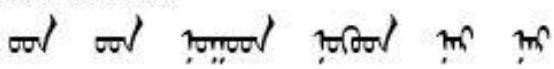


CASE-BOUND POSSESSIVE CASE

182C+1822 

182C+1822+1828 

Plural



PLURAL (first form may connect directly also)

1824+FVS1+1833 

1826+FVS1+1833 

1828+1824+182D+1824+1833 

1828+1826+FVS3+182D+1826+1833 

1828+1820+1837 

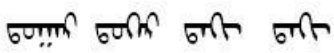
1828+1821+1837 

Negation

NEGATION (may or may not use NNBSP)

1826+182D+1821+1822 

Ordinal



ORDINAL

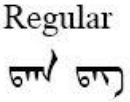
1833+FVS1+1824+182D+1820+1837 

1833+FVS1+1826+FVS3+182D+1821+1837 

1833+FVS1+1820+182C+1822 

1833+FVS1+1821+182C+1822 

Regular



REGULAR ACTION (may or may not use NNBSP)

1833+FVS1+1820+182D 

1833+FVS1+1821+182D 

OTHERS (may or may not use NNBSP)

1833+FVS1+1820 

1833+FVS1+1821 

1833+FVS1+1824+1828+1822 

1833+FVS1+1826+FVS3+1828+1822 

1833+ 1824+1828+1822 

1833+ 1826+FVS3+1828+1822 

有关废除蒙古文附加成分连接符 MSC/NNBSP 的字形控制属性的提案

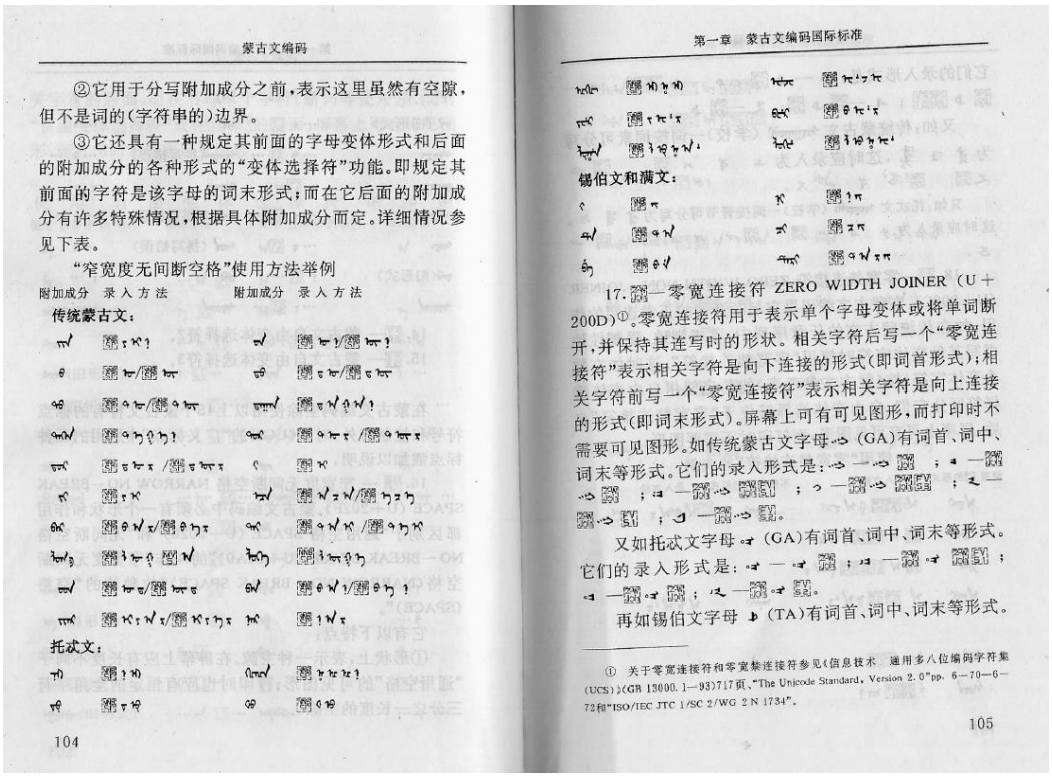
Jirimutu - jrmt@almas.co.jp
Siqin - siqin@almas.co.jp
Bao Haishan - baohaishan@delehi.com
Burigudu - burgood@vip.sina.com
Menghejiya - csmhjy@126.com

1. 背景

据我们所知为蒙古文 Unicode 编码提案引进【蒙古文附加成分连接符】是开始于 1998 年。正如确精扎布教授在呼和浩特市举办的 WG2 蒙古文编码专题会议上所解释，他定义了【蒙古文附加成分连接符】(或许在他的早期提案中称其为【蒙古文空格】)并提案为其申请新的专用码位,但是 WG2 组织却建议使用窄宽无间断空格 NNBSP 作为【蒙古文附加成分连接符】，从而相继的标准各级标准中用 NNBSP 作为【蒙古文附加成分连接符】至今。

在确精扎布教授 2002 年出版的《蒙古文编码》一书里可以找到他所提案的利用窄宽无间隔空格 NNBSP 作为【蒙古文附加成分连接符】，并附加成分字符变形模型的用例。如图 1-1 所示。

我们也在 Myatav Erdenechimeg，Richard Moore 与 Yumbayar Namsrai 著作的 TR170 文件的第 11 页里找到了完全相同的列表。在此省略列出其图标，请在网络上查找相关文件 (http://www.unicode.org/~asmus/mongolian/MD001-unu-tr170.html)。



蒙古文编码

②它用于分写附加成分之前，表示这里虽然有空隙，但不是词的(字符串的)边界。

③它还具有一种规定其前面的字母变体形式和后面的附加成分的各种形式的“变体选择符”功能。即规定其前面的字符是该字母的词末形式；而在它后面的附加成分有许多特殊情况，根据具体附加成分而定。详细情况参见下表。

“窄宽度无间断空格”使用方法举例

附加成分	录入方法	附加成分	录入方法
传统蒙古文：			
ᠮ	ᠮ+NNBSP+ᠠ	ᠮ	ᠮ+ZWJ+ᠠ/ᠮ+NNBSP+ᠠ
ᠪ	ᠪ+ᠠ/ᠪ+ZWJ+ᠠ	ᠪᠪ	ᠪ+ᠠ/ᠪ+ZWJ+ᠠ
ᠪᠪ	ᠪ+ᠠ/ᠪ+ZWJ+ᠠ	ᠭᠭ	ᠭ+ᠠ+NNBSP+ᠠ
ᠰᠰ	ᠰ+ᠠ+NNBSP+ᠠ	ᠪᠪ	ᠪ+ᠠ+ZWJ+ᠠ/ᠪ+ZWJ+ᠠ
ᠲᠲ	ᠲ+ᠠ+ZWJ+ᠠ/ᠲ+ZWJ+ᠠ	ᠴ	ᠴ+ZWJ
ᠨ	ᠨ+NNBSP+ᠠ	ᠰᠰ	ᠰ+ᠠ+ZWJ+ᠠ/ᠰ+ZWJ+ᠠ
ᠪᠪ	ᠪ+ᠠ+ᠠ/ᠪ+ZWJ+ᠠ	ᠬᠬ	ᠬ+ᠠ+ᠠ/ᠬ+ZWJ+ᠠ
ᠬᠬᠠ	ᠬ+ᠠ+ᠠ+ᠠ/ᠬ+ZWJ+ᠠ+ᠠ	ᠬᠬᠠ	ᠬ+ᠠ+ᠠ+ᠠ
ᠲᠲ	ᠲ+ᠠ+ᠠ/ᠲ+ZWJ+ᠠ	ᠪᠪ	ᠪ+ᠠ+ᠠ
ᠲᠲ	ᠲ+ᠠ+ᠠ+ZWJ+ᠠ/ᠲ+ZWJ+ᠠ+ᠠ	ᠮᠮ	ᠮ+ᠠ+ᠠ
托忒文：			
ᠳ	ᠳ+ᠠ	ᠳᠳ	ᠳ+ᠠ+ᠠ
ᠪᠪ	ᠪ+ᠠ	ᠪᠪ	ᠪ+ᠠ

ᠳᠳᠠ	ᠳ+ᠠ+ᠠ	ᠳᠳ	ᠳ+ᠠ+ᠠ
ᠲᠲ	ᠲ+ᠠ+ᠠ	ᠪᠪ	ᠪ+ᠠ+ᠠ
ᠰᠰᠠ	ᠰ+ᠠ+ᠠ+ᠠ	ᠪᠪ	ᠪ+ᠠ+ᠠ+ᠠ
锡伯文和满文：			
ᠰ	ᠰ+ᠠ	ᠰ	ᠰ+ᠠ
ᠮᠮ	ᠮ+ᠠ+ᠠ	ᠰᠰ	ᠰ+ᠠ+ᠠ
ᠪᠪ	ᠪ+ᠠ+ᠠ	ᠮᠮᠠ	ᠮ+ᠠ+ᠠ+ᠠ

17. ZWJ—零宽连接符 ZERO WIDTH JOINER (U+200D)①。零宽连接符用于表示单个字母变体或将单词断开，并保持其连写时的形状。相关字符后写一个“零宽连接符”表示相关字符是向下连接的形式(即词首形式)；相关字符前写一个“零宽连接符”表示相关字符是向上连接的形式(即词末形式)。屏幕上可有可见图形，而打印时不需要可见图形。如传统蒙古文字母 ᠭ (GA)有词首、词中、词末等形式。它们的录入形式是：ᠭ → ᠭ+ZWJ → ZWJ+ᠭ ；ᠭᠭ → ᠭ+ZWJ+ᠭ ；ᠭᠭᠭ → ZWJ+ᠭ+ZWJ+ᠭ ；ᠭ → ZWJ+ᠭ ；ᠭᠭ → ZWJ+ᠭ+ZWJ。

又如托忒文字母 ᠭ (GA)有词首、词中、词末等形式。它们的录入形式是：ᠭ → ᠭ+ZWJ ；ᠭᠭ → ᠭ+ZWJ+ᠭ ；ᠭᠭ → ᠭ+ZWJ ；ᠭᠭ → ZWJ+ᠭ。

再如锡伯文字母 ᠲ (TA)有词首、词中、词末等形式。

① 关于零宽连接符和零宽禁连接符参见《信息技术 通用多八位编码字符集(UCS)》(GB 13000.1—93)717页，“The Unicode Standard，Version 2.0”pp. 6—70—6—72和“ISO/IEC JTC 1/SC 2/WG 2 N 1734”。

图 1-1

我们还在 GB/T 26226-2010 找到如图 1-2 所示的类似窄宽无间断空格 NNBSP【蒙古文附加成分连接符】的用例。此列表包含了比《蒙古文编码》更多的蒙古文附加成分。

o) ——蒙古文自由变体选择符 3。

在蒙古文编码里除使用以上蒙古文特有的标点符号和控制字符外，还使用多个通用标点，见附录 E。另有从 GB 13000—2010 的“广义标点”中引用的 3 种标点，说明如下：

a) ——窄无间断空格 NARROW NO-BREAK SPACE (U+202F)。其形状和作用区别于“通用空格 SPACE (U+0020)”和“无间断空格 NO-BREAK SPACE (U+00A0)”，具有以下特点：

1) 其高度是恒定的，为全角字符的三分之一；
2) 用于分写附加成分之前，但不是词的边界；
3) 规定其前后字母的变体形式。

传统蒙古文窄无间断空格的使用方法见表 10。

表 10 传统蒙古文窄无间断空格的使用方法

显现形式	字符序列	显现形式	字符序列

图 1-2

2007 年之后，使用 OpenType 技术在字体与操作系统级别实现蒙古文 Unicode 编码成为可能。微软公司率先在 Windows Vista 系统提供了蒙古文 Unicode 字体。历经多年的努力，多个本地企业也在提供很多蒙古文字体。虽然这些字体在实现细节上各有所不同，正如您所知，我们一直努力统一蒙古文 Unicode 编码标准的实现规则。

在当前蒙古文 Unicode 编码标准中，蒙古文附加成分连接符 MSC/窄宽无间断空格 NNBSP 的应用方面，我们发现了一个重大问题。作为蒙古文附加成分连接使用的蒙古文附加成分连

接符 MSC/窄宽无间断空格 NNBSP 的方案依赖于字体渲染引擎的实现逻辑。当我们为蒙古文附加成分连接使用窄宽无间断空格 NNBSP 的时候（其实它还在控制后续附加成分的显现字形变化），我们都面临蒙古文附加成分在现存的字体渲染引擎（例如 Harfbuzz 与 Uniscribe）下显示字形错误。在微软 Uniscribe 字体渲染引擎下，最新版能够正常渲染常用正文中的蒙古文附加成分，但仍存在在特殊情况下的附加成分不能正确显示蒙古文附加成分。即使是常规正文中的蒙古文附加成分，在微软 Office2013 及 2016 等产品上失败了将近 1-2 年。最为显著的影响是目前所有蒙古文附加成分均不能在安卓，iOS 等移动手机系统的默认字体上正确显示。还有 Mac OS X 和 Linux 等个人电脑草系统的默认字体也显示不正确的蒙古文附加成分。

2. 介绍目前的解决方案

我们在几年前已发现应用窄宽无间断空格 NNBSP 作为蒙古文附加成分连接模型的问题，并且在几年以来一直在讨论如何改进用于蒙古文 Unicode 编码标准的窄宽无间断空格 NNBSP 作为蒙古文附加成分连接符的模型与规则。

在 2014 年 11 月至 2015 年 12 月的 W3C 论坛上，我们讨论了窄宽无间断空格 NNBSP 的蒙古文附加成分连接模型与其改进方案。在该论坛上讨论的蒙古文附加成分连接符 MSC/窄宽无间断空格 NNBSP 的蒙古文附加成分连接模型的定义如图 2-1。

目前的蒙古文附加成分连接的问题，其实在于使用窄宽无间断空格 NNBSP 作为蒙古文附加成分连接符。因为窄宽无间断空格 NNBSP 属性不能满足所有蒙古文附加成分连接与控制要求。我们于 2016 年提案在码位 U+180F 增设一个蒙古文附加成分连接符（MSC）。然而此提案尚未得到 UTC 与 WG2 的批准。

MSC 模型解决方案的所有蒙古文附加成分连接符请见附录 1。

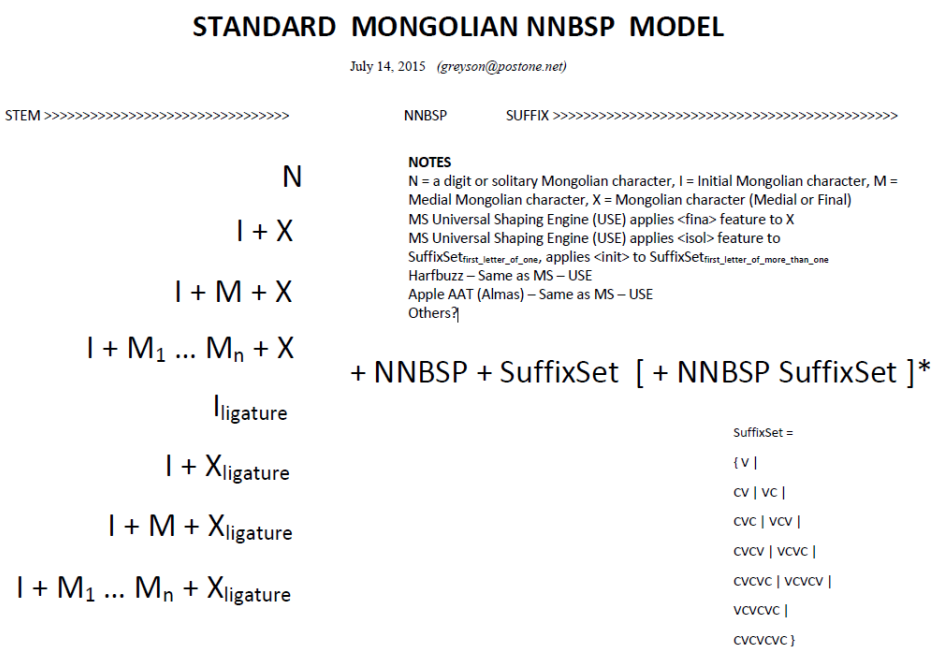


图 2-1

3. MSC/NNBSP 模型解决方案的现存问题

1) 使用 NNBSP 作为 MSC 时发生的问题

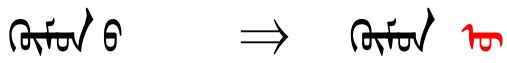
如下 3 例附加成分符。

<U+1832><U+1820><U+182F><U+180E><U+1820><U+202F><U+1836><U+1822><U+1828>

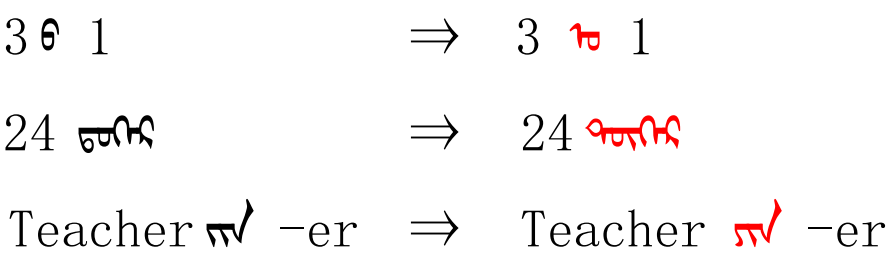
<U+1828><U+1823><U+182E><U+202F><U+1824><U+1828>

<U+182C><U+1826><U+182E><U+1826><U+1828><U+202F><U+1826>

在 Harfbuzz 字体渲染引擎上都被错误地显示为下例形式。移动手机系统安卓，iOS 以及 Mac OS X 与 Linux 系统现在都遇到这个问题。 只有一些版本的微软 Uniscribe 字体渲染引擎正确显示上述附加成分。



2) 在作为 MSC 的 NNBSP，在蒙古文附加成分符前面出现数字，标点符号或其他语种文字时大多数现存系统上都得不到正确显示的蒙古文附加成分。例如：



3) 当从 PDF 文件复制蒙古文粘贴时 NNBSP 都会被丢失。

4) 作为这些问题的解决方案，我们提案的新码位 U+180F 仍未被 UTC 与 WG2 接受和批准。仍需要等待相当长的时间才能在蒙古文 Unicode 编码标准中解决这些问题。

5) MSC 模型本身存在问题。即使定义了 MSC 新的码位 U+180F，我们仍旧面临下列问题。

   A) 蒙古文附加成分符受限于预定义的附加成分列表。将来我们很难增加或扩充附加成分列表。蒙古文附加成分列表仍受制于包含或剔出哪些附加成分符的未明确的议论。附加成分列表会依照语法定义与语言学的发展而改变。

   B) 从编码角度看，MSC 或 NNBSP 蒙古文附加成分连接模型方案是不合理的。根据我们多年的经验，该方案可以说是一个即不稳定又不充分的，具有很多问题的方案。我们需要废除 MSC 或 NNBSP 的附加成分连接符的附加成分控制属性，只把它用作一个连接符。附随 MSC 或 NNBSP 的蒙古文附加成分，应由相应蒙古文字符自由变体选择符明确地指定特殊显现字形。

   C) 目前，在不用 MSC 的情况下，我们不能显示或打印大多数蒙古文附加成分。但是在现实情况中，我们需要单独地显示或打印蒙古文附加成分符。

   D) 据一些专家在呼和浩特蒙古文编码专题会议上的解释，出版社与新闻机构按各自的喜好，分别选择蒙古文附加成分可换行或不可换行模式。若蒙古文附加成分的可换行模式很常见，我们无须如此受限，并且采用牺牲灵活性的方法定义蒙古文附加成分模型。作为蒙古文字使用者，我们熟知我们在日常生活中使用可换行模式与不可换行模式二者并存。

6) 蒙古文附加成分显示错误问题是多年以来一直阻碍我们在计算机与移动手机系统上使用蒙古文 Unicode 编码的最大的问题。为此我们在内蒙古已启用下列的替代方案，在移动手机终端广泛使用蒙古文 Unicode 编码。

4. 替代方案的实践考验

2017 之夏以来，我们已在各自的移动手机应用程序，Mac OS X 以及 Linux 系统上，为目前的 MSC 蒙古文附加成分连接模型启用了，像其他语言用 NNBSP 只作为窄宽无间断空格的替

代方案。经过半年的实践，我们认为这是最合理的蒙古文附加成分解决方案。

替代方案的内容为如下。

1) 如平常一样，用 NNBSP 作为蒙古文附加成分连接符。即使在一些附加成分上存在错误显示，在目前替代方案实施阶段为了与之前的方案保持兼容，我们忽略了存在的问题。NNBSP 仍旧运作为当前的 MSC/NNBSP 蒙古文附加成分模型。
2) 其次，我们为特定的字符定义了，蒙古文附加成分中蒙古文字一些不规则书写形式的新变形显现字形（或者用已有的特定变形显现字形），并且使用该字符自由变体选择符 1-3（FVS1-3）强行显示附加成分所需的变形显现字形。
3) 利用此规则，我们修改了所有的字体的显示逻辑。(事实上 Almas 字体从一开始包含了此逻辑。为内蒙古自治区民族事务委员会制定的共享工程转换规则也包含了此逻辑)。
4) 我们修改了我们在不同系统上的所有输入法，并为最终用户提供了此替代方案半年之久。根据众多用户的反馈，该方案即可行又稳定，并深受欢迎。
5) 目前我们使用该替代方案克服了 MSC/NNBSP 蒙古文附加成分控制与连接的大问题，并且在我们日常生活中开始广泛使用了蒙古文 Unicode 编码标准。
6) 我们应从这样的实践经验中学习与分析，以此来改善蒙古文附加成分编码方案。这样改善方案将帮助我们突破在各种操作系统，各种应用程序中使用蒙古文 Unicode 编码标准的巨大限制。

5. 关于蒙古文附加成分的提案

根据上述实践经验，我们为 UTC 蒙古文专题会议准备了如下的蒙古文附加成分编码模型的提案。

1) 修改目前的 MSC/NNBSP 蒙古文附加成分编码模型。
2) 继续为蒙古文使用窄宽无间断空格 NNBSP 作为蒙古文附加成分。恢复 NNBSP 属性与其在其他语种中的完全一致。
3) NNBSP 在蒙古文中可作为不可换行的蒙古文附加成分连接符使用。
4) 蒙古文附加成分的特殊字形的显示使用 FVS1-3 控制符。
5) 蒙古文附加成分显现字形不依赖于其前置的 NNBSP,而是依赖于正常的蒙古文字体渲染规则。蒙古文附加成分的特殊显现字形使用 FVS1-3 字形选择符。
6) 蒙古文附加成分可以采用可以换行的书写模式。该模式下我们使用普通空格作为附加成分前置符(而非 NNBSP) 。不论使用与不使用 NNBSP,蒙古文的附加成分都可以正确显示指定正确字形。
7) 换言之，该提案可理解为“废除 MSC/NNBSP 蒙古文附加成分控制属性的提案”。

关于详细的蒙古文附加成分编码列表，请参照附录 II。在附录 II 列举了现有的所有蒙古文附加成分。根据编码规则，我们可以将任何单词可视为蒙古文附加成分，并在一个文本

行里当成一个整体词汇以不可换行模式进行排版显示。

6. 评估与总结

该提案的优势

1) 获得批准后可以被迅速实施。
2) 无需为 MSC 新码位再等待 2-3 年。
3) 该提案为蒙古文内部的编码完善方案，对其他语种无任何影响。
4) 无需对字体渲染引擎做任何的修改。
5) 可以正确运行于当前主流个人电脑及移动设备操作系统。
6) 类似于旧方案，无需再次培训使用用户。
7) 对供应商及最终使用用户产生很小的影响。
8) 此蒙古文附加成分编码模型是基于规则，易于理解。
9) 此蒙古文附加成分编码模型的规则简单且、易于共享、易于交流和易于标准化。
10) 该方案具有灵活性、可扩展性和较强的稳定性。

该提案的缺点

1) 供应商需要为蒙古文附加成分连接模型修改字体逻辑。但是在目前应用问题存在情况下，无论如何供应商都需要修改字体逻辑才能解决存在的问题。
2) 供应商需修改输入法工具以支持新的蒙古文附加成分编码。同样，如果选择了其他解决方案，如：增加新的 MSC 码位，也需要修改输入法工具中的附加成分编码。
3) 现有的电子文本和语料库会产生不正确的附加成分显示。我们需要提供蒙古文附加成分自动修正工具以简化修正工作。
   如果我们采用新定义 MSC 码位的方案,同样需要修正现有的电子文本和语料库中的附加成分编码。还得需要做 NNBSP 与 MSC 的自动转换等。

附录 I

现有 MSC/NNBSP 模型的

蒙古文附加成分编码表

Vocative Case

᠋᠋

呼格

NNBSP+1820 ᠋

NNBSP+1821 ᠋

Genetive Case

ᡂ ᡂ ᡂ ᡃ ᡃ

定格

NNBSP+1836+1822+1828 ᡂ

NNBSP+1824+1828 ᡂ

NNBSP+1826+1828 ᡂ

NNBSP+1824 ᡃ

NNBSP+1826 ᡃ

Accusative Case

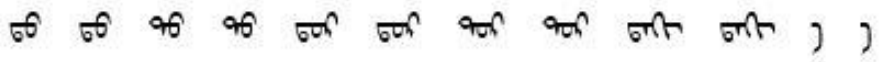
ᡄ ᡅ

宾格

NNBSP+1822 ᡄ

NNBSP+1836+1822 ᡅ

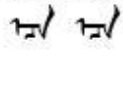
Dative-Locative Case



向位格

NNBSP+1833+1824	
NNBSP+1833+1826	
NNBSP+1832+1824	
NNBSP+1832+1826	
NNBSP+1833+1824+1837	
NNBSP+1833+1826+1837	
NNBSP+1832+1824+1837	
NNBSP+1832+1826+1837	
NNBSP+1833+1820+182C+1822	
NNBSP+1833+1821+182C+1822	
NNBSP+1820	
NNBSP+1821	

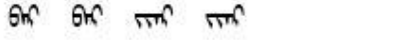
Ablative Case



从比格

NNBSP+1820+1834+1820	
NNBSP+1821+1834+1821	

Instrumental Case



凭借格

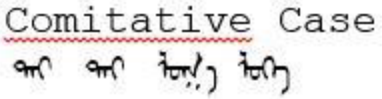
NNBSP+182A+1820+1837

NNBSP+182A+1821+1837

NNBSP+1822+1836+1820+1837

NNBSP+1822+1836+1821+1837

Comitative Case



和同联合格

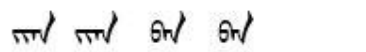
NNBSP+1832+1820+1822

NNBSP+1832+1821+1822

NNBSP+182F+1824+182D+180E+1820

NNBSP+182F+1826+182D+1821

Reflexive Case



领格

NNBSP+1822+1836+1820+1828

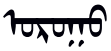
NBSP+1822+1836+1821+1828

NNBSP+182A+1820+1828

NNBSP+182A+1821+1828

Directive

导向格(或许用 NNBSP)

NNBSP+1824+1837+1824+182D+1824 

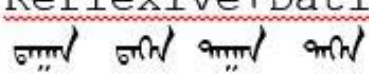
Reflexive+Accusative

领格 + 宾格

NNBSP+1836+1824+182D+1820+1828 

NNBSP+1836+1826+182D+1821+1828 

Reflexive+Dative-Locative



领格+向位格

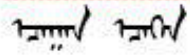
NNBSP+1833+1820+182D+1820+1828 

NNBSP+1833+1821+182D+1821+1828 

NNBSP+1832+1820+182D+1820+1828 

NNBSP+1832+1821+182D+1821+1828 

Reflexive+Ablative

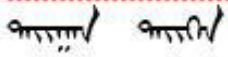


领格+从比格

NNBSP+1820+1834+1820+182D+1820+1828 

NNBSP+1821+1834+1821+182D+1821+1828 

Reflexive+Comitative



领格+和同格

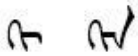
NNBSP+1832+1820+1836+1822+182D+1820+1828 

NNBSP+1832+1821+1836+1822+182D+1821+1828 

NNBSP+1832+1820+1822+182D+1820+1828 

NNBSP+1832+1821+1822+182D+1821+1828 

CASE-BOUND POSSESSION

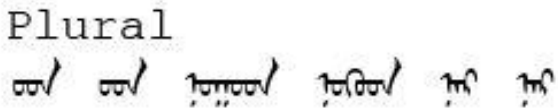


所有格

NNBSP+182C+1822 

NNBSP+182C+1822+1828 

Plural



复数助词(第一助词可以连写)

NNBSP+1824+1833 

NNBSP+1826+1833 

NNBSP+1828+1824+182D+1824+1833 

NNBSP+1828+1826+182D+1826+1833 

NNBSP+1828+1820+1837 

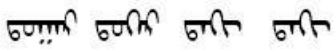
NNBSP+1828+1821+1837 

Negation

否定词(或许用 NNBSP)

NNBSP+1826+182D+1821+1822 

Ordinal



序数助词

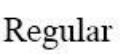
NNBSP+1833+1824+182D+1820+1837 

NNBSP+1833+1826+182D+1821+1837 

NNBSP+1833+1820+182C+1822 

NNBSP+1833+1821+182C+1822 

Regular



常态助词(或许用 NNBSP)

NNBSP+1833+1820+182D 

NNBSP+1833+1821+182D

附录 II

使用新提议方案的蒙古文附加成分编码表

Vocative Case

᠋ ᠋

呼格

1820+FVS2 ᠋

1821+FVS1 ᠋

Genetive Case

ᡂ ᡂ ᡂ ᡃ ᡃ

定格

1836+FVS1+1822+1828 ᡂ

1824+FVS1+1828 ᡂ

1826+FVS1+1828 ᡂ

1824+FVS1 ᡃ

1826 +FVS2 ᡃ

Accusative Case

ᡄ ᡄ

宾格

1822+FVS1 ᡄ

1836+FVS1+1822 ᡄ

Dative-Locative Case

向位格

1833+FVS1+1824 

1833+FVS1+1826 

1832+1824 

1832+1826 

1833+FVS1+1824+1837 

1833+FVS1+1826+FVS3+1837 

1832+1824+1837 

1832+1826+FVS3+1837 

1833+FVS1+1820+182C+1822 

1833+1821+182C+1822 

1820+FVS1 

1821+FVS1 

Ablative Case

从比格

1820+FVS1+1834+1820 

1821+1834+1821 

Instrumental Case

凭借格

182A+1820+1837 

182A+1821+1837 

1822+FVS1+1836+FVS1+1820+1837 

1822+FVS1+1836+FVS1+1821+1837 

Comitative Case

和同联合格

1832+1820+1822 

1832+1821+1822 

182F+1824+182D+180E+1820 

182F +1826+FVS3+182D+1821 

Reflexive Case

领格

1822+FVS1+1836+FVS1+1820+1828 

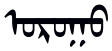
1822+FVS1+1836+FVS1+1821+1828 

182A+1820+1828 

182A+1821+1828 

Directive

导向格(或许用 NNBSP)

1824+1837+1824+182D+1824 

Reflexive+Accusative

领格 + 宾格

1836+1824+182D+1820+1828 

1836+1826+FVS3+182D+1821+1828 

Reflexive+Dative-Locative

领格+向位格

1833+FVS1+1820+182D+1820+1828 

1833+FVS1+1821+182D+1821+1828 

1832+1820+182D+1820+1828 

1832+1821+182D+1821+1828 

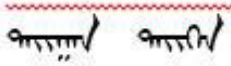
Reflexive+Ablative

领格+从比格

1820+FVS1+1834+1820+182D+1820+1828 

1821+1834+1821+182D+1821+1828 

Reflexive+Comitative



领格+和同格

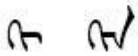
1832+1820+1836+FVS1+1822+182D+1820+1828 

1832+1821+1836+FVS1+1822+182D+1821+1828 

1832+1820+1822+182D+1820+1828 

1832+1821+1822+182D+1821+1828 

CASE-BOUND POSSESSION

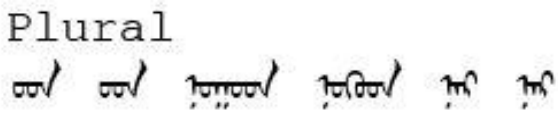


所有格

182C+1822 

182C+1822+1828 

Plural



复数助词(第一助词可以连写)

1824+FVS1+1833 

1826+FVS1+1833 

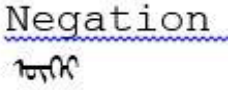
1828+1824+182D+1824+1833 

1828+1826+FVS3+182D+1826+1833 

1828+1820+1837 

1828+1821+1837 

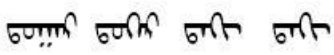
Negation



否定词(或许用 NNBSP)

1826+182D+1821+1822 

Ordinal



序数助词

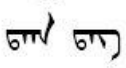
1833+FVS1+1824+182D+1820+1837 

1833+FVS1+1826+FVS3+182D+1821+1837 

1833+FVS1+1820+182C+1822 

1833+FVS1+1821+182C+1822 

Regular



常态助词(或许用 NNBSP)

1833+FVS1+1820+182D 

1833+FVS1+1821+182D 

其他(或许用 NNBSP)

1833+FVS1+1820 

1833+FVS1+1821 

1833+FVS1+1824+1828+1822 

1833+FVS1+1826+FVS3+1828+1822 

1833+ 1824+1828+1822 

1833+ 1826+FVS3+1828+1822