Discussion from Jirimutu@Almas Inc.
(jrmt@almas.co.jp)
July 12, 2015

1. NNBSP – Model

STEM>>>>>>>>>>>>>>>>   NNBSP   Suffix >>>>>>>>>>>>>>>>>>>

$$N$$
$$I$$
$$I + X$$
$$I + M + X$$

$I + M_1+…+M_n+X$ + NNBSP + SuffixSet + [NNBSP + SuffixSet]

I ligature

$I + X$ ligature

$I + M + X$ ligature   Suffix = { v | cv | vc | cvc |vcv | cvcv | vcvc |

$I + M_1 +…..+ M_n + X$ ligature   cvcvc | vcvcv |cvcvcv | vcvcvc | …. }

Greg Eck's Response – Jul 10 2015
Let's delete the solitary "I" and leave your N. Your "N" will include my former "I".

        [NNBSP + SuffixSet] *

with the Kleene asterisk designating 0 or more occurrences of the bracketed item.
Yes. It is the best. I agree these two definitions.

NOTE:

N=any character, I = Initial or Isolate, M = Medial, X = Mongolian Character (Medial or Final)

Greg Eck's Response – Jul 10 2015
Let's make this "N = any solitary Mongolian Character, M = Medial Mongolian Character, X = Mongolian Character (Medial or Final) ". That way, each category is mutually exclusive.
I agree this.

MS Universal Shaping Engine (USE) applies <fina> feature to X

MS Universal Shaping Engine (USE) applies <init> feature to the suffix set first letter.

Harfbuzz – Same as MS - USE

Apple – AAT Font applies <init> feature to the suffix set first letter, applies <isol> feature to the suffix if it is only one character.

Others?

Greg Eck's Response – Jul 10 2015
This has already been changed to:
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 (Almas) – Same as MS – USE
Is that acceptable on the Apple AAT line?   It is Ok for our AAT font.
One question remains is whether another Apple AAT font will render the same as the Almas implementation. Sorry, I do not know much about how the AAT fonts/rendering systems work.
No problem others can do the same with Almas Implementation. The font developer can define the character's positional properties in their state machine logic.

Here I would like to add another option for NNBSP (Mongolian Suffix Separator-MSS) Model.

That is

N + NNBSP + Mongolian Suffix Set + [NNBSP Mongolian Suffix Set]*

Where N is any solitary Mongolian Character (I am still insist onto other language characters, digital characters or punctuation will be included)

For Example,

23 ᠊ ᠬᠤᠭᠤᠷᠠᠭ᠎ᠠᠨ 24 ᠪᠣᠯᠣᠨ᠎ᠠ ··

3 ᠊ 2 ᠊ᠶ ᠠᠴᠠᠭᠠᠷ ᠬᠢᠭᠡᠳ ᠪᠣᠯᠣᠨ᠎ᠠ ··

5 ᠊ᠨ 5 ᠪᠤ 25 ··

a ᠊ᠨ ᠡᠭᠡᠰᠢᠭ ᠠᠪᠢᠶ᠎ᠠᠨ {ei, ʌ, æ, ə} ᠭᠡᠳᠡᠭ ᠪᠣᠯᠣᠨ᠎ᠠ ··

《᠊ᠤᠭᠨ ᠠᠭᠤᠯᠠᠯ》 ᠊ᠨ ᠣᠷᠤᠰᠬᠠᠯ ᠭ ᠤᠨ ᠪᠦ ᠊᠊·   《᠊ᠤᠭᠨ ᠠᠭᠤᠯᠠᠯ》 is name of one book.

Greg Eck's Response – Jul 10 2015
Could we restrict the discussion at first to non-numerals and agree on the 1820-18AA range first? Numerals will bring in another new dimension to the puzzle.
This kind of usage is very common in Mongolian actually.

My discussion on the possibility of replacement of NNBSP with other character:

1. The NNBSP is a space, not control character as it is defined in Unicode. we need here is one Mongolian Suffix Joiner (special control character) actually.

2. Some of the system software handles NNBSP as a space (narrow non-break space) and ignore or filters out this character before come into MS-USE or Harfbuzz. In that case, we lost the important control characters in the text and could not get correct Mongolian forms actually.   For example when we copy text from adobe PDF reader, the text will lost the NNBSP.

3. NNBSP had been filtered by the previous version browser like chrome and safari till last year and Mongolian text all get mistaken forms related with NNBSP.

4. If it is possible to replace it with other character to define the function, I would like to propose to use U+180F, we can name it as Mongolian Suffix Joiner.

Greg Eck's Response – Jul 8 2015

Another possible name is Mongolian Suffix Separator to pair it with the Mongolian Vowel Separator. Both actually join two parts of a word. Both are not spaces. Both are non-joining in the terms of ArabicSpacing.txt. One value of excluding the word "join" from the name is that it might reduce future confusion - a person might ask the question why is called a "Joiner" if it is actually "non-joining". We don't want to include the term "space" as it is not a space. If we use the term "separator" although it is still deficient, it does match the MVS terminology.

Mongolian Vowel Separator ......, Mongolian Suffix Separator.

Just a thought, but then we are getting ahead of ourselves.

I agree this, the "Mongolian Suffix Separator" is better than "Mongolian Suffix Joiner".

My discussion on STANDARD NNBSP MONGOLIAN MODEL:

1. What is the NNBSP definition? According to Unicode definition, it is Narrow No-Break Space, narrow form of a no-break space, typically the width of a thin space or a mid space.

2. If we use it as Mongolian Suffix Joiner (no-break space), we should add the definition into the Unicode standard. In the case, all of system or developer will pay their attention to this is used in Mongolian and have some special definition not just a narrow no-break space.

3. When we use NNBSP in Mongolian, we should include the usage of the Character.

   ① Is it only used for Mongolian Suffix? If so, can we define full collection of the Mongolian Suffix? Or We just consider the followed word after NNBSP is suffix ?

   Greg Eck's Response – Jul 8 2015
   I suggest that such a character be used only as the suggested name implies to "separate suffixes" or "join suffixes" to the stem or to each other. If there is another function that the character carries, then it should be reflected in a more generic name.
   I think it should be up to the font designer to decide how tight he/she wishes to restrict the characters following this potential new character.

   I agree this in this point. I will explain in the following content and get back to here.

   ② Is it can be used as other purpose in Mongolian like other Language. For example, we just use NNBSP as a "Narrow no-break Space", it can be used as connection of two words we do not want to separate them to two line or just want handle the two word as one word.

   In this case, it is better to consider our NNBSP model will not impact our normal word forms. It is just impact the specific Mongolian Suffix.

   Almas Font s are considered this possibility and our Font handle the Suffix only have been changed their first letter in special rule (Suffix Rule) and other followed normal words will not be changed the first letter.

For example:

Our Font shows like bellow

ᠵᠣᠭᠡ ᠣᠪ -- (1828+1823+182E+NNBSP+1824+1828)

ᠵᠣᠭᠡ ᠣᠭᠠᠷᠠᠭ -- (1828+1823+182E+NNBSP+1824+1837+1822+182C+1824)

But Mongolian Baiti shows this like bellow

ᠹᠥᠷ ᠳᠥ -- (1828+1823+182E+NNBSP+1824+1828)

ᠹᠥᠷ ᠣᠷᠢᠣᠬ -- (1828+1823+182E+NNBSP+1824+1837+1822+182C+1824)

PS. My colleagues suggested me that we should use this word for examples.

I would like to ask all members to consider remaining this possibility for the further usages. The other languages are using this character as this kind of usage as its definition.

Greg Eck's Response – Jul 8 2015
I do not consider this to be a mistake, but rather a difference in implementation design. One font designer has decided to tightly restrict the characters following the NNBSP. Another font designer has decided to form the grammatically correct suffixes correctly but not spell-check the text after the NNBSP.
I consider these decisions to be independent of the model and left to the choices of the font designer.

I don't agree this in this point as well. Let me include my opinion in the following discussion.

Greg Eck's Response – Jul 10 2015
I would also ask that we use another word as ORIQU is a verb and therefore outside of the scope of the discussion. Baiti designers would say that as it is not a suffix, it is not a properly typed word, therefore an anomoly giving unpredictable shaping.

It is the result of Mongolian Baiti only checking one character after NNBSP and reshaping all U|UE as ᠤ In <init>, But I want say is that we just reshape the U|UE after NNBSP when the suffix is ᠤ᠎ ᠤ᠎, Other words will be normal display form like ᠣᠷᠢᠣᠬ , ᠳᠤᠷᠢ᠎ .

Maybe the example listed above is not sufficient to explain what I want to discuss with all.

Additional example.

Our Font shows like bellow

ᠹᠥᠷ ᠳᠥ -- (1828+1823+182E+NNBSP+1824+1828)

ᠹᠥᠷ ᠤᠷᠢᠣᠳᠤ -- (1828+1823+182E+NNBSP+1824+1837+1824+182D+1824)

But Mongolian Baiti shows this like bellow

ᠬᠣᠧ ᠣᠧ -- (1828+1823+182E+NNBSP+1824+1828)

ᠬᠣᠧ ᠣᠨᠣᠶᠣᠺ -- (1828+1823+182E+NNBSP+1824+1837+1824+182D+1824)

Greg Eck's Response – Jul 10 2015
As I understand the URUGU, it is probably a postposition in transition to being a case suffix. Ulaanbaatar dictionaries will list it as уруу. Ulaanbaatar grammars will also list it as the directive case suffix -руу/-рүү // -луу/-лүү. Therefore, my team considers the implementation in the vertical script as appropriate as either an NNBSP-connected suffix OR as a separate word. Either way, the form should be as you show under your font with the full initially-ornamented "U".   The display as you show for Baiti is indeed incorrect. I will put in a bug report for this. Thanks for noticing this. Is there further discussion here?

What I want to say is that we reshape NNBSP connected U|UE as ᠧ only when the followed suffix is ᠧ{202F+1824+1828, 202F+1826+1828} ᠧ {202F+1824+1833, 202F+1826+1833}. All other U|UE started word will   leave it as normal Initial form of U | UE ᠊ ᠊.

The (1828+1823+182E+NNBSP+1824+1837+1824+182D+1824) is Mongolian suffix listed in DS05. But it is not displaying same in our font and Microsoft Baiti Font (It is hard to be accepted)

The solution will be listed in the following contents.

My discussion on MONGOLIAN NNBSP-CONNECTED SUFFIXES:

1. I have checked the DS05 listed Mongolian Suffixes list. It is covered most of the Mongolian Suffix well, compared with the our suffixes list find a little bit difference. For example, we included like bellow

ᠬᠢᠨ   ᠠᠴᠠ   ᠤᠨ

But I think it is Ok for us if we list out the irregular first letter writing suffixes, it is enough for this Model document. Others will follow the normal word glyph rendering rules applied

Greg Eck's Response – Jul 8 2015
Our grammar considers these forms to be pronouns and therefore not connected by the NNBSP. I believe this is the same view as espoused by Professor Quejingzhabu also. Are there other words that are questionable?

Oh sorry, It is not pronoun actually, if it is spelled as ᠬᠢᠨ ᠠᠴᠠ ᠂ it is pronoun. This three is Mongolian Suffix, and it is what we had been asked by the users to add to our Suffix on our Mongolian IME. The name of the Suffix I will confirm with Linguistic Expert and come back to you.

Greg Eck's Response – Jul 10 2015

Sorry, I had not looked at the word closely enough. My school of linguistics (following the late Dandy-Yadam Tserenpil - Ulaanbaatar) calls these Case Bound Particles rather than suffixes. I would add a few more (MINI, MANI, CINI, TANI, NI, ANO, INO).   I am not sure that they need to be NNBSP-connected, but can see some reasonings therein. Let's go on.

Then there are the CaseBound Suffixes (QI, QIN) also. They are not included in the DS05 listing, but should be listed.

Ok. I have confirmed with linguistic expert here. The Mongolian Name is ᠬᠣᠯᠪᠣᠭᠠᠨ ᠤ ᠳᠠᠭᠠᠪᠤᠷᠢ ᠦᠭᠡᠰ ᠂ It is Ok for us only if it is listed in DS05.

2. For the listed suffixes, would you please list out the Unicode in the Notes. Because, we find there significant difference in the encoding some of the suffixes. For example, we find following encoding difference

ᠠᠴᠠ

NNBSP + 1822 + 1828

NNBSP + 1822 + 1822 + 1828

NNBSP + 1835 + 1822 + 1828

NNBSP + 1836 + 1822 + 1828

We should define which is the correct encoding.

4. >Our grammar considers these forms to be pronouns and therefore not
   > connected by the NNBSP. I believe this is the same view as espoused
   > by Professor Quejingzhabu also. Are there other words that are questionable?

I have not listed all of the Suffixes what we are handling here, it is because we have
just considered that the Mongolian linguistic Experts have their own consideration
of which is suffix and which is not suffix. We do not need to roll into the endless
argument of the linguistic opinion here. For example,

ᠥᠭᠡ which is used in ᠰᠠᠢᠨ ᠪᠠᠢᠨᠠ᠂ ᠥᠭᠡ ᠤᠤ is not included in DS05,

and some part says it is suffix and some part says it is not suffix. But it is
really need a suffix character to handle it. But I can list out our additional
suffix list here, but it is still not enough for all of the Mongolian Suffixes
required by linguistic peoples.

1. ᠥ - how do we spell this ? {0020+1824+1824, 0020+1826+1826} or {0020+1824+180B+1824, 0020+1826+180B+1826} or {202F+1824+1824, 202F+1826+1826}. ?

   The {0020+1824+1824, 0020+1826+1826} cannot work.

   {0020+1824+180B+1824, 0020+1826+180B+1826} is not defined in any of the Fonts in http://r12a.github.io/scripts/mongolian/variants

   {202F+1824+1824, 202F+1826+1826} only this is the solution now.

   ᠬᠣᠨ ᠣᠨᠠᠠᠲᠣᠥ ? ᠣᠨᠠᠢ ᠬᠣᠨᠠᠠᠲᠣᠥ ?   Mongolian Bai

   ᠬᠣᠨ ᠣᠨᠠᠠᠲᠣ ᠥ᠂ ᠣᠨᠠᠢ ᠬᠣᠨᠠᠠᠲᠣ ᠥ᠂   Mongolian Universal White

   How do you think of it?

   What I want say here is that we need to declare **the basic rule of Mongolian Encoding.**

   1. *The encoding rule follows the modern normal and publicly accepted display form comes first.*
   2. *Abnormal display forms which are using in modern Mongolian will come to second. What I mention here modern Mongolian include the modern Todo and Shibe and Manchu.*
   3. *Abnormal display forms which are not using in modern Mongolian will come to third.*
   4. *Should consider the entire special requirement from linguistic experts and leave the implementation possibility for all of these requirements.*

   PS. Please add your consideration here for getting quick solution for following part of the discussion includes the linguistic argument of which is which. Maybe Unicode have exact definition of this. If anyone know of that please let us know.

5. Let me list out our suffix list addition to DS05. The suffix name I can provide after refer to linguistic expert if necessary or appreciate anyone can contribute to this.

ᠣᠥ ---- ᠪᠠᠷᠢ ᠪᠠᠷᠢ᠌᠎᠌ᠣ ᠣᠥ (202F+1824+1824, 202F+1826+1826)

> Greg Eck's Response – Jul 10 2015
> Discussed already.
> Check again discussion added in above

ᠢᠢ ---- ᠴᠠᠷᠢ ᠢᠢ   (202F+1828+1822)
ᠰᠢᠢ ---- ᠭᠤᠥ ᠰᠢᠢ   (202F+182E+1822+1828+1822)
ᠵᠢᠢ ---- ᠱᠢᠷᠢ ᠵᠢᠢ   (202F+1834+1822+1828+1822)

> Greg Eck's Response – Jul 10 2015
> Discussed already.
> Accept your opinion.

ᠣᠥ ᠢᠢ - ᠳᠤᠢᠢ   (202F+1833+1824+1828+1822, 202F+1833+1826+1828+1822)
ᠥᠥ ᠢᠢ - ᠶᠤᠢᠢ   (202F+1832+1824+1828+1822, 202F+1832+1826+1828+1822)

> Greg Eck's Response – Jul 10 2015
> Interesting. I have never seen this before! This is the dative suffix followed by the case-bound particle? If it is a suffix combining with a particle, then that puts the whole discussion of the entire case-bound particles back on the table to be potentially added to NNBSP-connection.
> It is exist in Mongolian really.

ᠳᠠ ---- ᠳᠤᠤᠤ᠎᠌ᠣ ᠳᠠ (202F+1833+1820, 202F+1833+1821)

> Greg Eck's Response – Jul 10 2015
> As with the question particle UU/UE, the interjection particle DA/DE is semantically tied to the entire clause/sentence and not to the word. I do not see a compelling reason to require NNBSP connectivity.
> Yes, I agree your opinion. I have just listed here for some user requested us to add it to suffixes list in our Mongolian IME.

ᠯᠠ ---- ᠴᠠᠷᠢ ᠯᠠ ᠪᠠᠷᠢ ᠣᠥ (202F+182F+1820, 202F+182F+1821)

> Greg Eck's Response – Jul 10 2015
> While LA/LE //   CU/CUE are tied to probably the phrase level, I still do not see a compelling reason to require NNBSP connectivity.
> Yes, I agree your opinion. I have just listed here for some user requested us to add it to suffixes list in our Mongolian IME.

ᠬᠠᠨ ---- ᠰᠠᠢᠨ ᠵᠡ ᠬᠠᠨ ᠰᠧᠨ ᠬᠠᠨ (202F+1830+1820+1828, 202F+1830+1821+1828)

Greg Eck's Response – Jul 10 2015
I am not familiar with SAN/SEN. Can you give more information on this one?

It is a emphasize suffix(?) Let me check with linguistic expert and feedback to you.

ᠵᠡ ---- ᠰᠠᠢᠨ ᠵᠡ ᠬᠠᠨ (202F+1834+1824, 202F+1834+1826)

Greg Eck's Response – Jul 10 2015
Follows LA/LE exactly.

Yes, I agree

ᠬᠠᠨ ---- ᠲᠤᠷ ᠬᠠᠨ (202F+1830+1822+182D)
ᠬᠠᠩ ---- ᠲᠤᠷᠢ ᠬᠠᠩ (202F+1830+1822+182D+FVS2)

Greg Eck's Response – Jul 10 2015
Although there is gender agreement, I see no compelling argument for an NNBSB connect.
Yes, I agree.

ᠳᠠᠨ ---- ᠳᠠ (202F+1833+1820+182D)
ᠳᠠᠩ ---- ᠳᠠ (202F+1833+1820+182D+FVS2)

Greg Eck's Response – Jul 10 2015
Is this a mis-spelling for the ordinal numeral suffix DAQI/DEXI? It should not be the verb suffix as the verb suffix is always connected directly to the stem - is that not correct? If it is the ordinal suffix, I think we have grounds for NNBSP connectivity.
Let me check with my team, how it is come into our suffixes list. And give example a little bit later.

ᠳᠤᠭᠠᠷ ---- 2 ᠳᠤᠭᠠᠷ (202F+1833+1824+182D+1820+1837)
ᠳᠡᠭᠡᠷ ---1 ᠳᠡᠭᠡᠷ (202F+1833+1826+182D+1821+1837)

Greg Eck's Response – Jul 10 2015
No doubt that this is the ordinal numeral suffix. In my opinion, both DUGAR/DUEGER // DAQI/DEQI should be connected by NNBSP.
Ok.

5. Let me bring following discussion here for more detailed discussion.

Our Font shows like bellow

ᠨᠣᠷ ᠤᠪ -- (1828+1823+182E+NNBSP+1824+1828)

ᠨᠣᠷ ᠤᠵᠠᠷᠢᠤ -- (1828+1823+182E+NNBSP+1824+1837+1824+182D+1824)

This is the correct display according to DS05 document.

But Mongolian Baiti shows this like bellow

ᠨᠣᠷ ᠤᠪ -- (1828+1823+182E+NNBSP+1824+1828)

ᠨᠣᠷ ᠤᠵᠠᠷᠢᠤ -- (1828+1823+182E+NNBSP+1824+1837+1824+182D+1824)

What I want to say here is that we should not just check the following character after NNBSP (or MSS) to select the form of the character. We should check if it is followed by the specific Suffixes or not and change the form of the character display. We will not impact first letter of other words except Suffixes.

What I want to say here is for example,

NNBSP (or MSS) + 1824 | 1826 get ᠦ

NNBSP (or MSS) + 1824 | 1826 + 1828 get ᠤᠪ

NNBSP (or MSS) + 1824 | 1826 + 1833 get ᠦᠦᠪ

NNBSP (or MSS) + 1824 + other Mongolian character get ᠡ

NNBSP (or MSS) + 1824 + other Mongolian character get ᠡᠷ

What I want to say here is that the NNBSP (or MSS) will only need to distinguish and handle the following display forms for 1824 or 1826 and display forms for 1836 and 1822 as well as display form of 1820. All of the other words followed after NNBSP (or MSS) will be displayed same with original normal words form.

ᠠ (202F+ 1820)

ᠡ (202F+ 1821)

ᠤ (202F+ 1824)

ᠦ (202F+ 1826)

Greg Eck's Response – Jul 10 2015
Should include 1833 here also as it is the variant that is always used as the first letter of the suffix. That would include DU/DUE, DUR/DUER, DAQI/DEQI(DAHI/DEHI), DUGAR/DUEGER, DAGAN/DEGEN. TU/TUE, TUR/TUR, TAQI/DEQI(TAHI/TEHI), TUGAR/TUEGER, TAGAN/TEGEN
Yes. I Agree this.

1820 forms would include ACA/ECE, ACAGAN/ECEGEN.

1836 forms should include YUGAN/YUEGEN.
Yes. I Agree this. But all of the following words no display issue.

ᠥᠨ (202F+ 1826+1828)

ᠤᠤᠨ (202F+ 1824+1833)

ᠤᠤᠨ (202F+ 1826+1833)

ᢗ (202F+ 1824+1824)

ᢗ (202F+ 1826+1826)

ᠮ (202F+ 1836+1822+1828)

ᠮ (202F+ 1822+1836+1820+1837)

ᠮ (202F+ 1822+1836+1821+1837)

ᠮ (202F+ 1822+1836+1820+1828)

ᠮ (202F+ 1822+1836+1821+1828)

ᠮ (202F+ 1836+1822)

ᢉ (202F+ 1822)

All of the other Mongolian suffixes after NNBSP (or MSS) will be displayed as their original/normal forms and it is constructed with their prefix word together to be just like one word.

5. Let me list some suffixes what I want to discuss from the DS05 here.

ᠮ (202F+ 1836+1822+1828)

ᠮ (202F+ 1822+1836+1820+1837)

ᠭᠡᠷ (202F+ 1822+1836+1821+1837)

ᠭᠡᠪ (202F+ 1822+1836+1820+1828)

ᠭᠡᠪ (202F+ 1822+1836+1821+1828)

ᠡᠭ (202F+ 1836+1822)

ᠡᠭᠳᠡᠪ (202F+ 1836+1824+182D+1820+1828)

ᠡᠭᠳᠡᠪ (202F+ 1836+1822+182D+1821+1828)

ᠡ+other Mongolian Characters (202F+ 1836+ other Mongolian Characters)

   The NNBSP(Or MSS) + 1836 model will acts same with the NNBSP(Or MSS) + U | UE Model listed above.

Our fonts and our IME are same with this. Anyone have any different spells for these suffixes?