Proposal to establish NARROW NO-BREAK SPACE as a definitely usable avatar of THIN SPACE
For consideration by Unicode Technical Committee
For consideration by Mongolian Working Group
For consideration by CLDR Technical Committee

2019-04-24 (revision 3)
Marcel Schneider (charupdate@orange.fr)

“We must always say what we see.
Above all we must always
— which is the most difficult — see what we see.”
Alain Finkielkraut quoting Charles Péguy

Proposal history

This paper is triggered by the 17 Mongolian documents L2/19-126 through L2/19-142 (listed as [1] through [17] in the References section) that Unicode posted to the UTC Document Register in the wake of Mongolian Working Group meeting #3 (MWG/3). It is submitted in extremis straddling the deadline of UTC meeting #159 scheduled to start on April 30, 2019, because completing L2/19-116 Proposal to clarify the purpose of U+202F NARROW NO-BREAK SPACE, L2/19-115 Proposal to ensure usability of fixed-width spaces, and L2/19-112 Proposal to define a space character as a group separator — where we proposed to “[w]ait until a decision is made about the fate of NARROW NO-BREAK SPACE in Mongol script” — it attempts to help settle issues that will be vital to CLDR survey #36 scheduled to start on May 23, 2019.

With respect to the meeting deadline, an early version was submitted on April 23, 2019 at 17:37 Pacific Time. However, due to various defects, submitting revisions turned out being desirable.

However, I just got aware that Unicode version 12.1.0 does not wait for UTC #159 decisions, and that NamesList.txt has already been posted in final version 12.1.0 on April 23, 2019 (i.e. yesterday, 18 hours ahead when adding this last comment). Due to other urgencies, adding more proposed actions in this paper is delayed.

Problem

In L2/19-112 we proposed to “[w]ait until a decision is made about the fate of NARROW NO-BREAK SPACE in Mongol script.” But Mongolian Working Group meeting #3 resulted in an unanimous preference among experts from both Mongolia and China / Inner Mongolia Autonomous Region for not keeping NNBSP as a format control in Mongol script. NNBSP should preferably be reduced to its whitespace role, for use at the discretion of the end-user, while its glyph shape controlling semantics is to be transferred to an existing dedicated format control. Also, the total number of format controls in use in Mongol script is wished to be drastically reduced.
Proposal to establish NARROW NO-BREAK SPACE as a definitely usable avatar of THIN SPACE

In front of the situation, maintaining the status quo seems to be no option. The replacement of NNBSP with U+180E MONGOLIAN VOWEL SEPARATOR (MVS) is currently investigated, and MVS is being re-engineered for that purpose. As a consequence, NNBSP shall be left as nothing else than a general-use whitespace with certain properties, in accordance with the way that Unicode intended it to be used in non-Mongolian scripts (see L2/98-268R, quoted in L2/19-112). For details, please check the compilation of quotations in the Background section below.

The new situation completely changes the deal. Instead of making U+2009 THIN SPACE usable by correcting its Line_Break property value from "BA" to "GL", Unicode now only needs to establish U+202F NNBSP, and to actively promote it as being indispensable to virtually all locales in virtue of the BIPM recommendation to use space as a group separator (see L2/19-112).

This outcome has the advantage of not breaking existing practice nor keyboard layouts such as NF Z71-300:2019 of the French NB, where the BÉPO variant (Annex B) has NNBSP in the Shift shift-state of the Space bar in synergy with big punctuation marks and digits for French-specific punctuation spacing and SI-conformant representation of numbers (while in Annex A it is on level 4 thereof).

That late move obviously brings the need to reword some parts of the Standard. In particular, the very name "NARROW NO-BREAK SPACE" is misleading, given that it suggests a proximity to U+00A0 NO-BREAK SPACE. Tailoring in word processors aside, NBSP is justifying like U+0020 SPACE. Therefore, while NO-BREAK SPACE is truly a non-breaking variant of SPACE, the same is false for NARROW NO-BREAK SPACE with respect to NO-BREAK SPACE. In reality, NNBSP is not a narrow variant of NBSP. It is not, for the single reason that it has not the same behavior in horizontal justification. A number of errors need to be fixed in the Standard in order to remove that ambiguity.

Background

The Mongolian Working Group (MWG) held meeting #3 in Ulaanbaatar on April 3–5, 2019. The comprehensive set of meeting documents is posted at: Documents for Mongolian Working Group Meeting 3.

Lisa Moore, Meeting Moderator, reported in MWG/3-N18R2, L2/19-139 [14], in the list of areas of significant interest, NNBSP as #1. How exactly it will be replaced has not yet been agreed, though:

These areas of significant interest need more detail before implementation can begin:

SI1. NNBSP. Inner Mongolia and Mongolia think NNBSP is very problematic and should not be used.

[...]

The working group endorsed the following resolutions:

MR1. Format control characters. To address concerns that were raised regarding Mongolian format control characters as currently implemented, clearer specifications for the behavior and use of Mongolian format control characters need to be agreed to and documented.

[...]
Several issues were raised that require further study. The following was agreed:

[...]

2. **MVS.** Liang Hai and Liang Jinbao will investigate new rules for MVS, to establish if it can be used in place of NNBSP.

**Roozbeh Pournader** summed up the MWG/3 proposals in MWG/3-N20, [L2/19-141](#) [16]. Four proponents among the eight referenced in the document addressed NNBSP:

**B. Proposals made by CITA-WG**

[...]

2. NNBSP should not be used as a suffix separator. MVS should be used as a suffix separator instead.

[...]

**C. Proposals made by Bolorsoft**

[...]

1. The use of FVS2, FVS3, NNBSP, and ZWNJ in the Mongolian encoding model should be eliminated.

[...]

7. MVS should be used instead of NNBSP in Mongolian text.

[...]

**E. Proposals made by Liang Jinbao**

[...]

3. The use of NNBSP should be abandoned. MVS should be renamed to MSC and used in places where currently either MVS or NNBSP is used.

[...]

**G. Proposals made by Anand Orkhonbaatar**

[...]

4. Special shaping behavior of NNBSP should be removed. Suffixes that need special shaping should use FVS or ZWJ. Users should be able to use either a space or NNBSP to separate suffixes.

The **Mongolian Academy of Sciences** issued *General comments and suggestions on the Mongolian Working Group 3 Meeting, MWG/3-N22, [L2/19-142](#) [17].* It is confident that the NNBSP format control issue can be resolved as a result of the meeting:

1. **The issues which can be resolved as a result of the meeting**
3. To exclude NNBSP character or integrate its role into MVS character operations (B.Enkhdalai, Shen Yilei, Liang Jinbao, Bolorsoft, Roozbeh Pournader)

The following is a review of other MWG/3 documents included in the UTC Document Register. Not all submissions do address or even mention the NNBSP issue. For the purpose of this paper, those that do not are blanked out here, but are still listed in the References section below.

Let’s start with a specification documenting the actual state of the art. This work sets a precedent for further development, and defines a benchmark methodology:

Liang Hai drafted a specification for the Hudum style Mongol script as a template for the necessary design process of a comprehensive Mongol specification, which is the basis of interoperability: *Towards a well-formed Mongolian specification that allows interoperable implementations*, MWG/3-N8R, L2/19-130 [5]. Liang Hai inventories NNBSP among the format controls in the *Minimal character set* (§ 3.1) for *Text representation* (§ 3):

**Vowel Separator (MVS) and Narrow No-Break Space (NNBSP).** MVS is a Mongolian-specific format control for requesting the splash variation. It is transcribed as “·” (a middle dot). See “The splash” in Section 3.5, *Additional variation patterns*. NNBSP is a Mongolian-specific whitespace and format control for marking and shaping enclitics, and note that it is also used as a general whitespace by other scripts. See “Enclitics” in Section 3.5. In terms of cursive joining behavior, MVS and NNBSP are both non-joining inline characters, like an ordinary space.

Later in the document, Liang Hai carves out the distinction between the use of MVS and the use of NNBSP, as the former controls shaping of the splash, the latter, of first letters in enclitic particles, also referred to as suffixes:

**The splash**

This non-joining grapheme of a and ä is usually observed after n, masculine x, masculine g, m, l, y, r, w, and only occasionally after j. The format control MVS is used to break cursive joining between the preceding consonant and a/ä, as well as to request the special forms required for certain consonants:

- Letters n, x, and g take what appear to be their special, onset final forms. .................. Predictive with MVS
- Letters y, w, and j take their under-differentiated forms (i.e., yodh, waw, and yodh, respectively). Predictive with MVS
- Writing in the form of splash is orthographically mandatory for a to appear after a word-medial x/g.
- Note that x and y normally do not have final forms as they are not used as codas.

**Enclitics**

The special whitespace between an enclitic and its preceding word (either the modified word or a preceding enclitic) is sometimes considered an internal gap in a grammatical word, and thus is sometimes preferred to be non-line-breaking, non-word-breaking, and narrower than an ordinary space. Typical enclitics exhibit at least one of the following variations:
Proposal to establish NARROW NO-BREAK SPACE as a definitely usable avatar of THIN SPACE

- Absence of onset placeholder and initial-body variation (because the beginning of an enclitic it is not the beginning of a stem). 

- An initial d takes its disambiguating lamedh form. 

- The under-differentiated, historical form of y. 

- Letters x and g show gender harmony with the preceding word. The gender features are only observable on these two letters because the first feature has neutralized gender-distinguishing features of all vowel letters.

NNBSP is used both for representing this whitespace and for requesting the special variations shown in enclitics. As the set of enclitics and usage of NNBSP are decided grammatical, an enclitic may or may not exhibit special variations but is still encoded with a preceding NNBSP. See Appendix A for a reference list and comparison.

The following table shows how the shaping process is more complex for Hudum Mongolian than for other cursive scripts (source: page 18 of [5]). NNBSP intervenes in shaping step 4 of shaping phase III:

**Table 6. Minimal shaping process**

<table>
<thead>
<tr>
<th>Shaping phase</th>
<th>Shaping step</th>
</tr>
</thead>
<tbody>
<tr>
<td>IA. General</td>
<td>- Basic character-to_glyph mapping</td>
</tr>
<tr>
<td>II. Cursive script</td>
<td>- Initiation of positional forms</td>
</tr>
<tr>
<td>III. Hudum-specific</td>
<td></td>
</tr>
<tr>
<td>Reduction of phonetic letters to graphemes</td>
<td></td>
</tr>
<tr>
<td><strong>Phonetic:</strong></td>
<td></td>
</tr>
<tr>
<td>1. Onset and coda</td>
<td></td>
</tr>
<tr>
<td>2. Gender-specific</td>
<td></td>
</tr>
<tr>
<td>3. MVS-involving</td>
<td></td>
</tr>
<tr>
<td>4. NNBSP-involving</td>
<td></td>
</tr>
<tr>
<td><strong>Graphemic:</strong></td>
<td></td>
</tr>
<tr>
<td>5. Offglide</td>
<td></td>
</tr>
<tr>
<td>6. Graphemic post-bowed</td>
<td></td>
</tr>
<tr>
<td><strong>Uncaptured:</strong></td>
<td>7. FVS-selected</td>
</tr>
</tbody>
</table>

The following detailed overview of steps 3 and 4, from Table 7 on page 20 of [5], shows that the use cases of MVS and of NNBSP seem to have a non-empty intersection, as “a” may occur after either of them:
In a different approach, the Working Group of the Communications and Information Technology Authority, Mongolia (CITA-WG) addressed format control streamlining in general, and NNBSP in particular, in its presentation about Improvement of Mongolian Unicode Block, MWG/3-N5, L2/19-128 [3]. The first step is to design fonts in a way that they do not use U+180E MONGOLIAN VOWEL SEPARATOR (MVS). Then, MVS is to take over the formatting role of NNBSP, so that in the future, NNBSP may or may not be used, lifting any constraint on the end-user as of typing a non-breaking space rather than a normal space. This new flexibility is appreciable especially where suffixes or suffix concatenations are long:

- MVS – used not as a Vowel Separator, but as a Suffix Separator (solved problems caused by NNBSP)

By the means of a re-engineered font, the CITA-WG demonstrates how it is possible to write in Mongol script without needing MVS, and as a consequence, how MVS can be used to replace NNBSP in its shape controlling role (slides 4 and 5 of [3]):

$$
\begin{array}{|c|c|}
\hline
3. \text{ MVS-involving} & \\
\hline
n/y/y/w & \text{if precedes MVS:} & \text{pre-splash} \\
\hline
x/g & \text{if precedes MVS that does not precede } d: & \text{pre-splash} \\
\hline
a/\bar{a} & \text{if follows MVS:} & \text{splash} \\
\hline
4. \text{ NNBSP-involving} & \\
\hline
a/i/u/\bar{u}/d & \text{if follows NNBSP:} & \text{enclitic-initial} \\
\hline
o/u/o/\bar{u} & \text{if follows an initial consonant that does not follow NNBSP:} & \text{initial-body} \\
\hline
Existing data is always a drawback for making changes. But Mongol script using communities, under the leadership of the Communications and Information Technology Authority, Mongolia, are ready to convert the data to the new rule by applying a very easy-to-implement process, presented on slide 10 of [3]:

---

**MVS – never used as a Vowel Separator!**

<table>
<thead>
<tr>
<th>Standard Font</th>
<th>Custom New Font</th>
</tr>
</thead>
<tbody>
<tr>
<td>НУТУГУН, ХУМУСИ,</td>
<td>НУТУГУН, ХУМУСИ,</td>
</tr>
<tr>
<td>ХИЛАГАЙ, НИГУРТЭЙ,</td>
<td>НУТУГУН, ХУМУСИ,</td>
</tr>
<tr>
<td>НИГУРТЭЙ,</td>
<td>НИГУРТЭЙ,</td>
</tr>
<tr>
<td>МУНГУР</td>
<td>МУНГУР</td>
</tr>
</tbody>
</table>

| MVS is used as vowel separator | There is no any format control used as vowel separator. |

---

**MVS – used as a Suffix Separator only!**

<table>
<thead>
<tr>
<th>Standard Font</th>
<th>Custom New Font</th>
</tr>
</thead>
<tbody>
<tr>
<td>НУТУГУН, ХУМУСИ,</td>
<td>НУТУГУН, ХУМУСИ,</td>
</tr>
<tr>
<td>ХИЛАГАЙ, НИГУРТЭЙ,</td>
<td>НУТУГУН, ХУМУСИ,</td>
</tr>
<tr>
<td>НИГУРТЭЙ,</td>
<td>НИГУРТЭЙ,</td>
</tr>
<tr>
<td>МУНГУР</td>
<td>МУНГУР</td>
</tr>
</tbody>
</table>

| NNBSP is used as Suffix Separator | No NNBSP, but MVS used instead of NNBSP as Suffix Separator |

Proposal to establish NARROW NO-BREAK SPACE as a definitely usable avatar of THIN SPACE

How to migrate existing data to new rule

- It is very easy.
  1. Replace all occurrences of MVS within whole text with empty string, i.e. delete all MVS.
  2. Replace all NNBSP with MVS.
  3. May be some small editing is required

The Bolosoft team, in turn, pointed the architectural mistakes in its presentation of *The refined phonetic model: The most convenient solution for recovering the current Mongolian encoding*, MWG/3-N7R2, L2/19-129 [4], recalling also its previously submitted *Solution for NNBSP Issues*, L2/18-293:

II. Problems: The fine-grained problems [cont.]
Architectural mistakes

NNBSP issues
The NNBSP is broken in Unicode standard. It never works flawlessly.
This character is defined as a space character and has additional trio functions.
It is used to display a narrow space.
It is involved in Mongolian shaping.
No reaction for our proposal L2/18-293.

[...]

Complex design and poor specification

- Semantic encoding
- Cursive joining
- Too many special characters such as NNBSP, MVS, FVS1, FVS2, FVS3, ZWJ, ZWNJ

[...]

IV. Solution: Replacing NNBSP by MVS
About proposal

- In 2018, we have already proposed to solve NNBSP issues introducing the MSC - Mongolian Suffix Connector.
- In “Proposed solution” on page 14 we have already mentioned that in meantime the MVS could be used as NNBSP.
- However, we didn’t receive any approval or constructive feedbacks to make decision.

Explanatory statement

Similarity:

- The functionality of MVS and NNBSP is very similar.
- MVS joins disjoint A and E.
- NNBSP joins the suffixes to its stem word or preceding suffix.

Differences:

MVS is a format control character, NNBSP is a space separator.

[...]
• The majority of users was new.
• Almost every new user asked about how to write suffixes.
• Almost 70% of users didn’t distinguish between MVS and NNBSP. NNBSP is simply used instead of MVS.
• Few users confused between KE and GE.
• The existing users are rather critical. They are still testing our solution.
• The majority of users were happy for the simple writing.
• NNBSP is still not supported well by major vendors like Webkit for MacOS and iOS. (Older versions were flawless. It means NNBSP implementation instable.)

**Can MVS fulfill all requirements of MSC?**

As mentioned in our proposal:
• In encoding level just one marker is enough for suffix joining.
• MVS solves all word boundary problems and word selection problems.
• For the shaping, MVS is far reliable than NNBSP.
• All other functionality solved by fonts and space is manipulated by fonts.
• The only deficit is line breaking. Anyways, NNBSP has also line breaking Issue.

**Decision and Release plan**
• From our study, we have decided to replace NNBSP by MVS.
• We have planned to release the Linux version in next week.
• Mobile versions are in demand. Release comes in four weeks.

**IV. Solution: Specification & documentation**
• Cleaning up unnecessary variants.
• KE and GE letters are disenfied.
• Reducing format control characters: FVS1, MVS, ZWI, [black print = these are kept]
  NNBSP, ZWNJ, FVS3, FVS2 [gray print = these are dropped]
• Basic rules for writing
• Documentation: [https://wiki.mngl.net](https://wiki.mngl.net)

[...]

• We have released our solution as “proof of concept” (only for Windows and Mac OS) to collect end user feedbacks (for MVS vs NNBSP, FVS reduction, KE, GE).

**Liang Jinbao** proposed in the same vein *An introduction to the execution of our Solution with Minimal Modifications, Subtitle: for achieving a stable and unified Mongolian encoding as soon as possible, MWG/3-N11R, L2/19-133 [8]:*

**Solve problems caused by NNBSP with MSC model**

**Status:** When using NNBSP in Mongolian, there will always be cases where the extra component is displayed incorrectly in all or part of the version of IOS, MacOS, Android, Office, Browser. This phenomenon has not been solved for a long time. There is only one way to solve this problem quickly and completely: abandon the use of NNBSP.

**Solution: Combine MVS&NNBSP into MSC**
There are two separators used in Mongolian: MVS and NNBSP. The function is
1. continuous word,
Proposal to establish NARROW NO-BREAK SPACE as a definitely usable avatar of THIN SPACE

2. split word vowel A/E,
3. split additional component.
We renamed MVS as MSC and used MSC wherever there is MVS and NNBSP requirements, which is a complete replacement for these two control symbols.

Anand Orkhonbaatar presented on his side Some possible improvement ideas to the current Mongolian encoding, MWG/3-N17, L2/19-138 [13], getting the shaping right by the use of MONGOLIAN FREE VARIATION SELECTOR (FVS) ONE, TWO or THREE (U+180B, U+180C, U+180D), or simply by a leading U+200D ZERO WIDTH JOINER (ZWJ):

3. Possible ways to write separated suffixes without NNBSP

Using FVSx to write suffixes
• NNBSP and SP would be the choice of the person who is using.
• In many documents, there are lines starting with suffixes.

Using ZWJ to write suffixes
• ZWJ is used to show the joining variation, just what exactly Mongolian separated suffixes nature is

Proposal which already made on the usage of FVSx
• https://www.unicode.org/L2/L2018/18099-mwg2-1-mong-proposal.pdf

Using ZWJ to write suffixes
q+o+t+a+n(2 nd v.) +SP/NNBSP+ZWJ+a(2 nd v.)

n+o+m+SP/NNBSP+ZWJ+u+n

n+a+r+a+SP/NNBSP+ZWJ+y(2 nd v.)+i

The above demonstration of the ZWJ solution is copy-pasted from final slide 16 of [13]:
We see that there is an overwhelming demand to free NNBSP from its compound semantics. Bundling a whitespace and a format control into an all-in-one character is not helpful in Mongol script, because it artificially restricts the layout options and makes the rendering unstraightforward when text is broken into narrow columns, while Mongolian suffixes may be long and can also concatenate together. Additionally, the actual proliferation of format control characters in Mongol script is very confusing for the end-user.

For all those reasons, the Mongolian encoding model is now being streamlined. One result of the process would be that the NARROW NO-BREAK SPACE falls fully into the common domain, for use with Mongol script and with any other script, like the character that it was intended to be when Unicode proposed to move the upcoming *MONGOLIAN SPACE into the General Punctuation block: A replacement of what U+2009 THIN SPACE should have been when Unicode started up.

In other words: Thanks to the Mongol Script and to Unicode, NARROW NO-BREAK SPACE has become the definitely usable avatar of the damaged THIN SPACE.

Proposed Actions

In this paper, the Proposed Actions section focuses on those data that can actually be updated for the upcoming version 12.1.0 of the Unicode Standard. That excludes all changes to the Core Specification and to the Standard Annexes. Only those changes are actionable (in the short term) that affect files in the Unicode Character Database (UCD). Changes beyond that limited scope are adjourned for now, with respect to the submission deadline of UTC meeting #159 (please refer to the Proposal history section above).
Proposal to establish NARROW NO-BREAK SPACE as a definitely usable avatar of THIN SPACE

1. In NamesList.txt, before U+202F NARROW NO-BREAK SPACE: Add the subheading “Space”.
2. Ibid., at U+202F: Add the informative alias "no-break thin space".
3. Ibid.: Change the second comment line from “a narrow form of a no-break space, typically the width of a thin space or a mid space” to “a non-breaking form of a thin space, with fixed width”, and in a new comment line, add: "some fonts give it the width of a mid space (2005)".
4. Ibid.: Add the following comment lines:
   4.1. “used to append suffixes in Mongolian” [This needs to be adjusted in the near future.];
   4.2. “used as a group separator in SI-compliant locales”;
   4.3. “used for French punctuation spacing”;
   4.4. “used in English to improve legibility of consecutive single and double quotation marks”.
5. Ibid., at U+2009 THIN SPACE: Make clear that this is unusable on the internet, i.e. in untailored environments (InDesign and TeX tailor it to be non-breaking), and urge users to resort to using the (already x-refed) NNBSP instead to achieve all intended effects.

[Further suggestions are adjourned, as NamesList.txt 12.1.0 has been posted in final version on April 23, 2019 (18 hours before writing this...).]

References


Proposal to establish NARROW NO-BREAK SPACE as a definitely usable avatar of THIN SPACE


**Acknowledgments**

Thanks to everyone who directly or indirectly helped put this paper together.

Thanks to Microsoft for Word Online and OneDrive.

Thanks to Google for Google Search.