

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	Proposal to encode one Manchu format character		
2. Requester's name:	MA Xudong (xudong.ma@gmail.com) WANG Shuo (150374704@qq.com) JIN Biao (63316836@qq.com) WANG Jun (sahalin@qq.com)		
3. Requester type (Member body/Liaison/Individual contribution):	Individual contribution		
4. Submission date:	2017-01-16		
5. Requester's reference (if applicable):	N/A		
6. Choose one of the following:			
This is a complete proposal:			YES
(or) More information will be provided later:			

B. Technical – General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	NO	
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	YES	
Name of the existing block:	MONGOLIAN	
2. Number of characters in proposal:	1	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary	B.1-Specialized (small collection)	X
C-Major extinct	D-Attested extinct	
F-Archaic Hieroglyphic or Ideographic	G-Obscure or questionable usage symbols	
B.2-Specialized (large collection)	E-Minor extinct	
4. Is a repertoire including character names provided?	YES	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	YES	
b. Are the character shapes attached in a legible form suitable for review?	YES	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	MA Xudong	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):		
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	NO	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	YES	
7. Special encoding issues:		
Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?	YES	
	See Further Information	

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database (<http://www.unicode.org/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)


C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	NO
If YES explain	
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?	YES
If YES, with whom?	
If YES, available relevant documents:	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	NO
Reference:	
4. The context of use for the proposed characters (type of use; common or rare)	COMMON
Reference:	
5. Are the proposed characters in current use by the user community?	NO
If YES, where? Reference:	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	YES
If YES, is a rationale provided?	
If YES, reference:	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	YES
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	NO
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	NO
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character?	YES
If YES, is a rationale for its inclusion provided?	
If YES, reference: See Further Information	
11. Does the proposal include use of combining characters and/or use of composite sequences?	NO
If YES, is a rationale for such use provided?	
If YES, reference:	
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	
If YES, reference:	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	Yes
If YES, describe in detail (include attachment if necessary)	
See Further Information	
13. Does the proposal contain any Ideographic compatibility characters?	NO
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	

Further Information

1. Proposed Character

One character is proposed to replace U+202F NARROW NO-BREAK SPACE (NNBSP) in order to present Manchu and Sibe correctly.

Code Point	Proposed Character Name	Representative Glyph
181A	MONGOLIAN MANCHU WORD JOINER	

2. Character Properties

General Properties

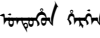
181A;MONGOLIAN MANCHU WORD JOINER;Cf;0;BN;;;;;N;;;;;

Line Break Properties

181A;GL # Cf MONGOLIAN MANCHU WORD JOINER

3. Details and Attestations

The Manchu alphabet was converted from the Mongolian alphabet. However, the Manchu script is very different from the Mongolian script in most aspects including the letters, variants, transforming rules, syllables, pronunciations, whiting styles, etc.

In Manchu, certain function words (Manchu:  *untuhun hergen*; Chinese: 虚字 *xūzì*, 虚词 *xūcí*) are usually written separated from the previous word by a seemingly ordinary whitespace, and in certain situations some of them can also be attached to the previous word. When written separated, these function words are considered to be independent words, although a line break opportunity does not occur before them. These certain function words when written separated hereinafter are referred to as the separated special function words.

The separated special function words include the case markers (ᠶ -*i*, ᠨᠢ -*ni*, ᠪᠡ -*be*, ᠳᠡ -*de*, ᠳᠡᠷᠢ -*deri*, ᠴᠢ -*qi*), the modal particles (ᠬᠠᠢ -*kai*, ᠨᠢ -*ni*, etc.), etc. Altogether there are six case markers, so there are more than six separated special function words in Manchu.

The whitespace is not a narrow gap. Its width is the same as an ordinary space, and also flexible as an ordinary space.

In Sibe, according to the current orthography, only the case marker ᠶ -*i* follows the same way as the separated special function words in Manchu. A line break opportunity can occur before all the other words in Sibe.

In Manchu and Sibe, only one of the separated special function words, i.e. the case marker ᠶ -*i*, is irregular, which is written the form like a small stand-alone final form of ᠶ *i*. All the other separated special function words are regular and written the same as the content words that share the same pronunciation.¹

¹ In early New Manchu (ᠲᠣᠩᠭᠢᠰᠢᠳᠠᠬᠤᠬᠡᠷᠭᠡᠨ *tongki fuka sindaha hergen*), ᠳᠡ -*de* was sometimes in irregular form as ᠳᠡᠳᠡ.

The separated special function words in Manchu and Sibe must be marked manually as the separated suffixes in Mongolian and Todo.

The following is about NNBSP in page 533, *The Unicode Standard Version 9.0 – Core Specification*:

Narrow No-Break Space. In Mongolian, Todo, Manchu, and Sibe, certain grammatical suffixes are separated from the stem of a word or from other suffixes by a narrow gap. There are many such suffixes in Mongolian, usually occurring in masculine and feminine pairs (for example, the dative suffixes *-dur* and *-dür*), and a stem may take multiple suffixes. In contrast, there are only six separated suffixes for Manchu and Sibe, and stems do not take more than one suffix at a time.

As any suffixes are considered to be an integral part of the word as a whole, a line break opportunity does not occur before a suffix, and the whitespace is represented using U+202F NARROW NO-BREAK SPACE (NNBSP). For a Mongolian font it is recommended that the width of NNBSP should be one-third the width of an ordinary space (U+0020 SPACE).

NNBSP affects the form of the preceding and following letters. The final letter of the stem or suffix preceding the NNBSP takes the final positional form, whereas the first letter of the suffix following NNBSP may take the normal initial form, a variant initial form, a medial form, or a final form, depending on the particular suffix.

Obviously, the above descriptions about Manchu and Sibe are incorrect, and U+202F NARROW NO-BREAK SPACE (NNBSP), which is currently used in Mongolian, Todo, Manchu and Sibe to separate the separated suffixes or the separated special function words, is not correct in Manchu and Sibe.

1. The width of NNBSP is narrower than that Manchu and Sibe need.
2. The width of NNBSP is fixed, while what Manchu and Sibe need is a flexible one.
3. The suffixes following NNBSP are considered to be an integral part of the word as a whole in Mongolian and Todo. But in Manchu and Sibe the separated special function words are considered to be independent words.

U+00A0 NO-BREAK SPACE (NBSP), U+2060 WORD JOINER (WJ), U+FEFF ZERO WIDTH NO-BREAK SPACE (ZWNBS) (use as an indication of non-breaking is deprecated), or the proposed “U+180F MONGOLIAN SUFFIX CONNECTOR (MSC)” are not suitable for the purpose in Manchu and Sibe, either. One of the reasons is that we cannot distinguish between the case marker ᠰ *-i* and the subjective non-breaking normal ᠰ *i* with any of these characters.

The character should be an objective must in Manchu and Sibe. A proper format character should be encoded for the special purpose in Manchu and Sibe.

1. It has no visual form. (→ WJ)
2. It indicates that there should be no line break at the point. (→ WJ, NBSP, NNBSP, MSC)
3. The first letter of the separated special function words following it takes the proper form depending on the particular separated special function words. (→ NNBSP, MSC)
4. It is used in Manchu before the separated special function words.
5. It is used in Sibe before the separated special function word, i.e. the case marker ᠰ *-i*.
6. The suggested name of the character is Mongolian Manchu Word Joiner (MMWJ). This character is also used in Sibe, but mainly used in Manchu, so the name Mongolian Manchu Word Joiner is

more proper than the name Mongolian Sibe Word Joiner.

7. The suggested code point is U+181A.

8. The space before it is an ordinary space (U+0020 SPACE).

Use in Manchu

Input: 1820 182A 1874 1820 0020 181A 1869 185D 0020 ...
ᡤᡠᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ...

Align Left: ᡤᡠᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ
Distributed: ᡤᡠᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ
Transliteration: *abka-de deyere gasha bi.*

Input: 182A 1873 1868 1865 185D 0020 181A 1873 0020 ...
ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ...

Align Left: ᡳᡳᡳᡳ ᡳᡳᡳᡳ
Distributed: ᡳᡳᡳᡳ ᡳᡳᡳᡳ
Transliteration: *bithe-i dobton*

Input: 182A 1873 1868 1865 185D 1873 0020 ...
ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ...

Align Left: ᡳᡳᡳᡳ ᡳᡳᡳᡳ
Distributed: ᡳᡳᡳᡳ ᡳᡳᡳᡳ
Transliteration: *bithei dobton*

Use in Sibe

Input: 182A 185E 1868 1865 185D 0020 181A 185E 0020 ...
ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ...

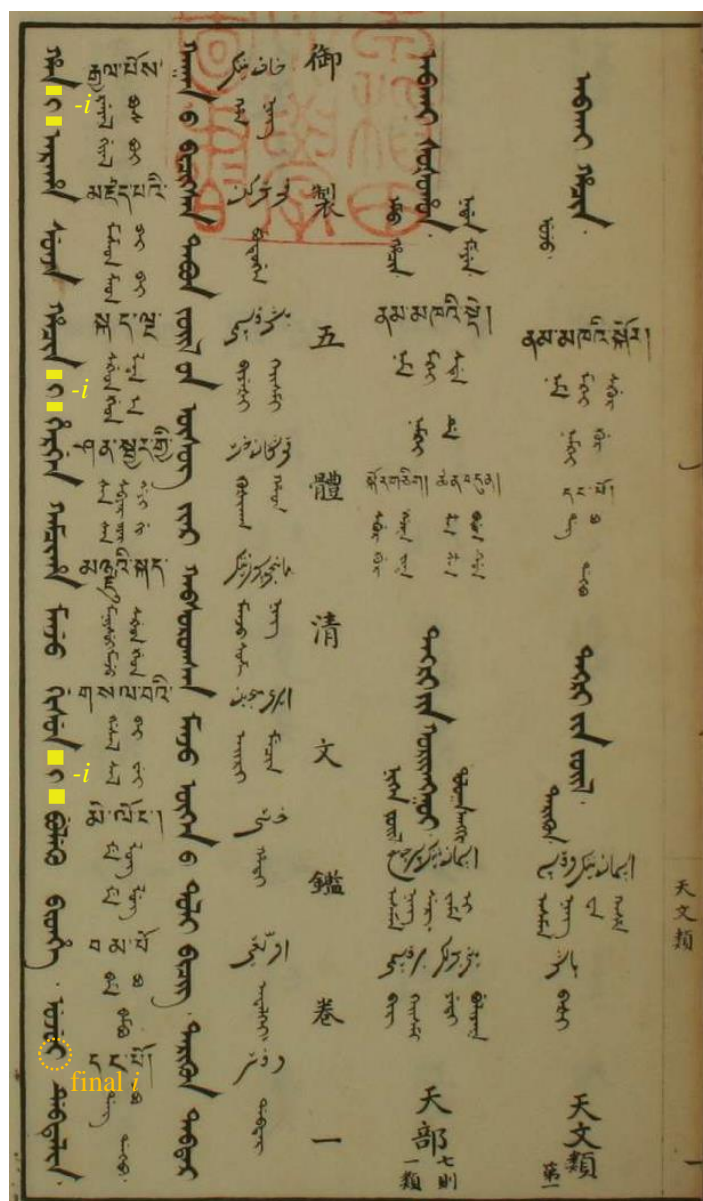
Align Left: ᡳᡳᡳᡳ ᡳᡳᡳᡳ
Distributed: ᡳᡳᡳᡳ ᡳᡳᡳᡳ
Transliteration: *bithe-i dobton*

Input: 182A 185E 1868 1865 185D 185E 0020 ...
ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ᡳᡳᡳᡳ ...

Align Left: ᡳᡳᡳᡳ ᡳᡳᡳᡳ
Distributed: ᡳᡳᡳᡳ ᡳᡳᡳᡳ
Transliteration: *bithei dobton*



4. Text Examples

4.1 Example 1

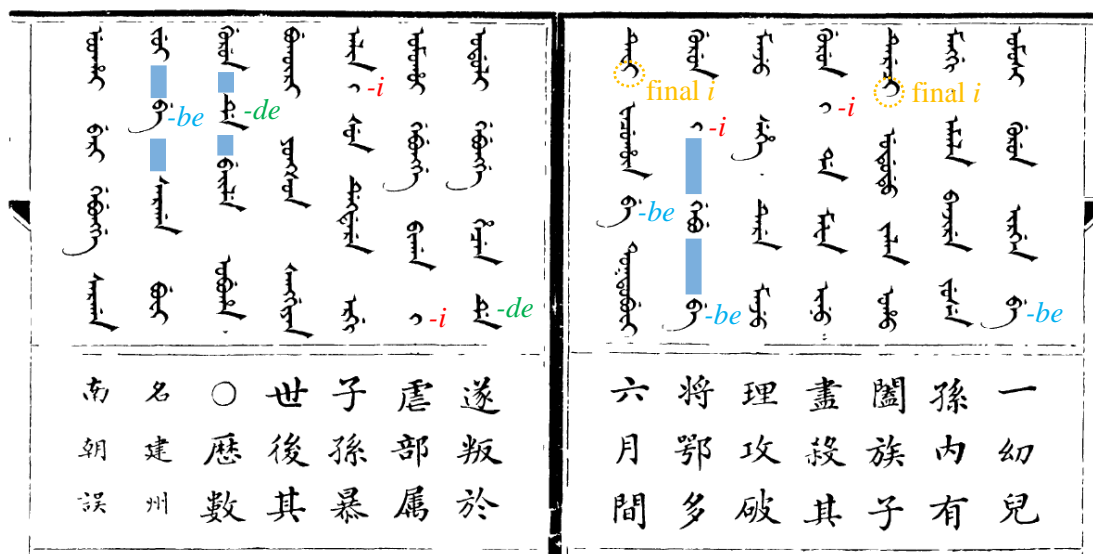


《御制五体清文鑑》 (Yu Zhi Wu Ti Qing Wen Jian): p. 1, vol. 1

The whitespace above and below the case marker -i are in the same size. The case marker -i is smaller than the final form of *i* as in *ujui*.

the case marker -i	the word <i>ujui</i>
	

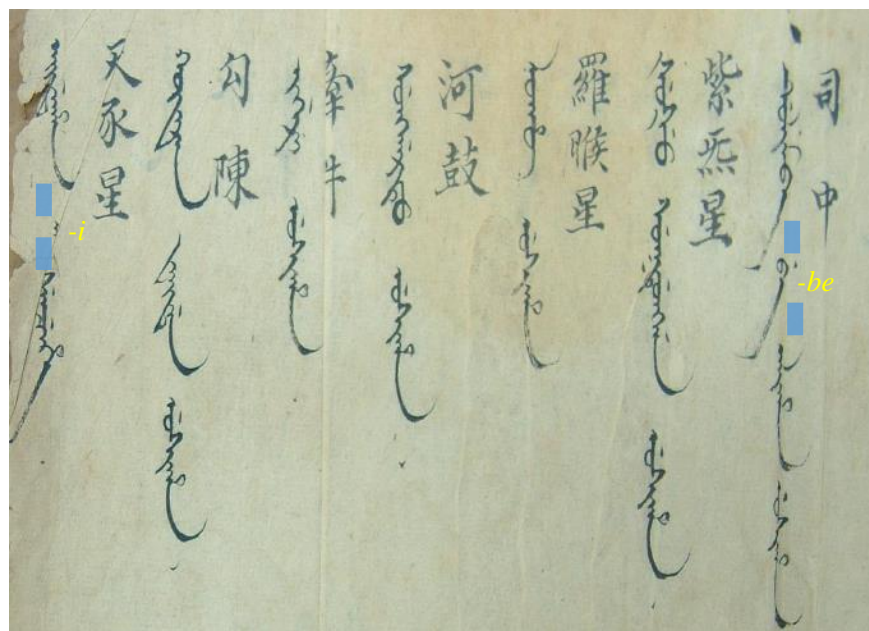
4.2 Example 2



《满洲实录》 (Man Zhou Shi Lu): vol. 1

The spaces above the case markers *-i*, *-be*, *-de* are flexible and in the same size as the other spaces in the same line. The case marker *-i* is smaller than the final form of *i* as in other words.

4.3 Example 3



《钦定新清语》 (Qin Ding Xin Qing Yu): p. 2

The above Manchu is written in the semi-cursive script (running script). The case marker *-i* and *-be* are treated in the same way as that in the regular script.

4.4 Example 4



《圆觉经》(Yuan Jue Jing): p. 25, vol. 1

5. References

5.1 Reference 1

The Unicode Standard Version 9.0 – Core Specification, p. 269:

Narrow No-Break Space. U+202F NARROW NO-BREAK SPACE (NNBSP) is a narrow version of U+00A0 NO-BREAK SPACE, which except for its display width behaves exactly the same in its line breaking behavior. It is regularly used in Mongolian in certain grammatical contexts (before a particle), where it also influences the shaping of the glyphs for the particle. In Mongolian text, the NNBSP is typically displayed with 1/3 the width of a normal space character. The NNBSP can be used to represent the narrow space occurring around punctuation characters in French typography, which is called an “espace fine insécable.”

5.2 Reference 2

《GB 26226-2010-T 信息技术 蒙古文变形显现字符集和控制字符使用规则》, p. 22:

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。

5.3 Reference 3

《清文启蒙》(*Qing Wen Qi Meng*), vol. 3:

凡如 ㄣ、ㄥ、ㄨ、ㄩ、ㄨ、ㄩ 等虚字，不可提写在行首，若系实解，或作汉话用者，方可提起写得。

5.4 Reference 4

《现代锡伯文学语言正字法》(*Xian Dai Xi Bo Wen Xue Yu Yan Zheng Zi Fa*), pp. 90-91:

移行时“i”和标点符号不能移至行首。应写在行末。