JTC1/SC2/WG2 N4717
2016-04-21

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type:	**Working Group Document**
Title:	**Proposal to encode one Bopomofo letter**
Source:	**Andrew West**
Status:	**Individual Contribution**
Action:	**For consideration by JTC1/SC2/WG2 and UTC**
Date:	**2016-04-21**

1. Introduction

This is a proposal to encode one Bopomofo letter.

312F	BOPOMOFO LETTER NN	ㄯ

Mandarin Chinese has a few characters that are pronounced as a syllabic nasal, *m*, *n*, or *ng*. In the PRC (e.g. in *Xiàndài Hànyǔ Cídiǎn* 现代汉语词典 [Contemporary Chinese Dictionary]) these syllabic nasals are represented using U+3107 BOPOMOFO LETTER M ㄇ, U+310B BOPOMOFO LETTER N ㄋ, and U+312B BOPOMOFO LETTER NG ㄫ.

However, in early sources, notably the 1937 *Guóyǔ cídiǎn* 國語辭典 [Dictionary of the National Language], and in some Taiwan books, the syllabic nasals are represented by a set of special Bopomofo letters derived from the Bopomofo letters M, N, and NG, but with a vertical stroke in the middle (see Fig. 1, Fig. 2, and Fig. 3). The syllabic letters M and NG with a vertical stroke are already encoded as U+31AC BOPOMOFO LETTER IM ㆬ and U+31AD BOPOMOFO LETTER NGG ㆭ. However, the corresponding syllabic letter N, derived from U+310B BOPOMOFO LETTER N with a vertical stroke in the middle, is not currently encoded. This is problematic, as explained in the Chinese Wikipedia article on Bopomofo (see Fig. 4).

Fig. 1 shows the unencoded Bopomofo letter NN with an attached falling tone mark. This mark is not part of the character.

Table 1: Different Representations of Syllabic Nasals

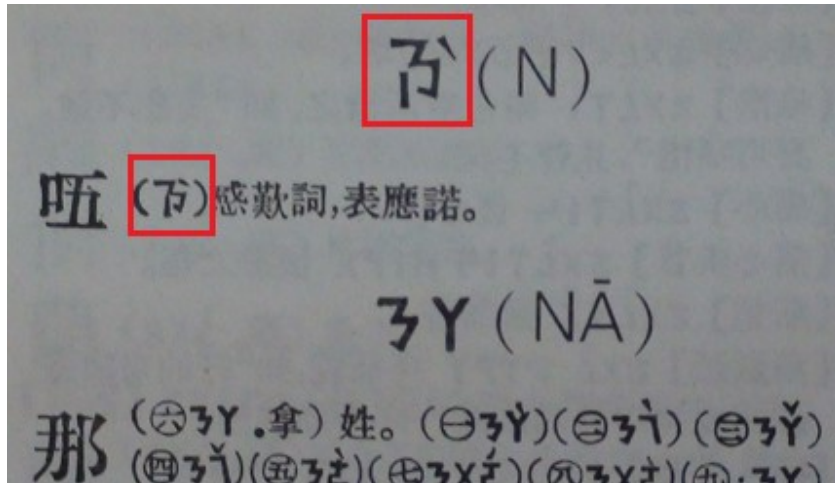| Syllable (pinyin)	*Xiàndài hànyǔ cídiǎn* 现代汉语词典	*Guóyǔ cídiǎn* 國語辭典
hm	ㄏㄇ	ㄏ||||
hng	ㄏㄫ	ㄏㄫ
m	ㄇ	|||
n	ㄋ	ㄋ
ng	ㄫ	ㄫ

2. Unicode Properties

312F;BOPOMOFO LETTER NN;Lo;0;L;;;;;N;;;;;

Other properties are the same as for other Bopomofo letters.

3. Figures

Fig. 1: *Guóyǔ cídiǎn* 國語辭典 p. 249

Showing the unencoded character with a falling tone mark
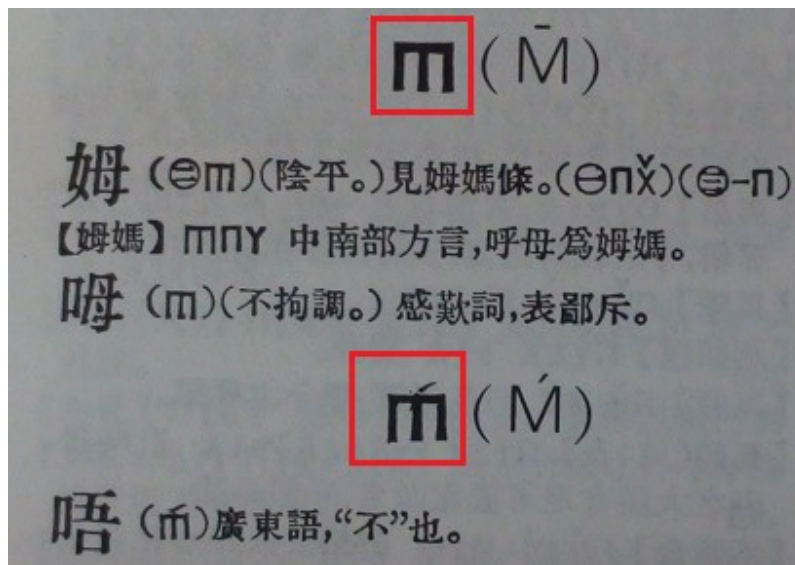
Fig. 2: *Guóyǔ cídiǎn* 國語辭典 p. 81 (U+31AC ㆬ)

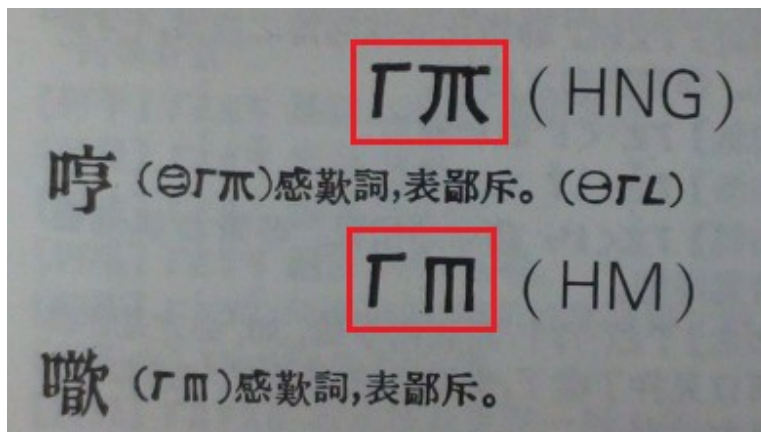Fig. 3: *Guóyǔ cídiǎn* 國語辭典 p. 417 (U+31AC ㋬ and U+31AD ㋭)

Fig. 4: https://zh.wikipedia.org/wiki/注音符號 (2016-02-05)

Chinese Wikipedia article discussing the proposed character and the inability to represent it in text as it is not encoded in Unicode.

4. *Bibliography*

Guóyǔ cídiǎn 國語辭典 [Dictionary of the National Language]. Shanghai: Commercial Press, 1937. Beijing: Commercial Press, 2011 facsimile reprint. ISBN 978-7-80103-788-6

5. *Proposal Summary Form*

<table>
<tr><td colspan="2">
SO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://www.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://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest Roadmaps.
</td></tr>
</table>

A. Administrative

1. **Title:**	*Proposal to encode one Bopomofo letter*
2. Requester's name:	*Andrew West*
3. Requester type (Member body/Liaison/Individual contribution):	*Individual contribution*
4. Submission date:	*2016-04-21*
5. Requester's reference (if applicable):	
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: *Bopomofo*
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**	B.2-Specialized (large collection)
C-Major extinct	D-Attested extinct	E-Minor extinct
F-Archaic Hieroglyphic or Ideographic		G-Obscure or questionable usage symbols

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?
     Andrew West
   - b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
     Andrew West
6. References:
   - a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?   *YES*
   - 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)?   *NO*

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: N4102-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)	*rare*
Reference:	
5. Are the proposed characters in current use by the user community?	*YES*
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?	*NO*
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?	*NO*
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
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?	*NO*
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility characters?	*NO*
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	