

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	ㄩ
------	--------------------	---

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 ㄻ, and U+312B BOPOMOFO LETTER NG ㄽ.

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 ㄩ and U+31AD BOPOMOFO LETTER NGG ㄽ. 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)	<i>Xiàndài hànnyǔ cídiǎn</i> 现代汉语词典	<i>Guóyǔ cídiǎn</i> 國語辭典
hm	ㄏㄢ	ㄏㄢ
hng	ㄏㄤ	ㄏㄤ
m	ㄇ	ㄇ
n	ㄋ	ㄋ
ng	ㄤ	ㄤ

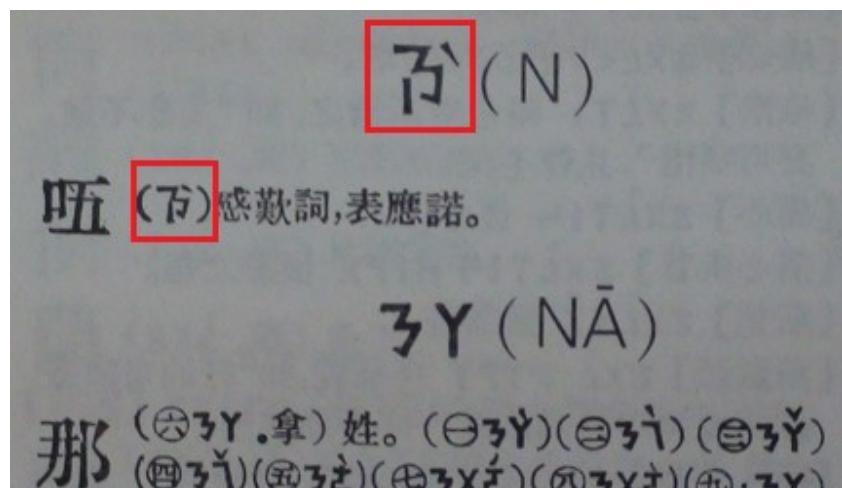
2. Unicode Properties

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

Other properties are the same as for other Bopomofo letters.

3. Figures

Fig. 1: *Guoyu cidiǎn* 國語辭典 p. 249



Showing the unencoded character with a falling tone mark

Fig. 2: *Guoyu cidiǎn* 國語辭典 p. 81 (U+31AC 呀)

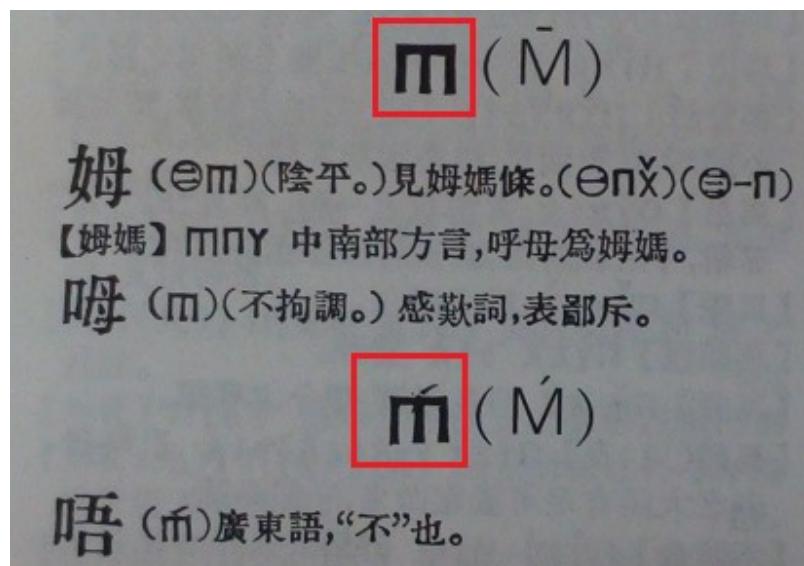


Fig. 3: *Guoyu cidiǎn* 國語辭典 p. 417 (U+31AC 婆 and U+31AD 婦)

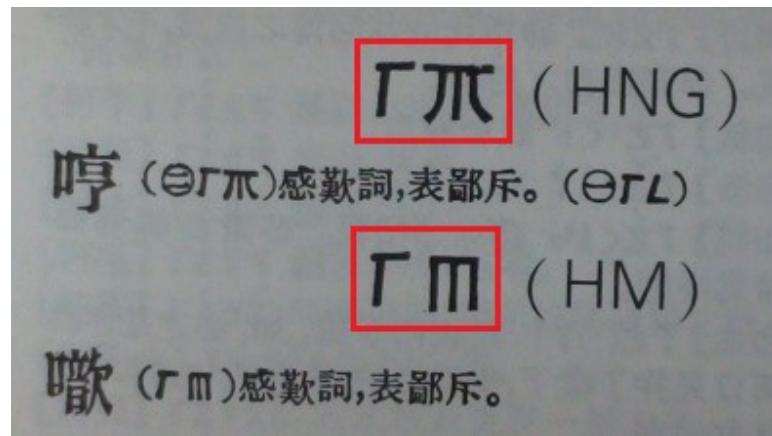


Fig. 4: <https://zh.wikipedia.org/wiki/注音符號> (2016-02-05)

1932年教育部在「編定《國音常用字彙》特組會議」時決定，為了便利說明，添補一個注音符號「ㄞ」（ㄞ），作為「ㄓㄔㄕㄔㄕㄔㄕㄔㄕ」七個聲母單獨成音節時的省略韻母（即虛母）。另外有三個注音符號ㄞ (-ng)、ㄞ (-m)、ㄞ (-n)，用作解釋聲隨韻母（ㄞ、ㄞ、ㄞ）時使用，(-n)的字形是ㄞ多加一筆直豎，尤解作ㄚ+ㄞ、ㄥ解成ㄜ+ㄞ、ㄞ為ㄚ+ (-n)、ㄞ為ㄞ+ (-n)；同理，複韻母ㄞ解為ㄚ+ㄧ、ㄞ為ㄚ+ㄨ。ㄞ、ㄞ、(-n)絕少單獨使用，「嗯」常唸作「ㄣ」，也有人唸成「ㄣ (-n)」。但以上符號只作為發音輔助說明用未列入正式符號表中。

1986年，中華民國教育部公佈羅馬字拼寫的「注音符號第二式」，注音符號因此又稱為「第一式」。

使用現況 [編輯]

在臺灣，小學生學習漢字前，必須上十週的注音符號教學課，也有不少幼稚園亦已教授。日常生活中，注音符號既用來標注生僻字，亦是常用的漢字輸入法，其推廣相當普遍，多數電腦使用者均熟練使用。

在臺灣話、客家話的教學上，臺灣省國語推行委員會另外增添新符號以能拼讀這些「鄉土語言」，稱之為「臺灣話方音符號」，比如臺灣話音節的ㄞ、ㄞ，被正式訂為臺灣話方音符號，載入Unicode注音符號擴展區用字，(-n)則沒有收入，故無法顯示。

在中國大陸及港澳，学校不教注音符号，但是一些字典、词典都通过汉语拼音与注音符号并列使用的方式给字词注音，如《新华字典》、《现代汉语词典》等。

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

4. *Bibliography*

Guoyu cidian 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

SO/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://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.	
A. Administrative	
1. Title:	<i>Proposal to encode one Bopomofo letter</i>
2. Requester's name:	<i>Andrew West</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>
4. Submission date:	<i>2016-04-21</i>
5. Requester's reference (if applicable):	
6. Choose one of the following: This is a complete proposal: (or) More information will be provided later:	<i>YES</i>
B. Technical – General	
1. Choose one of the following: a. This proposal is for a new script (set of characters): Proposed name of script:	<i>NO</i>
b. The proposal is for addition of character(s) to an existing block: Name of the existing block:	<i>YES</i> <i>Bopomofo</i>
2. Number of characters in proposal:	<i>1</i>
3. Proposed category (select one from below - see section 2.2 of P&P document): A-Contemporary B.1-Specialized (small collection) 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	<input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
4. Is a repertoire including character names provided? a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document? b. Are the character shapes attached in a legible form suitable for review?	<i>YES</i> <i>YES</i> <i>YES</i>
5. Fonts related: a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Andrew West</i> <i>Andrew West</i>
6. References: a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<i>YES</i> <i>YES</i>
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)?	<i>NO</i>
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 [redacted]	
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? [redacted]	
If YES, available relevant documents: [redacted]	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	NO
Reference: [redacted]	
4. The context of use for the proposed characters (type of use; common or rare)	rare
Reference: [redacted]	
5. Are the proposed characters in current use by the user community?	YES
If YES, where? Reference: [redacted]	
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? [redacted]	
If YES, reference: [redacted]	
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? [redacted]	
If YES, reference: [redacted]	
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? [redacted]	
If YES, reference: [redacted]	
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? [redacted]	
If YES, reference: [redacted]	
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? [redacted]	
If YES, reference: [redacted]	
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? [redacted]	
If YES, reference: [redacted]	
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) [redacted] [redacted]	
13. Does the proposal contain any Ideographic compatibility characters?	NO
If YES, are the equivalent corresponding unified ideographic characters identified? [redacted]	
If YES, reference: [redacted]	