

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 for additions to the Miao script****Author: Adrian Cheuk****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2017-10-03**

1. Introduction. This proposal requests the addition of 16 Miao characters used by 4 language groups in the provinces of Yunnan, Guizhou, and Sichuan, China. The following is a demographic summary:

Ethnologue Name	Name in Country	ISO 639-3	Population	Language Area
Aluo	Gan Yi	yna	25,000~40,000 (2007)	North Yunnan, South Sichuan
Gepo	Bai Yi	ygp	100,000 (2007)	East Yunnan
Miao, Small Flowery	Xiaohua Miao	sfm	84,000 (1995)	Northwest Guizhou
Yi, Wuding-Luquan	Hei Yi	ywq	250,000 (2007)	Northeast Yunnan, South Sichuan

2. New Characters. Each of the 4 language groups uses a subset of the characters in the Miao block plus a few not found in the block. We propose to encode these additional characters as explained below:

2.1 Consonant Letters

Representative Glyph	Code Point	Character Name	Remarks
𑜀	U+16F45	MIAO LETTER BRI	• used in Xiaohua Miao
𑜁	U+16F46	MIAO LETTER SYI	• used in Xiaohua Miao
𑜂	U+16F47	MIAO LETTER DZYI	• used in Xiaohua Miao
𑜃	U+16F48	MIAO LETTER TE	• used in Bai Yi
𑜄	U+16F49	MIAO LETTER TSE	• used in Bai Yi
𑜅	U+16F4A	MIAO LETTER RTE	• used in Bai Yi

While U+16F45 is already evidenced in Xiaohua Miao publications (see Figure 14), U+16F46..U+16F47 are characters the community has recently decided to add to their alphabet to further distinguish sounds. Figure 11 shows an alphabet chart containing these characters (as well as Chinese prose describing the particulars of a March community meeting to confirm character identities). There are currently no publications yet showing either, but at the time of writing one is already being printed. Figure 15 shows an excerpt of the final draft sent to the printer. Since their addition is a result of community decision (see Figure 12), these two characters are included in this proposal to be encoded together with U+16F45.

Concerning U+16F48..16F4A, until now they have been treated as variants of U+16F0B, U+16F38, and U+16F0F respectively. They were not addressed in N3877, somewhat surprisingly. For the reasons listed below it is felt that keeping them as variants is not appropriate:

- (1) Linguistic difference: U+16F48..U+16F4A are all voiceless letters. Compared to U+16F0A, U+16F37, and U+16F0E, the addition of the left serif is seen to mark **devoicing** (resulting in [t], [ts], and [t] respectively). The wart, on the other hand, is often seen to mark **voicing** (so U+16F01, U+16F0B, U+16F0F, etc.) or, when applied to nasals, **voiced aspiration** (so U+16F05, U+16F11, U+16F24). As such, the left serif and the wart are actually distinct linguistic devices with opposite functions and relationships with the grapheme base. It is therefore unreasonable to consider them the same. Furthermore, the wart is generally not used in Yi languages, so it simply does not fit Bai Yi linguistically.
- (2) Legibility criterion: Per TUS 10.0 (p.19), “*Plain text must contain enough information to permit the text to be rendered legibly, and nothing more.*” Bai Yi speakers will not be able to understand a default rendering (per the code chart) of U+16F0B, U+16F38, and U+16F0F as representing U+16F48..U+16F4A. Therefore, it would be necessary to add variation selection sequences for these characters to resolve the confusion in a plain text environment like Facebook. Rather than proposing those sequences, it would be better simply to disunify the characters and give them their own code points.

2.2 Nukta

Representative Glyph	Code Point	Character Name	Remarks
	U+16F4F	MIAO SIGN NUKTA	• used in Gan Yi

This combining mark was originally written like a comma on the lower-left of the base character (see Figure 9). When made into a font, it now assumes the shape of a short vertical bar as shown above. Currently it is only found to be applied to U+16F0E and U+16F10, marking contrast in the place of articulation, but it is not unlikely that other groups will employ the same device on other characters. The Gan Yi actually do not have a name for it, but since it functions similarly as the Indic *nukta*, this name is proposed for encoding purpose.

With the addition of the nukta, it has become necessary to define different classes of marks to ensure unambiguous representation given the potential co-occurrence of marks in all the lower-left, upper-right, and lower-right positions. Compared to the others, the nukta is the most closely bound to the base character, followed by the aspiration mark. Using the notation in N3789, the revised syllable structure is: (N)C(M₁)(M₂)(M₃)V⁺(S/T), where M₁ is U+16F4F, M₂ is either U+16F51 or U+16F52, and M₃ is U+16F53. Note that while a font may choose to support fewer classes based on available orthographic evidence, the shaping engine should not be forced to do so, as structurally all 3 classes can co-occur.

2.3 Vowel Signs

Representative Glyph	Code Point	Character Name	Remarks
	U+16F7F	MIAO VOWEL SIGN UOG	• used in Hei Yi
	U+16F80	MIAO VOWEL SIGN YUI	• used in Hei Yi
	U+16F81	MIAO VOWEL SIGN OG	• used in Gan Yi and Bai Yi
	U+16F82	MIAO VOWEL SIGN OER	• used in Gan Yi and Xiaohua Miao
	U+16F83	MIAO VOWEL SIGN VW	• used in Xiaohua Miao and Bai Yi
	U+16F84	MIAO VOWEL SIGN IG	• used in Bai Yi
	U+16F85	MIAO VOWEL SIGN EA	• used in Bai Yi
	U+16F86	MIAO VOWEL SIGN IONG	• used in Bai Yi
	U+16F87	MIAO VOWEL SIGN UI	• used in Bai Yi

All proposed vowel signs contrast in some way with existing characters in the Miao block in a given language group, so there is no question to their distinct identities. Note, however, that some of them represent different sounds across language groups. In this proposal, naming priority is first given to usage in Miao language groups and then to usage in other groups, and within each category, to the group with the most population. Note also that there are not yet any publications that show U+16F86..U+16F87. However, since two groups of Bai Yi speakers from different language areas at different points in time (2012 and 2016) confirmed their usage in their language (see alphabet chart in Figure 2 and letters from community leaders acknowledging their use in Figures 3 and 4), it is included in this proposal to be encoded together with the other vowels. As with other vowel signs, the proposed characters can be positioned off the baseline via positioning tone marks (U+16F8F..U+16F92) to indicate different tones.

3. Collating Order. Within each language group there is a preferred sort order according to their alphabet charts (Figures 2, 5, 10, and 11). However, there is no consistent order across groups. The default order proposed in N3789 will be used as a basis in this proposal. The new characters being proposed and their attested combinations with modifiers are inserted below and underlined to show their positions relative to existing combinations in the collating order. An exception is the nukta (U+16F4F), which has no position of its own other than being part of attested combinations. Combinations unattested at the time of N3789 but now attested with new data are also added and underlined. Short character names are shown after encoded and proposed characters only; combinations are not named.

... < 𐄀 PLA < 𐄁 BRI < 𐄂' < (𐄃 < (𐄄 < (𐄅' <) MA < ... < 𐄆 YI TA < 𐄇 TE < 𐄈' < ... < 𐄉 DDA < 𐄊 RTE < 𐄋 < 𐄌' < ... < 𐄍' < 𐄎 < 𐄏 NNA < ... < 𐄐' < 𐄑 REFORMED TSHA < 𐄒 DZYI < 𐄓' < (𐄔 < (𐄕 < (𐄖' < 𐄗 NYHA < ... < 𐄘 SHA < 𐄙 SYI < 𐄚 SSA < ... < 𐄛 YI TSA < 𐄜 TSE < 𐄝' < ... < 𐄞 ANG < 𐄟 OG < 𐄠 O < ... < 𐄡 W < 𐄢 VW < 𐄣 E < ... < 𐄤 OEY < 𐄥 IG < 𐄦 I < ... < 𐄧 IO < 𐄨 IONG < 𐄩 IE < ... < 𐄪 UANG < 𐄫 UOG < 𐄬 UI < 𐄭 UU < ... < 𐄮 Y < 𐄯 YUI < 𐄰 AE < 𐄱 AEE < 𐄲 EA < 𐄳 OER < 𐄴 ER < ...

4. Unicode Character Properties. All proposed consonant letters have a general category of Lo while all proposed vowel signs, Mc. The nukta has a general category of Mn. None of these characters can be confused with other characters in the Miao block. Detailed character properties are as follows:

- 16F45;MIAO LETTER BRI;Lo;0;L; ; ; ; ;N; ; ; ; ;
- 16F46;MIAO LETTER SYI;Lo;0;L; ; ; ; ;N; ; ; ; ;
- 16F47;MIAO LETTER DZYI;Lo;0;L; ; ; ; ;N; ; ; ; ;
- 16F48;MIAO LETTER TE;Lo;0;L; ; ; ; ;N; ; ; ; ;
- 16F49;MIAO LETTER TSE;Lo;0;L; ; ; ; ;N; ; ; ; ;
- 16F4A;MIAO LETTER RTE;Lo;0;L; ; ; ; ;N; ; ; ; ;
- 16F4F;MIAO SIGN NUKTA;Mn;0;L; ; ; ; ;N; ; ; ; ;
- 16F7F;MIAO VOWEL SIGN UOG;Mc;0;L; ; ; ; ;N; ; ; ; ;
- 16F80;MIAO VOWEL SIGN YUI;Mc;0;L; ; ; ; ;N; ; ; ; ;
- 16F81;MIAO VOWEL SIGN OG;Mc;0;L; ; ; ; ;N; ; ; ; ;
- 16F82;MIAO VOWEL SIGN OER;Mc;0;L; ; ; ; ;N; ; ; ; ;
- 16F83;MIAO VOWEL SIGN VW;Mc;0;L; ; ; ; ;N; ; ; ; ;
- 16F84;MIAO VOWEL SIGN IG;Mc;0;L; ; ; ; ;N; ; ; ; ;
- 16F85;MIAO VOWEL SIGN EA;Mc;0;L; ; ; ; ;N; ; ; ; ;
- 16F86;MIAO VOWEL SIGN IONG;Mc;0;L; ; ; ; ;N; ; ; ; ;
- 16F87;MIAO VOWEL SIGN UI;Mc;0;L; ; ; ; ;N; ; ; ; ;

5. Bibliography

Anonymous. n.d. 白彝文大小字母. ms.

Anonymous. n.d. 柏格林拼音文字母. ms.

Anonymous. n.d. Gan Yi alphabet chart. ms.

Anonymous. n.d. Xiaohua Miao alphabet chart. ms.

Anonymous. n.d. 新约圣经（小花苗）. n.p.

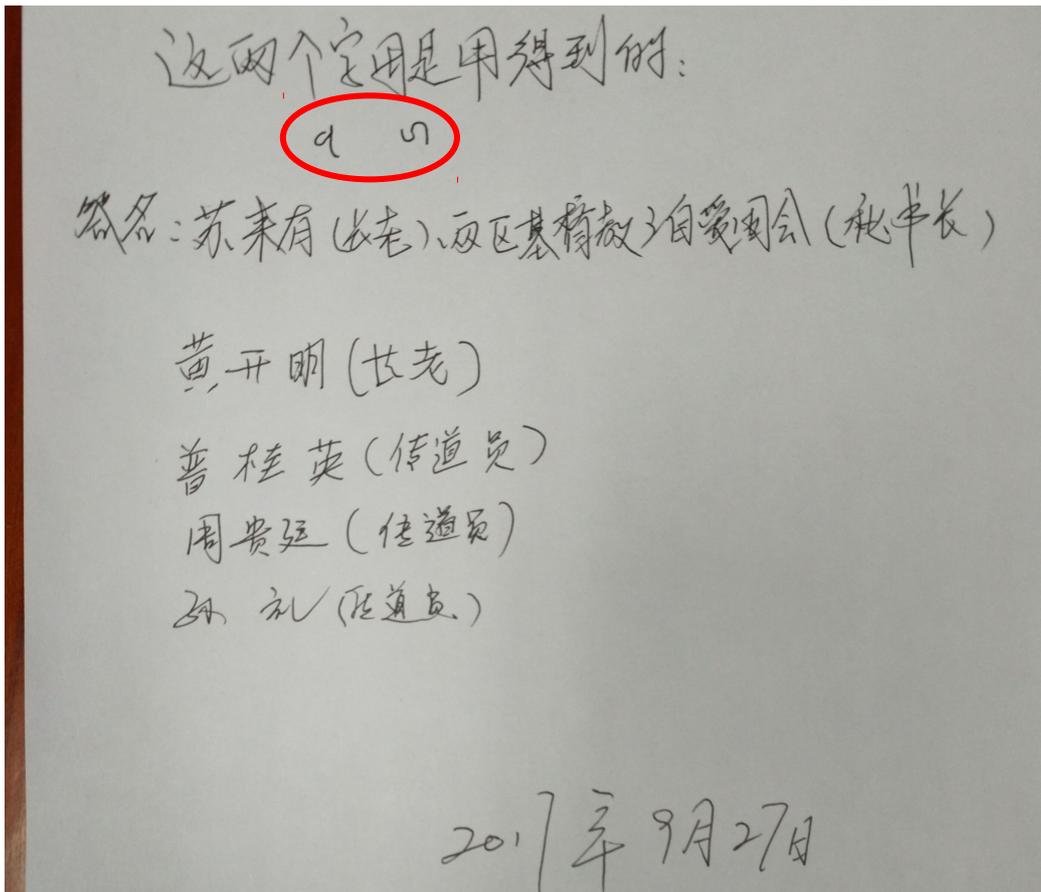


Figure 3: First letter (in Chinese) showing the Bai Yi community's acknowledgement of usage of \odot_a MIAO VOWEL SIGN IONG and \odot_u MIAO VOWEL SIGN UI.

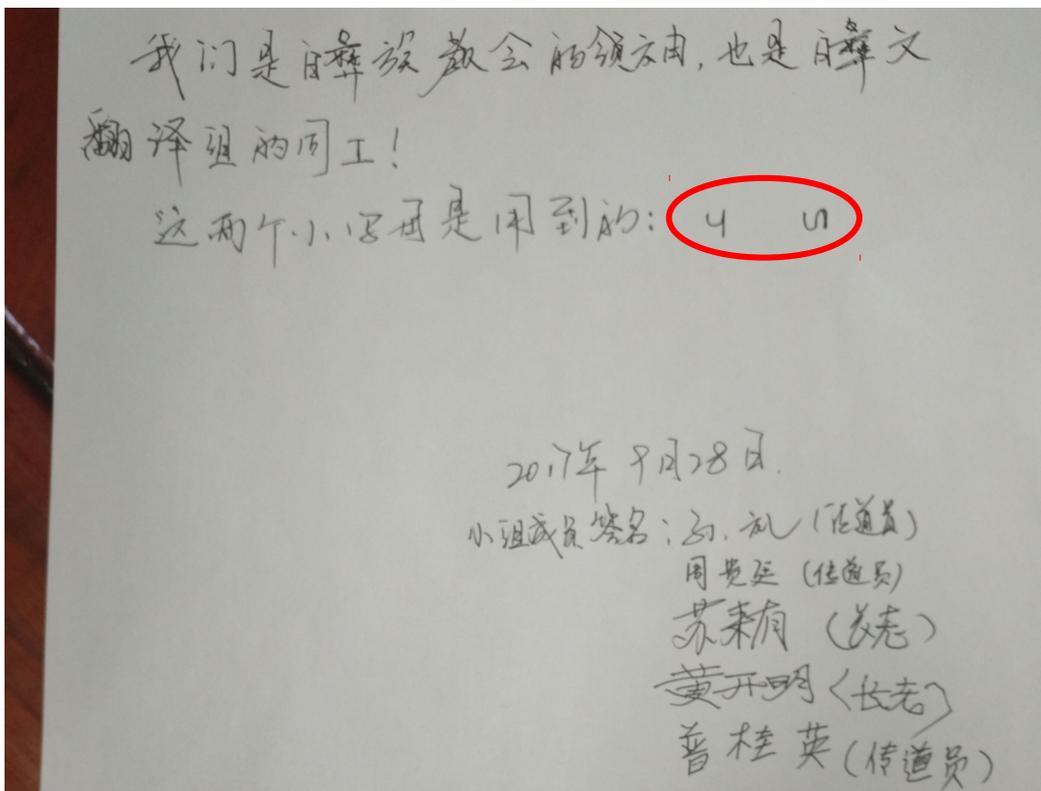


Figure 4: Second letter (in Chinese) showing the Bai Yi community's acknowledgement of usage of \odot_a MIAO VOWEL SIGN IONG and \odot_u MIAO VOWEL SIGN UI.

柏格林拼音文字母

声母:

Y I J L C J' T T' L L C C' J C

啊 (鸡) (肉) 八 (骂) 爬 法 (鼠) 寒 兰 (烫) 甲 (棧道) 强 (聪明) (加上)

J' J C C' L C L' t t C t' T C T

扛 嘎 知 (增) 痴 (坏) (滑) (损坏) 杂 (小走) (滴) 擦 (上去) (堂的鼻音)

T' I T E z C) (z e J J G S U V A 3

塔 达 (扁) (凉) (窥探) 忙 南 (拿) 娘 瞎 杀 (鸟) 撒 王 亡 羊 (下去)

R 日

韵母:

- I = II 5 0 0 P r r n u ü

啊 艾 (鸭子) 凹 欧 (水泼声) (起泡) (窝囊) (一) (日) 衣 乌 淤

p l ɔ c n II u5 no nɬ r o5 n= uɬ uo

(呻吟) 诶 (肾) 儿 腰 (骨) 雍 耶 鸦 (官) 银 威 窝

v n w c5 uc u= ɬ n0 uɬ nɬ

蛙 (头) (谷短声) 额 (碗的南方音) 围 昂 哟 玉 羊

Figure 5: Hei Yi alphabet chart, showing P MIAO VOWEL SIGN UOG and ü MIAO VOWEL SIGN YUI.

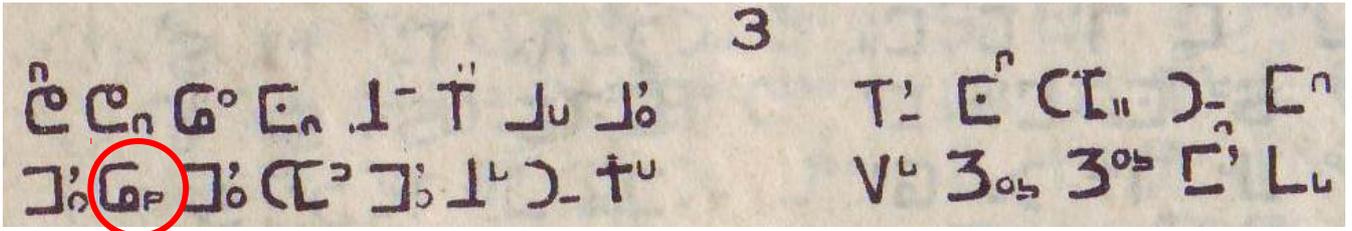


Figure 6: Sample from a Hei Yi song book, showing \circ_p MIAO VOWEL SIGN UOG.

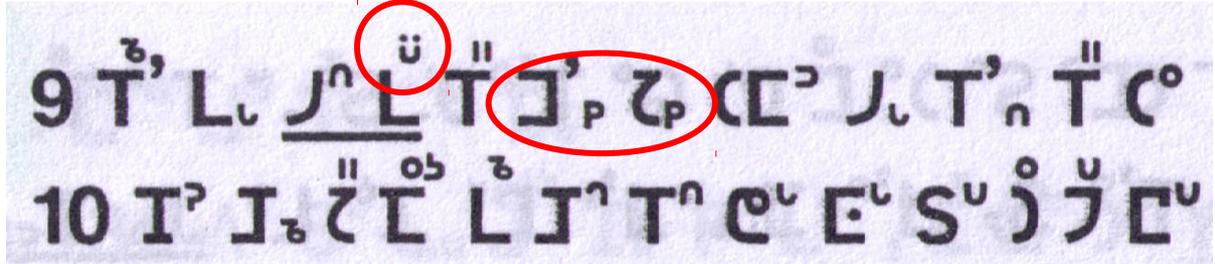


Figure 7: Sample from a Hei Yi scripture, showing $\circ_{\ddot{u}}$ MIAO VOWEL SIGN YUI and \circ_p MIAO VOWEL SIGN UOG.

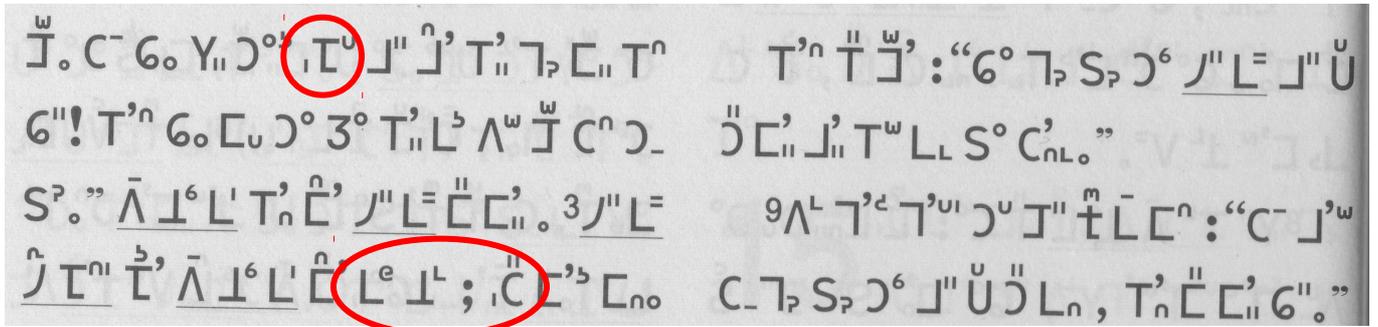


Figure 8: Sample from a Gan Yi scripture, showing \circ_{dot} MIAO SIGN NUKTA, \circ_e MIAO VOWEL SIGN OG, and \circ_L MIAO VOWEL SIGN OER.

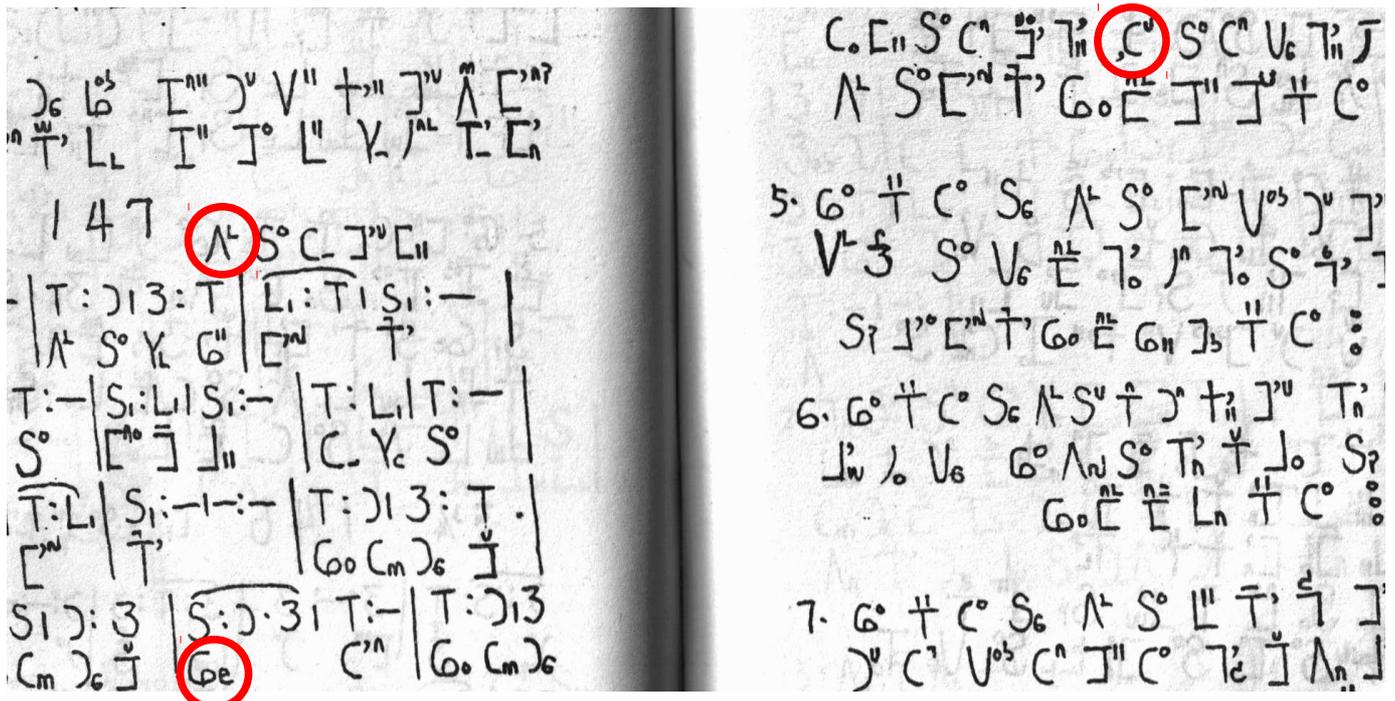


Figure 9: Sample from a Gan Yi song book, showing \circ_L MIAO VOWEL SIGN OER, \circ_e MIAO VOWEL SIGN OG, and \circ_{dot} MIAO SIGN NUKTA.

S^U C' _{nu} ɔ⁶ (S'' t' ^L ɲ' ɔ^o)

ɿ ɿ' ɿ'' T I T' ɿ ɿ' ɿ'' C ɿ
 C' t t' v ɿ ɿ' ɿ ɿ C C'
 ɿ C L ɿ ɿ U R ɿ ɿ S ɿ Y
 ɿ ɿ C ɿ'

S^U C' _{nu} ɔ^o (ɿ' t' ^L S'' ɔ^o)

- ɿ ɿ' ɿ''
 ɿ ɿ' ɿ'' ɿ ɿ' ɿ'' ɿ ɿ' ɿ'' ɿ ɿ' ɿ'' ɿ ɿ' ɿ''
 ɿ ɿ' ɿ'' ɿ ɿ' ɿ'' ɿ ɿ' ɿ'' ɿ ɿ' ɿ'' ɿ ɿ' ɿ''
 ɿ ɿ' ɿ'' ɿ ɿ' ɿ'' ɿ ɿ' ɿ'' ɿ ɿ' ɿ''

Figure 10: Gan Yi alphabet chart, showing ɿ C MIAO SIGN NUKTA, ɿ L MIAO VOWEL SIGN OER, and ɿ e MIAO VOWEL SIGN OG.

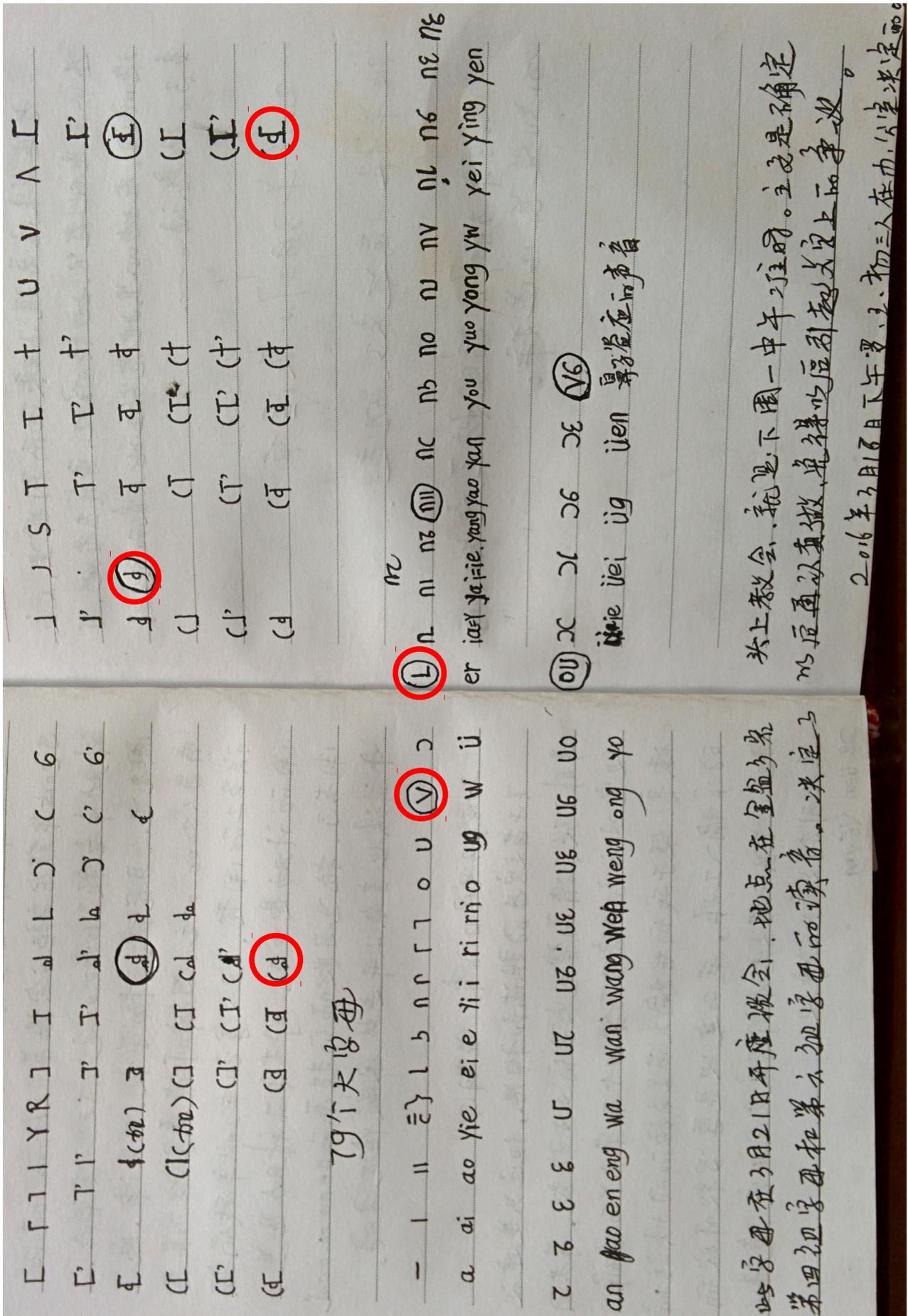


Figure 11: Xiaohua Miao alphabet chart (rotated), showing **d** MIAO LETTER BRI, **s** MIAO LETTER SYI, **l** MIAO LETTER DZYL, **vw** MIAO VOWEL SIGN VW, and **ler** MIAO VOWEL SIGN OER.

关于小花苗支字母加增的信

尊敬的牧长，同志们：

感谢您们在我们小花苗族苗文文字事业上有所付出。由于我们小花苗族文字的差异性，与其它苗族支系的文字要多使用声母，而原苗文软件（柏格里文）里没有体现出来，这给我们的翻译工作和软件使用方面带来极大的不便。为了更好地把小花苗族文字语音表达清楚，经全体小花苗族教牧同工代表会议通过，必须加添“**g**”和“**z**”这两个字母。此“g”字与“~~g~~”混用而不能表达苗族的语音。比如在经中“摩西”的“西”字，拼出后不“西”是“施”；又如“西门”的“西”，拼出后同样成了“施门”等。

为此，必须加添以上两个字母。

以马内利！

特此。

周光忠 牧师（神召会基召会西台和书台）
 罗学明 牧师（外城基召会西台会长）

2017.9.27.

Figure 12: Letter (in Chinese) showing the Xiaohua Miao community's decision to add **J** MIAO LETTER SYI and **z** MIAO LETTER DZYI to their alphabet.

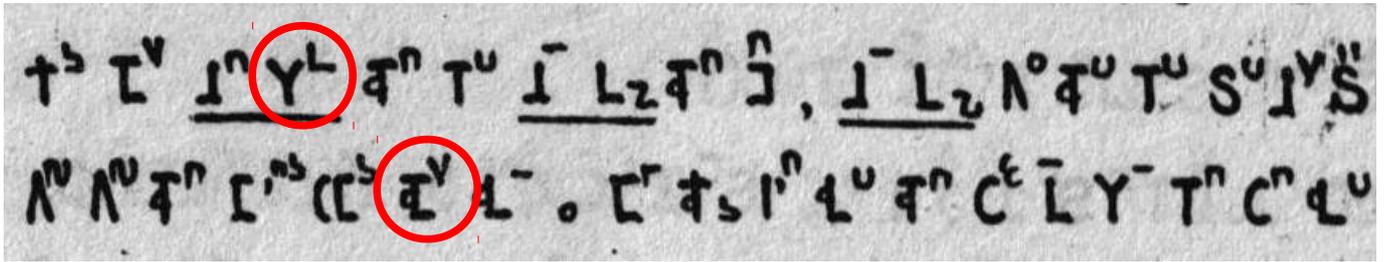


Figure 13: Sample from a Xiaohua Miao scripture, showing \circ_L MIAO VOWEL SIGN OER and \circ_V MIAO VOWEL SIGN VW.

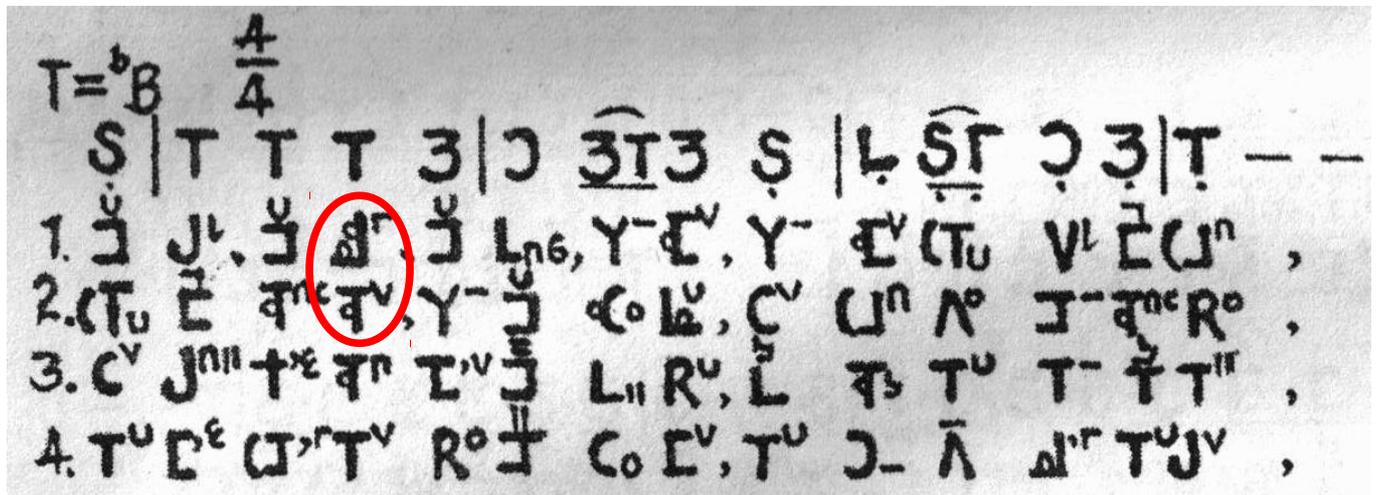


Figure 14: Sample from a Xiaohua Miao song book, showing \downarrow MIAO LETTER BRI and \circ_V MIAO VOWEL SIGN VW.

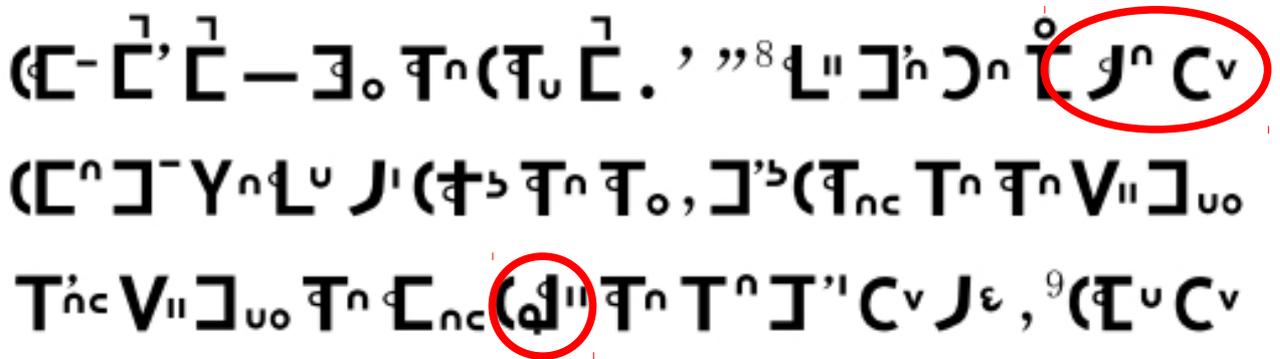


Figure 15: Sample from a Xiaohua Miao scripture printing draft, showing \downarrow MIAO LETTER SYI, \circ_V MIAO VOWEL SIGN VW, and \downarrow MIAO LETTER BRI.

**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 for additions to the Miao script

2. Requester's name: Adrian Cheuk

3. Requester type (Member body/Liaison/Individual contribution): Individual contribution

4. Submission date: 2017-10-03

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: Miao

2. Number of characters in proposal: 16

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

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? SIL International

b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): SIL International (nrsl@sil.org)

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: 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? If YES explain	<i>No</i>
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.)? If YES, with whom? If YES, available relevant documents:	<i>Yes</i> <i>Yunnan Minority Language Commission and language experts from Hei Yi, Bai Yi, Gan Yi, and Xiaohua Miao</i> <i>See Jackson 2016a-c.</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>Yes</i> <i>Ethnologue</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>Common</i> <i>Songs and scriptures</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>Yes</i> <i>The provinces of Yunnan, Guizhou, and Sichuan, China</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	<i>No</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<i>Yes</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>No</i>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>No</i>
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? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>No</i>
11. Does the proposal include use of combining characters and/or use of composite sequences? 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:	<i>Yes</i> <i>Yes</i> <i>Proposal includes combining marks</i> <i>No</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	<i>No</i>
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	<i>No</i>