

Universal Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная Организация по Стандартизации

Doc Type: Working Group Document
Title: Proposal to encode the Tai Xaaü script in UCS
Source: China National Body
Status: National Body Contribution
Action: For consideration by ISO/IEC JTC1/SC2/WG2
Date: 2025-06-16

0 Executive summary

This proposal is a revised version of:

- China National Body: “Proposal to encode the Tai Xaaü script in UCS” (WG2 N5310)

and it incorporates comments made by the Ethnic Affairs Commission (EAC) of Honghe Prefecture, the EAC of Yunnan Province, China Electronics Standardization Institute (CESI), the Institute of Software, China Academy of Sciences (ISCAS), and other individual experts. Compared to the previous version, this proposal rewrite [Section 2.1, “Relations with Tai Viet,”](#) specifically:

- This proposal conducts a survey of font styles and explains the stylistic differences between Tai Xaaü and Tai Viet.
- This proposal solicits feedback from the Tai Xaaü community on each Tai Viet glyph in all the available Tai Viet fonts, provides a glyph comparison table, and explains the glyphic differences between Tai Xaaü and Tai Viet.
- This proposal investigates similar cases of Brahmic scripts in UCS and explains the differences in ethnic identity among user communities of Tai Xaaü and Tai Viet based on recently published papers.
- Based on the differences in font styles, glyph shapes and user identity, this proposal recommends disunifying Tai Xaaü to Tai Viet.

As in the previous version, this proposal proposes adding 76 Tai Xaaü characters to UCS.

1 Introduction

The Tai Xaaü (ᨲᨶᨳᨱ, tai⁴³xa:u³³, *tai xaaü*) script is an Abugida script written by the Dai people (傣族) of Yunnan Province (云南省), PRC, to record the Jinping dialect of the Dai language (傣语金平方言). The total population of the Dai people in China is over 1.32 million (2021 statistics), mainly distributed in Yunnan Province and a few in Sichuan Province (四川省).

The Dai people have a long history. Like the Zhuang (壮族), Buyei (布依族), Dong (侗族), etc., who are close in origin, the Dai people are also descendants of the ancient Baiyue (百越), of which the Dai people are part of the Dianyue (滇越). According to the records of the Dianyue in *Shiji* (史记) and *Hanshu* (汉书), during the Han Dynasty (汉朝), the ancestors of the Dai people lived mainly in present-day south-western Yunnan Province.

The Dai people can be divided into several branches according to their endonyms, such as Tai Lue (傣仂) for those living in Xishuangbanna (西双版纳), Tai Taue (傣德) and Tai Le (傣哪) for those living in Dehong (德宏), Gengma (耿马), Menglian (孟连) and so on, Tai Xaau (傣皓) for those living in Jinping (金平) county of Honghe (红河), and so on. These endonyms have the same first syllable and different second syllable, inferring that they originate from the same ethnic group. After the establishment of PRC, they are collectively known as the Dai people.

The Dai language (傣语) belongs to the Zhuang–Dai (or Tai) languages of the Zhuang–Dong (or Dong–Tai) languages of the Sino–Tibetan language family, with more than one million speakers in Yunnan Province and Sichuan Province, PRC. According to the difference in phonology and vocabulary, there are four dialects of the Dai language in China: Dehong dialect (德宏方言, ISO 639–3: [tdd](#)), Xishuangbanna dialect (西双版纳方言, ISO 639–3: [khh](#)), Honghe–Jinsha dialect (红金方言, ISO 639–3: [tiz](#)) and Jinping dialect (金平方言, ISO 639–3: [twh](#)).

The Tai Xaau script is written by the Tai Xaau people, mainly in Jinping Miao, Yao, and Dai Autonomous County (金平苗族瑶族傣族自治县), Honghe Hani and Yi Autonomous Prefecture (红河哈尼族彝族自治州), and a few in Jiangcheng Hani and Yi Autonomous County (江城哈尼族彝族自治县), Pu'er City (普洱市), both in Yunnan Province (云南省), PRC, to record the Jinping dialect of the Dai language, so it is also known as Jinping Dai writing system (金平傣文).

The Tai Xaau script is an Abugida script in which consonant letters behave as base units with inherited vowels, while vowel letters behave as additional units and cannot appear alone. Similar to Tai Tham and New Tai Lue, the Tai Xaau script contains two sets of consonant letters, constructed as oppositions of voiceless and voiced consonants in Proto–Tai, and expressed as higher and lower tone classes. However, unlike Tai Tham and New Tai Lue, the Tai Xaau script has no subjoined consonants, that is, all consonant letters behave identically.

2 Script details

2.1 Relations with Tai Viet

The Tai Xaau script is related to the Tai Viet script (ISO 15924: [Tavt](#)), which is encoded in ISO/IEC 10646:2003/Amd.5:2007 (and in Unicode Standard Version 5.2), as it can be noted that some letters are similar in their glyphs and sounds. According to [L2/06–041](#) (Ngô & Brase, 2006) and [L2/07–039R](#) (Brase, 2007), the Tai Viet script was said to be constructed from several scripts, namely Black Tai, White Tai, Red Tai and Lao Song, to form a universal script for use in Vietnam, Laos and Thailand. The result of the Tai Viet character set, “is focused on the use of the script by the Tai Dam of Sơn La Province, Vietnam”, “contains the traditional Sơn La character set as a whole, plus three pairs of

aspirated consonants (note: high/low kho, cho and pho) required by the Tai Dón language of Lai Châu Province”. [L2/08–217 \(Brase, 2008\)](#) further analyzed the Tai Don writing systems and concluded that the 13 extra characters should be disunified and encoded, and mentioned that, “Other readers may wish to argue for the disunification of other glyphs, or for the disunification of the entire Tai Don/Jinping script from Tai Viet. All such contributions and advice are earnestly sought”.

From China’s perspective, we agree with the latter opinion: to disunify Tai Xaa (the “Jinping script”) to Tai Viet. The main reasons are:

Style difference. The digital fonts that we have collected and which support Tai Viet include *Unifont* (and *Unifont Smooth*), *Noto Sans Tai Viet*, *Tai Heritage Pro*, and *TaiViet*. We have also collected digital fonts that support Tai Xaa; see [Appendix A.3](#) and [this website](#) for more information. As can be seen, these font styles do not overlap; the style of Tai Xaa’s digital font is not found in Tai Viet’s, and vice versa. It has been noted that the digital fonts for Tai Xaa are trending towards the handwriting styles using hard and soft pens. There are similar handwriting styles in Tai Viet, as seen in the [Origin Myth in Tai Dam](#), suggesting a potential common style based on handwriting.

As there is currently no common font style for China and Vietnam users, we are aware of potential font implementation issues that could arise from unifying Tai Viet and Tai Xaa. As Tai Viet was encoded earlier, users have expressed concerns about the lack of localization of common characters.

Glyph difference. We refer to the following three materials: (1) [L2/08–217 \(Brase, 2008\)](#) analyzed the differences between Tai Viet (“Existing Tai Viet”) and Tai Xaa (“Jinping Dai”) using a sans-serif font. There are some issues with the glyphs, such as loops appearing at the turns of strokes, which are not recognized by the Tai Xaa community. (2) The Tai Xaa community has conducted glyph comparisons; see [this website](#). The main issue with this comparison lies in the description and comparison of the phonemes of Tai Dam and Tai Xaa languages. (3) [Dao and Luo \(2025\)](#) provide a brief description of the unique letters of Tai Viet and Tai Xaa.

We then provide a comparison between Tai Viet and Tai Xaa based on these three materials. The first and the second columns show the Tai Viet glyphs in *TaiViet* (also in the Code Charts) and [L2/08–217 \(Brase, 2008\)](#). The third and the fourth columns show the Tai Xaa glyphs in *Xiaorui* and [L2/08–217 \(Brase, 2008\)](#). The fifth column provides the explanation. Gray glyphs indicate that the glyphs listed by [L2/08–217 \(Brase, 2008\)](#) in the “Jinping Dai” column are not credible.

Please note that the comparison in this table is purely glyph-dependent. The focus is on whether two glyphs are written the same way rather than on whether they correspond to the same phoneme. For instance, the Tai Viet letter  TAI VIET LETTER HIGH PO and the Tai Xaa letter  LOW MO represent unrelated phonemes, yet they are placed on the same line because they are written in the same way. Additionally, ‘No Tai Viet correspondence’ does not imply the absence of Tai Viet letters that record the same phoneme; rather, it signifies the absence of letters written in the similar way.

Tai Viet		Tai Xaaü		Note
<i>TaiViet</i>	Brase	<i>Xiaorui</i>	Brase	
ə	ə	ə	ə	Different. Tai Xaaü requires an open counter.
ɣ	ɣ	ɣ	ɣ	Different. Tai Xaaü requires an open counter and a separate stroke.
ɯ	ɯ	ɯ	ɯ	Same.
ɔ	ɑ	ɰ	ɰ	Different. The Tai Viet shape is unrecognizable.
		ɱ	ɱ	There is no Tai Viet correspondence.
ʒ	ʒ	ɲ	ɲ	Different. Tai Xaaü requires a vertical stroke start.
ʋ	ʋ	ɶ	ɶ	Different. Tai Xaaü requires a loop stroke start and no loop stroke end.
ɔ	ɔ	ɔ	ɔ	Similar. Tai Xaaü requires a closed counter.
ɹ	ɹ	ɹ	ɹ	Different. The Tai Viet shape is unrecognizable.
ɻ	ɻ	ɻ	ɻ	Different. The Tai Viet shape is unrecognizable.
		ɻ	ɻ	There is no Tai Viet correspondence.
		ɻ	ɻ	There is no Tai Viet correspondence.
		ɻ	ɻ	There is no Tai Viet correspondence.
ʈ	ʈ	ʈ	ʈ	Similar. The circumflex shape can be found in the Tai Xaaü font, but is not standard.
ɯ	ɯ	ɯ	ɯ	Different. Tai Xaaü requires a vertical stroke start and a leftward stroke end.
u	u	u	u	Different. Tai Xaaü requires a vertical stroke start and a leftward stroke end.
ɹ	ɹ	ɹ	ɹ	Different. Tai Xaaü requires a vertical stroke start.
		ɹ	ɹ	There is no Tai Viet correspondence.
ʋ	ʋ	ʋ	ʋ	Different. Tai Xaaü requires no loop shape.

u	u			Different. Tai Xaa requires no loop shape.
ɯ	ɯ			Different. Tai Xaa requires a turn stroke. The Tai Viet shape will be misrecognized.
				There is no Tai Viet correspondence.
ɑ	ɑ			Different. Tai Xaa requires no loop shape. They also record unrelated phonemes.
				There is no Tai Viet correspondence.
				There is no Tai Viet correspondence.
ɰ	ɰ			Same. The shapes evolve to be the same, despite the different phonemes.
ɱ	ɱ			Different. Tai Xaa requires no loop shape. They also record unrelated phonemes.
ɯ	ɯ			Different. Tai Xaa requires no loop shape.
ɰ	ɰ			Same.
ɱ	ɱ			Same.
ɲ	ɲ			Different. Tai Xaa requires a loop shape.
ɳ	ɳ			Different. Tai Xaa requires the descender.
ɴ	ɴ			Different. Tai Xaa requires a loop shape.
ɵ	ɵ			Different. Tai Xaa requires a curved stroke start.
				There is no Tai Viet correspondence.
ɶ	ɶ			Different. Tai Xaa requires no loop shape.
ɷ	ɷ			Different. Tai Xaa requires a loop shape and a vertical stroke.
ɸ	ɸ			Same.
				There is no Tai Viet correspondence.

				Same.
				Different. Tai Xaa requires no loop shape.
				Different. Tai Xaa requires a comma shape below the letter.
				There is no Tai Viet correspondence.
				There is no Tai Viet correspondence.
				Similar. Tai Xaa requires a smaller shape.
				There is no Tai Viet correspondence.
				There is no Tai Viet correspondence. This mark appears in and it is different from / . Tai Xaa requires no loop shape.
				Similar. Tai Xaa requires the caron shape.
				Same.
				Same.
				Same. Though they record different phonemes.
				Different. Tai Xaa requires thw loop shapes, the <i>Tai Viet</i> glyph is preferred.
				There is no Tai Viet correspondence.
				Different.
				Different. Tai Xaa requires a dot shape.
				Same.
				There is no Tai Viet correspondence.
				Different. Tai Xaa requires no loop shape.
				Same. Though they record different phonemes.

				Similar. Tai Xaa requires smaller shape, the <i>TaiViet</i> glyph is preferred.
				Same. Though they record different phonemes.
				Same.
				Same.
				Different. Tai Xaa requires hook shape.
				Different.
				Different. Tai Xaa requires a vertical stroke start and no turn stroke.
				There is no Tai Viet correspondence.
				There is no Tai Viet correspondence.
				There is no Tai Viet correspondence.
				There is no Tai Viet correspondence.
				Different.
				Different.
				There is no Tai Viet correspondence.
				There is no Tai Viet correspondence.
				There is no Tai Viet correspondence.
				There is no Tai Viet correspondence.

Of the 44 consonant letters, five are identical and two are similar to those in the Tai Viet alphabet. Of the 18 vowel letters and marks, seven are identical and three are similar to those in Tai Viet. Of the two tone marks, one is identical to those used in Tai Viet. All the symbols and punctuation differ from those used in Tai Viet. The above comparison table shows that the vast majority of Tai Xaa letters are written differently from Tai Viet. These different spellings are not found in Tai Xaa fonts and are not accepted by users.

Similar cases. The Tai Xaau community and researchers believe that the differences between Tai Viet and Tai Xaau are similar to those between Thai and Lao. At the glyph level, such differences are greater than those between Kannada and Telugu, and also greater than some of the differences between Grantha and Tamil. For scripts with different shapes, the identity of the user group is more important, and surveys in China show that there are differences in identity between users of Tai Xaau and Tai Viet, as described in [Dao and Luo \(2025\)](#).

Based on the differences in font styles, glyph shapes and user identity, this proposal recommends disunifying Tai Xaau to Tai Viet.

2.2 Encoding model

The presence of pre-base vowel letters in the Tai Xaau script, similar to those found in Tai Tham and New Tai Lue, allows for the consideration of different encoding models, namely those based on visual and logical order. Following several meetings and discussions, this document proposes a model based on visual order.

The main reason for choosing visual order is that all consonants in the Tai Xaau script behave identically to base characters. Unlike many Brāhmī model-based scripts, the Tai Xaau script is not rich in consonant conjuncts. This means that even when a control character is introduced as a vowel killer, it plays no role and causes confusion.

Additionally, as will be discussed later, the vowel letters and tone marks used as combining marks can be placed more flexibly because their only function is to ensure that the whole syllable can be distinguished from the others. Using visual order simplifies the support of the many possible positions of these combining marks.

2.3 Character names

The Latinization of the Tai Xaau phonology has been determined for use in the character names, as shown below.

/b/	/d/	/f/	/h/	/k/	/k ^h /	/l/	/m/	/n/	/ŋ/	/ɲ/
B	D	F	H	K	KH	L	M	N	NG	NY
/p/	/p ^h /	/ʔ/	/s/	/t/	/t ^h /	/ts/	/ts ^h /	/v/	/x/	/j/
P	PH	Q	S	T	TH	TS	TSH	V	X	Y
/k ^w /	/x ^w /	/ŋ ^w /	/h ^w /	/k ^{hw} /	/ʔ ^w /					
KW	XW	NGW	HW	KHW	QW					
/a/	/a:/	/i/	/e/	/ɛ/	/ɔ/	/o/	/u/	/ʉ/	/ə/	
A	AA	I	E	AE	O	OA	U	UE	EI	

A and AA. For closed syllables, the phoneme /a/ has long and short oppositions. In some documents it is written as /a/ and /a:/, in others as /ǎ/ and /a/, and in this document, they are transcribed as A and AA. For open syllables, the phoneme /a/ has no long and short oppositions and is transcribed as A.

2.4 Consonant letters

There are 44 consonant letters:

๑ HIGH QO	๑๕ HIGH HO YO	๑๗ HIGH HO MO	๑๗ HIGH HO VO
๑ LOW QO	๑๖ LOW YO	๑๘ LOW MO	๑๘ LOW VO
๑๑ HIGH KO	๑๑ HIGH TO	๑๑ HIGH FO	๑๑ HIGH KHO
๑๑ LOW KO	๑๑ LOW TO	๑๑ LOW FO	๑๑ LOW KHO
๑๑ HIGH XO	๑๑ HIGH THO	๑๑ HIGH BO	๑๑ HIGH TSHO
๑๑ LOW XO	๑๑ LOW THO	๑๑ LOW BO	๑๑ LOW TSHO
๑๑ HIGH HO NGO	๑๑ HIGH HO NO	๑๑ HIGH HO LO	๑๑ HIGH HO NYO
๑๑ LOW NGO	๑๑ LOW NO	๑๑ LOW LO	๑๑ LOW NYO
๑๑ HIGH TSO	๑๑ HIGH PO	๑๑ HIGH HO	
๑๑ LOW TSO	๑๑ LOW PO	๑๑ LOW HO	
๑๑ HIGH SO	๑๑ HIGH PHO	๑๑ HIGH DO	
๑๑ LOW SO	๑๑ LOW PHO	๑๑ LOW DO	

High consonant letters with “ho”. There are seven high consonant letters with “ho” in their character names, because the characters are named according to how people pronounce these letters. For example, ๑ HIGH QO is pronounced /to³³ʔo³³/, ๑ LOW QO is pronounced /to³³ʔo⁴³/, ๑๑ HIGH KO is pronounced /to³³kɔ³³/, ๑๑ LOW KO is pronounced /to³³kɔ⁴³/, and so on. And for the seven letters, for example, ๑๑ HIGH HO NGO is pronounced as /hɔ³³ʔo³³/, ๑๑ HIGH HO YO is pronounced as /hɔ³³jɔ³³/, and so on. /to³³/ refers to a classifier, and therefore does not appear in character names. /hɔ³³/ indicates that the first part of these letters comes from the letter ๑ HIGH HO.

Labialization indicator. The consonant letter ๑ LOW VO is also used for indicating labialized consonant. For example, in the word ๑๑๑๑ (kwaen) “loosen”, ๑๑ is the consonant cluster for /k^w/ (if the inherited vowel is not considered). For a Brāhmī model, this consonant cluster would need to be encoded as <๑๑ HIGH KO, virama, ๑ LOW VO>, but it is indistinguishable from <๑๑ HIGH KO, ๑ LOW VO>. People pronounce this consonant cluster as /k^wa³³/ (kwa), because when people pronounce it with /ɔ/ (o), they find that the labialization is not reflected (/k^wɔ³³/ ~ /kɔ³³/), so they switch to using /a/.

High Qo. The consonant letter ๑ HIGH QO is also used as a vowel letter for /ɔ/. For example, in the word ๑๑๑๑ (pod) “lung”, ๑ HIGH QO is used for /ɔ/.

Consonants for coda. There are eight consonant letters that are used as a coda in syllables, they are: ๑ LOW YO, ๑ LOW VO, ๑ LOW NGO, ๑ LOW NO, ๑ LOW MO, ๑ HIGH KO, ๑ HIGH DO, and ๑ HIGH BO. They are used in syllables that end with /i/ (or /j/), /u/ (or /w/), /ɲ/, /n/, /m/, /k/, /t/, and /p/.

Initials. There are two types of initials in Tai Xaa, those using non-labialized single consonants, which use the 44 consonants mentioned above, and those using labialized single consonants, which do not appear in the modern texts but do appear in pre-modern literature. Eleven consonants (high/low ko, xo, ngo, ho and kho, plus high qo) have been identified that can be followed by a labialization mark, but it should be assumed that all consonants with a labialization mark constitute legal strings.

2.5 Vowel letters and vowel marks

There are 7 vowel letters and 11 vowel marks (presented with dotted circle):

MAAI KA-1	MAAI KI	MAAI KU-1	MAAI KAUE
MAAI KA-2	MAAI KE	MAAI KU-2	MAAI KAU
MAAI KA-3	MAAI KAE	MAAI KU-3	MAAI KAI
MAAI KA-4	MAAI TSAM	MAAI KUE	
MAAI KA-5	MAAI KOA	MAAI KEI	

Maai Ka. There are one vowel letter and four vowel marks that are pronounced as *maai ka*, and they can be grouped into two categories. MAAI KA-1 and MAAI KA-2 can be used in open syllables. Note that the vowel mark MAAI KA-2 can only be used in open syllables with tone 3, and the vowel letter MAAI KA-1 can also be used in closed syllables for long /a/. The other three vowel marks MAAI KA-3, MAAI KA-4, and MAAI KA-5 can only be used in closed syllables. Since there is no opposition of long and short /a/ in open syllables in Jinping dialect, they are all pronounced *maai ka*.

The three vowel marks MAAI KA-3, MAAI KA-4, and MAAI KA-5 are used with specific and complementary coda letters. MAAI KA-3 is combined with LOW NO or HIGH DO to form /an/ or /at/. MAAI KA-4 is combined with LOW NGO or HIGH KO to form /aŋ/ or /ak/. MAAI KA-5 is combined with LOW MO or HIGH BO to form /am/ or /ap/. Note that the two vowel marks MAAI KA-4 and MAAI KA-5 are rather flexible, as they can be positioned above the consonant letters in initial and in final. It should be assumed that whatever consonant letter the two vowel marks are placed on, the character sequences corresponding to the syllables they form are equivalent. For example, the character sequences *ŋn* /lak⁵⁵/ is equivalent to *ŋñ*.

Maai Kae and Maai Kau. The pre-base vowel letter MAAI KAE is used for vowel /ɛ/, as in *unə* /kɛu³³/, and the pre-base and post-base vowel letter MAAI KAU is used for vowels /au/, as in *ɪɪ* /ʔau³³/. Because the glyphs of these two characters are very similar, it should be assumed that the character sequence of two MAAI KAU is confusable to one MAAI KAE.

Maai Ku. There are one vowel letter and two vowel marks that are pronounced as *maai ku*. MAAI KU-1 can be used in any syllables. MAAI KU-2 can only be used with LOW NGO, for example, in *ɥɛ* /luŋ⁴³/. MAAI KU-3 can only be used with LOW MO, for example, in *ɤw* /tsum³³/.

Finals. All the finals in the modern text are listed in Fig. 2, which includes 9 finals with monophthongs, 14 finals with diphthongs, and 62 finals with consonant coda. Further finals appear in pre-modern literature, e.g. the single letter MAAI KU-2 and the single letter MAAI KA-5 are also used as finals.

2.6 Tone marks

There are 2 tone marks:

MAI SAK	MAI HOI
---------	---------

There are eight tone classes in Jinping dialect, according to Dao et al. (2019):

	A	B	C	D
high	1 / ³³ /	2 / ⁵⁵ /	3 / ²⁵ /	7 / ⁵⁵ /
low	4 / ⁴³ /	5 / ⁵² /	6 / ³¹ /	8 / ⁴³ /

For the eight tone classes:

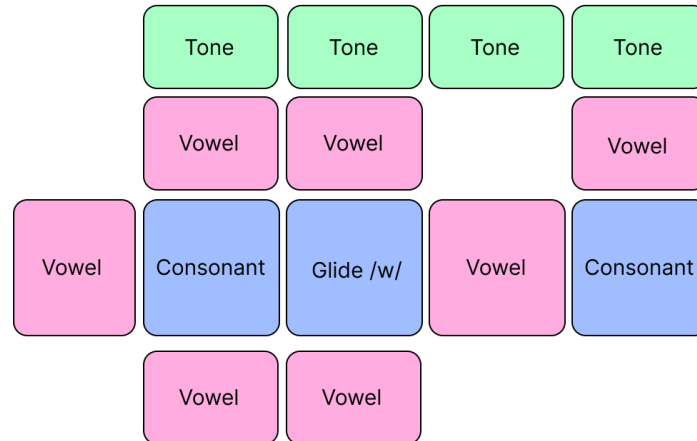
$\dot{n}$, /ka ³³ /	$\dot{n}$ HIGH KO, $\dot{\cdot}$ MAAI KA-1
$\dot{n}$, /ka ⁵⁵ /	$\dot{n}$ HIGH KO, $\dot{\circ}$ MAI SAK, $\dot{\cdot}$ MAAI KA-1
$\dot{n}$, /ka ²⁵ /	$\dot{n}$ HIGH KO, $\dot{\circ}$ MAI HOI, $\dot{\cdot}$ MAAI KA-1
$\dot{n}$, /ka ⁴³ /	$\dot{n}$ LOW KO, $\dot{\cdot}$ MAAI KA-1
$\dot{n}$, /ka ⁵² /	$\dot{n}$ LOW KO, $\dot{\circ}$ MAI SAK, $\dot{\cdot}$ MAAI KA-1
$\dot{n}$, /ka ³¹ /	$\dot{n}$ LOW KO, $\dot{\circ}$ MAI HOI, $\dot{\cdot}$ MAAI KA-1
$\dot{n}\check{\cdot}$ /kak ⁵⁵ /	$\dot{n}$ HIGH KO, $\dot{n}$ HIGH KO, $\dot{\circ}$ MAAI KA-4
$\dot{n}\check{\cdot}$ /kak ⁴³ /	$\dot{n}$ LOW KO, $\dot{n}$ HIGH KO, $\dot{\circ}$ MAAI KA-4

As can be seen above, $\dot{\circ}$ MAI SAK is used for tone 2 and tone 5, and $\dot{\circ}$ MAI HOI is used for tone 3 and tone 6.

In the Tai Xaaü script, the tone is used to modify the whole syllable, so its position is relatively flexible, and in the literature, it can be added above any grapheme cluster. Therefore, it should be assumed that the resulting character sequence is equivalent regardless of the grapheme cluster to which the tone mark is added. For example, the character sequence $\dot{\cdot}\dot{\cdot}\dot{\cdot}$ /sin⁵⁵/ is equivalent to $\dot{\cdot}\dot{\cdot}\dot{\cdot}$.

2.7 Syllable structure

The structure of each syllable in the Tai Xaaü script can be summarized as:



The structure of a Tai Xaaü syllable can be described as **OW?VC?T?**, where:

- **O** is a required onset, and it can be any of 44 consonant letters in section 2.4;
- **W** is an optional labialization indicator $\dot{\cdot}$ LOW VO;
- **V** is a required vowel letter or vowel mark, where $\dot{\cdot}$ MAAI KA-1 and $\dot{\cdot}$ MAAI KOA are post-base vowel letters, $\dot{\cdot}$ MAAI KAU should be used twice (one pre-base and one post-base), the other four vowel letters are pre-base;
- **C** is an optional coda, and it can be any of the eight consonant letters as described in section 2.4; the C and V will form a ligature $\dot{\cdot}$ or $\dot{\cdot}$, as described in section 2.5;
- **T** is an optional tone mark, and it can be any of two as described in section 2.6.

For examples:

word	input
ໄຢໄ /ʔau ³³ / “want”	ໄ MAAI KAU (V), ອ HIGH QO (O), ໄ MAAI KAU (V)
ໄ໓ໄ /kau ²⁵ / “nine”	ໄ MAAI KAU (V), ນ HIGH KO (O), ອ MAI HOI (T), ໄ MAAI KAU (V)
ໄ໓ໄ /pau ⁵² / “who”	ໄ MAAI KAUE (V), ນ LOW PHO (O), ອ MAI SAK (T)
ໄ໓ໄ /neu ³³ / “sticky”	ໄ HIGH HO NO (O), ອ MAAI KE (V), ອ LOW VO (C)
ໄ໓ໄ /sɔŋ ³³ / “two”	ໄ HIGH SO (O), ອ HIGH QO (V), ອ LOW NGO (C)
ໄ໓ໄ /mak ⁴³ / “love”	ໄ LOW MO (O), ນ HIGH KO (C), ອ MAAI KA-4 (V)
ໄ໓ໄ /mwi ⁵⁵ / “drizzle”	ໄ HIGH HO MO (O), ອ MAAI KUE (V), ອ MAI SAK (T), ອ LOW YO (C)
ໄ໓ໄ /huŋ ⁵² / “daybreak”	ໄ MAAI KU-2 (V), ອ LOW HO (O), ອ MAI SAK (T), ອ LOW NGO (C)
ໄ໓ໄ /x ^w a ³³ / “right”	ໄ HIGH XO (O), ອ LOW VO (W), ອ MAAI KA-1 (V)
ໄ໓ໄ /x ^w a:ŋ ³³ / “horizontal”	ໄ HIGH XO (O), ອ LOW VO (W), ອ MAAI KA-1 (V), ອ LOW NGO (C)
ໄ໓ໄ /k ^w ɛŋ ³³ / “soup”	ໄ MAAI KAE (V), ນ HIGH KO (O), ອ LOW VO (W), ອ LOW NGO (C)
ໄ໓ໄ /k ^w in ³³ / “eat”	ໄ HIGH KO (O), ອ LOW VO (W), ອ MAAI KI (V), ອ LOW NO (C)
ໄ໓ໄ /k ^{hw} an ³³ / “scarf”	ໄ HIGH KHO (O), ອ LOW VO (W), ອ LOW NO (C), ອ MAAI KA-3 (V)

2.8 Symbols and punctuation

There are 5 symbols with specific meanings:

ໄ NUENG “one”	ໄ BEN “moon”
ໄ KUN “person”	ໄ MAN “he”
ໄ MA “horse”	

Note that one should use ໄ MAN and should not use <ໄ LOW MO, ອ MAAI KA-3>. More symbols will be discovered during the subsequent literature compilation process, so we hope to reserve some codepoints for them.

There are 7 Tai Xaaú-specific punctuation that are used for indicating the end of a sentence or an article:

ໄ MAAI SUT	ໄ LAN KAAAN-1
ໄ MAAI SAM	ໄ LAN KAAAN-2
ໄ MAAI SAM KOAK	ໄ LAN KAAAN-3
ໄ MAAI SAM MAET	

The rest of the punctuation is identical to that used in Han ideographs.

2.9 Collation

The syllables in the Tai Xaaú script are ordered by initials, finals and tones.

- The order of initials is from ๑ HIGH QO to ๓๓ LOW NYO, then from ๑ HIGH QO, ๑ LOW VO to ๓๓ LOW NYO, ๑ LOW VO.
- The order of finals is (where “-” refers to an initial):
a (๑ < ๑), i (๑), e (๑), ae (๑), o (๑), oa (๑), u (๑), ue (๑), ei (๑);
ai (๑), aai (๑), oi (๑), oai (๑), ui (๑), uei (๑), eii (๑);
au (๑-๑), aau (๑-๑), iu (๑-๑), eu (๑-๑), aeu (๑-๑), eiu (๑-๑);
aue (๑);
ang (๑), aang (๑), ing (๑), eng (๑), aeng (๑-๑), ong (๑), oang (๑),
ung (๑ < ๑-๑), ueng (๑), eing (๑);
an (๑), aan (๑), ..., ein (๑);
am (๑), aam (๑), ..., um (๑ < ๑), ..., eim (๑);
ak (๑), aak (๑), ..., eik (๑);
at (๑), aat (๑), ..., eit (๑);
ap (๑), aap (๑), ..., eip (๑).
• The order of tones is: no tone, ๑ MAI SAK, ๑ MAI HOI.

Note that, the syllables recorded in the Tai Xaaui script are basically the original form of the morphemes, but in spoken language, the same morphemes change when they form different words, depending on their contexts, resulting in inconsistencies between the reading forms and the written forms. For examples, when /ma⁵⁵/ “gain” appears in the first syllable of a compound word, it should be read as /m⁵⁵/, thus changing its final; or, when a syllable in tone 4 follows a syllable in tone 3, the tone value /⁴³/ will change to /⁴⁵/, as /kun⁴³/ “person” in /ha²⁵kun⁴⁵/ “five persons”.

The order of Tai Xaaui characters is:

๑ HIGH QO < ๑ LOW QO < ๓ HIGH KO < ๓ LOW KO < ๑๖ HIGH XO < ๑๑ LOW XO <
๑๑ HIGH HO NGO < ๑ LOW NGO < ๑ HIGH TSO < ๑ LOW TSO < ๑ HIGH SO <
๑ LOW SO < ๑๑ HIGH HO YO < ๑ LOW YO < ๑ HIGH TO < ๑ LOW TO <
๑ HIGH THO < ๑ LOW THO < ๑ HIGH HO NO < ๑ LOW NO < ๑ HIGH PO <
๑ LOW PO < ๑ HIGH PHO < ๑ LOW PHO < ๑ HIGH HO MO < ๑ LOW MO <
๑ HIGH FO < ๑ LOW FO < ๑ HIGH BO < ๑ LOW BO < ๑ HIGH HO LO <
๑ LOW LO < ๑ HIGH HO < ๑ LOW HO < ๑ HIGH DO < ๑ LOW DO <
๑ HIGH HO VO < ๑ LOW VO < ๑ HIGH KHO < ๑ LOW KHO < ๑ HIGH TSHO <
๑ LOW TSHO < ๑๑ HIGH HO NYO < ๑ LOW NYO <
๑ MAI SAK < ๑ MAI HOI <
๑ MAAI KA-1 < ๑ MAAI KA-2 < ๑ MAAI KA-3 < ๑ MAAI KA-4 < ๑ MAAI KA-5 <
๑ MAAI KI < ๑ MAAI KAE < ๑ MAAI KE < ๑ MAAI TSAM < ๑ MAAI KOA <
๑ MAAI KU-1 < ๑ MAAI KU-2 < ๑ MAAI KU-3 < ๑ MAAI KUE < ๑ MAAI KEI <
๑ MAAI KAUE < ๑ MAAI KAU < ๑ MAAI KAI <
๑ NUENG < ๑ KUN < ๑ MA < ๑ BEN < ๑ MAN <
๑ MAAI SUT < ๑ MAAI SAM < ๑ MAAI SAM KOAK < ๑ MAAI SAM MAET <
๑ LAN KAAAN-1 < ๑ LAN KAAAN-2 < ๑ LAN KAAAN-3

3 Proposed characters

3.1 Repertoire

The proposed repertoire is shown as below.

Codepoint	Glyph	Character Name
U+XX00		TAI XAAU LETTER HIGH QO
U+XX01		TAI XAAU LETTER LOW QO
U+XX02		TAI XAAU LETTER HIGH KO
U+XX03		TAI XAAU LETTER LOW KO
U+XX04		TAI XAAU LETTER HIGH XO
U+XX05		TAI XAAU LETTER LOW XO
U+XX06		TAI XAAU LETTER HIGH HO NGO
U+XX07		TAI XAAU LETTER LOW NGO
U+XX08		TAI XAAU LETTER HIGH TSO
U+XX09		TAI XAAU LETTER LOW TSO
U+XX0A		TAI XAAU LETTER HIGH SO
U+XX0B		TAI XAAU LETTER LOW SO
U+XX0C		TAI XAAU LETTER HIGH HO YO
U+XX0D		TAI XAAU LETTER LOW YO
U+XX0E		TAI XAAU LETTER HIGH TO
U+XX0F		TAI XAAU LETTER LOW TO

U+XX10	ŋ	TAI XAAU LETTER HIGH THO
U+XX11	ɲ	TAI XAAU LETTER LOW THO
U+XX12	ɲʰ	TAI XAAU LETTER HIGH HO NO
U+XX13	ɲ̥	TAI XAAU LETTER LOW NO
U+XX14	ʋ	TAI XAAU LETTER HIGH PO
U+XX15	ɲʰ	TAI XAAU LETTER LOW PO
U+XX16	ɲʰ	TAI XAAU LETTER HIGH PHO
U+XX17	ɲ̥	TAI XAAU LETTER LOW PHO
U+XX18	ɲʰ	TAI XAAU LETTER HIGH HO MO
U+XX19	ɲ̥	TAI XAAU LETTER LOW MO
U+XX1A	ɲʰ	TAI XAAU LETTER HIGH FO
U+XX1B	ɲ̥	TAI XAAU LETTER LOW FO
U+XX1C	ʋ	TAI XAAU LETTER HIGH BO
U+XX1D	ɲ̥	TAI XAAU LETTER LOW BO
U+XX1E	ɲ	TAI XAAU LETTER HIGH HO LO
U+XX1F	ɲ̥	TAI XAAU LETTER LOW LO
U+XX20	ɲ	TAI XAAU LETTER HIGH HO
U+XX21	ɲ̥	TAI XAAU LETTER LOW HO
U+XX22	ɲ	TAI XAAU LETTER HIGH DO
U+XX23	ɲ̥	TAI XAAU LETTER LOW DO

U+XX24		TAI XAAU LETTER HIGH HO VO
U+XX25		TAI XAAU LETTER LOW VO
U+XX26		TAI XAAU LETTER HIGH KHO
U+XX27		TAI XAAU LETTER LOW KHO
U+XX28		TAI XAAU LETTER HIGH TSHO
U+XX29		TAI XAAU LETTER LOW TSHO
U+XX2A		TAI XAAU LETTER HIGH HO NYO
U+XX2B		TAI XAAU LETTER LOW NYO
U+XX30		TAI XAAU MAAI KA-1
U+XX31		TAI XAAU MAAI KA-2
U+XX32		TAI XAAU MAAI KA-3
U+XX33		TAI XAAU MAAI KA-4
U+XX34		TAI XAAU MAAI KA-5
U+XX35		TAI XAAU MAAI KI
U+XX36		TAI XAAU MAAI KE
U+XX37		TAI XAAU MAAI KAE
U+XX38		TAI XAAU MAAI TSAM
U+XX39		TAI XAAU MAAI KOA
U+XX3A		TAI XAAU MAAI KU-1
U+XX3B		TAI XAAU MAAI KU-2

U+XX3C		TAI XAAU MAAI KU-3
U+XX3D		TAI XAAU MAAI KUE
U+XX3E		TAI XAAU MAAI KEI
U+XX3F		TAI XAAU MAAI KAUE
U+XX40		TAI XAAU MAAI KAU
U+XX41		TAI XAAU MAAI KAI
U+XX42		TAI XAAU TONE MAI SAK
U+XX43		TAI XAAU TONE MAI HOI
U+XX44		TAI XAAU SYMBOL NUENG
U+XX45		TAI XAAU SYMBOL KUN
U+XX46		TAI XAAU SYMBOL MA
U+XX47		TAI XAAU SYMBOL BEN
U+XX48		TAI XAAU SYMBOL MAN
U+XX49		TAI XAAU SYMBOL MAAI SUT
U+XX4A		TAI XAAU SYMBOL MAAI SAM
U+XX4B		TAI XAAU SYMBOL MAAI SAM KOAK
U+XX4C		TAI XAAU SYMBOL MAAI SAM MAET
U+XX4D		TAI XAAU SYMBOL LAN KAAAN-1
U+XX4E		TAI XAAU SYMBOL LAN KAAAN-2
U+XX4F		TAI XAAU SYMBOL LAN KAAAN-3

3.2 Character properties

The information in NamesList.txt is as follows.

@@	XX00	Tai Xaau	XX5F
@		Consonants	
XX00	TAI XAAU LETTER HIGH QO		
	x (tai viet letter low o - AAAE)		
XX01	TAI XAAU LETTER LOW QO		
	x (tai viet letter high o - AAAF)		
XX02	TAI XAAU LETTER HIGH KO		
	x (tai viet letter low ko - AA80)		
XX03	TAI XAAU LETTER LOW KO		
XX04	TAI XAAU LETTER HIGH XO		
XX05	TAI XAAU LETTER LOW XO		
XX06	TAI XAAU LETTER HIGH HO NGO		
XX07	TAI XAAU LETTER LOW NGO		
	x (tai viet letter high ngo - AA89)		
XX08	TAI XAAU LETTER HIGH TSO		
	x (tai viet letter low co - AA8A)		
XX09	TAI XAAU LETTER LOW TSO		
	x (tai viet letter high co - AA8B)		
XX0A	TAI XAAU LETTER HIGH SO		
XX0B	TAI XAAU LETTER LOW SO		
XX0C	TAI XAAU LETTER HIGH HO YO		
XX0D	TAI XAAU LETTER LOW YO		
	x (tai viet letter high yo - AAA5)		
XX0E	TAI XAAU LETTER HIGH TO		
	x (tai viet letter low to - AA94)		
XX0F	TAI XAAU LETTER LOW TO		
	x (tai viet letter high to - AA95)		
XX10	TAI XAAU LETTER HIGH THO		
	x (tai viet letter low tho - AA96)		
XX11	TAI XAAU LETTER LOW THO		
XX12	TAI XAAU LETTER HIGH HO NO		
	x (tai viet letter low no - AA98)		
XX13	TAI XAAU LETTER LOW NO		
	x (tai viet letter high no - AA99)		
XX14	TAI XAAU LETTER HIGH PO		
XX15	TAI XAAU LETTER LOW PO		
XX16	TAI XAAU LETTER HIGH PHO		
XX17	TAI XAAU LETTER LOW PHO		
XX18	TAI XAAU LETTER HIGH HO MO		
XX19	TAI XAAU LETTER LOW MO		
XX1A	TAI XAAU LETTER HIGH FO		
XX1B	TAI XAAU LETTER LOW FO		
XX1C	TAI XAAU LETTER HIGH BO		
	x (tai viet letter low bo - AA9A)		
XX1D	TAI XAAU LETTER LOW BO		
	x (tai viet letter high bo - AA9B)		
XX1E	TAI XAAU LETTER HIGH HO LO		
	x (tai viet letter low lo - AAA8)		
XX1F	TAI XAAU LETTER LOW LO		
XX20	TAI XAAU LETTER HIGH HO		
	x (tai viet letter low ho - AAAC)		
XX21	TAI XAAU LETTER LOW HO		
	x (tai viet letter high ho - AAAD)		
XX22	TAI XAAU LETTER HIGH DO		
XX23	TAI XAAU LETTER LOW DO		
	x (tai viet letter high do - AA93)		
XX24	TAI XAAU LETTER HIGH HO VO		
	x (tai viet letter low vo - AAAA)		

XX25 TAI XAAU LETTER LOW VO
 x (tai viet letter high vo - AAAB)
 XX26 TAI XAAU LETTER HIGH KHO
 XX27 TAI XAAU LETTER LOW KHO
 x (tai viet letter high kho - AA83)
 XX28 TAI XAAU LETTER HIGH TSHO
 x (tai viet letter low cho - AA8C)
 XX29 TAI XAAU LETTER LOW TSHO
 x (tai viet letter high cho - AA8D)
 XX2A TAI XAAU LETTER HIGH HO NYO
 XX2B TAI XAAU LETTER LOW NYO
 @ Vowels
 XX30 TAI XAAU MAAI KA-1
 x (tai viet vowel aa - AAB1)
 XX31 TAI XAAU MAAI KA-2
 XX32 TAI XAAU MAAI KA-3
 XX33 TAI XAAU MAAI KA-4
 x (tai viet mai kang - AAB0)
 XX34 TAI XAAU MAAI KA-5
 x (tai viet vowel am - AABE)
 XX35 TAI XAAU MAAI KI
 x (tai viet vowel i - AAB2)
 XX36 TAI XAAU MAAI KE
 x (tai viet vowel ia - AAB8)
 XX37 TAI XAAU MAAI KAE
 * precedes consonant in visual order
 x (tai viet vowel e - AAB5)
 XX38 TAI XAAU MAAI TSAM
 XX39 TAI XAAU MAAI KOA
 x (tai viet vowel ua - AABA)
 XX3A TAI XAAU MAAI KU-1
 x (tai viet vowel u - AAB4)
 XX3B TAI XAAU MAAI KU-2
 * precedes consonant in visual order
 x (tai viet vowel o - AAB6)
 XX3C TAI XAAU MAAI KU-3
 XX3D TAI XAAU MAAI KUE
 x (tai viet vowel ue - AAB3)
 XX3E TAI XAAU MAAI KEI
 x (tai viet mai khít - AAB7)
 XX3F TAI XAAU MAAI KAUE
 * precedes consonant in visual order
 x (tai viet vowel aue - AAB8)
 XX40 TAI XAAU MAAI KAU
 * precedes and follows consonant in visual order
 x (tai viet vowel uea - AAB9)
 XX41 TAI XAAU MAAI KAI
 * precedes consonant in visual order
 x (tai viet vowel ay - AABC)
 @ Tones
 XX42 TAI XAAU TONE MAI SAK
 x (tai viet tone mai ek - AABF)
 x (combining acute accent - 0301)
 XX43 TAI XAAU TONE MAI HOI
 x (tai viet tone mai tho - AAC1)
 x (combining hook above - 0309)
 @ Word ligature symbols
 XX44 TAI XAAU SYMBOL NUENG
 = one
 x (tai viet symbol nueng - AADC)
 XX45 TAI XAAU SYMBOL KUN
 = person
 x (tai viet symbol kon - AADB)

```

XX46      TAI XAAU SYMBOL MA
          = horse
XX47      TAI XAAU SYMBOL BEN
          = moon
XX48      TAI XAAU SYMBOL MAN
          = the third person singular
@         Punctuation
XX49      TAI XAAU SYMBOL MAAI SUT
XX4A      TAI XAAU SYMBOL MAAI SAM
          x (tai viet symbol koi koi - AADF)
XX4B      TAI XAAU SYMBOL MAAI SAM KOAK
          x (tai viet symbol ho hoi - AADE)
XX4C      TAI XAAU SYMBOL MAAI SAM MAET
XX4D      TAI XAAU SYMBOL LAN KAAAN-1
XX4E      TAI XAAU SYMBOL LAN KAAAN-2
XX4F      TAI XAAU SYMBOL LAN KAAAN-3

```

The information in UnicodeData.txt is as follows.

```

XX00;TAI XAAU LETTER HIGH QO;Lo;0;L;;;;N;;;;;
XX01;TAI XAAU LETTER LOW QO;Lo;0;L;;;;N;;;;;
XX02;TAI XAAU LETTER HIGH KO;Lo;0;L;;;;N;;;;;
XX03;TAI XAAU LETTER LOW KO;Lo;0;L;;;;N;;;;;
XX04;TAI XAAU LETTER HIGH XO;Lo;0;L;;;;N;;;;;
XX05;TAI XAAU LETTER LOW XO;Lo;0;L;;;;N;;;;;
XX06;TAI XAAU LETTER HIGH HO NGO;Lo;0;L;;;;N;;;;;
XX07;TAI XAAU LETTER LOW NGO;Lo;0;L;;;;N;;;;;
XX08;TAI XAAU LETTER HIGH TSO;Lo;0;L;;;;N;;;;;
XX09;TAI XAAU LETTER LOW TSO;Lo;0;L;;;;N;;;;;
XX0A;TAI XAAU LETTER HIGH SO;Lo;0;L;;;;N;;;;;
XX0B;TAI XAAU LETTER LOW SO;Lo;0;L;;;;N;;;;;
XX0C;TAI XAAU LETTER HIGH HO YO;Lo;0;L;;;;N;;;;;
XX0D;TAI XAAU LETTER LOW YO;Lo;0;L;;;;N;;;;;
XX0E;TAI XAAU LETTER HIGH TO;Lo;0;L;;;;N;;;;;
XX0F;TAI XAAU LETTER LOW TO;Lo;0;L;;;;N;;;;;
XX10;TAI XAAU LETTER HIGH THO;Lo;0;L;;;;N;;;;;
XX11;TAI XAAU LETTER LOW THO;Lo;0;L;;;;N;;;;;
XX12;TAI XAAU LETTER HIGH HO NO;Lo;0;L;;;;N;;;;;
XX13;TAI XAAU LETTER LOW NO;Lo;0;L;;;;N;;;;;
XX14;TAI XAAU LETTER HIGH PO;Lo;0;L;;;;N;;;;;
XX15;TAI XAAU LETTER LOW PO;Lo;0;L;;;;N;;;;;
XX16;TAI XAAU LETTER HIGH PHO;Lo;0;L;;;;N;;;;;
XX17;TAI XAAU LETTER LOW PHO;Lo;0;L;;;;N;;;;;
XX18;TAI XAAU LETTER HIGH HO MO;Lo;0;L;;;;N;;;;;
XX19;TAI XAAU LETTER LOW MO;Lo;0;L;;;;N;;;;;
XX1A;TAI XAAU LETTER HIGH FO;Lo;0;L;;;;N;;;;;
XX1B;TAI XAAU LETTER LOW FO;Lo;0;L;;;;N;;;;;
XX1C;TAI XAAU LETTER HIGH BO;Lo;0;L;;;;N;;;;;
XX1D;TAI XAAU LETTER LOW BO;Lo;0;L;;;;N;;;;;
XX1E;TAI XAAU LETTER HIGH HO LO;Lo;0;L;;;;N;;;;;
XX1F;TAI XAAU LETTER LOW LO;Lo;0;L;;;;N;;;;;
XX20;TAI XAAU LETTER HIGH HO;Lo;0;L;;;;N;;;;;
XX21;TAI XAAU LETTER LOW HO;Lo;0;L;;;;N;;;;;
XX22;TAI XAAU LETTER HIGH DO;Lo;0;L;;;;N;;;;;
XX23;TAI XAAU LETTER LOW DO;Lo;0;L;;;;N;;;;;
XX24;TAI XAAU LETTER HIGH HO VO;Lo;0;L;;;;N;;;;;
XX25;TAI XAAU LETTER LOW VO;Lo;0;L;;;;N;;;;;
XX26;TAI XAAU LETTER HIGH KHO;Lo;0;L;;;;N;;;;;
XX27;TAI XAAU LETTER LOW KHO;Lo;0;L;;;;N;;;;;
XX28;TAI XAAU LETTER HIGH TSHO;Lo;0;L;;;;N;;;;;
XX29;TAI XAAU LETTER LOW TSHO;Lo;0;L;;;;N;;;;;
XX2A;TAI XAAU LETTER HIGH HO NYO;Lo;0;L;;;;N;;;;;
XX2B;TAI XAAU LETTER LOW NYO;Lo;0;L;;;;N;;;;;

```

```

XX30;TAI XAAU MAAI KA-1;Lo;0;L;;;;N;;;;;
XX31;TAI XAAU MAAI KA-2;Mn;220;NSM;;;;N;;;;;
XX32;TAI XAAU MAAI KA-3;Mn;230;NSM;;;;N;;;;;
XX33;TAI XAAU MAAI KA-4;Mn;230;NSM;;;;N;;;;;
XX34;TAI XAAU MAAI KA-5;Mn;230;NSM;;;;N;;;;;
XX35;TAI XAAU MAAI KI;Mn;230;NSM;;;;N;;;;;
XX36;TAI XAAU MAAI KE;Mn;230;NSM;;;;N;;;;;
XX37;TAI XAAU MAAI KAE;Lo;0;L;;;;N;;;;;
XX38;TAI XAAU MAAI TSAM;Mn;230;NSM;;;;N;;;;;
XX39;TAI XAAU MAAI KOA;Lo;0;L;;;;N;;;;;
XX3A;TAI XAAU MAAI KU-1;Mn;220;NSM;;;;N;;;;;
XX3B;TAI XAAU MAAI KU-2;Lo;0;L;;;;N;;;;;
XX3C;TAI XAAU MAAI KU-3;Mn;230;NSM;;;;N;;;;;
XX3D;TAI XAAU MAAI KUE;Mn;230;NSM;;;;N;;;;;
XX3E;TAI XAAU MAAI KEI;Mn;230;NSM;;;;N;;;;;
XX3F;TAI XAAU MAAI KAUE;Lo;0;L;;;;N;;;;;
XX40;TAI XAAU MAAI KAU;Lo;0;L;;;;N;;;;;
XX41;TAI XAAU MAAI KAI;Lo;0;L;;;;N;;;;;
XX42;TAI XAAU TONE MAI SAK;Mn;230;NSM;;;;N;;;;;
XX43;TAI XAAU TONE MAI HOI;Mn;230;NSM;;;;N;;;;;
XX44;TAI XAAU SYMBOL NUENG;Lo;0;L;;;;N;;;;;
XX45;TAI XAAU SYMBOL KUN;Lo;0;L;;;;N;;;;;
XX46;TAI XAAU SYMBOL MA;Lo;0;L;;;;N;;;;;
XX47;TAI XAAU SYMBOL BEN;Lo;0;L;;;;N;;;;;
XX48;TAI XAAU SYMBOL MAN;Lo;0;L;;;;N;;;;;
XX49;TAI XAAU SYMBOL MAAI SUT;Po;0;L;;;;N;;;;;
XX4A;TAI XAAU DIGIT MAAI SAM;Po;0;L;;;;N;;;;;
XX4B;TAI XAAU DIGIT MAAI SAM KOAK;Po;0;L;;;;N;;;;;
XX4C;TAI XAAU DIGIT MAAI SAM MAET;Po;0;L;;;;N;;;;;
XX4D;TAI XAAU DIGIT LAN KAAAN-1;Po;0;L;;;;N;;;;;
XX4E;TAI XAAU DIGIT LAN KAAAN-2;Po;0;L;;;;N;;;;;
XX4F;TAI XAAU DIGIT LAN KAAAN-3;Po;0;L;;;;N;;;;;

```

The Tai XaaU script is a newly proposed script for ISO 15924. The suggested code is `Taxa`.

Script=`Taxa`

The information in `IndicPositionalCategory.txt` is as follows.

```

# Indic_Positional_Category=Right
XX30          ; Right          # Lo      TAI XAAU MAAI KA-1
XX39          ; Right          # Lo      TAI XAAU MAAI KOA
XX40          ; Right          # Lo      TAI XAAU MAAI KAU

# Indic_Positional_Category=Visual_Order_Left
XX37          ; Visual_Order_Left # Lo      TAI XAAU MAAI KAE
XX3A          ; Visual_Order_Left # Lo      TAI XAAU MAAI KU-2
XX3E..XX40    ; Visual_Order_Left # Lo [3]  TAI XAAU MAAI KAUE..TAI XAAU MAAI KAI

# Indic_Positional_Category=Top
XX32..XX36    ; Top            # Mn [5]  TAI XAAU MAAI KA-3..TAI XAAU MAAI KE
XX38          ; Top            # Mn      TAI XAAU MAAI TSAM
XX3C..XX3E    ; Top            # Mn [3]  TAI XAAU MAAI KU-3..TAI XAAU MAAI KEI
XX42..XX43    ; Top            # Mn [2]  TAI XAAU TONE MAI SAK..TAI XAAU TONE
          MAI HOI

# Indic_Positional_Category=Bottom
XX31          ; Bottom         # Mn      TAI XAAU MAAI KA-2
XX3A          ; Bottom         # Mn      TAI XAAU MAAI KU-1

```

The information in `IndicSyllableCategory.txt` is as follows.

```

# Indic_Syllabic_Category=Vowel_Dependent
XX30          ; Vowel_Dependent # Lo      TAI XAAU MAAI KA-1

```

```

XX31      ; Vowel_Dependent      # Mn      TAI XAAU MAAI KA-2
XX32..XX36 ; Vowel_Dependent      # Mn      [5] TAI XAAU MAAI KA-3..TAI XAAU MAAI KE
XX37      ; Vowel_Dependent      # Lo      TAI XAAU MAAI KAE
XX38      ; Vowel_Dependent      # Mn      TAI XAAU MAAI TSAM
XX39      ; Vowel_Dependent      # Lo      TAI XAAU MAAI KOA
XX3A      ; Vowel_Dependent      # Mn      TAI XAAU MAAI KU-1
XX3B      ; Vowel_Dependent      # Lo      TAI XAAU MAAI KU-2
XX3C..XX3E ; Vowel_Dependent      # Mn      [3] TAI XAAU MAAI KU-3..TAI XAAU MAAI KEI
XX3F..XX41 ; Vowel_Dependent      # Lo      [3] TAI XAAU MAAI KAUE..TAI XAAU MAAI KAI

# Indic_Syllabic_Category=Consonant
XX00..XX2B ; Consonant            # Lo      [44] TAI XAAU LETTER HIGH QO..TAI XAAU LETTER
LOW NYO

# Indic_Syllabic_Category=Tone_Mark
XX42..XX43 ; Tone_Mark           # Mn      [2] TAI XAAU TONE MAI SAK..TAI XAAU TONE
MAI HOI

```

The information in LineBreak.txt is as follows.

```

XX00..XX2B ; SA      # Lo [44] TAI XAAU LETTER HIGH QO..TAI XAAU LETTER LOW NYO
XX30      ; SA      # Lo      TAI XAAU MAAI KA-1
XX31      ; SA      # Mn      TAI XAAU MAAI KA-2
XX32..XX36 ; SA      # Mn      [5] TAI XAAU MAAI KA-3..TAI XAAU MAAI KE
XX37      ; SA      # Lo      TAI XAAU MAAI KAE
XX38      ; SA      # Mn      TAI XAAU MAAI TSAM
XX39      ; SA      # Lo      TAI XAAU MAAI KOA
XX3A      ; SA      # Mn      TAI XAAU MAAI KU-1
XX3B      ; SA      # Lo      TAI XAAU MAAI KU-2
XX3C..XX3E ; SA      # Mn      [3] TAI XAAU MAAI KU-3..TAI XAAU MAAI KEI
XX3F..XX41 ; SA      # Lo      [3] TAI XAAU MAAI KAUE..TAI XAAU MAAI KAI
XX42..XX43 ; SA      # Mn      [2] TAI XAAU TONE MAI SAK..TAI XAAU TONE MAI HOI
XX44..XX48 ; SA      # Lo      [5] TAI XAAU SYMBOL NUENG..TAI XAAU SYMBOL MAN
XX49..XX4F ; SA      # Po      [7] TAI XAAU SYMBOL MAAI SUT..TAI XAAU SYMBOL LAN KAN-3

```

The information in PropList.txt is as follows.

```

XX49..XX4F ; Terminal_Punctuation # Po      [7] TAI XAAU SYMBOL MAAI SUT..TAI XAAU
SYMBOL LAN KAN-3
XX49..XX4F ; Sentence_Terminal    # Po      [7] TAI XAAU SYMBOL MAAI SUT..TAI XAAU
SYMBOL LAN KAN-3

XX31      ; Other_Alphabetic      # Mn      TAI XAAU MAAI KA-2
XX32..XX36 ; Other_Alphabetic      # Mn      [5] TAI XAAU MAAI KA-3..TAI XAAU MAAI KE
XX38      ; Other_Alphabetic      # Mn      TAI XAAU MAAI TSAM
XX3A      ; Other_Alphabetic      # Mn      TAI XAAU MAAI KU-1
XX3C..XX3E ; Other_Alphabetic      # Mn      [3] TAI XAAU MAAI KU-3..TAI XAAU MAAI KEI

XX42..XX43 ; Diacritic            # Mn      [2] TAI XAAU TONE MAI SAK..TAI XAAU TONE
MAI HOI

XX37      ; Logical_Order_Exception # Lo      TAI XAAU MAAI KAE
XX3A      ; Logical_Order_Exception # Lo      TAI XAAU MAAI KU-2
XX3E..XX40 ; Logical_Order_Exception # Lo      [3] TAI XAAU MAAI KAUE..TAI
XAAU MAAI KAI

```

The information in VerticalOrientation.txt is as follows.

```

XX00..XX2B ; R      # Lo [44] TAI XAAU LETTER HIGH QO..TAI XAAU LETTER LOW NYO
XX30      ; R      # Lo      TAI XAAU MAAI KA-1
XX31      ; R      # Mn      TAI XAAU MAAI KA-2
XX32..XX36 ; R      # Mn      [5] TAI XAAU MAAI KA-3..TAI XAAU MAAI KE
XX37      ; R      # Lo      TAI XAAU MAAI KAE
XX38      ; R      # Mn      TAI XAAU MAAI TSAM

```

```

XX39      ; R      # Lo      TAI XAAU MAAI KOA
XX3A      ; R      # Mn      TAI XAAU MAAI KU-1
XX3B      ; R      # Lo      TAI XAAU MAAI KU-2
XX3C..XX3E ; R      # Mn [3]  TAI XAAU MAAI KU-3..TAI XAAU MAAI KEI
XX3F..XX41 ; R      # Lo [3]  TAI XAAU MAAI KAUE..TAI XAAU MAAI KAI
XX42..XX43 ; R      # Mn [2]  TAI XAAU TONE MAI SAK..TAI XAAU TONE MAI HOI
XX44..XX48 ; R      # Lo [5]  TAI XAAU SYMBOL NUENG..TAI XAAU SYMBOL MAN
XX49..XX4F ; R      # Po [7]  TAI XAAU SYMBOL MAAI SUT..TAI XAAU SYMBOL LAN KAN-3

```

The information in DoNotEmit.txt is as follows.

```

# Tai_Xaau_Word_Ligature:
#   Sequences that look like a Tai Xaau word ligature but should be avoided
#   in representing that word ligature.

XX19 XX32; XX48; Tai_Xaau_Word_Ligature # TAI XAAU LETTER LOW MO, TAI XAAU MAAI KA-3;
      TAI XAAU SYMBOL MAN

```

The information in confusables.txt (UTR #39) is as follows.

```

XX37 ; XX40 XX40 ; MA #* ( ㄥ → ㄥ ) TAI XAAU MAAI KAE → TAI XAAU MAAI KAU, TAI XAAU
      MAAI KAU #

```

Reference

Compilation Office of Honghe Hani and Yi Autonomous Prefecture 红河哈尼族彝族自治州民族志编写办公室, (1989). *Ethnography of Honghe Hani and Yi Autonomous Prefecture in Yunnan Province* / 云南省红河哈尼族彝族自治州民族志. Kunming / 昆明: Yunnan University Press / 云南大学出版社.

DAO Jie 刀洁, (2016). “Phonological Difference between Jinping Tai Script and Jinping Tai Language” / 金平傣文和金平傣语的音系差异, in: *Protection and Inheritance of Language Resources* / 语言资源的保护与传承. Beijing / 北京: Minzu Publishing House / 民族出版社.

DAO Jie 刀洁, ZHOU Yan 周焱, WANG Yuqing 王玉清, (2019). “The standardization of the phonetics of Jinping Dai writing system in Yunnan” / 云南金平傣文语音的规范化问题, in: *Language Plannology Studies* / 语言规划学研究, No. 2: pp. 57–68.

DAO Jie 刀洁, LUO Wenbin 罗文斌, (2025). “On the significance of the Tai Xaau encoding in forging a strong sense of community for the Chinese nation” / 论傣皓文字符编码对铸牢中华民族共同体意识的意义, in: *Ethnic Today* / 今日民族, No. 5: pp. 68–70.

HE Bingkun 何炳坤, (1999). “Jinping Dai Script” / 金平傣文, in: HE Lifeng 和丽峰, XIONG Yuyou 熊玉有. *Outline of Yunnan Minzu Scripts* / 云南少数民族文字概要. Kunming / 昆明: Yunnan Minzu Publishing House / 云南民族出版社.

HE Shaoyin 和少英, DAO Jie 刀洁, (2007). *The Catchers in their Homeland on the Boundary Line — the Society and Culture of Jinping Dai Minzu* / 守望国境线上的家园——金平傣族的社会文化. Kunming / 昆明: Yunnan University Publishing House / 云南大学出版社.

LI Mengdie 李梦蝶, (2020). *A Comparative Study on the Phonetics of Thai and Jinping Dai Languages* / 泰语和金平傣语的语音比较研究. Kunming / 昆明: Yunnan Minzu University / 云南民族大学, Master Thesis.

LUO Wenbin 罗文斌 (chief editor), (2017).  / 金平傣文教程. Honghe / 红河: Cultural Center of Jinping Miao, Yao and Dai Autonomous County / 金平苗族瑶族傣族自治县文化馆.

Pittayawat PITTAYAPORN, (2009). *The Phonology of Proto-Tai*. New York: Cornell University, PhD Dissertation.

Planning Office for the Collation and Publication of Minzu Ancient Literature in Yunnan Province 云南省少数民族古籍整理出版规划办公室, (2020). *Collection of Yunnan Minzu Rare Ancient Literature* / 云南少数民族古籍珍本集成. Kunming / 昆明: Yunnan People's Publishing House / 云南人民出版社, Vol. 86.

<http://www.kaom.net/min_z_kamtai.php>

<<https://zhongguoyuyan.cn/province/62/侗台语族>>

Appendix A Glyphs comparison

A.1 Glyphs in ancient literatures

Below is a list of the glyphs used in Tai Xaa script, where “—” means unattested. Glyphs in the second to fourth columns are from [Dao et al. \(2019\)](#) [Dao], [Luo \(2017\)](#) [Luo] and [COHAP \(1989\)](#) [COHAP]. And glyphs in the remaining five columns are from ancient literatures in [POCP \(2020\)](#), they are:

- **DD:** *Dōngdēng Gùshì* 东登故事, “Story of Dongdeng (person name)”;
- **HL:** *Hūnlúo Yǔ Nánwō* 婚罗与喃窝, “Hunluo (person name) and Nanwo (person name)”;
- **ZL:** *Zhúlóu Shíwù* 竹楼实务, “Practice of Building Bamboo Houses”;
- **AXL:** *Àixiànglún* 艾相伦, “Aixianglun (person name)”;
- **Others:** *Cǎiyào Pǔjì* 采药谱记, “Records of Herb Collection” and *Qínggē* 情歌, “Love Song”.

Glyph	Dao	Luo	COHAP	DD	HL	ZL	AXL	Others
9	9	9	9					
9	9	9	9					—
n	n	n	n					
n	n	n	n					
16	16	16	16					
7	7	7	7					
916	916	916	916					
6	6	6	6					
5	5	5	5					
m	m	m	m					
m	m	m	m					
m	m	m	m					
96	96	96	96					

				—	—	—		—
			—	—	—	—		—
			—					
			—					
			—			—		—
			—	—	—	—		—
	—	—	—	—	—	—		—
			—	—	—	—		—
			—	—	—	—		—
			—					—
			—					—
			—					
	—	—	—		—	—		—
			—	—	—	—		—
			—	—	—	—		—

A.2 Glyphs in modern literatures

Below is a list of the glyphs used in Tai Xaa script. Glyphs in the second to sixth columns are from:

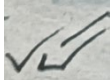
- **DD:** *Dōngdēng* 冬灯, “Dongdeng (person name)”;
- **CS:** Capitaine SILVESTRE, (1918), “*Les Thai Blancs de Phong-Tho*”, in: *Bulletin de l’École Française d’Extrême-Orient*, Vol. 18, No. 4: pp. 1–56.
- **GM:** Georges MINOT, (1940), “*Dictionnaire Tay Blanc Français*”, in: *Bulletin de l’École française d’Extrême-Orient*, Vol. 40, No. 1: pp. 1–237.

- **MF:** Michel FERLUS, (2006), “*Les écritures thai du Vietnam*”, in: *Cahiers de Linguistique – Asie Orientale*, Vol. 35, No. 2: pp. 209–239.
- **HDL:** HE Bingkun 何炳坤, DAO Jiafan 刀家凡, LUO Jun 罗军, (1996), *Jinping Dai Textbook (for Literacy)* / 金平傣文课本 (扫盲用). Honghe / 红河: Education Management Committee of Mengla Town, Jinping Miao, Yao and Dai Autonomous County / 金平苗族瑶族傣族自治县勐拉乡教育管理委员会.

Glyph	DD	CS	GM	MF	HDL
๑					
๒					
๓					
๔					
๕					
๖					
๗					
๘					
๙					
๐					
๑๐					
๑๑					
๑๒					
๑๓					
๑๔					
๑๕					
๑๖					
๑๗					
๑๘					
๑๙					
๒๐					

91					
1					
2					
3					
4					
5					
6					
7					
8					
9					
0					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					
43					
44					
45					
46					
47					
48					
49					

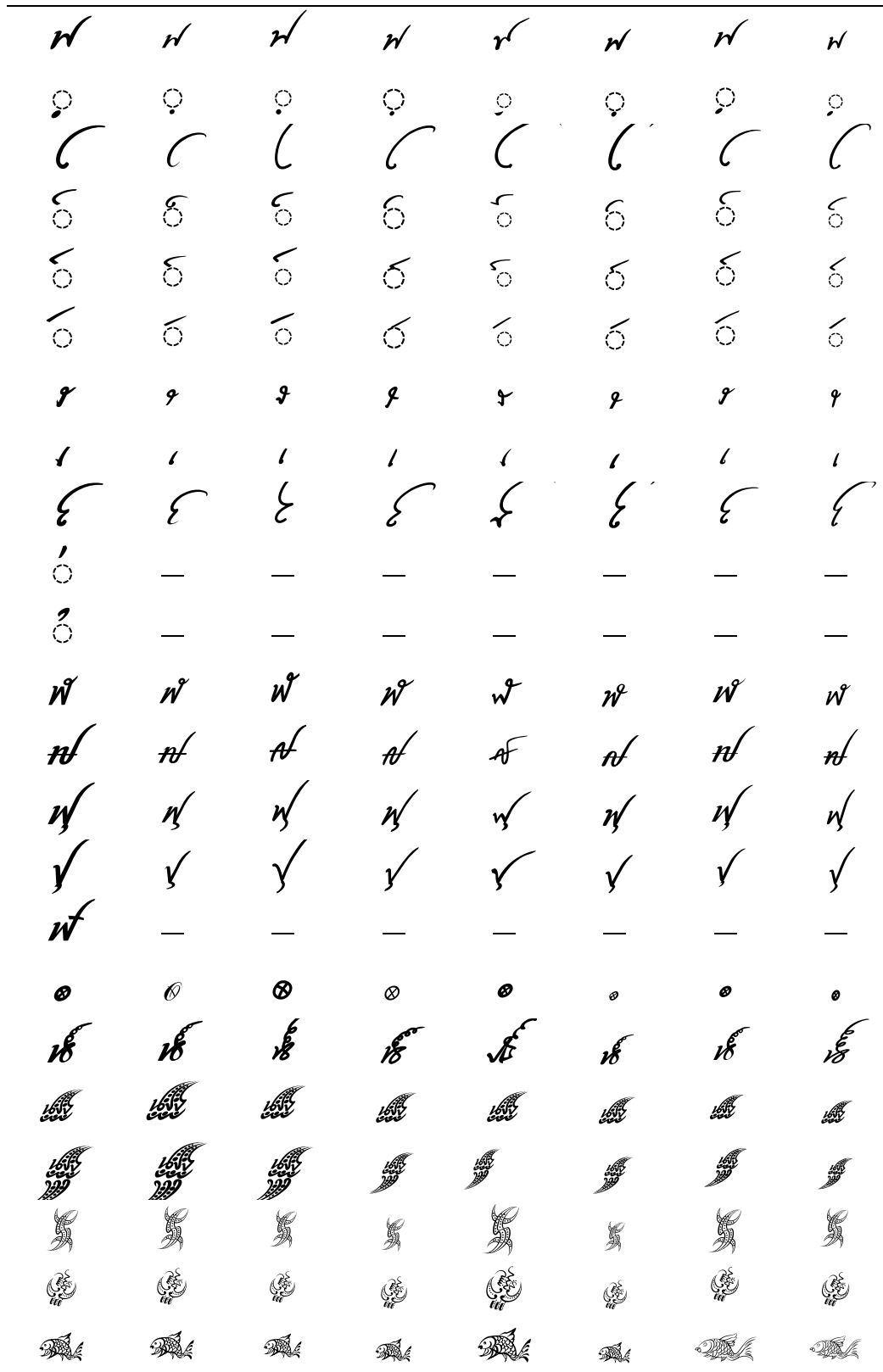
4					
913					
113					
,					
°					
17					
16					
✓					
○					
○					
○					
○					
○					
○					
11					
°					
W					
°					
○					
○					
○					
○					
○					
8					
1					
3					
○					

◌̇	—	—	—	—	—
ŋ		—	ŋ	—	
n		—	st	—	
ŋ		—	—	—	
ŋ		—	—	—	
ŋ		—	—	—	—
◌̇		—	—	—	
ŋ		—	ŋ	—	—
ŋ		—	ŋ	—	—
ŋ		—	—	—	—
ŋ		—	—	—	—
ŋ		—	—	—	—
ŋ		—	—	—	—

A.3 Glyphs in computerize fonts

Below is a list of the glyphs in several computerize fonts used in Tai Xaaü script. The names of fonts are: Xiaorui Regular (小睿体, made in 2018), Huangtao (黄韬体, made in 2019, HT), Huanghe (黄禾体, made in 2022, HH), Xiaosong (小宋体, made in 2022, XS), Huataiwen (华泰温体, made in 2022, HTW), Yinzong (尹宗体, made in 2023, YZ), Xiaorui Light (小睿体细体, made in 2024), Bobo (波波体, made in 2024, BB).

Glyph Xiaorui	HT	HH	XS	HTW	YZ	Xiaorui Light	BB
ŋ	ŋ	ŋ	ŋ	ŋ	ŋ	ŋ	ŋ
ŋ	ŋ	ŋ	ŋ	ŋ	ŋ	ŋ	ŋ
n	n	n	n	n	n	n	n
n	n	n	n	n	n	n	n



	xx0	xx1	xx2	xx3	xx4	xx5
0	 xx00	 xx10	 xx20	 xx30	 xx40	
1	 xx01	 xx11	 xx21	 xx31	 xx41	
2	 xx02	 xx12	 xx22	 xx32	 xx42	
3	 xx03	 xx13	 xx23	 xx33	 xx43	
4	 xx04	 xx14	 xx24	 xx34	 xx44	
5	 xx05	 xx15	 xx25	 xx35	 xx45	
6	 xx06	 xx16	 xx26	 xx36	 xx46	
7	 xx07	 xx17	 xx27	 xx37	 xx47	
8	 xx08	 xx18	 xx28	 xx38	 xx48	
9	 xx09	 xx19	 xx29	 xx39	 xx49	
A	 xx0A	 xx1A	 xx2A	 xx3A	 xx4A	
B	 xx0B	 xx1B	 xx2B	 xx3B	 xx4B	
C	 xx0C	 xx1C		 xx3C	 xx4C	
D	 xx0D	 xx1D		 xx3D	 xx4D	
E	 xx0E	 xx1E		 xx3E	 xx4E	
F	 xx0F	 xx1F		 xx3F	 xx4F	

Consonants

xx00	ᵛ	TAI XAAU LETTER HIGH QO → AAAE ᵛ tai viet letter low o
xx01	ᵛ	TAI XAAU LETTER LOW QO → AAAF ᵛ tai viet letter high o
xx02	ᵛ	TAI XAAU LETTER HIGH KO → AA80 ᵛ tai viet letter low ko
xx03	ᵛ	TAI XAAU LETTER LOW KO
xx04	ᵛ	TAI XAAU LETTER HIGH XO
xx05	ᵛ	TAI XAAU LETTER LOW XO
xx06	ᵛ	TAI XAAU LETTER HIGH HO NGO
xx07	ᵛ	TAI XAAU LETTER LOW NGO → AA89 ᵛ tai viet letter high ngo
xx08	ᵛ	TAI XAAU LETTER HIGH TSO → AA8A ᵛ tai viet letter low co
xx09	ᵛ	TAI XAAU LETTER LOW TSO → AA8B ᵛ tai viet letter high co
xx0A	ᵛ	TAI XAAU LETTER HIGH SO
xx0B	ᵛ	TAI XAAU LETTER LOW SO
xx0C	ᵛ	TAI XAAU LETTER HIGH YO
xx0D	ᵛ	TAI XAAU LETTER LOW YO → AAA5 ᵛ tai viet letter high yo
xx0E	ᵛ	TAI XAAU LETTER HIGH TO → AA94 ᵛ tai viet letter low to
xx0F	ᵛ	TAI XAAU LETTER LOW TO → AA95 ᵛ tai viet letter high to
xx10	ᵛ	TAI XAAU LETTER HIGH THO → AA96 ᵛ tai viet letter low tho
xx11	ᵛ	TAI XAAU LETTER LOW THO
xx12	ᵛ	TAI XAAU LETTER HIGH HO NO → AA98 ᵛ tai viet letter low no
xx13	ᵛ	TAI XAAU LETTER LOW NO → AA99 ᵛ tai viet letter high no
xx14	ᵛ	TAI XAAU LETTER HIGH PO
xx15	ᵛ	TAI XAAU LETTER LOW PO
xx16	ᵛ	TAI XAAU LETTER HIGH PHO
xx17	ᵛ	TAI XAAU LETTER LOW PHO
xx18	ᵛ	TAI XAAU LETTER HIGH HO MO
xx19	ᵛ	TAI XAAU LETTER LOW MO
xx1A	ᵛ	TAI XAAU LETTER HIGH FO
xx1B	ᵛ	TAI XAAU LETTER LOW FO
xx1C	ᵛ	TAI XAAU LETTER HIGH BO → AA9A ᵛ tai viet letter low bo
xx1D	ᵛ	TAI XAAU LETTER LOW BO → AA9B ᵛ tai viet letter high bo
xx1E	ᵛ	TAI XAAU LETTER HIGH LO → AAA8 ᵛ tai viet letter low lo
xx1F	ᵛ	TAI XAAU LETTER LOW LO
xx20	ᵛ	TAI XAAU LETTER HIGH HO → AAAC ᵛ tai viet letter low ho
xx21	ᵛ	TAI XAAU LETTER LOW HO → AAAD ᵛ tai viet letter high ho
xx22	ᵛ	TAI XAAU LETTER HIGH DO
xx23	ᵛ	TAI XAAU LETTER LOW DO → AA93 ᵛ tai viet letter high do
xx24	ᵛ	TAI XAAU LETTER HIGH VO → AAAA ᵛ tai viet letter low vo
xx25	ᵛ	TAI XAAU LETTER LOW VO → AAAB ᵛ tai viet letter high vo
xx26	ᵛ	TAI XAAU LETTER HIGH KHO
xx27	ᵛ	TAI XAAU LETTER LOW KHO → AA83 ᵛ tai viet letter high kho
xx28	ᵛ	TAI XAAU LETTER HIGH TSHO → AA8C ᵛ tai viet letter low cho
xx29	ᵛ	TAI XAAU LETTER LOW TSHO → AA8D ᵛ tai viet letter high cho
xx2A	ᵛ	TAI XAAU LETTER HIGH HO NYO
xx2B	ᵛ	TAI XAAU LETTER LOW NYO

Vowels

xx30	ᵛ	TAI XAAU MAAI KA-1 → AAB1 1 tai viet vowel aa
xx31	ᵛ	TAI XAAU MAAI KA-2
xx32	ᵛ	TAI XAAU MAAI KA-3
xx33	ᵛ	TAI XAAU MAAI KA-4 → AAB0 ᵛ tai viet mai kang
xx34	ᵛ	TAI XAAU MAAI KA-5 → AABE ᵛ tai viet vowel am
xx35	ᵛ	TAI XAAU MAAI KI → AAB2 ᵛ tai viet vowel i
xx36	ᵛ	TAI XAAU MAAI KE → AAB8 ᵛ tai viet vowel ia
xx37	ᵛ	TAI XAAU MAAI KAE • precedes consonant in visual order → AAB5 ᵛ tai viet vowel e
xx38	ᵛ	TAI XAAU MAAI TSAM
xx39	ᵛ	TAI XAAU MAAI KOA → AABA ᵛ tai viet vowel ua
xx3A	ᵛ	TAI XAAU MAAI KU-1 → AAB4 ᵛ tai viet vowel u
xx3B	ᵛ	TAI XAAU MAAI KU-2 • precedes consonant in visual order → AAB6 ᵛ tai viet vowel o
xx3C	ᵛ	TAI XAAU MAAI KU-3
xx3D	ᵛ	TAI XAAU MAAI KUE → AAB3 ᵛ tai viet vowel ue
xx3E	ᵛ	TAI XAAU MAAI KEI → AAB7 ᵛ tai viet mai khít
xx3F	ᵛ	TAI XAAU MAAI KAUE • precedes consonant in visual order → AABB ᵛ tai viet vowel aue
xx40	ᵛ	TAI XAAU MAAI KAU • precedes and follows consonant in visual order → AAB9 ᵛ tai viet vowel uea
xx41	ᵛ	TAI XAAU MAAI KAI • precedes consonant in visual order → AABC ᵛ tai viet vowel ay

Tones

xx42	ᵛ	TAI XAAU TONE MAAI SAK → AABF ᵛ tai viet tone mai ek → 0301 ᵛ combining acute accent
xx43	ᵛ	TAI XAAU TONE MAAI HOI → AAC1 ᵛ tai viet tone mai tho → 0309 ᵛ combining hook above

Word ligature symbols

xx44	ᵛ	TAI XAAU SYMBOL NUENG = one → AADC ᵛ tai viet symbol nueng
xx45	ᵛ	TAI XAAU SYMBOL KUN = person → AADB ᵛ tai viet symbol kon
xx46	ᵛ	TAI XAAU SYMBOL MA = horse
xx47	ᵛ	TAI XAAU SYMBOL BEN = moon
xx48	ᵛ	TAI XAAU SYMBOL MAN = the third person singular

Punctuation

xx49	ᵛ	TAI XAAU SYMBOL MAAI SUT
xx4A	ᵛ	TAI XAAU SYMBOL MAAI SAM → AADF ᵛ tai viet symbol koi koi
xx4B	ᵛ	TAI XAAU SYMBOL MAAI SAM KOAK → AADE ᵛ tai viet symbol ho hoi
xx4C	ᵛ	TAI XAAU SYMBOL MAAI SAM MAET
xx4D	ᵛ	TAI XAAU SYMBOL LAN KAAAN-1
xx4E	ᵛ	TAI XAAU SYMBOL LAN KAAAN-2
xx4F	ᵛ	TAI XAAU SYMBOL LAN KAAAN-3

Figures

⊞⊞ 44 辅音字母表

⊞			⊞		
/ʔ/			/ʔ/		
⊞	⊞	⊞	⊞	⊞	⊞
/k/	/x/	/hŋ/	/ʔk/	/ʔx/	/ʔŋ/
⊞	⊞	⊞	⊞	⊞	⊞
/tɕ/	/s/	/hɿ/	/ʔtɕ/	/ʔs/	/ʔɿ/
⊞	⊞	⊞	⊞	⊞	⊞
/t/	/th/	/hn/	/ʔt/	/ʔth/	/ʔn/
⊞	⊞	⊞	⊞	⊞	⊞
/p/	/ph/	/hm/	/ʔp/	/ʔph/	/ʔm/
⊞	⊞	⊞	⊞	⊞	⊞
/f/	/hv/	/hl/	/ʔf/	/ʔv/	/ʔl/
⊞	⊞	⊞	⊞	⊞	⊞
/h/	/rr/	/vb/	/ʔh/	/ʔrr/	/ʔvb/
⊞	⊞	⊞	⊞	⊞	⊞
/kh/	/ʔ/	/hŋ/	/ʔkh/	/ʔ/	/ʔŋ/

⊞⊞⊞ 18 元音字母表

前标元音: (4个)

上标元音: (7个)

后标元音: (5个)

下标元音: (2个)

前标元音: (1个)

说明: 在金平傣文元音字母中前标元音 "⊞" [ong], "⊞" [on], "⊞" [an] 不常使用, 在使用中必须与其它字母搭配使用。



第二课 复合元音 ⊞⊞⊞ 2 ⊞⊞⊞

一、韵尾辅音字母 (8个)

⊞	⊞	⊞	⊞	⊞	⊞	⊞	⊞
'yo	'vo	'go	'no	'mo	ko	rr	po
y	v	g	n	m	k	rr	p

在学习复合元音之前请大家熟记以上八个辅音字母。八个作为韵尾使用的辅音字母中前四个为高音组后、四个为低音组。

二、单元音 (9个)

⊞	⊞	⊞	⊞	⊞	⊞	⊞	⊞	⊞
/a/	/i/	/e/	/ɛ/	/u/	/ɿ/	/o/	/ol/	/u/

三、拼读规则

单元音

mai ka

a

→

韵尾辅音

me yo

'kay 'yo

→

mai ay

Figure 1 Tai Xaaui textbook, from top to bottom:
(a) Consonant letters and vowel letters; (b) a list of finals
Source: Luo (2017), pages 1–2, 7–8

表 8 规范声母

高音组			低音组		
p			p		
n	h	h	n	h	h
h	h	h	h	h	h
m	h	h	m	h	h
h	h	h	h	h	h
h	h	h	h	h	h
h	h	h	h	h	h
h	h	h	h	h	h

4. 金平傣文规范韵母

以 83 个韵母作为金平傣文的规范韵母。包含 9 个单元音、14 个双元音、60 个带辅音韵尾的韵母（简称“辅音尾韵母”）。列表 9。

表 9 规范韵母

单元音		o	o	o	u	u	u	u	u	u
双元音	uo									
	uo	uo				uo	uo	uo	uo	uo
	uo	uo	uo	uo	uo					uo
辅音尾韵母	uo	uo	uo	uo	uo	uo	uo	uo	uo	uo
	uo	uo	uo	uo	uo	uo	uo	uo	uo	uo
	uo	uo	uo	uo	uo	uo	uo	uo	uo	uo
	uo	uo	uo	uo	uo	uo	uo	uo	uo	uo
	uo	uo	uo	uo	uo	uo	uo	uo	uo	uo
	uo	uo	uo	uo	uo	uo	uo	uo	uo	uo

5. 金平傣文规范声调

以六个调位作为金平傣文的 6 个声调，3 个声调符号作为规范调符。列表 10。

表 10 规范声调

调位	一/33/	二/55/	三/25/	四/43/	五/52/	六/31/
调值	33	55	25	43	52	31
调符						
例词	h	h	h	h	h	h

Figure 3 Tai Xaau orthography:
 Initials [table 8], finals [table 9] and tones [table 10]
 Source: [Dao et al. \(2019\)](#), page 67

/55/、/25/、/43/、/52/、/31/。其中调位/55/包含了舒声调第2调的调值[55]和入声调第7调的调值[55]；调位/43/包含了舒声调第4调的调值[43]和入声调第8调的调值[43]。

四、特殊符号

金平傣文有一些符号较为特殊，既不是声母，也不是韵母，而是一种固定的特殊符号，专门用来书写某个词或表示句末停顿的符号。这样的符号不多，可分两类：一种是固定字符；另一种是句末符号。

(一) 固定字符

固定字符是指用一个特定的符号来表示某一个词的固定书写形式，构成独立使用的独体字。这种固定字符不多，传统文字中只出现在“一、人、月亮、马”4个词中。即

𐄣 [nɯŋ⁵²] 一 𐄤 [kɯŋ⁴³] 人
𐄥 [bən³⁸] 月亮 𐄦 [ma³¹] 马

计算机字体根据需要增加了2个类似的独体字。即

𐄧 [lu⁴³] 孩子 𐄨 [la²⁵] 迟

上述的独体字也可以用声韵调拼合的方式书写，如：𐄣 (一)、𐄤 (人)、𐄥 (月亮)、𐄦 (马)、𐄧 (孩子)、𐄨 (迟)。使用时可根据需要选择使用独特字或声韵调拼合的字。

(二) 句末符号

金平傣文一般不使用标点符号，每一句话结束时就在后面标注一个句末符号。但并没有固定用哪一个字符，大多因人而异。有的字符与其说是字，不如说是图画，有的字形类似象形文字。句末符号有如下6个。即



其中，前5个均统称为[ma: i³³ sam³¹]，有“结束符号”之意，通常用在长篇的故事或诗歌里。最后1个称为[ma: i³³ mɛt⁵⁵]，有“完成符号”之意，用在句子之后，表示一句话的完成。

五、结语

综上所述，金平傣文是一种源于印度梵文字母系统的拼音文字，也是跨境自称“傣端”或“傣皓”的傣族白傣支系使用的民族文字。该文字由于创制的年代久远，与现行口语在声母、韵母、声调等方面都出现了一定的差异。加之长期以来均通过手抄方式流传于民间，字体形式因人而异，既没有规范的字体，也没有计算机字体，这对跨境傣族语言文化的传承和发展带来了不利的影响。文字是一个民族进步的标志性特征之一，它对推动民族文化的发展起到积极的作用。尤其是当今数字化、信息化的时代，文字所发挥的作用更加具有不可替代性。为了能更好保护与传承优秀的民族文化，本文通过对金平傣文的实地调查，并根据现行口语的语音特征，对其语音的规范化问题进行了探讨，并提出金平傣文语音规范化的基本原则和标准。具体归纳总结表8。

1. 金平傣文规范字体

以金平县融媒体中心罗文斌同志所设计的金平傣文计算机字体小睿体作为金平傣文数字化、信息化的规范字体。该字体包含了在传统文字符号基础上所设计的字体和为完整记录现行语言而新增的文字符号字体。

2. 金平傣文标准音

以金平县勐拉镇旧勐村的白傣语语音作为金平傣文的标准音。该村落的历史文化、地理位置、语音特点，符合标准选择的一般原则。

3. 金平傣文规范声母

以22个辅音音位44个傣文字母作为金平傣文的规范声母。列表8。

Figure 4 Tai Xaa orthography:
Symbols [section 4.1], punctuation [section 4.2]
Source: Dao et al. (2019), page 68

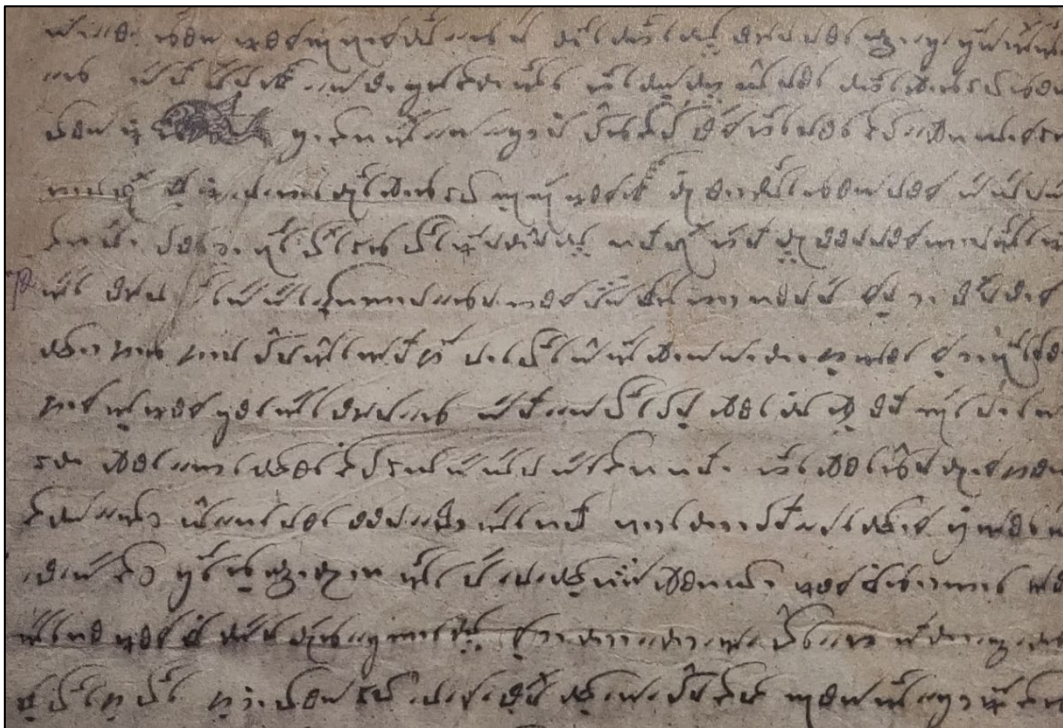
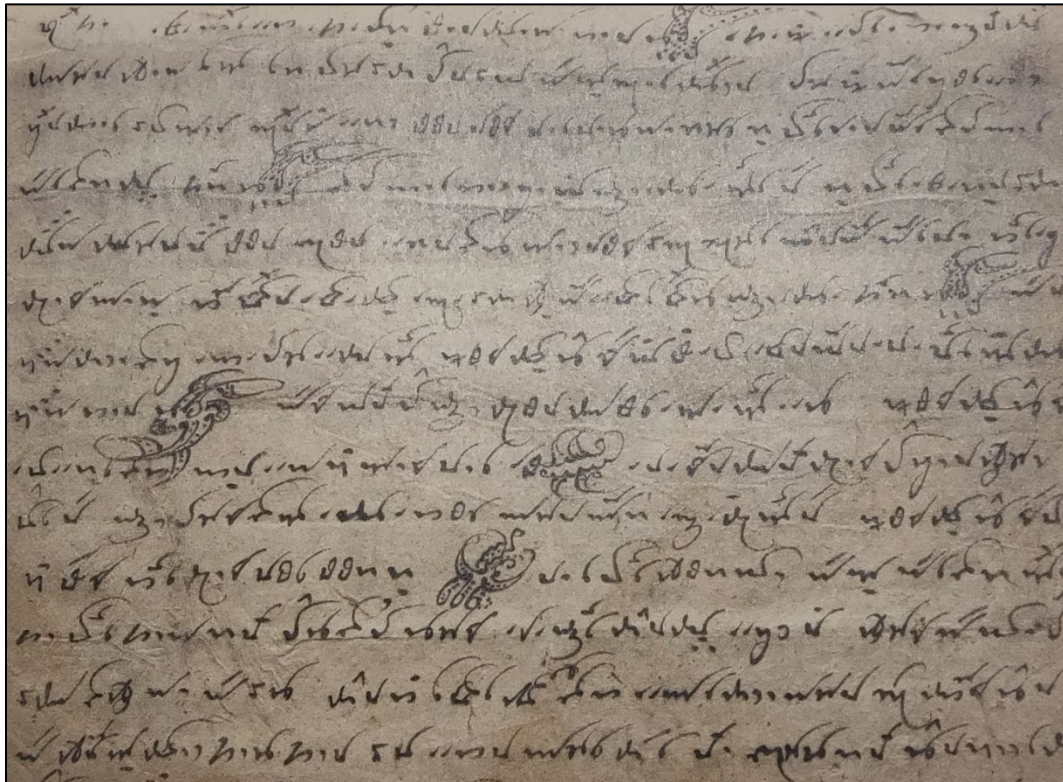


Figure 5 Tai Xaau literature, from top to bottom:
 (a) Punctuation 𐄢, 𐄣 and 𐄤; (b) Punctuation 𐄥
 Source: “Aixianglun”, in: [POCP \(2020\)](#), pages 329–330

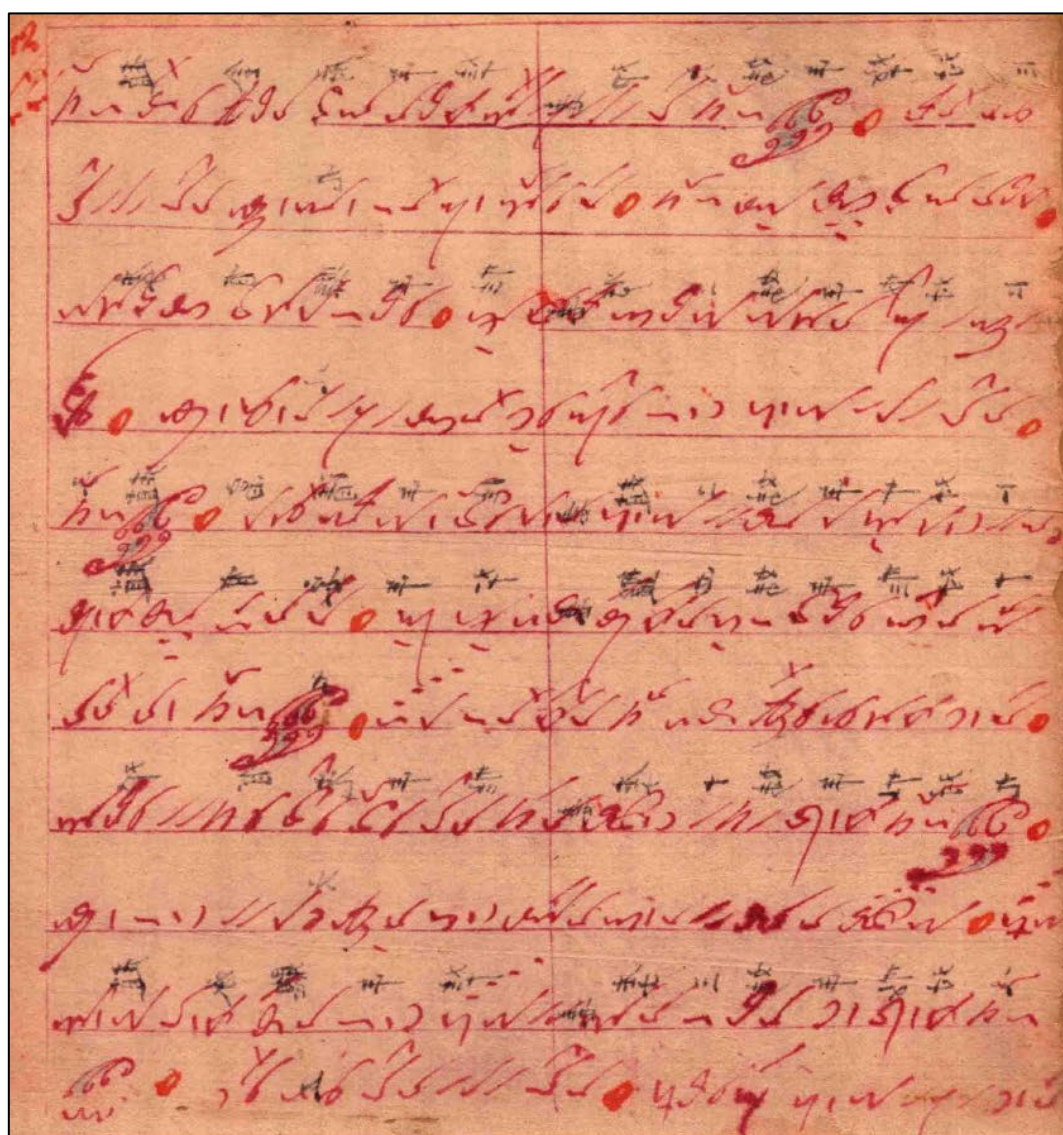


Figure 6 Tai Xau literature: Punctuation 𑜋 and 𑜌
 Source: “Hunluo and Nanwo”, in: [POCP \(2020\)](#), page 61



Figure 7 Teaching Tai Xau script, from top to bottom:
 (a) Teaching the Tai Xau script, note that the blackboard shows six consonant letters with “ho” in their character names; (b) Students copy the letters from the textbook
 Source: Provided by LUO Wenbin



Figure 8 Tai Xaa script on slogans, from top to bottom:
 (a) Slogan in the school; (b) Slogan in the garden; (c) Slogan on the wall
 Source: Provided by LUO Wenbin



Figure 9 On September 10, 2024, the Ethnic Affairs Commission (EAC) of Honghe Prefecture and the EAC of Yunnan Province held a seminar meeting for the purpose of developing and submitting the proposal for the Tai Xaa character set. Approved by the EAC of Yunnan Province, the minutes of the meeting have been made public
Source: Yunnan EAC

Summary Form

ISO/IEC JTC 1/SC 2/WG 2 PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646 Please fill all the sections A, B and C below. Please read Principles and Procedures Document (P&P) from unicode.org/wg2/principles.html for guidelines and details before filling this form. Please ensure you are using the latest Form from unicode.org/wg2/summaryform.html. See also unicode.org/roadmaps for latest Roadmaps.

A. Administrative

1. Title:	<i>Proposal to encode the Tai Xaa script in UCS</i>
2. Requester's name:	<i>China National Body</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Member body contribution</i>
4. Submission date:	<i>June 16, 2025</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<i>yes</i>
(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):	<i>yes</i>
Proposed name of script:	<i>Tai Xaa</i>
b. The proposal is for addition of character(s) to an existing block:	<i>no</i>
Name of the existing block:	
2. Number of characters in proposal:	<i>76</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary ☒	B.1-Specialized (small collection)
C-Major extinct	B.2-Specialized (large collection)
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?	<i>yes</i>
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<i>yes</i>
b. Are the character shapes attached in a legible form suitable for review?	<i>yes</i>
5. Fonts related:	
a. Who will provide the appropriate font to the Project Editor of 10646 for publishing the standard?	<i>Chen Zhuang</i>
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Chen Zhuang, chenzh-zhuang@163.com</i>
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<i>yes</i>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<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 behavior information such as line breaks, widths etc., Combining behavior, Spacing behavior, Directional behavior, Default Collation behavior, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at www.unicode.org for such information on other scripts. Also see Unicode Character Database (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.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	<i>no</i>
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.)?	<i>yes</i>
If YES, with whom? <i>Dai people lived in Yunnan Province, PRC</i>	
If YES, available relevant documents: <i>this document</i>	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	<i>yes</i>
Reference: <i>this document</i>	
4. The context of use for the proposed characters (type of use; common or rare)	<i>common</i>
Reference: _____	
5. Are the proposed characters in current use by the user community?	<i>yes</i>
If YES, where? Reference: <i>Jinping Miao, Yao, and Dai Autonomous County, Honghe Hani and Yi Autonomous Prefecture, Yunnan Province, PRC.</i> <i>(中华人民共和国云南省红河哈尼族彝族自治州金平苗族瑶族傣族自治县)</i>	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	<i>no</i>
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)?	<i>yes</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	<i>no</i>
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?	<i>no</i>
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?	<i>yes</i>
If YES, is a rationale for its inclusion provided? <i>yes</i>	
If YES, reference: <i>this document, section 2.1 (relations with Tai Viet)</i>	
11. Does the proposal include use of combining characters and/or use of composite sequences?	<i>yes</i>
If YES, is a rationale for such use provided? <i>yes</i>	
If YES, reference: <i>combining characters are an inherent part of the script</i>	
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? <i>no</i>	
If YES, reference: _____	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	<i>no</i>
If YES, describe in detail (include attachment if necessary): _____	

13. Does the proposal contain any Ideographic compatibility characters?	<i>no</i>
If YES, are the equivalent corresponding unified ideographic characters identified? _____	
If YES, reference: _____	