Tai scripts in Viet Nam

Introduction

There are various Brahmic writing systems used for the *Tai Dam*, *Tai Don*, *Tai Daeng* and *Tai Yo* languages in Viet Nam. Users of these scripts are also found in Laos, Thailand and China. Although these languages are tonal, none of these writing systems traditionally have any tone indicators. Recently, the modernized scripts have borrowed the tone mark system from the *Lao* and *Thai* alphabets.

A unified and standardized alphabet mainly based on the *Tai Dam* script was approved for encoding the Unicode Standard as *Tai Viet*. Currently in 2021, the Tai Viet Unicode block does not include glyph variants which are found in *Tai Don* and *Tai Daeng* and should be disunified. Before being officially allotted to new code points in Unicode, these variants can be represented as their *Tai Viet* code points typed along with the special character ZWJ (zero width joiner) as long as these features are supported by the font.

For example: <uAAAB> = ∂ (*Tai Dam*, *Tai Don*)   vs   <uAAAB> + <ZWJ> = ɣ (*Tai Daeng*)

Letters	*Tai Dam*	*Tai Don* (Mường So)	*Tai Daeng* (Thanh Hoá)
High KO <uAA81>	ຄ	ᴎ	ຖ <uAA81>+<ZWJ>
Low KHHO <uAA84>	ຉ	ଛ	ଃ <uAA84>+<ZWJ>
Low NGO <uAA88>	ຕ	௦	ຬ <uAA88>+<ZWJ>
High NGO <uAA89>	໐	໕	໑ <uAA89>+<ZWJ>
Low SO <uAA8E>	✕	₩ <uAA8E>+<ZWJ>	ⱨ <uAA8E>+<ZWJ>
High SO <uAA8F>	ຝ	₩ <uAA8F>+<ZWJ>	ຑ
High THO <uAA97>	ຢ	ຖ <uAA97>+<ZWJ>	ຘ <uAA87>+<ZWJ>
High LO <uAAA9>	ჳ	ຖ <uAAA9>+<ZWJ>	ຉ
Low YO <uAAA4>	ຟ	ຜ <uAAA4>+<ZWJ>	ຜ <uAAA4>+<ZWJ>
Low VO <uAAAA>	ຜ	ຜ	ຜ <uAAAA>+<ZWJ>
High VO <uAAAB>	ວ	ວ	ɣ <uAAAB>+<ZWJ>

Table 1. Tai Viet Unicode letters and their regional variants in pink cells

The *Tai Viet* Unicode block is not suitable for the *Lai Tay* and *Lai Pao* scripts which are traditionally used for the *Tai Yo* language in different regions. In contrast to other Tai scripts, the *Yo Lai Tay* script is written vertically from top to bottom and ordered from right to left like the traditional Chinese writing. Both of these Brahmic alphabets should be proposed for new Unicode blocks.
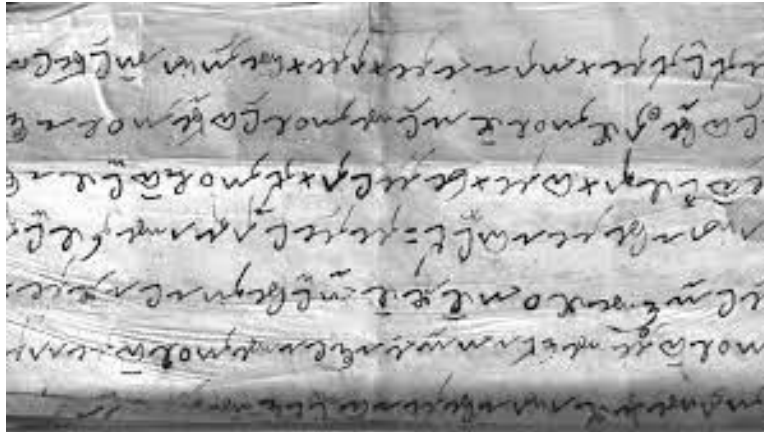

Figure 1. A *Tai Yo* manuscript using the traditional *Lai Pao* script

Figure 2. A *Tai Yo* manuscript using the traditional *Lai Tay* script

Initial consonants

In Table 2, modernized letters are written in red beside their traditional counterparts in blue (if available). Letters in pink cells require ZWJ to render. Letter in yellow cells use different code points.

Letters	*T. Dam*	*T. Don*	*T. Daeng*	*T. Yo Lai Tay*	*T. Yo Lai Pao*
Low KO	ꫀ	ꫀ	ꫀ	ꫝ	ꫝ
High KO	ꪄ	ꪁ	ꪁ <uAA81>+<ZWJ>	ꪂ	ꪁ
Low KHO		ꪃ		ꪄ	ꪅ
High KHO		ꪆ		ꪄ < ꪅ	ꪆ < ꪇ
Low KHHO	ꪈ	ꪉ	ꪄ <uAA84>+<ZWJ>		
High KHHO	ꪊ	ꪇ <uAA87>	ꪋ		
Low GO	ꪌ			ꪍ	ꪎ
High GO	ꪏ			ꪐ	ꪑ
Low NGO	ꪒ	ꪓ	ꪈ <uAA88>+<ZWJ>	ꪔ < ꪕ	ꪖ
High NGO	ꪗ	ꪘ	ꪉ <uAA89>+<ZWJ>	ꪙ	ꪚ
Low CO	ꪛ	ꪜ	ꪝ	ꪞ	ꪟ
High CO	ꪠ	ꪡ	ꪢ	ꪣ < ꪤ	ꪥ < ꪦ
Low CHO		ꪧ		ꪨ	ꪩ
High CHO		ꪪ		ꪫ	ꪬ
Low SO	ꪭ	ꪮ <uAA8E>+<ZWJ>	ꪯ <uAA8E>+<ZWJ>	ꪰ	ꪱ
High SO	ꪲ	ꪳ <uAA8F>+<ZWJ>	ꪴ	ꪵ	ꪶ
Low NYO	ꪷ	ꪸ	ꪹ	ꪺ	ꪻ
High NYO	ꪼ	ꪽ	ꪾ	꪿	ꫀ
Low DO	꫁	ꫂ	꫃	꫄	꫅
High DO	꫆	꫇	꫈	꫉ < ꫊	꫋ < ꫌
Low TO	꫍	꫎	꫏	꫐	꫑

High TO	ꨓ	ꨓ	ꨓ		
Low THO	ꨔ	ꨔ	ꨔ		
High THO	ꨕ	ꨕ <uAA97>+<ZWJ>	ꨕ <uAA97>+<ZWJ>	<	<
Low TRO					
High TRO					
Low NO	ꨮ	ꨮ	ꨮ	<	
High NO	ꨮ	ꨮ	ꨮ		
Low BO	ꨝ	ꨝ	ꨝ		
High BO	ꨝ	ꨝ	ꨝ	<	<
Low PO	ꨚ	ꨚ	ꨚ		
High PO	ꨝ	ꨝ <uAA9D>+<ZWJ>	ꨝ		
Low PHO		ꨟ			
High PHO		ꨟ	ꨟ <uAA9F>+<ZWJ>	<	<
Low FO	ꨠ	ꨠ	ꨠ <uAA8C>		
High FO	ꨡ	ꨡ	ꨡ		
Low MO	ꨫ	ꨫ	ꨫ	<	
High MO	ꨩ	ꨩ	ꨩ		
Low YO	ꨬ	ꨬ <uAAA4>+<ZWJ>	ꨬ <uAAA4>+<ZWJ>		
High YO	ꨬ	ꨬ	ꨬ	<	
Low RO	ꩁ				
High RO	ꩁ				
Low LO	ꩈ	ꩈ	ꩈ	<	
High LO	ꩉ	ꩉ <uAAA9>+<ZWJ>	ꩉ		
Low VO	ꩂ	ꩂ	ꩂ ⁰ <uAAAA>+<ZWJ>		

Letters	T. Dam	T. Don	T. Daeng	T. Yo Lai Tay	T. Yo Lai Pao
High VO	ꪉ	ꪀ	ꪹꪮ [0] <uAAAB>+<ZWJ>	ꪹꪮ < ꪮ	ꪮ
Low HO	ꪬ	ꪬ	ꪬ	ꪬ	ꪬ
High HO	ꪖ	ꪖ	ꪖ	ꪖ	ꪖ
Low O	ꪱ	ꪱ	ꪱ	ꪱ	ꪱ
High O	ꪲ	ꪲ	ꪲ	ꪲ < ꪲ	ꪲ < ꪲ
Low KVO			ꪶ [1]	ꪶ	
KN Ligature			ꪻ [2]		

Table 2. Initial consonants

[0]: ꪉ and ꪮ of *Tai Daeng* in Mường Mùn are respectively ꪮ and ꪮ like in *Tai Dam* and *Tai Don*.

[1]: ꪶ = <uAA80> + <ZWJ> + <uAAAB> + <ZWJ>

[2]: ꪻ = <uAA80> + <ZWJ> + <uAA99>

Rimes

In Table 3, ◌ stands for the initial consonant.

Letters	T. Dam	T. Don	T. Daeng	T. Yo Lai Tay	T. Yo Lai Pao
Yo Lai Tay AY				◌ꪸ [3]	
Low VO				◌ꪫ [3]	
High VO	◌ꪉ	◌ꪫ	◌ꪹꪮ [0] <uAAAB>+<ZWJ>		◌ꪮ

Table 3. Medial consonant -w-

[3]: In *Yo Lai Tay*, ◌ꪫ and ◌ꪸ are both used for the medial -w- after all consonants except Low KO, which has its own ligature Low KVO ꪶ.

In Table 4, ꪀ is used in place of the initial consonant, ◌ stands for the optional final consonant. Rimes without ◌ are complete rimes carrying an inherent final consonant. A tone mark can be added right after (or on top of) the blue part and before the black part of the rime.

Letters	T. Dam	T. Don	T. Daeng	T. Yo Lai Tay	T. Yo Lai Pao
ANG	ꪀꪰ	ꪀꪲ	ꪀꪉ	ꪚ	ꪖꪉ

AN	ꪀꪮ	ꪀꪮ	ꪀꪮ	ꪶꪀꪮ	ꪀꪮꪮ
Sign AN					ꪰ
AM	ꪀ	ꪀ	ꪀ	ꪶꪀ	ꪀꫝ
AK	ꪀꪀ	ꪀꪀ	ꪀꪀ	ꪶꪀꪀ	ꪀꪀ
AT	ꪀꪮ	ꪀꪮ [4]	ꪀꪮ	ꪶꪀꪮ	ꪀꪮ
AP	ꪀꪜ	ꪀ√	ꪀꪜ	ꪶꪀꪜ	ꪀꪜ
AA	ꪀꪱꫛ	ꪀꪱꪮ	ꪀꪱꫛ	ꪶꪀꫛ	ꪀꪱꫛ
AA tone C1		ꪁꪮ [5] \<uAACF\>			
I	ꪀꫛ	ꪀꪮ	ꪀꫛ	ꪶꪀꫛ	ꪀꫛ
II			ꪀꫛ [6]		
UE	ꪀꫛ	ꪀꪮ	ꪀꫛ	ꪶꪀꫛ	ꪀꫛ
UEE			ꪀꫛ [6]		
U	ꪀꫛ	ꪀꪮ	ꪀꫛ	ꪶꪀꫛ	ꪀꫛ
UU			ꪀꫛ [6]		
IA	ꪀꫛ	ꪀꪮ	ꪀꫛ	ꪶꪀꫛ	ꪀꫛ
E	ꪶꪀꫛ		ꪶꪀꫛ	ꪶꪀꫛ	ꪀꫛ
EE			ꪀꫛ [6]		
UEA	ꪶꪀꫛ	ꪀꪮ \<uAAB7\>	ꪶꪀꫛ \<uAAB9\>+…+\<uAAB0\>	ꪶꪀꫛ	ꪀꫛ
EO	ꪶꪀꫛ				
UA	ꪀꪺꫛ	ꪀꪺꪮ	ꪀꪺꫛ	ꪶꪀꪺꫛ	ꪀꪺꫛ
OE	ꪶꪀꫛ		ꪶꪀꫛ	ꪶꪀꫛ	ꪀꫛ
OENG		ꪶꪀꪉ [7]			ꪀꪉ [7]
OEM		ꪶꪀꪣ [7]	ꪀꪣ [7]	ꪶꪀ [7]	ꪀꪣ [7]
AE	ꪶꪶꪀꫛ	ꪶꪶꪀꪮ	ꪶꪶꪀꫛ	ꪶꪀꫛ	ꪀꫝꫛ

AY	ꪵꪀ	ꪹꪀ	ꪶꪀ	ꪶ	ꪖꪽ
AUE	ꪹꪀ	ꪼꪀ	ꪹꪀꫛ	ꪹꪣ	ꪖꪮꫛ
AU	ꪹꪀꪱ	ꪹꪀꪱ	ꪹꪀꪱ	ꪹ	ꪖꪮ
Single O	ꪀꪲ		ꪀꪰ	ꪹ 8	ꪖ
Low O	ꪀꪮꫛ	ꪀꪮ	ꪀꪮꫛ	ꪹꪮꫛ	ꪖꪮꫛ

Table 4. Vowels

[4]: Vowel letter AT ꪀ꪿ in *Tai Don* has not been assigned in Unicode.

$$ꪀ꪿ = <uAA80> + <uAAB0> + <ZWJ> + <uAA92>$$

[5]: The unassigned ꪁ in *Tai Don* is a variant of vowel AA with the inherent *Proto-Tai* tone C1.

[6]: Before being officially supported by Unicode, these vowel signs are formed as follows:

$$ꪀ̃ = <uAA80> + <uAAB2> + <uAAB2>$$

$$ꪀ̈ = <uAA80> + <uAAB3> + <uAAB3>$$

$$ꪀ̬ = <uAA80> + <uAAB4> + <uAAB4>$$

$$ꪀ̓ = <uAA80> + <uAAB8> + <uAAB8>$$

[7]: Positional variants of vowel OU before final consonants -NG and/or -M

$$ꪀꪣ = <uAA80> + <uAAB6> + <ZWJ> + <uAAA3>$$

[8]: Standalone initial consonants along with an optional tone mark in *Yo Lai Tay* have the inherent vowel O with no final consonant.

Letters	*T. Dam*	*T. Don*	*T, Daeng*	*T. Yo Lai Tay*	*T. Yo Lai Pao*
Visarga			ꫛ 9		
Low KO as -K	ꪀ	ꪀ	ꪀ	ꪹ	ꪖ
Low DO as -T	ꪍ	ꪀ꪿	ꪝ	�	ꪣ
Low BO as -P	ꪴ	ꪼ	ꪴ	ꪒ	ꪞ
High NGO as -NG	�	ꪶ	ꪉ <uAA89>+<ZWJ>	ꪒ	ꪆ
High NO as -N	ꪌ	ꪌ	ꪌ	ꪒ	ꪜ

High MO as -M	ม	ฟ	ม	ᩉᩉ	ᨾ
Low NYO as -Y				ᩈ	
High YO as -Y	ย	ฝ	ฝ		ᨿ
Low VO as -W				ᩅ	
High VO as -W	ว	ว	ɤ <uAAAB>+<ZWJ>		ᩅ

Table 5. Final consonants

[9]: Visarga ะ in Tai Daeng represents ending glottal stop.

Tone marks

Proto-Tai has 4 phonemic tones: A, B, C and D. D tone in modern Tai languagese may have different phonemic indicators depending on the length of the rime (S for short and L for long) and the voicing of the initial consonant in *Proto-Tai*: DL1/DS1 for voiceless ones and DL2/DS2 for voiced ones.

Tone A of *Proto-Tai* is not indicated by any marks in all of these modernized scripts. Tone B and tone C of *Proto-Tai* are indicated respectively by tone mark/letter 1 (*Tai Viet Unicode MAI EK* or *MAI NUENG*) and tone mark/letter 2 (*Tai Viet Unicode MAI THO* or *MAI SONG*) as in the *Thai* alphabet. Tone ending letters, which now is old-fashioned, are written at the end of the word.

Letters	*T. Dam*	*T. Don*	*T. Daeng*	*T. Yo Lai Tay*	*T. Yo Lai Pao*
Tone A	◌	◌	◌	◌	◌
Tone B	ń	ń	ń	ᩈ	ᨶ
Tone B (old)	◌ℓ	◌ℓ	◌ℓ		
Tone C	ń̃	ń̃	ň	ᩈ	ᨶ
Tone C (old)	◌]	◌]	◌]		
Tone DL1	ŋ̣n	ŋ̣n	ŋ̣n	ᩈᩈ [10]	ᨶᨶ [9]
Tone DL2	ọn	ḿn	ŋ̣n	ᩈᩈ [10]	ᨿᨶ [9]
Tone DS1	ŋ̣́n [11]	ŋ̣́n [11]	ŋ̣n	ᩈᩈ	ᨶᨶ
Tone DS2	ọ̌n [11]	m̌n [11]	ŋ̣n	ᩈᩈ	ᨿᨶ

Table 6. Tone marks

10: Only rimes with final -k ꪙ in *Tai Don* and *Tai Dam* require tone mark for short rimes and no tone mark for long rimes since the final consonant becomes a glottal stop in long rimes.

11: All dead long rimes in the modernized *Yo Lai Tay* and *Yo Lai Pao* require tone mark 1 ꪒ while all dead short rimes require no tone mark.

Symbols

Letters	T. Dam	T. Don	T. Daeng	T. Yo Lai Tay	T. Yo Lai Pao
Person	ꪤ	ꪮ	ꪭ		
Number One	ꪤ	ꪮ	ꪭ		
Et cetera	ꪬ	ꪬ			
Ho Hoi	ꪮ	ꪮ	ꪮ		
Repetition	ꪱ	ꪱ	ꪱ	ꪹꪒ/ꪹ	ꪹꪒ
Shortened repetition			ꪫ <uAADF>+<ZWJ>		
Mueang				◎	

Table 7. Special symbols