Tai scripts in Viet Nam using Tai Viet Unicode

# 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* (including *Tai Khao* in Mường Mùn) 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> = ꪫ (*Tai Dam*, *Tai Don*) vs <uAAAB> + <ZWJ> = ꪫ‍ (*Tai Daeng*)

# Initial consonants

In Table 1, modernized letters are written in red. Letters in pink cells require ZWJ to render. Letter in yellow cells use different code points.

|  |  |  |  |
| --- | --- | --- | --- |
| ***Letters*** | ***Tai Dam*** | ***Tai Don*** | ***Tai Daeng*** |
| Low KO | ꪀ | ꪀ | ꪀ |
| High KO | ꪁ | ꪁ | ꪁ‍ <uAA81>+<ZWJ> |
| Low KHO |  | ꪂ |  |
| High KHO |  | ꪃ |  |
| Low KHHO | ꪄ | ꪄ | ꪄ‍ <uAA84>+<ZWJ> |
| High KHHO | ꪅ | ꪇ <uAA87> | ꪅ |
| Low GO | ꪆ |  |  |
| High GO | ꪇ |  |  |
| Low NGO | ꪈ | ꪈ | ꪈ‍ <uAA88>+<ZWJ> |
| High NGO | ꪉ | ꪉ | ꪉ‍ <uAA89>+<ZWJ> |
| Low CO | ꪊ | ꪊ | ꪊ |
| High CO | ꪋ | ꪋ | ꪋ |
| Low CHO |  | ꪌ |  |
| High CHO |  | ꪍ |  |
| Low SO | ꪎ | ꪎ‍ <uAA8E>+<ZWJ> | ꪎ‍ <uAA8E>+<ZWJ> |
| High SO | ꪏ | ꪏ‍ <uAA8F>+<ZWJ> | ꪏ |
| Low NYO | ꪐ | ꪐ | ꪐ |
| High NYO | ꪑ | ꪑ | ꪑ |
| Low DO | ꪒ | ꪒ | ꪒ |
| High DO | ꪓ | ꪓ | ꪓ |
| Low TO | ꪔ | ꪔ | ꪔ |
| High TO | ꪕ | ꪕ | ꪕ |
| Low THO | ꪖ | ꪖ | ꪖ |
| High THO | ꪗ | ꪗ‍ <uAA97>+<ZWJ> | ꪗ‍ <uAA97>+<ZWJ> |
| Low NO | ꪘ | ꪘ | ꪘ |
| High NO | ꪙ | ꪙ | ꪙ |
| Low BO | ꪚ | ꪚ | ꪚ |
| High BO | ꪛ | ꪛ | ꪛ |
| Low PO | ꪜ | ꪜ | ꪜ |
| High PO | ꪝ | ꪝ‍ <uAA9D>+<ZWJ> | ꪝ |
| Low PHO |  | ꪠ |  |
| High PHO |  | ꪟ | ꪟ‍ <uAA9F>+<ZWJ> |
| Low FO | ꪠ | ꪞ | ꪌ <uAA8C> |
| High FO | ꪡ | ꪡ | ꪡ |
| Low MO | ꪢ | ꪢ | ꪢ |
| High MO | ꪣ | ꪣ | ꪣ |
| Low YO | ꪤ | ꪤ‍ <uAAA4>+<ZWJ> | ꪤ‍ <uAAA4>+<ZWJ> |
| High YO | ꪥ | ꪥ | ꪥ |
| Low RO | ꪦ |  |  |
| High RO | ꪧ |  |  |
| Low LO | ꪨ | ꪨ | ꪨ |
| High LO | ꪩ | ꪩ‍ <uAAA9>+<ZWJ> | ꪩ |
| Low VO | ꪪ | ꪪ | ꪪ‍ 0 <uAAAA>+<ZWJ> |
| High VO | ꪫ | ꪫ | ꪫ‍ 0 <uAAAB>+<ZWJ> |
| Low HO | ꪬ | ꪬ | ꪬ |
| High HO | ꪭ | ꪭ | ꪭ |
| Low O | ꪮ | ꪮ | ꪮ |
| High O | ꪯ | ꪯ | ꪯ |
| Low KVO |  |  | ꪀ‍ꪫ‍ 1 |
| KN Ligature |  |  | ꪀ‍ꪙ 2 |

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

0: ꪫ‍ and ꪪ‍ of *Tai Daeng* in Mường Mùn (*Tai Khao*) are respectively ꪫ and ꪪ like in *Tai Dam* and *Tai Don*.

1: ꪀ‍ꪫ‍ = <uAA80> + <ZWJ> + <uAAAB> + <ZWJ>

2: ꪀ‍ꪙ = <uAA80> + <ZWJ> + <uAA99>

# Rimes

In Table 2, ◌ stands for the initial consonant.

|  |  |  |  |
| --- | --- | --- | --- |
| ***Letters*** | ***Tai Dam*** | ***Tai Don*** | ***Tai Daeng*** |
| High VO | ◌ꪫ | ◌ꪫ | ◌ꪫ‍ 0 <uAAAB>+<ZWJ> |

**Table 2. Medial consonant -w-**

In Table 3, 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*** | ***Tai Dam*** | ***Tai Don*** | ***Tai Daeng*** |
| ANG | ꪀꪰꪉ | ꪀꪰꪉ | ꪀꪰꪉ‍ |
| AN | ꪀꪽ | ꪀꪽ | ꪀꪰꪙ |
| Sign AN |  |  |  |
| AM | ꪀꪾ | ꪀꪾ | ꪀꪾ |
| AK | ꪀꪰꪀ | ꪀꪰꪀ | ꪀꪰꪀ |
| AT | ꪀꪰꪒ | ꪀꪰ‍ꪒ 3 | ꪀꪰꪒ |
| AP | ꪀꪾꪚ | ꪀꪾꪚ | ꪀꪰꪚ |
| AA | ꪀꪱ◌ | ꪀꪱ◌ | ꪀꪱ◌ |
| AA tone C1 |  | ꪀ꫏◌ 4 <uAACF> |  |
| I | ꪀꪲ◌ | ꪀꪲ◌ | ꪀꪲ◌ |
| II |  |  | ꪀꪲꪲ◌ 5 |
| UE | ꪀꪳ◌ | ꪀꪳ◌ | ꪀꪳ◌ |
| UEE |  |  | ꪀꪳꪳ◌ 5 |
| U | ꪀꪴ◌ | ꪀꪴ◌ | ꪀꪴ◌ |
| UU |  |  | ꪀꪴꪴ◌ 5 |
| IA | ꪀꪸ◌ | ꪀꪸ◌ | ꪀꪸ◌ |
| E | ꪹꪀꪸ◌ |  | ꪹꪀꪸꪸ◌ |
| EE |  |  | ꪀꪸꪸ◌ 5 |
| UEA | ꪹꪀ◌ | ꪀꪷ◌ <uAAB7> | ꪹꪀꪰ◌ <uAAB9>+...+<uAAB0> |
| EO | ꪹꪀꪷ◌ |  |  |
| UA | ꪀꪺ◌ | ꪀꪺ◌ | ꪀꪺ◌ |
| OE | ꪶꪀ◌ |  | ꪶꪀ◌ |
| OENG |  | ꪶꪀꪉ 6 |  |
| OEM |  | ꪶꪀꪣ 6 | ꪀꪶ‍ꪣ 6 |
| AE | ꪵꪀ◌ | ꪵꪀ◌ | ꪵꪀ◌ |
| AY | ꪼꪀ | ꪼꪀ | ꪼꪀ |
| AUE | ꪻꪀ | ꪻꪀ | ꪻꪀ◌ |
| AU | ꪹꪀꪱ | ꪹꪀꪱ | ꪹꪀꪱ |
| Single O | ꪀꪷ |  | ꪀꪷ |
| Low O | ꪀꪮ◌ | ꪀꪮ◌ | ꪀꪮ◌ |

**Table 3. Vowels**

3: Vowel letter AT ꪀꪰ‍ꪒ in *Tai Don* has not been assigned in Unicode.

ꪀꪰ‍ꪒ = <uAA80> + <uAAB0> + <ZWJ> + <uAA92>

4: The unassigned ꪀ꫏◌ in *Tai Don* is a variant of vowel AA with the inherent *Proto-Tai* tone C1.

5: Before being officially supported by Unicode, these vowel signs are formed as follows:

ꪀꪲꪲ = <uAA80> + <uAAB2> + <uAAB2>

ꪀꪳꪳ = <uAA80> + <uAAB3> + <uAAB3>

ꪀꪴꪴ = <uAA80> + <uAAB4> + <uAAB4>

ꪀꪸꪸ = <uAA80> + <uAAB8> + <uAAB8>

6: Positional variants of vowel OU before final consonants -NG and/or -M

ꪀꪶ‍ꪣ = <uAA80> + <uAAB6> + <ZWJ> + <uAAA3>

|  |  |  |  |
| --- | --- | --- | --- |
| ***Letters*** | ***Tai Dam*** | ***Tai Don*** | ***T, Daeng*** |
| Visarga |  |  | ꫊ 7 |
| Low KO as -K | ꪀ | ꪀ | ꪀ |
| Low DO as -T | ꪒ | ꪒ | ꪒ |
| Low BO as -P | ꪚ | ꪚ | ꪚ |
| High NGO as -NG | ꪉ | ꪉ | ꪉ‍ <uAA89>+<ZWJ> |
| High NO as -N | ꪙ | ꪙ | ꪙ |
| High MO as -M | ꪣ | ꪣ | ꪣ |
| Low NYO as -Y |  |  |  |
| High YO as -Y | ꪥ | ꪥ | ꪥ |
| Low VO as -W |  |  |  |
| High VO as -W | ꪫ | ꪫ | ꪫ‍ 0 <uAAAB>+<ZWJ> |

**Table 4. Final consonants**

7: Visarga ꫊ 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 is outdated, are written at the end of the word.

|  |  |  |  |
| --- | --- | --- | --- |
| ***Letters*** | ***Tai Dam*** | ***Tai Don*** | ***Tai Daeng*** |
| Tone A | ◌ | ◌ | ◌ |
| Tone B | ꪀ꪿ | ꪀ꪿ | ꪀ꪿ |
| Tone B (old) | ◌ꫀ | ◌ꫀ | ◌ꫀ |
| Tone C | ꪀ꫁ | ꪀ꫁ | ꪀ꫁ |
| Tone C (old) | ◌ꫂ | ◌ꫂ | ◌ꫂ |
| Tone DL1 | ꪀꪴꪀ | ꪀꪴꪀ | ꪀꪴꪴꪀ |
| Tone DL2 | ꪁꪴꪀ | ꪁꪴꪀ | ꪁ‍ꪴꪴꪀ |
| Tone DS1 | ꪀꪴ꪿ꪀ 8 | ꪀꪴ꪿ꪀ 8 | ꪀꪴꪀ |
| Tone DS2 | ꪁꪴ꫁ꪀ 8 | ꪁꪴ꫁ꪀ 8 | ꪁ‍ꪴꪀ |

**Table 5. Tone marks**

8: Only rimes with final -k ꪀ 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.

# Symbols

|  |  |  |  |
| --- | --- | --- | --- |
| ***Letters*** | ***Tai Dam*** | ***Tai Don*** | ***Tai Daeng*** |
| Person | ꫛ | ꫛ | ꫛ |
| Number One | ꫜ | ꫜ | ꫜ |
| Et cetera | ꫝ | ꫝ |  |
| Ho Hoi | ꫞ | ꫞ | ꫞ |
| Repetition | ꫟ | ꫟ | ꫟ |
| Shortened repetition |  |  | ꫟‍ <uAADF>+<ZWJ> |

**Table 6. Special symbols**