

# Tay Viet Script for Unicode

**Jim Brase**

SIL International

**February 6, 2007**

Referred to as Việt Thái, row AA in the latest Roadmap

## Sociolinguistic background

The Tay Viet script is used by three Tai languages spoken primarily in northwestern Vietnam, northern Laos, and central Thailand—Tai Dam (also Black Tai or Tai Noir), Tai Dón (White Tai or Tai Blanc), and Thai Song (Lao Song or Lao Song Dam). The Thai Song of Thailand are geographically removed from, but linguistically related to the Tai people of Vietnam and Laos. There are also populations in Australia, China, France, and the United States. The script is related to other Thai scripts used throughout Southeast Asia.

The *Ethnologue* estimates the total population of the three languages, across all countries, at 1.3 million. (Tai Dam 764,000, Tai Dón 490,000, Thai Song 32,000.)

The script is still used by the Tai people in Vietnam, and there is a desire to introduce it into formal education there (Cầm Trọng 2005). On the other hand, it is not known whether it is in current use in Laos, Thailand, or China.

A fourth language, Tai Daeng (Red Tai or Tai Rouge, 165,000), uses a very closely related script. But the differences in the vowel structure of Tai Daeng are significant enough that it will probably require encoding as a separate script.

## Script name

Several different spellings have been employed for the name of the language and script. In linguistic circles, it is common to use “Thai” to indicate the language of central Thailand, and “Tai” to indicate the language family. But even that usage is followed not consistently.

At a workshop on the Tay script, held in Dien Bien Phu, Vietnam, in November 2006, native speakers indicated that they prefer the spelling “Tay”, because it more closely reflects their own pronunciation for the name of their language<sup>1</sup>. We use “Tay Viet” here, to distinguish it from other Tai, or Tay, languages and scripts.

---

<sup>1</sup> The Vietnamese alphabet uses ‘t’ for /t/, ‘th’ for /tʰ/, ‘ai’ for /a:j/, and ‘ay’ for /aj/. The Tay of Vietnam call themselves /taj<sup>4</sup>/, which transcribes to ‘tay’ in the Vietnamese alphabet.

## Basic features

The Tay Viet script shares many features with other Tai alphabets:

- It is written left to right.
- There is a double set of initial consonants, one for low tone class and one for high tone class.
- Vowel marks are positioned before, after, above, or below the syllable's initial consonant, depending on the vowel. Some vowels are written with digraphs.
- The consonants *do not* carry an implicit vowel. The vowel must always be written explicitly.

## Storage order

Characters will be stored in visual order, for the following reasons:

- Established keyboarding practices use visual order, as does the handwriting. While input methods can be developed to support reordering of the input stream, those currently available are not advanced enough to provide a transparent editing environment after the input stream has been reordered. When the typist wants to edit text that has already been reordered, he finds that it is not stored in the order he sees and expects.
- Experience gained by SIL with this script in the 1990s on a Macintosh based system revealed that the user experiences considerable confusion when both the input and output streams are reordered. Thus, visual order will result in a much improved user experience.
- The Tay make up a relatively small user community. Consequently, vendors may not produce the necessary software to render phonetic order, while there is a strong and urgent need to get the script into users' hands as soon as possible. Complex systems will be difficult and expensive to maintain for a small community. Simplifying the input model will help to ensure that they have a system that works well.
- Visual order is also used by the Lao script, to which the Tay is closely related.

## Word and Syllable Structure

The Tay languages are almost exclusively monosyllabic. A very small number of words have an unstressed initial syllable, and loan words may be polysyllabic. The practice followed in Baccam, et. al. (1989) was to write polysyllabic words without space between the syllables. No tone is written on loan words or on the unstressed initial syllable of a native word.

There are two different systems of tone marks in use, one using combining marks written over the initial consonant, the other using spacing marks written on the baseline at the end of the syllable. See **Tone classes and tone marks**, below, for a discussion of the two different tone systems.

Depending on the tone system that is used, the written syllable may have any of the following structures:

**Using combining tone marks**

V<sub>1</sub> C W? T? F?  
 C W? V<sub>2</sub> T? F?  
 C W? T? V<sub>3</sub> F?  
 V<sub>1</sub> C W? V<sub>2</sub> T? F?  
 V<sub>1</sub> C W? T? V<sub>3</sub> F?

**Using spacing tone marks**

V<sub>1</sub> C W? F? T?  
 C W? V<sub>2</sub> F? T?  
 C W? V<sub>3</sub> F? T?  
 V<sub>1</sub> C W? V<sub>2</sub> F? T?  
 V<sub>1</sub> C W? V<sub>3</sub> F? T?

An initial consonant C is always written. Even when the initial consonant is null (phonetically a glottal stop), it is written with the symbol ʘ or ʘ̣.

Initial velar consonants may be labialized, indicated by W?. The labialization is marked by the high-series letter ‘v’, ʘ, following the consonant.

V<sub>1</sub> indicates a pre-vowel that is rendered before the consonant. V<sub>2</sub> is a combining vowel rendered above or below the consonant, and V<sub>3</sub> is a post-vowel rendered after the consonant. Vowel digraphs can be formed from sequences V<sub>1</sub> + V<sub>2</sub> or V<sub>1</sub> + V<sub>3</sub>.

T indicates an optional tone mark which, as already noted, may be either a combining tone mark over the initial consonant, or a spacing mark at the end of the syllable.

F indicates an optional final consonant.

In handwriting, styles vary as to where combining marks are placed. They are typically placed over (or under) the initial consonant, but when there is a final consonant present, they often drift to the right, sometimes being in the gap between the two consonants, and sometimes being over the final consonant.

When there is a labialized consonant, the placement of a combining mark can be important in resolving ambiguity. Combining marks are written over the second part of the labialized consonant. Thus:

ᵏᵑ̣ /kiw/ vs. ᵏᵑ̣̣ /k<sup>w</sup>i/, and  
 ᵏᵑ̣̣̣ /kɛw<sup>3</sup>/ vs. ᵏᵑ̣̣̣̣ /k<sup>w</sup>ɛ<sup>3</sup>/

**Vowels**

Vowel symbols can be classified according to where they are written relative to the initial consonant.

Vowels written before the consonant:

ᵏ /ɛ/    ᵏᵏᵏ /kɛn<sup>2</sup>/ ‘seed’  
 ᵏ /o/    ᵏᵏᵏ /xon<sup>1</sup>/ ‘fur, feather’  
 ᵏ /iə/    ᵏᵏ /sia<sup>1</sup>/ ‘tiger’  
 ʘ /əw/    ʘᵏᵏ /nəw<sup>2</sup>/ ‘large, big’  
 ʘ /aj/    ʘᵏ /daj<sup>3</sup>/ ‘to attain’

Vowels written above the consonant:

◌̃ /a/	ຕັຕ /tat <sup>2</sup> / ‘to cut’
◌̂ /i/	ຕັ໊ /t <sup>h</sup> iw <sup>1</sup> / ‘to whistle’
◌̃̄ /iə/	ມັ /miə <sup>4</sup> / ‘wife’
◌̂̄ /i/	ປັ /pin <sup>1</sup> / ‘arrow’
◌̃̄ /ɔʰ/	ພັ /pɔ <sup>4</sup> / ‘enough, sufficient’ (only used in open syllables)
◌̃̄̄ /am/	ຄັ /kam <sup>4</sup> / ‘gold’

Vowels written below the consonant:

◌̣ /u/	ຸນ /xun <sup>2</sup> / ‘dust’
--------	-------------------------------

Vowels written after the consonant:

າ /a:/	ອັາ /ʔa:η <sup>2</sup> / ‘basin, tub’
ອ /ɔ/	ອອ /ʔɔ <sup>2</sup> / ‘to go out’ (used in closed syllables)
î (only used in Vietnamese loan words)	
ຮ /uə/	ສຮ /suə <sup>3</sup> / ‘trousers’
ຳ /an/	ປຳ /pan <sup>3</sup> / ‘to squeeze’

Digraph vowels:

These are combining sequences which do not need to be encoded as separate units.

◌̃̄̄ = ◌̃ + ◌̃̄ /e/	ເັມ /tem <sup>1</sup> / ‘full’
◌̃̄̄̄ = ◌̃̄ + ◌̃̄̄ /ə/	ເັມັ /həη <sup>1</sup> / ‘long’
◌̃̄̄̄̄ = ◌̃̄̄ + ◌̃̄̄̄ + າ /a:w/	ເັມັາ /ɲa:w <sup>1</sup> / ‘reflection’
◌̃̄̄̄̄̄ = ◌̃̄̄̄ + ັ + ◌̃̄̄̄̄̄ /ap/	ເັມັັ /kap <sup>2</sup> / ‘with, and’

## Tone classes and tone marks

In the Tay Viet script each consonant has two forms. The low form of the initial consonant indicates that the syllable uses tone 1, 2, or 3. The high form of the initial consonant indicates that the syllable uses tone 4, 5, or 6. This is sufficient by itself to define the tone of checked syllables (those ending /p/, /t/, /k/, or /ʔ/), in that these syllables are restricted to tones 2 and 5.

Traditionally, the Tay Viet script did not use any further marking for tone. The reader had to determine the tone of unchecked syllables from the context. In recent times, however, several groups have introduced tone marks into Tay Viet writing. Tai speakers in the United States begin using Lao tone marks with their script about 30 years ago, and those marks are included in SIL’s Tai Heritage font (developed for the Tai Dam language). These symbols are written as combining marks above the initial consonant, and are identified by their Laotian names, *mai ek* and *mai tho*.

The Song Petburi font (a Thai Song font) includes Thai style tone marks, which are identical to the Lao.

The Tai community in Vietnam, however, invented their own tone marks written on the base line at the end of the syllable, which they name *mai nwnɡ*<sup>2</sup> and *mai song*.

When combined with the consonant class, two tone marks are sufficient to unambiguously mark the tone. Thus, depending on which system one uses, tones may be written as follows on unchecked syllables:

	no mark	◌ <sup>ˊ</sup>	◌ <sup>ˋ</sup>
low class consonant	tone 1	tone 2	tone 3
high class consonant	tone 4	tone 5	tone 6

Marking tones with symbols *mai ek* and *mai tho*

	no mark	◌.e	◌.j
low class consonant	tone 1	tone 2	tone 3
high class consonant	tone 4	tone 5	tone 6

Marking tones with symbols *mai nwnɡ* and *mai song*

It is recognized that the existence of two distinct sets of tone marks is a disadvantage to the script. However, they cannot be unified, because both their combining classes and their storage order are different. For example:

$$\overset{\sim}{\text{㗗}}\text{㗗} = \text{㗗} + \overset{\sim}{\text{㗗}} + \text{㗗} + \text{㗗}$$

$$\text{㗗}\text{㗗}\text{㗗} = \text{㗗} + \text{㗗} + \text{㗗} + \text{㗗}$$

(/xɔŋ<sup>3</sup>/, ‘to trip over’)

Perhaps in time, one system will become dominant and the other will die out. But for now, both are in use.

## Final consonants

In written form, the low-tone class symbols for ‘b’ ( 㗗 ) and ‘d’ ( 㗗 ) are used for syllable final /p/ and /t/, respectively, as is the practice in many Thai scripts.

The low-tone class symbol for ‘k’ ( 㗗 ) is used for both final /k/ and final /ʔ/.

The high-tone class symbols are used for writing final /j/ ( 㗗 ) and the final nasals, /m/ ( 㗗 ), /n/ ( 㗗 ), and /ŋ/ ( 㗗 ). High-tone /v/ ( 㗗 ) is used for final /w/.

There are a number of exceptions to the above rules in the form of Vowel + Final Consonant ligatures. These vary from region to region. The ones included in this proposal are the ones

<sup>2</sup> The letter ‘w’ is used here to write a high-central vowel, /i/.

with the broadest usage: /-aj/ ( 𑜁𑜂 ), /-am/ ( 𑜁𑜃 ), /-an/ ( 𑜁𑜄 ), and /-əw/ ( 𑜁𑜅 ). Another ligature, /-ap/ ( 𑜁𑜆 ), is composed of the –am ligature plus ‘b’, but does not need to be encoded.

## Symbols

There are five non-alphabetic symbols:

Symbol	Name	Tay name/ pronunciation	meaning
𑜁𑜂	TAY VIET SYMBOL KON	/kon <sup>4</sup> /	‘person’
𑜁𑜃	TAY VIET SYMBOL NWNĠ	/niŋ <sup>5</sup> /	‘one’
𑜁𑜄	TAY VIET SYMBOL SAM	<i>sam</i>	signals repetition of the previous word
𑜁𑜅	TAY VIET SYMBOL HO HOI	<i>ho hoi</i>	beginning of text (used in songs and poems)
𑜁𑜆	TAY VIET SYMBOL KOI KOI	<i>koi koi</i>	end of text (used in songs and poems)

## Do the ligatures need to be encoded?

As noted above, there are four Vowel + Final Consonant ligatures. Two others, the symbols TAY VIET SYMBOL KON and TAY VIET SYMBOL NWNĠ, are ligatures of entire words. In addition, the ligature TAY VIET VOWEL AN can be used by itself as a contraction for the word /nan<sup>6</sup>/, ‘that’. Do these ligatures need to be encoded, or can they be treated as combining sequences?

In at least two cases, the ligatures are used to distinguish between homophonous words:

𑜁𑜂 = /kon<sup>4</sup>/ ‘person’                      𑜁𑜅 = /kon<sup>4</sup>/ ‘to stir’

and:

𑜁𑜃 = /nan<sup>6</sup>/ ‘that’                              𑜁𑜆 = /nan<sup>4</sup>/ ‘loud, noisy’

In the latter case, the tone does distinguish the words. But not all of the user community has learned to write tones. If the TAY VIET VOWEL AN is not encoded, and if /nan<sup>6</sup>/ and /nan<sup>4</sup>/ are written without tone, a rendering engine will not be able to distinguish between them.

It is not known if there is any homonym for TAY VIET SYMBOL NWNĠ.

Keyboarding order and the use of visual storage order makes it more practical to encode the ligatures TAY VIET VOWEL AW and TAY VIET VOWEL AY. If they are not encoded, the input method and the rendering must both handle complicated reordering.

So there is some reason to encode the majority of the ligatures. It is felt that they should all be encoded and handled the same for the sake of consistency.

## Word spacing

In the last 30 years, users in both Vietnam and the United States have started writing spaces between words. This is contrary to the traditional practice of Tai scripts. When the author asked

representatives of the Tai community whether it was necessary to define a line breaking algorithm in order to support the old form with no spaces, there was no support for it. Everyone prefers to use interword spacing.

## Sort order

The Tai scripts do not have an established standard for sorting. Sequences have sometimes been borrowed from neighboring languages. Baccam, et. al. (1989) is a bilingual dictionary that uses the Lao order, adjusted for differences between the Tai Dam and Lao character sets. Cầm Trọng (2005) prefers an order based on the Vietnamese alphabet (Quốc ngữ).

We had to select an order for our code chart, and chose the one used by Baccam, largely because of the relationship of the Tay Viet script to Lao. This is recommended as the default sort order at this time.

More discussion with the Tay community is needed on this matter, but it is possible that communities in different countries will want to use different orders. I describe below the orders used by Baccam and Cầm Trọng.

### Sort order derived from Lao, Baccam, et. al. 1989

The Lao sort order is based on pronunciation: the initial consonant is evaluated, then the vowel or final vowel-consonant ligature, then the final consonant, and last of all the tone.

The initial consonants are sorted in the order listed in the code chart. If the initial consonant is a labialized velar, it comes after the corresponding unlabialized consonant. The symbols TAY VIET SYMBOL KON and TAY VIET SYMBOL NWNNG are sorted as if they were fully spelled out.

The vowel order is modified from the code chart by inserting the digraph vowels and the character TAY VIET LETTER O LOW (ᨀ) into their proper places:

- ◌̣    TAY VIET MAY KANG
- ᨀ    TAY VIET VOWEL AA
- ◌̂    TAY VIET VOWEL AA WITH CIRCUMFLEX
- ◌̄    TAY VIET VOWEL I
- ◌̆    TAY VIET VOWEL W
- ◌̈    TAY VIET VOWEL U
- ◌̉    TAY VIET VOWEL WA + TAY VIET VOWEL IA
- ◌̋    TAY VIET VOWEL E
- ◌̌    TAY VIET VOWEL O
- ◌̍    TAY VIET MAY KHIT
- ᨀ    TAY VIET LETTER O LOW
- ◌̎    TAY VIET VOWEL WA + TAY VIET MAY KHIT
- ◌̏    TAY VIET VOWEL IA

- ⊘ TAY VIET VOWEL WA
- ⊘ TAY VIET VOWEL UA
- ⊘ TAY VIET VOWEL AW
- ⊘ TAY VIET VOWEL AY
- ⊘ TAY VIET VOWEL WA + TAY VIET VOWEL AA
- ⊘ TAY VIET VOWEL AN
- ⊘ TAY VIET VOWEL AM
- ⊘ TAY VIET LETTER BO LOW + TAY VIET VOWEL AM

Any final consonants are then sorted in the code chart order.

Finally, tones are sorted with *mai ek* equivalent to *mai nwnng*, and *mai tho* equivalent to *mai song*. Thus:

- ⊘ = ⊘..e
- ⊘ = ⊘..j

**Sort order derived from Quốc ngữ, Chăm Trọng 2005**

The Quốc ngữ order would also be based on pronunciation. A number of questions remain to be answered, but I show here the charts from Chăm Trọng (2005) without further comment.

N.	Quốc ngữ	Reading	Low gr.	High gr.
1	b	bo	∕	∕∕
2	c,k,q	co	n	ɳ
3	ch, tr	cho	∕	∕
4	kh	kho	n	ɳ
5	đ,gi,r	đo	∕	∕
6	đ	đo	o	u
7	h	ho	∕	∕
8	l	lo	∕	∕
9	m	mo	∕∕	∕
10	n	no	∕∕	∕
11	nh	nho	∕∕	∕
12	ng, ngh	ngo	∕∕	∕

N.	Quốc ngữ	Reading	Low gr.	High gr.
13	o	o	o	o
14	p	po	∕	∕
15	ph,f	pho	∕	∕
16	s,x	xo	x	n
17	t	to	∕	∕
18	th	tho	∕	∕
19	v	vo	∕∕	∕
20	g,gh	go	∕	∕
21	r	ro	∕	∕
22	ts	tso	∕	∕
23	kh'	kh'o	∕	∕
24	f	fo	∕	∕



Nu m.	Qu. ngữ	Rea đing	Com b 1 cons	COMBINING WITH 8 LAST CONSONANTS							
				n: c, ch	ư: p	o: t	ư: m	ư: n	g:ng, nh	ư:i, y	ư:o, u
1	a	ca, ka	əɪ	əɪn əɪne	əɪv	əɪn	əɪɲ	əɪɳ	əɪɣ	əɪy	əɪu
2	ă	câng	əɪ̃	əɪ̃n əɪ̃ne	əɪ̃v	əɪ̃n	əɪ̃ɲ	əɪ̃ɳ	əɪ̃ɣ		
3	â	câng	ɨə	ɨəɳ ɨəne	ɨəv	ɨən	ɨəɲ	ɨəɳ	ɨəɣ	ɨəy	
4	o	o	əə	əəɳ əəne	əəv	əən	əəɲ	əəɳ	əəɣ	əəy	
5	ô	cô, công	ə̄	ə̄ɳ ə̄ne	ə̄v	ə̄n	ə̄ɲ	ə̄ɳ	ə̄ɣ	ə̄y	

Nu m.	Qu. ngữ	Rea đing	Com b 1 cons	COMBINING WITH 8 LAST CONSONANTS							
				n: c, ch	ư: p	o: t	ư: m	ư: n	g:ng, nh	ư:i, y	ư:o, u
6	ơ, â	cơ	ɨə	ɨəɳ ɨəne	ɨəv	ɨən	ɨəɲ	ɨəɳ	ɨəɣ	ɨəy	
7	e	ke	ɛə	ɛəɳ ɛəne	ɛəv	ɛən	ɛəɲ	ɛəɳ	ɛəɣ		ɛəu
8	ê	kê	fə	fəɳ fəne	fəv	fən	fəɲ	fəɳ	fəɣ		fəu
9	i	ki	əɳ	əɳɳ əɳne	əɳv	əɳn	əɳɲ	əɳɳ	əɳɣ		əɳu
10	u	cu	əɣ	əɣɳ əɣne	əɣv	əɣn	əɣɲ	əɣɳ	əɣɣ	əɣy	
11	ư	cư	ət	ətɳ ətne	ətv	ətɳ	ətɲ	ətɳ	ətɣ	ətɳ	ətɳ
12	oa	ka-vo	əɳɳ	əɳɳɳ əɳɳne	əɳɳv	əɳɳn	əɳɳɲ	əɳɳɳ	əɳɳɣ	əɳɳy	
13	ia, iê	kia	ət	ətɳ ətne	ətv	ətɳ	ətɲ	ətɳ	ətɣ		ətɳ
14	ua, uô	cua	ət	ətɳ ətne	ətv	ətɳ	ətɲ	ətɳ	ətɣ	ətɳ	
15	ua, uơ	cua	ɨə	ɨəɳ ɨəne	ɨəv	ɨən	ɨəɲ	ɨəɳ	ɨəɣ	ɨəy	
16	oe	que	ə̄ɳ	ə̄ɳɳ ə̄ɳne		ə̄ɳn		ə̄ɳɳ	ə̄ɳɣ		
17	uê	quê	fəɳ	fəɳɳ fəɳne		fəɳn		fəɳɳ	fəɳɣ		
18	au	cau lom	ɨəɳ								
19	ay	kay	ə̄								
20	aur	caur	fə								
21	uya, uyê	quya	ətɳ	ətɳɳ ətɳne					ətɳɣ		
22	uy	quy	ətɳ	ətɳɳ ətɳne							

## Reserved characters

It is recommended that character codes xx44..xx5A be reserved for future expansion of the Tai Viet character set.

The current proposal is focussed on the use of the script by the Tai Dam of Son La province, Vietnam. It contains the traditional Son La character set, plus three pairs of aspirated consonants required by the Tai Don language of Lai Chau province.

TAY VIET LETTER KHO LOW	TAY VIET LETTER KHO HIGH
TAY VIET LETTER CHO LOW	TAY VIET LETTER CHO HIGH
TAY VIET LETTER PHO LOW	TAY VIET LETTER PHO HIGH

Tai Don can be written with the resulting character set, but only if one uses the orthographic conventions of Son La. If one wishes to write Tai Don in one of their traditional styles, some additional characters will probably be needed. The author has identified four consonants and two vowels which have definite or probably contrast with characters from the Son La tradition, and 12-14 characters which use significantly different glyphs, although not in a contrastive way. However, these require additional study, and hopefully the input of someone who is an expert in Tai Don, before they can be proposed for the character set.

## Consideration of the Tai Daeng script

Finally, the committee is encouraged to consider whether Tai Daeng should be encoded as a separate script. About 50% of the Tai Daeng consonants have unique glyphs, and the vowel system is significantly different due to the fact that the language has length contrast on all vowels.

## Bibliography

- \_\_\_ . “Các mẫu tự Thái ở miền Tây Bắc Việt Nam.” Internet:  
<http://www.evertype.com/standards/tai/viet-thai-samples.pdf>
- \_\_\_ . Song Petburi font. Internet: <http://www.seasite.niu.edu/tai/TaiDam/index.htm>.
- Baccam Don, Baccam Faluang, Baccam Hung, Dorothy Fippinger. 1989. *Tai Dam – English, English – Tai Dam Vocabulary Book*. Summer Institute of Linguistics.
- Cầm Trọng. 2005. “Thai Scripts in Vietnam,” in *Workshop on the Preservation and Digitization of Tai Scripts*. Hanoi, Vietnam.
- Điêu Chính Nhìem and Jean Donaldson. 1970. *Pǎp San Khhâm Pák Tǎy-Keo-Eng* (Ngũ-Vụng Thái-Việt-Anh, Tai-Vietnamese-English Vocabulary). Saigon.
- Lò Văn Mười (𑜋𑜧 𑜏𑜢𑜤𑜰𑜫 𑜏𑜢𑜤𑜰𑜫). 1966. *Ép Sǔ Tǎy Piên Peng*. (𑜏𑜢𑜤𑜰𑜫 𑜏𑜢𑜤𑜰𑜫 𑜏𑜢𑜤𑜰𑜫 𑜏𑜢𑜤𑜰𑜫, Learning the Revised Tai Alphabet.)
- SIL. Tai Heritage font. Published by SIL International. Internet:  
[http://scripts.sil.org/cms/scripts/page.php?site\\_id=nrsi&item\\_id=SILTD\\_home](http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=SILTD_home)
- Workshop on the Preservation and Digitization of Tai Scripts*. Hanoi, Vietnam. 2005.
- Workshop on Encoding and Digitalizing Thai Scripts*. Điện Biên, Vietnam. 2006.

TAY VIET—ROW AA

	xx0	xx1	xx2	xx3	xx4	xx5
0	h	h̄	h̄	ȯ	ȯ	
1	o	ō	ō	o	e	
2	o	o	o	ȯ	ȯ	
3	o	o	o	ȯ	ȯ	
4	o	o	o	ȯ		
5	o	o	o	ȯ		
6	o	o	o	ȯ		
7	o	o	o	ȯ		
8	o	o	o	ȯ		
9	o	o	o	ȯ		
A	o	o	o	ȯ		
B	o	o	o	ȯ		o
C	o	o	o	ȯ		o
D	o	o	o	ȯ		o
E	o	o	o	ȯ		o
F	o	o	o	ȯ		o

## TAY VIET Names Table

### Consonants

xx00	ᨠ	TAY VIET LETTER KO LOW
xx01	ᨡ	TAY VIET LETTER KO HIGH
xx02	ᨢ	TAY VIET LETTER KHO LOW
xx03	ᨣ	TAY VIET LETTER KHO HIGH
xx04	ᨤ	TAY VIET LETTER KHHO LOW
xx05	ᨥ	TAY VIET LETTER KHHO HIGH
xx06	ᨦ	TAY VIET LETTER GO HIGH
xx07	ᨧ	TAY VIET LETTER GO LOW
xx08	ᨨ	TAY VIET LETTER NGO LOW
xx09	ᨩ	TAY VIET LETTER NGO HIGH
xx0A	ᨪ	TAY VIET LETTER CO LOW
xx0B	ᨫ	TAY VIET LETTER CO HIGH
xx0C	ᨬ	TAY VIET LETTER CHO LOW
xx0D	ᨭ	TAY VIET LETTER CHO HIGH
xx0E	ᨮ	TAY VIET LETTER SO LOW
xx0F	ᨯ	TAY VIET LETTER SO HIGH
xx10	ᨰ	TAY VIET LETTER NHO LOW
xx11	ᨱ	TAY VIET LETTER NHO HIGH
xx12	ᨲ	TAY VIET LETTER DO LOW

xx13	ᨳ	TAY VIET LETTER DO HIGH
xx14	ᨴ	TAY VIET LETTER TO LOW
xx15	ᨵ	TAY VIET LETTER TO HIGH
xx16	ᨶ	TAY VIET LETTER THO LOW
xx17	ᨷ	TAY VIET LETTER THO HIGH
xx18	ᨸ	TAY VIET LETTER NO LOW
xx19	ᨹ	TAY VIET LETTER NO HIGH
xx1A	ᨺ	TAY VIET LETTER BO LOW
xx1B	ᨻ	TAY VIET LETTER BO HIGH
xx1C	ᨼ	TAY VIET LETTER PO LOW
xx1D	ᨽ	TAY VIET LETTER PO HIGH
xx1E	ᨿ	TAY VIET LETTER PHO LOW
xx1F	ᩀ	TAY VIET LETTER PHO HIGH
xx20	ᩁ	TAY VIET LETTER FO LOW
xx21	ᩂ	TAY VIET LETTER FO HIGH
xx22	ᩃ	TAY VIET LETTER MO LOW
xx23	ᩄ	TAY VIET LETTER MO HIGH
xx24	ᩅ	TAY VIET LETTER YO LOW
xx25	ᩆ	TAY VIET LETTER YO HIGH

xx26	Ꞑ	TAY VIET LETTER RO LOW
xx27	ꞑ	TAY VIET LETTER RO HIGH
xx28	Ꞓ	TAY VIET LETTER LO LOW
xx29	ꞓ	TAY VIET LETTER LO HIGH
xx2A	ꞔ	TAY VIET LETTER VO LOW
xx2B	ꞕ	TAY VIET LETTER VO HIGH
xx2C	Ꞗ	TAY VIET LETTER HO LOW
xx2D	ꞗ	TAY VIET LETTER HO HIGH
xx2E	Ꞙ	TAY VIET LETTER O LOW
xx2F	ꞙ	TAY VIET LETTER O HIGH

### Vowels and Finals

xx30	Ꞛ	TAY VIET MAY KANG
xx31	ꞛ	TAY VIET VOWEL AA
xx32	Ꞝ	TAY VIET VOWEL AA WITH CIRCUMFLEX
xx33	ꞝ	TAY VIET VOWEL I
xx34	Ꞟ	TAY VIET VOWEL W
xx35	ꞟ	TAY VIET VOWEL U
xx36	Ꞡ	TAY VIET VOWEL E
xx37	ꞡ	TAY VIET VOWEL O

xx38	Ꞣ	TAY VIET MAY KHIT
xx39	ꞣ	TAY VIET VOWEL IA
xx3A	ꞣ	TAY VIET VOWEL WA
xx3B	Ꞥ	TAY VIET VOWEL UA
xx3C	ꞥ	TAY VIET VOWEL AW
xx3D	Ꞧ	TAY VIET VOWEL AY
xx3E	ꞧ	TAY VIET VOWEL AN
xx3F	Ꞩ	TAY VIET VOWEL AM

### Tones

xx40	ꞩ	TAY VIET TONE MAI EK
xx41	Ɦ	TAY VIET TONE MAI NUNG
xx42	Ɜ	TAY VIET TONE MAI THO
xx43	Ɡ	TAY VIET TONE MAI SONG

### Symbols

xx5B	Ɬ	TAY VIET SYMBOL KON
xx5C	Ɪ	TAY VIET SYMBOL NUNG
xx5D	ꞯ	TAY VIET SYMBOL SAM
xx5E	Ʞ	TAY VIET SYMBOL HO HOI
xx5F	Ʇ	TAY VIET SYMBOL KOI KOI

## Character Properties

code value/ range	Rep Glyph	Unicode Character Name	Gen Cat	Can Comb Class	Bidi Cat	Logical Order Exception
xx00 ...xx2F			Lo	0	L	
xx30	◌̈́	TAY VIET MAY KANG	Mn	0	NSM	
xx31	1	TAY VIET VOWEL AA	Lo	0	L	
xx32	î	TAY VIET VOWEL AA WITH CIRCUMFLEX	Lo	0	L	
xx33	◌̈́ <sup>^</sup>	TAY VIET VOWEL I	Mn	0	NSM	
xx34	◌̈́ <sup>~</sup>	TAY VIET VOWEL W	Mn	0	NSM	
xx35	◌̈́ <sup>u</sup>	TAY VIET VOWEL U	Mn	0	NSM	
xx36	“	TAY VIET VOWEL E	Lo	0	L	yes
xx37	(	TAY VIET VOWEL O	Lo	0	L	yes
xx38	◌̈́	TAY VIET MAY KHIT	Mn	0	L	
xx39	◌̈́ <sup>^</sup>	TAY VIET VOWEL IA	Mn	0	NSM	
xx3A	‘	TAY VIET VOWEL WA	Lo	0	L	yes
xx3B	⋈	TAY VIET VOWEL UA	Lo	0	L	
xx3C	ᵝ	TAY VIET VOWEL AW	Lo	0	L	yes
xx3D	Ϛ	TAY VIET VOWEL AY	Lo	0	L	yes
xx3E	ư	TAY VIET VOWEL AN	Lo	0	L	
xx3F	◌̈́ <sup>h</sup>	TAY VIET VOWEL AM	Mn	0	NSM	
xx40	◌̈́	TAY VIET TONE MAI EK	Mn	0	NSM	
xx41	◌̈́ <sup>e</sup>	TAY VIET TONE MAI NNG	Lo	0	L	
xx42	◌̈́ <sup>~</sup>	TAY VIET TONE MAI THO	Mn	0	NSM	
xx43	)	TAY VIET TONE MAI SONG	Lo	0	L	
xx5B	đ	TAY VIET SYMBOL KON	So	0	L	
xx5C	đ	TAY VIET SYMBOL NNG	So	0	L	
xx5D	h	TAY VIET SYMBOL SAM	So	0	L	
xx5E	ư	TAY VIET SYMBOL HO HOI	So	0	L	
xx5F	h	TAY VIET SYMBOL KOI KOI	So	0	L	

### Script Samples

# ພາປັ ມັ ຂັ ພີ ນອບ ຂາ ກຸ້ ທຸ້

ອັນ ຫາ ນີ້ ມາ ປູ້ ຫຼື ທີ່ ດ້ານ ນອກ, ກ່ອນ ຂາມ ຂັ້ ຂໍ ມະ ປີ້ ແອວ ດີ້. ປຸ່ນ ດີ້ ໃດ ເປັນ ແນວ ຫັ້ ຕາັ້, ຕອບ ແນວ ທາ ດັ້ ນີ້ ນຸ່ ຂັ້ ຫັ້ ປອບ, ດີ້ ຫັ້ ດ້ ຫັ້ ປອບ ແກ່ ເຫຼ່າ, ດາປັ ດາຍ ແກ່ ປີ້ ທີ່ ພັ້ ຫຸ່ນ ຫຸ່ ນັ້ ນຸ່ ທີ່ ເປັນ ດີ້. ອັນ ຫາ ດີ້ ໄດ້ ແກ່ ອາປັ ແກ່ ນອບ ກຸ້ ຂາ ປັ ເອາ ນີ້ ດີ້. ຫາ ຈາ ປຸ່ນ ປາ ໃ້ ຫມ ພັ້ ດີ້ ຫາ ຫມ ຂັ້ ນາ, ດາປັ ນັ້ ດ້ ນີ້ ກໍ ທີ່ ຕາ ອງ ໃຈ ນັ້ ຂໍ້ ແອວ, ອາບ ຂໍ້ ດັ້ ປູ້ ອັ້ ດຸ່ ດີ້ ເປັນ ຫາປັ. ຫອປັ ປອບ ໄດ້ ນັ້ ທາປັ ມາ ແມ່ນ ດັ້ ຕັ້ ນ ໄດ້ ນັ້ ແອ, ດັ້ ດີ້ ກໍ ນັ້ ດັ້ ດີ້ ທີ່ ດາປັ ແກ່ ນີ້ ນັ້ ດີ້ ຫັ້ ນັ້ ດີ້ ນັ້ ດີ້

Figure 1—A modern text from Son La.

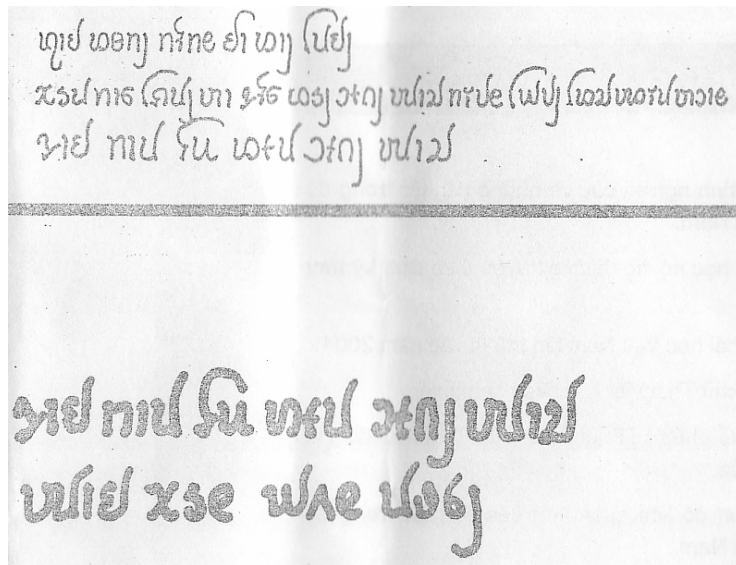


Figure 2—From *Giới Thiệu Chương Trình Thái Học Việt Nam*, 1999.



ကဝ်ၵ်	ကးຈိပ	ca-chóp	hoe
ကဝ်ၵ်	ကးຈဝပ	ca-chóp	hoe
ကဝ်ၵ်ၵ်	ကးၵ်ၵ်	ca-dǎng	stiffened, frozen
ကဝ်ၵ်	ကးၵ်	ca-dép	basket with strap
ကဝ်ကဝ်	ကးတက	ca-ta	even if
ကဝ်ကဝ်က	ကးတကက	ca-tác	to cackle
ကဝ်ၵ်ၵ်	ကးဗမ	ca-bem	coffer
ကံၵ်	ကံၵ်	cǎng	ape
ကံၵ်	ကံၵ်	cát	to gnaw

Figure 3—From Baccam et. al., p 13. The left hand column is Tai Dam.

နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ်  
 နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ်  
 နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ်  
နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ်  
 နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ်  
 နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ်  
 နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ်  
 နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ်  
 နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ်  
နဝ်ၵ်ၵ် နဝ်ၵ်ၵ် နဝ်ၵ်ၵ်

Figure 4—From *Khhâm Kháo Đì Châu Dê-su Seo Lũng Ók Mác Tèm*, 1983. A handwritten text in Tai Don.