

Towards a Unicode Proposal for the Unified Tai Script

last edited January 27, 2006

Contents

Objective

What is Unicode

Language Names and Locations

Data Sources

Key Language Features

Syllable Structure

Initial Consonant Charts

Summary of Initial Consonants

Final Consonants

Vowels

Tones

The Tai Writing System

Basic Features

Tone Classes and Tone Marks

Final Consonants

Word Spacing

Baseline

Sort Order

Correlation of the Character Sets Between Languages

The Unified Tai Alphabet

Traditional Writing Outside of the Unified Alphabet

History of Technology and the Tai Scripts

Selection of the Unicode Character Set

Key Issues

Is it sufficient to encode only the Unified Alphabet?

Should Tai Daeng be included?

Should Thai Song be included?

Criteria for Selecting Characters for the Standard

Comment on Character Images Used in This Paper

Transcription of Character Names

Rationale for Disunification of Characters and Other Comments

Characters Not Encoded

Bibliography

Unified Tai Character Set (Readers who downloaded the smaller, individual files should see the file TaiScriptCharSet.pdf)

Script Samples Listed by Phoneme (Readers who downloaded the smaller, individual files should see the files TaiScriptByPhoneme_Conson.pdf, TaiScriptByPhoneme_Vowels.pdf, TaiScriptByPhoneme_Ideogr.pdf, and TaiScriptByPhoneme_Notes.pdf)

Objective

The purpose of this paper is to explain the proposed Unicode character set for the Unified Tai Script, and to solicit feedback on how this proposal might be improved. Please send your comments to jim_brace@sil.org.

What is Unicode

If you are unfamiliar with Unicode, see <http://www.unicode.org/standard/WhatIsUnicode.html>.

Language Names and Locations

We are concerned with the traditional hand-written scripts of four Tai languages spoken primarily in Vietnam, Laos, and Thailand. There are also populations in Australia, China, France, and the United States. The *Ethnologue* (http://www.ethnologue.com/country_index.asp?place=Asia) gives their names as Tai Daeng (also Red Tai or Tai Rouge), Tai Dam (Black Tai or Tai Noir), Tai Dón (White Tai or Tai Blanc), and Thai Song (Lao Song or Lao Song Dam). The *Ethnologue* lists several other closely related languages that are spoken in Vietnam. I don't have any detailed information about most of them. One of them, Tày Tac, spoken in the Mường Tac district of eastern Sơn La province, has been documented by Donaldson and Edmondson (1997). But the writing samples that are available for this study from Mường Tac and nearby Mường Sang are not substantially different from the Tai Dam and Tai Daeng samples. Therefore, we will not deal with Tày Tac as a separate language.

The heartland of the Tai Dam is in Sơn La province of Vietnam. There are smaller Tai Dam populations in Houaphan and Luang Namtha provinces of Laos. The Tai Dón are generally located to the north of the Tai Dam, especially in Lai Chau and Dien Bien provinces. The Tai Daeng are located to the south, in Houaphan province of Laos and down into Thanh Hoa and Nghe An provinces of Vietnam. But there is considerable intermingling of all three languages with each other and with surrounding language groups.

The Thai Song are geographically removed from, but linguistically related to the Tai people. They were taken from the Tai Dam area two or three centuries ago and resettled in the Petchaburi and Kanchanaburi regions of central Thailand. (Whitehouse 1975, and <http://www.seasite.niu.edu/tai/TaiDam/article/a4.htm>). Although the heartland of the Thai Song is somewhat separated from the other languages, given their close linguistic relationship I thought it would be beneficial to compare their writing to that of the other three.

The *Ethnologue* estimates the total population of the four languages, across all countries, at 1.5 million. (Tai Daeng 165,000, Tai Dam 764,000, Tai Dón 490,000, Thai Song 32,000.)

Collectively, the people of the three northern language groups refer to themselves as Tai (with an unaspirated ‘t’), in contrast to “Thai” (with an aspirated ‘t’) by which they refer to Central (Bangkok) Thai. Some outsiders refer to them collectively as Viet Thai, but I am not aware of any of my Tai contacts (members of the Tai Dam language community who live in the United States) ever using this term in reference to themselves.

Based on the geographic distribution of these people, the number of languages involved, and their own name for themselves, I am proposing that their writing system be called the “Unified Tai Script.”

Data Sources

The data I used for my study is found in the file *Tai script by phoneme*, available in both spreadsheet and PDF formats. I collected 36 different data samples from 17 sources, representing about 20 different regions or dialects. It is difficult to get an accurate count of the number of regions, because authors identified the regions that their data represented differently. Some merely named a language, others a province, and others a village. Some names and political boundaries change over time. The table and map below identify the regions represented by the data.

In addition, I made a small intelligibility survey involving three Tai Dam readers in the U.S. The purpose was to see how readily they could read the script from the other languages. Unfortunately, I have not been able to do this with readers from the other languages.

Region	Sources	Description
Mường Tè	Ferlus 2003	District in the far northwest corner of Lai Chau province, northwest of the town of Lai Chau. Northern Tai Dón region. See http://vietnam.sawadee.com/laichau/map.htm
Mường So, Phong Thổ	—, <i>Các Mẫu Tư Thái Ở Miền Tây Bắc Việt Nam</i> Lo Văn Mười	Mường So is the Tai name for Phong Thổ, a district in the northeast corner of Lai Chau Province. (See http://www.fallingrain.com/world/VM/22/Phong_Tho.html and http://vietnam.sawadee.com/laichau/map.htm)
Mường Lay, Mường Cha	—, <i>Các Mẫu Tư Thái Ở Miền Tây Bắc Việt Nam</i> Lo Văn Mười	There are several possible uses for the name Mường Lay: <ul style="list-style-type: none"> • It is the Tai name for Lai Chau province. • It is the former name for Mường Cha, a district in northern Dien Bien Province. See http://vietnam.sawadee.com/laichau/map.htm. • It is a town on Hwy 12 south of the town of Lai Chau.
Lai Châu	Finot 1917, citing Pontalis	Province of northwestern Vietnam and town that is the capital of that province. Heart of the Tai Dón region. At the end of 2003 Lai Châu was divided into two parts: a smaller Lai Châu province and a newly created Dien Bien province. Data from a Lai Châu source could be from either of these two provinces.

Dien Bien		A new province created out of the southern half of Lai Châu in 2003. Although none of the data sources mention Dien Bien Province, data attributed to Lai Châu Province could be from Dien Bien.
Lống Liễn	Lo Văn Mười	I was unable to locate this area. The data attributed to Lống Liễn is clearly marked as Tai Dón by the presence of both aspirated stops and fricatives.
Nghĩa-lộ	Finot 1917, citing Diguet	Town in Yen Bai Prov., on Highway 32, approximately 104-27' E, 21-30'N (Microsoft Encarta)
Van-bu, Ta Bu	Finot 1917, citing Diguet	The location of Van-bu is uncertain. One contact says it no longer exists, but was near Ta Bu, a subdistrict of Mường La in Sơn La Province. I have shown Ta Bu on the map. (See http://www.traveljournals.net/explore/vietnam/map/m4956526/ta_bu.html)
Hưng-hoá	Finot 1917, citing Silvestre	In Tam Nông district, Phú Thọ province. Lat. 21°15'N, Long 105° 18'E. See http://www.traveljournals.net/explore/vietnam/map/m4938950/hung_hoa.html .
Sơn La	Ferlus 2003, citing Hưng Hóa Ký Lược (HHKL)	Province of Vietnam south of Lai Chau, west and slightly north of Hanoi. Also the town by the same name that is the provincial capital and a district in the center of the province. It is the heart of the Tai Dam region. The northern part of the province would be the southern edge of the Tai Dón region.
Thuan Chau, Mường Moui	Tai Heritage font	Thuan Chau is a district in northwest Sơn La province. The district capital of the same name is on Hwy 6 northwest of Sơn La. The artist and consultant who designed the Tai Heritage font came from this region. Mường Moui is the Tai name for the region.
Mường La	_, <i>Các Mẫu Tự Thái Ở Miền Tây Bắc Việt Nam</i>	District in north central Sơn La Province., north of the town of Sơn La. See http://vietnam.sawadee.com/sonla/map.htm Also the Tai name for the entire Sơn La province.
Phù Yên, Mường Tấc	_, <i>Các Mẫu Tự Thái Ở Miền Tây Bắc Việt Nam</i> Ferlus 2003 Lo Văn Mười	Phù Yên is a district on the eastern edge of Sơn La Province., north of Mốc Châu District. Mường Tấc is the Tai name.
Mốc Châu, Mường Sang	_, <i>Các Mẫu Tự Thái Ở Miền Tây Bắc Việt Nam</i> Ferlus 2003 Lo Văn Mười	Mốc Châu is a district in the southeast corner of Sơn La Province, south of Phù Yên, on opposite side of the Black River. The Tai name is Mường Sang.
Cau Chau	Lo Văn Mười	I was unable to locate this area. The script style appears to be Tai Dam.
Houaphan	Lafont	Laotian province south of Sơn La and north of Nghe An. Has both Tai Dam and Tai Daeng populations.
Luang Namtha	Lafont	Town and province in northwest Laos, just south of where the borders of China, Burma, and Laos meet. Has Tai Dam population in the area.
Kanchanaburi	Whitehouse	Northwest of Bangkok, Thailand.
Petchaburi	Song Petburi font	On the Gulf of Thailand southwest of Bangkok, Thailand.



Key Language Features

Syllable Structure

In Tai Daeng, Tai Dam, and Tai Dón, syllables are either CV or CVC. In Thai Song, syllables are either CV, CCV, CVC, or CCVC. Thai Song can form consonant clusters with /l/ as the second segment, or, if the first segment is a velar, with /w/. The other languages do not form clusters with /l/. They do combine velar consonants with /w/, but these are usually analyzed as a single-segment labialized consonant rather than a consonant cluster.

(Note: Whitehouse p. 28 lists CCCV and CCCVC as possible syllable patterns for Thai Song, but these are phonetic patterns in which the initial stop and its aspiration are considered separate segments. But on the phonemic level, the pattern is CCV and CCVC. Example, /k^hwak/ ‘draw out of’.)

Initial Consonant Charts

The following charts show which consonants occur in a syllable-initial position for each language.

Tai Daeng

	Labial	Alveolar	Alveo-palatal	Velar	Labio-velar	Glottal
Stop, vl. (affricate)	p	t	tɕ	k	kw	ʔ
Stop, vd.	b	d				
Asp. Stop		t ^h		k ^h	k ^h w	
Fricative	f	s				h
Liquid	v	l	j			
Nasal	m	n	ɲ	ŋ	ŋw	

Tai Dam

	Labial	Alveolar	Alveo-palatal	Velar	Labio-velar	Glottal
Stop, vl. (affricate)	p	t	tɕ	k	kw	ʔ
Stop, vd.	b	d				
Asp. Stop		t ^h				
Fricative	f	s		x ¹	xw ¹	h
Liquid	v	l	j			
Nasal	m	n	ɲ	ŋ	ŋw	

¹ Gedney lists these as aspirated stops, k^h and k^{hw}, but Fippinger analyzes them as fricatives. This may reflect a variation in dialects.

Tai Dón

	Labial	Alveolar	Alveo-palatal	Velar	Labio-velar	Glottal
Stop, vl. (affricate)	p	t	tɕ	k	kw	ʔ
Stop, vd.	b	d				
Asp. Stop (affricate)	p ^h	t ^h	tɕ ^h	k ^h	k ^h w	
Fricative	f	s		x	xw	h
Liquid	v	l	j			
Nasal	m	n	ɲ	ŋ	ŋw	

Thai Song

	Labial	Alveolar	Alveo-palatal	Velar	Labio-velar	Glottal
Stop, vl. (affricate)	p	t	tɕ	k		ʔ
Stop, vd.	b					
Asp. Stop	p ^h	t ^h		k ^h		
Fricative	f	s				h
Liquid	w	l	j			
Nasal	m	n	ɲ	ŋ		

Summary of Initial Consonants

Tai Dón has the largest inventory of initial consonants (26). Tai Dam omits four of the aspirated stops—it does not have /p^h/, /tɕ^h/, /k^h/, or /k^hw/. Tai Daeng does not have /p^h/, /tɕ^h/, /x/, or /xw/. Thai Song does not have /d/, /tɕ^h/, /x/ or /xw/. Although the labialized velar consonants /kw/, /k^hw/, and /ŋw/ do occur in Thai Song, they are analyzed as consonant clusters.

Final Consonants

Tai Dam, Tai Dón, and Thai Song have nine consonants which occur in syllable final position: /p/, /t/, /k/, /ŋ/, /w/, /j/, /m/, /n/, and /ŋ/. I have no information on the inventory of final consonants for Tai Daeng, but I assume it to be the same.

Vowels

Tai Dam Vowels

i	i ²	u
i ^{◌◌}	i ^{◌◌}	u ^{◌◌}
e	ə	o
ɛ	a	ɔ
	aa	

Tai Dón Vowels

i	i ²	u
e	ə	o
ɛ	a	ɔ
	aa	

Tai Dam has 13 vowels, including three glides, and a length contrast between /a/ and /aa/. Tai Dón lacks the three glides, giving it only 10 vowels. Words that use a glide in Tai Dam use the corresponding mid vowel in their Tai Dón cognates.

Tai Daeng Vowels

i	i ²	u
(ii) ?	(i ^h) ?	(uu) ?
i ^{◌◌}	i ^{◌◌}	u ^{◌◌}
(ii ^{◌◌}) ?	(i ^h ◌◌) ?	(uu ^{◌◌}) ?
e	ə	o
(ee) ?	(əə) ?	(oo) ?
æ	a	ɔ
(ææ) ?	aa	(ɔɔ) ?

Thai Song Vowels

i	i ²	u
ii	i ^h	uu
i ^{◌◌}	i ^{◌◌}	u ^{◌◌}
ii ^{◌◌}	i ^h ◌◌	uu ^{◌◌}
e	ə	o
ee	əə	oo
æ	a	ɔ
ææ	aa	ɔɔ
	aa ⁱ	

Thai Song has a length distinction on all vowels. Whitehouse also lists /aaⁱ/ as a vowel glide. A similar sounds exists in Tai Dam and Tai Dón, but is analyzed as a Vowel + final /w/.

According to Gedney, it is uncertain whether there is a length contrast between vowels other than /a/ and /aa/ in Tai Daeng. One of his sources had a clear length distinction for other vowels, but Gedney did not have enough data to determine if this was native to Tai Daeng or an influence of Lao. Aam and Aanu's 1974 Tai Daeng primer shows a length distinction for other vowels, but this could be a result of the influence of how Lao is written, and cannot be taken as proof that oral Tai Daeng has length contrast on all of its vowels. Other samples of Tai Daeng writing do not show a length contrast on the vowels³, but the samples may be

² Various Tai linguists have analyzed this vowel as /i/ high-central unrounded, /ɨ/ high-central rounded, or /u/ back-central unrounded

³ The most recent Tai Daeng document that I have aquired, an untitled, undated, 100-page hand written manuscript, appears to show length contrast on some vowels. I have only had time to analyze a few pages of it, but have been able to identify apparent contrasts in the written form between /i/ vs. /i^h/ and /i^{◌◌}/ vs. /i^h◌◌/.

incomplete, or there may be an oral contrast which is not reflected in the writing in those regions. Up to now, I have not found a linguistic study which provides a definitive answer to this question.

Strangely, although oral Thai Song has a complete set of long and short vowels, that distinction is not reflected in its written character inventory. In contrast, length distinction has not been determined for oral Tai Daeng, but Aam and Aanu's primer includes a complete set of long and short vowels for its written form.

Tones

Tai Daeng has five tones. The other languages all have six.

The Tai Writing System

Basic Features

The Tai scripts share many features common to most Thai alphabets. They are written left to right. (A related script, the Tai Do, is written vertically, but is beyond the scope of this study.) There is a double set of initial consonants, one for high tone class and one for low tone class. Vowel marks can be placed before, after, above, or below the syllable's initial consonant, depending on the vowel. Vowel digraphs are common.

Tone Classes and Tone Marks

In the Tai scripts each consonant has two forms. In Tai Dam and Tai Dón, the high form of the initial consonant indicates that the syllable uses tone 1, 2, or 3. The low form of the initial consonant indicates that the syllable uses tone 4, 5, or 6. I do not have complete information on tones for Tai Daeng or Thai Song, but assume the practice is similar.⁴

Traditionally, these scripts did not use any further marking for tone, and the reader had to determine the tone from the context. In recent times, however, several groups have made innovations which introduce tone marks into Tai writing. The Tai Heritage font borrowed tone marks from the Lao, and these are now widely used by the Tai Dam community in the U.S. The Unified Tai Alphabet (see below) invented a new set of spacing tone marks. The Song Petburi font includes Thai style tone marks. I do not know if these were innovative, or if they have been used by the Thai Song for some time. Aam and Aanu's Tai Daeng Primer has a unique set of tone marks. Again, I do not know how long these have been in use.

⁴ Some people reverse these labels. I.e., the form of the consonant which indicates tones 1-3 is labeled "low" and the form of the consonant which indicates tones 4-6 is labeled "high." This reversal may be based on the numbering system for the tones, or on the phonetic realization of the tones in some languages. But the phonetic expression of the tones varies from language to language, so that approach does not give a consistent result.

The usage of "high" for tones 1-3 and "low" for tones 4-6 is consistent with comparative Tai linguistics, in which system of tone boxes developed by Dr. William Gedney is often used to compare tones across languages and analyze their derivation from the tones of Proto Tai. It is also consistent with the names of corresponding characters in Lao and Standard Thai. E.g. Tai Dam ᨧ᩠ᩅ 'to seek', tone 1, is cognate with Lao ᨧ᩠ᩅ, the first character of which is LAO LETTER HO SUNG ('high'). Tai Dam ᨧ᩠ᩅ᩠ᩅ 'strength', tone 4, is cognate with Lao ᨧ᩠ᩅ᩠ᩅ, spelled with the LAO LETTER HO TAM ('low').

When combined with the consonant class, two tone marks are sufficient to unambiguously mark the tone. Thus, some authors mark tone in Tai Dam as follows:

	no mark	ˊ ...	ˋ ...
high class consonant	tone 1	tone 2	tone 3
low class consonant	tone 4	tone 5	tone 6

Note, however, that syllables ending with a stopped consonant (/p/, /t/, /k/, or /ʔ/) are restricted to tones 2 or 5, and that no marking other than the consonant class is necessary for those syllables.

The practice for the other languages would be similar to that for Tai Dam.

Final Consonants

In written form, the high-tone class symbols for ‘b’ and ‘d’ are used for syllable final /p/ and /t/, as is the practice in all Thai scripts. This usage should not mislead one into thinking that oral /b/ and /d/ occur syllable final.

The high-tone class symbol for ‘k’ is used for both final /k/ and final /ʔ/.

The low-tone class symbols are used for writing final /j/ and the final nasals, /m/, /n/, and /ŋ/. Low-tone /v/ is used for final /w/.

There are a number of exceptions to the above rules, in the form of Vowel + Final Consonant ligatures or digraphs. These vary from region to region, but the ones with the broadest usage are the ligatures for /-aj/, /-am/, /-an/, and /-əw/, and the digraph for /aw/. There are a number of other ligatures: /-aŋ/ and /-at/; and digraphs: /-um/ and /-uŋ/; whose usage is limited to some dialects of Tai Dón.

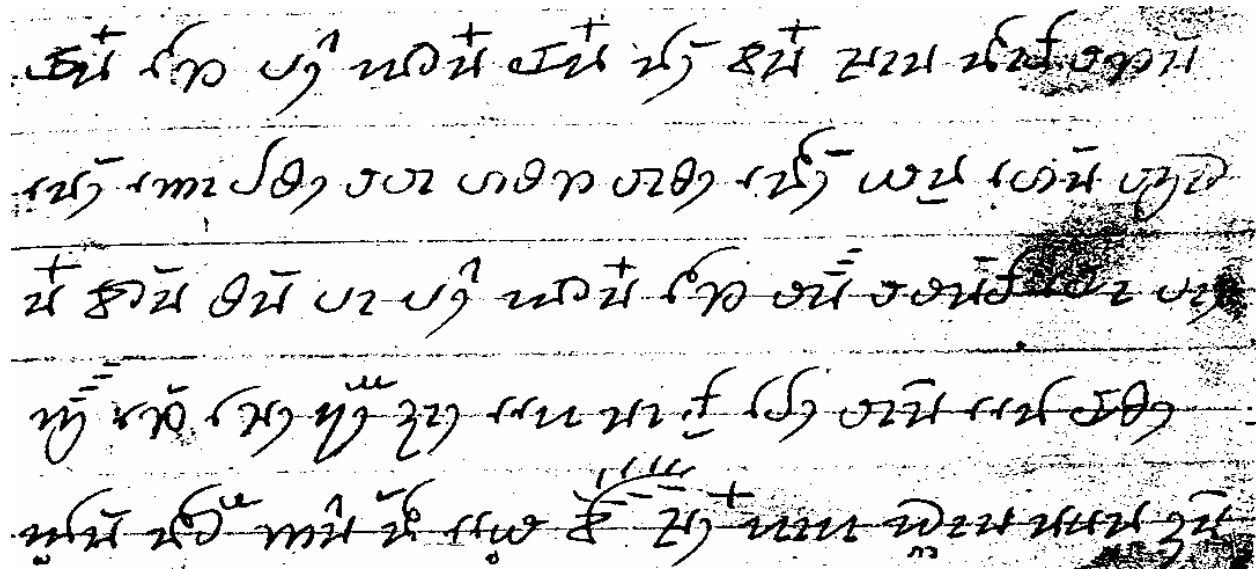
Word Spacing

Traditional Tai writing does not use space between words. However, in the last 20 years the Tai Dam community in the U.S. has adopted the practice of using word spacing, although the spaces are usually narrower than for Latin alphabets. A trilingual pamphlet published by the Hanoi National University in 1999, *Giới Thiệu Chương Trình Thái Học Việt Nam*, shows spacing between words in the Tai script.

The untitled Tai Daeng manuscript that I most recently acquired shows clear spacing between words. This is a surprise, as the manuscript appears to be rather old. See the sample on the next page.

Baseline

Tai scripts usually use a bottom baseline. However, the untitled Tai Daeng manuscript was written on lined paper, and has a center baseline—another surprise. See the sample on the next page.



A Tai Daeng manuscript showing word spacing and a center baseline

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) use an order borrowed from Lao. Since this is the most closely related language with an established standard, it seemed logical to follow it, which I have done as nearly as possible.

Correlation of the Character Sets Between Languages

The Tai languages have used their own traditional writing system for many centuries. Although related to the Lao alphabet, the character shapes and inventory are distinct from it. Tai Dam and Tai Dón script dialects have a high degree of correlation, with only a few characters unique to each language. However, there are a number of characters that have different uses in each language. For example,

Character	Tai Dam usage	Tai Dón usage
	low tone class /m/ 	
	low tone class /p/ 	low tone class /m/
	low tone class /f/ 	low tone class /p/
		low tone class /f/

Tai Daeng has less than a 50% correlation to the Tai Dam and Tai Dón—much lower than I expected. I identified 20 consonants and 15 vowel or tone marks that are unique to Tai Daeng. About 30 characters are shared with other languages.

Given the geographic separation and the likelihood of strong Thai influence, I expected the Thai Song writing to be different from the others. But in fact, I found the correlation between Thai Song and Tai Dam characters to be stronger than expected—on about the same level as the correlation between the Tai Dam and Tai Dón. To be sure, the style of the Song Petburi font, my best source for the Thai Song, was quite different from the typical style of Tai Dam writing, but the basic underlying shapes were similar, reflecting its Tai Dam origin.

The Unified Tai Alphabet

Anyone attempting to establish a standard for writing the Tai scripts must cope with the great diversity of symbols that are used by these languages. The diversity is the result of differences in phonology, isolation of communities from one another, and the fact that until very recently (the last quarter century) all versions of the scripts were only produced by hand writing. Undoubtedly, the cultural practice of passing the script from father to son also contributed to the diversification—over generations personal differences of style become embedded as communal differences, and eventually different communities do not recognize each other’s writing.

A central system of education can reverse this trend and establish standards, and the Vietnamese government made some efforts in this direction in the past. About 45 years ago, the Vietnamese government proposed that the Tai communities in their country should agree on a standard way of writing their languages. The result was the proposed *Thống Nhất*, or Unified Alphabet, of 1961. (See *Các Mẫu Tư Thái Ở Miền Tây Bắc Việt Nam*.) A revision of this proposal was published in 1966 (Lo Văn Mười, ໑໐ ພອ ວໍ າ໑໑). I am aware of some usage of the Unified Alphabet in Vietnam. (*Giới Thiệu Chương Trình Thái Học Việt Nam*, Hanoi National University, 1999). But I do not know whether any form of this alphabet is being taught in formal education.

I will henceforth refer to this version of the script as the “Unified Alphabet.”

Traditional Writing Outside the Unified Alphabet

Not everyone has had the opportunity to learn the Unified Alphabet. Some have adopted alphabets from surrounding cultures (Vietnamese, Lao, or Thai). But some still use their traditional writing. This would include the elderly, who learned to read and write before the Unified Alphabet was introduced, and those Tai communities outside of Vietnam, including the Tai Daeng of Laos, the Thai Song of Thailand, Tai Dam communities in Laos, the United States, and France, and many smaller communities. All writing samples that I have seen from communities outside of Vietnam continue to use the traditional forms of the writing. It is my desire to include all of these communities in this proposal.

History of Technology and the Tai Scripts

Around 1940 or 1950, Rev. Jean Fune had a typewriter made for the Tai Dam script. About 1980 Tai Dam immigrants to the United States also had a typewriter built for their script.

The first efforts to computerize the Tai Dam script employed bit map technology. Researchers at Northern Illinois University and SIL International made independent efforts in that direction in the early 1980s. SIL produced a bit-map font for laser printer between 1987 and 1988. In the mid 1990s they converted that font to their present Tai Heritage True Type set of fonts.

As noted above, in the 1950s researchers in Vietnam designed a standard Unified Alphabet intended for use by all of the Tai languages in that country. In the 1990s they became interested in computerizing that form of the alphabet, and have produced a PUA encoded font for it. (*Giới Thiệu Chương Trình Thái Học Việt Nam* and *Workshop on the Preservation and Digitization of Tai Scripts*.)

Selection of the Unicode Character Set




Key Issues

Is it sufficient to encode only the Unified Alphabet?

This is the most crucial question to be answered. My conclusion is “No, it is not sufficient,” for the following reasons.

1. As noted above, not everyone one can read the Unified Alphabet. Some communities will try to continue using their traditional form of the script.
2. One possible solution is to encode only the Unified Alphabet, and then to make language-specific fonts for each of the languages which reflect their traditional form. Thus, a Tai Dón person would use a Tai Dón font, and a Tai Dam person would use a Tai Dam font, but they would have the same encoding.

This would have the effect of introducing a phonemic-based encoding. It might appear attractive, because it could enhance communication between language communities. Something typed in the Tai Dam tradition could be converted to the Tai Dón tradition just by changing fonts. Difference in language would still remain, but differences in the alphabet could be removed.

Alas, I do not think it would work as advertised. Note the example cited above, under **Correlation of the Character Sets Between Languages**, showing the difference between Tai Dam and Tai Dón in how they use the low-tone characters /m/, /p/, and /f/. In a system using a phonemic-base encoding and a language-specific font, the same glyph, , would be used for the character UNIFIED TAI LETTER PO LOW in a Tai Dam font and for UNIFIED TAI LETTER MO LOW in a Tai Dón font. If a Tai Dón person wanted to type /me4/, ‘wife’, but accidentally used a Tai Dam font, the glyph  would be incorrectly entered and stored as the character UNIFIED TAI LETTER PO LOW. Then, if someone changed the document to use the correct font, the character would be displayed as , changing the meaning of the text. This is incorrect behavior. The bottom line is that in such a system, the encoding is interpreted by the font, which is a gross violation of the spirit of Unicode.

Therefore, I have concluded that it is necessary to include in the standard for the Unified Tai Script the Unified Alphabet of 1961 plus any characters that are unique to the traditional forms of the four languages with which we are concerned.

Should Tai Daeng be included?

As noted above, the Tai Daeng character set has less than a 50% correlation to those of the other languages. Should it be included as part of the Unified Tai Script, or should it be encoded as its own script?

Although it has many unique characters, the basic form and mechanics of the script are similar to that of the other languages. Consequently, I am proposing that it be considered part of the same script and be encoded with the other languages. I welcome other opinions of this matter, especially from those who have more experience with Tai Daeng.

Should Thai Song be included?

I feel that the differences between the Thai Song and the other script dialects are mostly stylistic. If the stylistic differences are set aside, the correlation to the Tai Dam is quite high. My conclusion is that it should be included in the Unified Tai Script.

Criteria for Selecting Characters for the Standard

Not all of the available data are of the same quality. Many are samples taken from various scholarly works. Some symbols appear only in a single sample. Unfortunately, it is impossible to tell how widely used such symbols are, whether they represent a character that is in general usage in a community or a personal style. Consequently, I do not think it is wise to include a form based on its use in a single data source. In the end, I focused my selections on those sources which reflect either published texts or a published font. This yielded five sources:

- The 1961 Unified Alphabet and the 1966 revision. (“Các Mẫu Từ Thái Ở Miền Tây Bắc Việt Nam,” 1961, and ‘Lo Văn Mười, 1966) The preliminary Viet Thai proposal made by Michael Everson was based on these.
- The SIL Tai Heritage font, representing the Tai Dam language. A large body of published text is available which uses this character set, in both computer and hand written form.
- The Nhim and Donaldson data, representing Tai Dón. This is based on the character chart at the beginning of their 1970 dictionary, and is supported by published text which uses the same character set.
- Aam and Aanu’s Tai Daeng Primer, representing Tai Daeng. This is the only book published in Tai Daeng that I have found.
- The Song Petburi font, representing Thai Song. I do not have any texts published in Thai Song, but the font is posted on the web at <http://www.seasite.niu.edu/tai/TaiDam/index.htm>.

Comment on Character Images Used in This Paper

The representative characters in the proposed character set (see the file *Tai Script – char set*) are a mixture of two different fonts and a variety of scanned images. The styles of the characters are consequently inconsistent. A comprehensive font showing representative characters in a consistent style is awaiting the services of a type designer, but will be included with the final proposal.

Transcription of Character Names

As far as possible, I have tried to assign Unicode names that reflect the native names. But Unicode names can only use the upper case Latin letters A-Z, the digits 0-9, SPACE, and HYPHEN. Transliterating the native names into this limited character set poses some challenges. Some of the more difficult points are naming:

- aspirated velar stop and velar fricatives—Vietnamese-based Romanizations will usually use ‘kh’ for either the aspirated velar stop or the velar fricative. Nhim and Donaldson used ‘kh’ for the aspirated stop and ‘khh’ for the fricative. I have followed their practice.
- unaspirated vs. aspirated palatal affricate—Nhim and Donaldson used ‘ch’ for the unaspirated and ‘ts’ for the aspirated, but Martini used ‘c’ for the unaspirated and ‘ch’ for the aspirated. I chose the latter, as it seems less confusing, and is also consistent with the Lao name, “LAO LETTER CO.”
- alveopalatal nasals—I chose to use ‘nh’ following the practice of the Vietnamese alphabet, Nhim and Donaldson, and Fippinger.
- palatal semivowel—Vietnamese uses ‘d’ for this sound, but this would be very confusing for westerners. Another option would be the phonetic symbol ‘j’. But I chose to follow the lead used for the name of the corresponding Lao character, “LAO LETTER YO.”
- vowels—It is particularly difficult to come up with a good transliteration scheme for the vowels. I use ‘ue’ for the high central vowel /i/, ‘eh’ for the low front /ɛ/, and ‘uh’ for the mid central /ə/.

I am open to any suggestions for improving the transliteration scheme.

Rationale for Disunification of Characters and Other Comments

◌ UNIFIED TAI LETTER KO LOW (xx01) vs. ◌ UNIFIED TAI LETTER ALTERNATE KO LOW (xx2F) vs. ◌ UNIFIED TAI LETTER TAI DON KO LOW (xx4A)

The UNIFIED TAI LETTER KO LOW contrasts with UNIFIED TAI LETTER ALTERNATE KO LOW, the former being used by the Unified Alphabet, the latter being used by the traditional forms of Tai Dam and Thai Song. They are distinct symbols used by distinct populations for the same value.

The former symbol was chosen by the Unified alphabet, but it was innovative, in that it does not appear in any of the other data samples. The 1966 Revised Alphabet seemed to recognize this, and chose to use the second, more traditional symbol. But later publications reverted back to the innovative symbol. (E.g. “Giới Thiệu Chương Trình Thái Học Việt Nam.”) In my intelligibility survey, only one of three Tai Dam readers recognized the innovative form as a KO. Since use of the innovative form seems to already be established, I include it among the symbol for the Unified Alphabet. The other symbol, UNIFIED TAI LETTER ALTERNATE KO LOW, is required for writing the traditional forms of Tai Dam and Thai Song.

UNIFIED TAI LETTER TAI DON KO LOW contrasts with UNIFIED TAI LETTER KO LOW. They are distinct symbols used by distinct populations for the same value.

UNIFIED TAI LETTER TAI DON KO LOW contrasts with UNIFIED TAI LETTER ALTERNATE KO LOW in the traditional Tai Don script. The former is used by traditional Tai Don for low /k/. The latter is used by traditional Tai Don for low /x/.

3 UNIFIED TAI LETTER KHO LOW (xx03) vs. 3 UNIFIED TAI LETTER KHHO LOW (xx05)

These two characters contrast in the Unified Alphabet. The contrast is formed by the down stroke at the beginning of the first character.

Most Tai Dam dialects and one Tai Don dialect (from Mường So) use the second symbol as a velar fricative (KHHO LOW). Other Tai Dam dialects or a Tai Don dialect from Lồng Liền use the first symbol for the same purpose. The Tai Daeng from Laos use a symbol similar to the first for its low class aspirated velar.⁵ But none of these demonstrate a contrast. In my own intelligibility survey, two of three Tai Dam readers regarded the symbols as equivalent to each other; the third did not identify the first symbol. Based on their usage in the traditional scripts, one could argue that the two symbols are a stylistic variation of the same character. But they are contrastive in the Unified Alphabet, the first representing an aspirated velar and the second a velar fricative. So it is necessary to include both symbols as separate characters.

“h plus low class consonant” ligatures

UNIFIED TAI LETTER NGO HIGH (xx06)


UNIFIED TAI LETTER NHO HIGH (xx0E)

UNIFIED TAI LETTER NO HIGH (xx16)

UNIFIED TAI LETTER MO HIGH (xx20)

UNIFIED TAI LETTER VO HIGH (xx26)

All of these high class consonants are formed from ligatures of “h” + “low class consonant”. Thus, these could be classified as presentation forms. However, I would disunify them and give them distinct code points for the following reasons:

- For backwards compatibility—these forms were given distinct code in previous fonts. See the Tai Heritage font.
- Encoding these consonants as “h” + “low class consonant” may create ambiguities in the data. Consider the sequence . Without considering semantic information, this sequence could be parsed as /hiə1 ma:4/ or as /maw1/.

⁵ The form of the Tai Daeng symbol is different enough that one might argue that it should also be disunified, but I have insufficient data to sustain that argument.

✓ UNIFIED TAI LETTER CO HIGH (xx08)

There are two common styles for this character, as shown by the form ✓ and the alternate form 𑜀 in the Tai Heritage font. My observation while working with the Tai Dam in Iowa, starting with the publishing hand-written books through the development of the Tai Heritage font, and confirmed in my intelligibility survey, is that these two glyphs are in free variation among the Tai Dam. Virtually every Tai Dam reader recognizes either one as a CO. I do not have this degree of experience with the other Tai languages, but I have never found these two glyphs to be in contrast.

While the two glyphs were given separate code points in the Windows Tai Heritage font, this was done primarily because of the limitations in technology. I see no reason to separate the two glyphs into separate characters. They should be unified as a single character.

𑜄 UNIFIED TAI LETTER YO HIGH and 𑜅 UNIFIED TAI LETTER YO LOW

Very unusual characters among the Tai Daeng, Tai Dam, and Tai Dón. Their use in the Unified Alphabet was innovative. But they are used in the traditional Thai Song.



𑜆 UNIFIED TAI LETTER O HIGH 𑜇 and UNIFIED TAI LETTER O LOW

The ...O HIGH has several variant symbols: 𑜆 𑜇 . The same variations exist for the ...O

LOW, along with additional variations in how the tail is formed: 𑜆 𑜇 𑜈 . In my intelligibility survey, none of the three Tai Dam readers recognized the variants. But my memory is that I have seen these variations in Tai Dam literature in the past. They are probably stylistic variations, but not widely recognized. They never occur in contrast, and I think they can be handled as alternate forms of the same character.

UNIFIED TAI SYMBOL NEUNG (One)

This character should not be called a digit. It is a ligature for the word “one”, and is only used in text, never with other digits to form decimal numbers.

Among the variations for this character, the Thai Song form, , was recognized by two out of three Tai Dam readers in my survey. The Tai Daeng form  was not recognized by any of them. But the glyphs are not used in contrast, and can be treated as alternate glyphs of the same character.

UNIFIED TAI SYMBOL SAM (Repetition)

There are many variations to this symbol. Each dialect seems to have its own version. I have seen variations within the Tai Dam community. I tested the five variations below in my intelligibility survey. The first is from the Tai Heritage font. The second is from the Everson Viet Thai preliminary Unicode proposal, and is based on the 1966 revision of the Unified Alphabet. It is also similar to four samples from Lo Văn Mười (Lóng Lién, Cau Chau, Mường Tắc, and Mường Sang). The third is from Minot’s 1940 work, and is similar to the Mường Lay and Mường So sample of Lo Văn Mười. The fourth is from Aam and Aanu’s Tai Daeng primer. And the last is from the Song Petuburi Thai Song font.





There were three Tai Dam readers in my survey. The first symbol is from the Tai Heritage font and represents the style all three would normally use. The second symbol was recognized by all three readers whether in isolation or in the context of the other repetition symbols. The third and fourth symbols were not recognized by any of them. Reader One misidentified the fourth as the UNIFIED TAI VOWEL UA. But when I told him what the symbols were, he said he thought he’d be able to figure them out from context if he saw them in a text. The fifth symbol was identified by Readers Two and Three as the Lao repetition symbol (LAO KO LA) when seen in the context of the other repetition symbols. (Although the origin is Thai Song, it is essentially identical in appearance to the Lao character.) But Reader Three misidentified it as the UNIFIED TAI LETTER NHO LOW when seen isolation.

I’m not sure if one form can be found which is recognized by everyone, but I am hesitant to try encoding all these forms. The second form above seems to be the most universal. It was use in the largest number of Lo Văn Mười’s samples, and was recognized by all of the Tai Dam readers in my survey. The other styles can be implemented where needed as alternate glyph forms or in different style fonts. But I’m not sure how adequate that solution will be for Tai Daeng regions in Laos or among the Thai Song.



 UNIFIED TAI LETTER ALTERNATE CO LOW (xx30) vs.  UNIFIED TAI LETTER CO LOW (xx09)

The UNIFIED TAI LETTER ALTERNATE CO LOW contrasts with UNIFIED TAI LETTER CO LOW, being a unique symbol used by a different population for the same value. Visually, the distinction between the two is in the placement of the hook that forms the transition between the body and tail.

Compare UNIFIED TAI LETTER ALTERNATE CO LOW also with UNIFIED TAI LETTER KHHO LOW (xx05), in which the tail curves in the other direction, and with UNIFIED TAI LETTER THO HIGH (xx14), which lacks the transitional hook between the body and tail.

 UNIFIED TAI LETTER ALTERNATE SO HIGH (xx31) vs.  UNIFIED TAI LETTER SO HIGH (xx0C)

These characters contrast in being different characters used by different populations for the same value.

 UNIFIED TAI LETTER ALTERNATE NHO LOW (xx32) vs.  UNIFIED TAI LETTER NHO LOW

These two characters contrast in the traditional Tai Dón script. The former is used for low /ɲ/, the latter for low /kʰ/. The former is also used in Thai Song for low /ɲ/.

 UNIFIED TAI LETTER TAI DAENG NHO LOW (xx33) vs.  UNIFIED TAI LETTER NHO LOW (xx0F)

Both of these symbols are formed by adding a tail to UNIFIED TAI LETTER TO HIGH. The direction of the tail is the primary distinctive feature between them.

Normally, the direction that a tail curves is not the sole distinctive feature that differentiates one character from another. The only other exception to this is the pair UNIFIED TAI LETTER THO HIGH (xx14), and UNIFIED TAI LETTER THO LOW (xx15), and the latter character may be considered to be innovative.

In my legibility survey with the Tai Dam, all three readers failed to recognize xx15. They consistently identified it as xx14, focusing on the body of the character rather than the tail. So the body of the character seems to be more important than the direction of the tail.

Thus one might argue that the direction of the tail is never contrastive, and that the two symbols in question are stylistic variations of each other.

On the other hand, the directions of the tails are pretty rigid--I've never seen anyone vary them from their prescribed directions. Again, in my legibility survey, none of the Tai Dam readers identified xx33 as a NHO LOW. It is for this reason that I have concluded that the

proposed UNIFIED TAI LETTER TAI DAENG NHO LOW and UNIFIED TAI LETTER NHO LOW are contrasting characters used by distinct populations for the same value.

𑜀 UNIFIED TAI LETTER ALTERNATE THO LOW (xx34) vs. 𑜁 UNIFIED TAI LETTER THO LOW (xx15)

The UNIFIED TAI LETTER ALTERNATE THO LOW is formed by adding a subscript mark to the UNIFIED TAI LETTER TO LOW. It is used in a variety of Tai Daeng and Tai Dón dialects. The two are clearly distinct symbols used by distinct populations for the same value.

𑜂 UNIFIED TAI LETTER ALTERNATE FO LOW (xx35) vs. 𑜃 UNIFIED TAI LETTER PO LOW (xx1B), 𑜄 UNIFIED TAI LETTER PHO LOW (xx1D), and 𑜅 UNIFIED TAI LETTER FO LOW (xx1F)

The UNIFIED TAI LETTER ALTERNATE FO LOW contrasts with the other three characters listed here. The Unified Alphabet uses the 2nd, 3rd, and 4th symbols to represent low /p/, low /p^h/ and low /f/. Some Tai Dón dialects use the 4th, 3rd, and 1st symbols, respectively, to write these three sounds, while the second is used for low /m/. (A variation of the 1st symbol is also used to represent low /f/ in Thai Song.) Thus a clear four-way contrast exists between these symbols in Tai Dón.

𑜆 UNIFIED TAI LETTER ALTERNATE YO HIGH (xx36) vs. 𑜇 UNIFIED TAI LETTER YO HIGH (xx22)

The use of the UNIFIED TAI LETTER YO HIGH is limited to the Unified Alphabet and the Thai Song. The Tai Dam are fairly consistent with the form of the UNIFIED TAI LETTER ALTERNATE YO HIGH shown here. These two symbols contrast as distinct characters used by distinct populations for the same value.

The Tai Daeng have their own character for this sound. See UNIFIED TAI LETTER TAI DAENG YO HIGH.

The Tai Dón have considerable variation in the way the high /j/ is written. The most common forms are the UNIFIED TAI LETTER TAI DAENG YO HIGH, or a stylistic variation of the UNIFIED TAI LETTER ALTERNATE YO HIGH in which the hat is flat rather than pointed.

𑜈 UNIFIED TAI LETTER ALTERNATE YO LOW (xx37) vs. 𑜉 UNIFIED TAI LETTER YO LOW (xx23)

The use of the UNIFIED TAI LETTER YO LOW is limited to the Unified Alphabet and the Thai Song. The Tai Dam and Tai Dón mostly use the UNIFIED TAI LETTER

ALTERNATE YO LOW. These two symbols contrast as distinct characters used by distinct populations for the same value.

𑜏 UNIFIED TAI LETTER ALTERNATE LO LOW (xx38) vs. 𑜏 UNIFIED TAI LETTER LO LOW (xx25)

Tai Dam readers survey did not recognize the style of the UNIFIED TAI LETTER ALTERNATE LO LOW shown about, which is a common Tai Daeng style. When shown a Thai Song style of this letter, 𑜏, they did recognize it. However, the Thai Song style is similar to the Lao, and all of the readers in my survey were familiar with Lao. There is doubt as to whether a Tai Dam reader who is not familiar with Lao would recognize the Thai Song character. Thus, there is some uncertainty in my mind whether these symbols should be unified or disunified. Pending input from others, I am going to propose they be disunified, as distinct characters used by distinct populations for the same value.

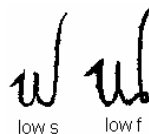
𑜏 UNIFIED TAI LETTER TAI DAENG KO LOW (xx39) vs. 𑜏 UNIFIED TAI LETTER KO LOW (xx01) and 𑜏 UNIFIED TAI LETTER ALTERNATE KO LOW (xx2F)

The UNIFIED TAI LETTER TAI DAENG KO LOW is a distinct character used by a distinct population with the same value as the other two.

It is also necessary to compare this character with the others that have a similar form. See the next paragraph.

𑜏 UNIFIED TAI LETTER TAI DAENG KO LOW (xx39) vs. 𑜏 UNIFIED TAI LETTER PO LOW (xx1B)

The rounding of the bowls in the form of UNIFIED TAI LETTER TAI DAENG KO LOW vs. the sharp angles of UNIFIED TAI LETTER PO LOW has never been found to be contrastive in any script samples. However, the presence of the head (loop) in the lower right corner of UNIFIED TAI LETTER TAI DAENG KO LOW is often a contrastive feature. Compare the samples below of a low /s/ and a low /f/ in Tai Daeng.



I believe this establishes that UNIFIED TAI LETTER TAI DAENG KO LOW is contrastive with UNIFIED TAI LETTER PO LOW in form as well as in use. If an attempt were made to unify UNIFIED TAI LETTER TAI DAENG KO LOW with UNIFIED TAI LETTER PO LOW, Tai Daeng readers would not recognize the resulting character because of the absence of the head.

𑜀 UNIFIED TAI LETTER TAI DAENG KO ALTERNATE (xx3A)

The UNIFIED TAI LETTER TAI DAENG KO ALTERNATE is formed by adding a subscript mark to UNIFIED TAI LETTER KO HIGH. Aam and Aanu's Tai Daeng primer (from Laos) lists this symbol in its character chart, but gives no phonemic value for it and doesn't use it in the text. It appears that the most likely reason for this is that it is used as an alternative form for different characters in different Tai Daeng regions. Thus, other authors list it either as a low /k/, a low /kh/, or a low /x/ in Mường Sang, Mốc Châu, Mường Tấc, and Houaphan. Thus, this is a distinct character used by a distinct population with various uses.

𑜁 UNIFIED TAI LIGATURE TAI DAENG KN (xx3B)

Aam and Aanu identify this character as a KN ligature. I understand this to mean an initial /k-/ and final /-n/. (The initial cluster /kn-/ never occurs in the Tai languages, and the character is used in the Tai Daeng word /kin1/, 'to eat'.) Although this character could be classified as a presentation form, leading to a decision not to encode it, I believe that it should be encoded. Implementing it as a presentation form would be difficult. The components that make up the ligature can occur in any of the following sequences:

- UNIFIED TAI LETTER KO HIGH + UNIFIED TAI LETTER NO LOW
- KO HIGH + vowel + NO LOW
- KO HIGH + tone + NO LOW
- KO HIGH + vowel + tone + NO LOW

Furthermore, if the author does not use inter-word spacing, the rendering system would have to determine whether the NO LOW was the final consonant of a syllable, or the initial consonant of the next syllable. This would require the rendering system to employ a very sophisticated parser, with a high possibility of error. There may be ambiguities that the parser could not resolve. I think it is better to just encode the character.

𑜂 UNIFIED TAI LIGATURE TAI DAENG KW (xx3C)

In most instances, the labialized consonants are written using the sequence “consonant + UNIFIED TAI LETTER VO LOW”. This is the only instance I know of that uses a ligature to write a labialize consonant. This form needs to be encoded as a character. The sequence 𑜀𑜂 is ambiguous. It can be read either as /kew/ or as /kwe/. Thus, if the glyph in question is categorized as a presentation form, a rendering engine would not be able parse this sequence and know whether or not to display the presentation form. (Readers can resolve the ambiguity from semantic context, but a parser could not.)

𑜃 UNIFIED TAI LETTER TAI DAENG KHO HIGH (xx3D) vs. 𑜄 UNIFIED TAI LETTER KHO HIGH (xx02)



These are two distinct characters used by distinct populations for the same value.

 UNIFIED TAI LETTER TAI DAENG NGO HIGH (xx3E) vs.  UNIFIED TAI LETTER NGO HIGH (xx06)

The UNIFIED TAI LETTER TAI DAENG NGO HIGH contrasts with other NGO HIGH forms as a unique character used by a distinct population for the same value.

 UNIFIED TAI LETTER TAI DAENG NGO LOW (xx3F) vs.  UNIFIED TAI LETTER NGO LOW (xx07)

The UNIFIED TAI LETTER TAI DAENG NGO LOW contrasts with other NGO LOW forms as a unique character used by a distinct population for the same value.

 UNIFIED TAI LETTER TAI DAENG SO LOW (xx40) vs.  UNIFIED TAI LETTER SO LOW (xx0D)

The heads, or loops, on Tai characters are distinctive features. Thus the absence of the head on UNIFIED TAI LETTER TAI DAENG SO LOW makes it a distinct character, regardless of the number of strokes in the body of the character.

UNIFIED TAI LETTER TAI DAENG SO LOW contrasts with UNIFIED TAI LETTER SO LOW as a distinct character used by a distinct population for the same value.

 UNIFIED TAI LETTER TAI DAENG NHO HIGH (xx41)

UNIFIED TAI LETTER TAI DAENG NHO HIGH is a ligature of UNIFIED TAI LETTER HO HIGH plus UNIFIED TAI LETTER TAI DAENG NHO LOW. It should be encoded for the same reason the other “h plus low class consonant” are encoded. See above.

 UNIFIED TAI LETTER TAI DAENG DO HIGH (xx42) vs.  UNIFIED TAI LETTER DO HIGH (xx10)

The UNIFIED TAI LETTER TAI DAENG DO HIGH contrasts with UNIFIED TAI LETTER DO HIGH because the distinctive features for the two characters are different. The UNIFIED TAI LETTER DO HIGH, used in Tai Dam and Thai Song, is distinguished from the UNIFIED TAI LETTER KO HIGH by the absence of the down stroke at the beginning of the letter. But the UNIFIED TAI LETTER TAI DAENG DO HIGH is distinguished from the UNIFIED TAI LETTER KO HIGH by the inward curl at the lower right.

Tai Dam readers misidentify the UNIFIED TAI LETTER TAI DAENG DO HIGH as a high /k/. I think there is a high probability that Tai Daeng readers would misidentify UNIFIED TAI LETTER DO HIGH.

 UNIFIED TAI LETTER TAI DAENG PO LOW (xx43) vs.  UNIFIED TAI LETTER PO LOW (xx1B)

The curve vs. the sharp form of the bottom of the characters is not significant—there are a number of characters which show this variation between dialects. The feature which is in question is the looped tail of UNIFIED TAI LETTER TAI DAENG PO LOW. The only place a looped tail is unambiguously contrastive is in distinguishing between the characters UNIFIED TAI LETTER NGO HIGH and UNIFIED TAI LETTER NO HIGH. The former is formed from a ligature of UNIFIED TAI LETTER HO HIGH and UNIFIED TAI LETTER NGO LOW, but a simple ligature of these two would look too much like the NO HIGH, so a looped tail is added to the top to disambiguate it.

A looped tail also occurs on the low /b/ and the low /p/ in several script dialects, including Tai Daeng, but never in contrast to another character without it. If the looped tail were omitted from Tai Daeng low /p/, it could still be distinguished from the low /s/ by the length (or absence) of the tail. However, some features are difficult to maintain in hand writing. It is easy for a scribe to make the tail of an /s/ too short, and the tail of a /p/ too long. It is possible that scribes started adding the looped tail to the low /p/ to prevent ambiguity between it and the low /s/. Thus the loop may be contrastive in the mind of Tai Daeng readers.

Finally, heads, loops, and points in Tai characters are usually distinctive features. Thus, I would not expect a Tai Daeng reader to recognize the version of the low /p/ from Unified Alphabet, because of the absence of the loop. Nor would I expect a Tai Dam reader to recognize the Tai Daeng version, because of the presence of the loop. The latter expectation was verified by my legibility survey with Tai Dam readers.

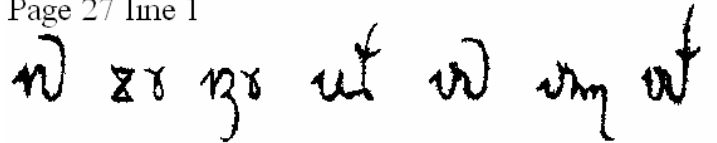
In conclusion, I consider these to be contrasting characters used by distinct populations for the same value.

 UNIFIED TAI LETTER TAI DAENG FO ALTERNATE (xx44)

I have not been able to establish the usage for this character. Aam and Aanu include it in their consonant chart, and gloss it with the Lao characters ‘ ’ (low /f/ + /j/). Such a cluster does not occur syllable initial in the Tai languages, so one possibility is that it is intended as an initial consonant + final consonant ligature.

However, from the text on page 27 of Aam and Aanu, it appears otherwise.

Page 27 line 1



 high high low high low low "ho yo"
 /kw/ /kʰw/ /kʰw/ /fj/ ? /v/ /ɲ/


Page 27 line 5



 /ɬaw3/ /luen1/
 Lord great

Text from Aam and Aanu's Tai Daeng Primer

On line 1, the TAI DAENG FO ALTERNATE is listed with several other complex consonants. Then on line 5 it forms a syllable with the vowel /aa/. This is followed by a three word phrase. If the middle word of this phrase is /faa6/, ‘sky’, the three word phrase is a name sometimes used for ‘God’. This indicates that the TAI DAENG FO ALTERNATE is an alternative symbol for writing high /f/.

This usage of the TAI DAENG FO ALTERNATE may be parallel to the usage of the UNIFIED TAI LETTER PHO LOW in traditional Tai Dam. There is no /p^h/ in the Tai Dam language, but UNIFIED TAI LETTER PHO LOW is used for writing just one word, , /fiə4/, ‘official’.

I cannot resolve the question of the usage of this character without the aid of a Tai Daeng reader. I think the suggested name is adequate. I have included the character here in an effort to establish a complete character set, even though I am not certain of the usage.

 UNIFIED TAI LETTER TAI DAENG YO (xx45)

 UNIFIED TAI LIGATURE TAI DAENG HO YO (xx46)

In the Ferlus 1988 Tai Daeng sample and in the *Các Mẫu Tư Thái...* from Mường Sang, the UNIFIED TAI LETTER TAI DAENG YO is used for a low /j/. In a number of other samples from Tai Dam, Tai Dón, and Tai Daeng regions, it is used for a high /j/.

In my best Tai Daeng manuscripts, the Tai Daeng primer from Aam and Aanu, and the untitled manuscript, there is some uncertainty over which of these characters is the high /j/ and which is the low /j/. The gloss on the UNIFIED TAI LETTER TAI DAENG YO, along with the long tail, would lead one to conclude that it is the high /j/. But the use of a “ho plus...” ligature points the other way. In all other cases, “ho plus...” ligatures are used to change a low consonant into a high consonant. Thus, the UNIFIED TAI LETTER


TAI DAENG YO should be the low consonant, and UNIFIED TAI LIGATURE TAI DAENG HO YO should be the high consonant. This latter position is reinforced by the fact that in both Aam and Aanu’s primer and in the untitled Tai Daeng manuscript, the UNIFIED TAI LETTER TAI DAENG YO is used for a syllable final /j/. In all other Tai dialects, the low /j/ is used syllable final.

Because of the variation in how the UNIFIED TAI LETTER TAI DAENG YO is used, I have chosen not to include the consonant class in the character names. The use of the “ho plus” ligature to distinguish the consonant class instead of the length of the tail sets these two characters off from other characters used to write /j/. Consequently, they should be disunified from UNIFIED TAI LETTER YO HIGH and UNIFIED TAI LETTER YO LOW, and from UNIFIED TAI LETTER ALTERNATE YO HIGH and UNIFIED TAI LETTER ALTERNATE YO LOW.

 UNIFIED TAI LETTER TAI DAENG VO HIGH (xx47) vs.  UNIFIED TAI LETTER VO HIGH (xx26)

Both of these characters are ‘h’ + ‘v’ ligatures, but the Tai Daeng version forms the ‘v’ in a counterclockwise direction and extends the height of the ‘v’. The result is that the character evolves into a significantly different form. None of the Tai Dam readers in my survey recognized the Tai Daeng form as a high /v/.

These are two unique characters used by distinct populations for the same value.

 UNIFIED TAI LETTER TAI DAENG VO LOW (xx48) vs.  UNIFIED TAI LETTER VO LOW (xx27)

The UNIFIED TAI LETTER TAI DAENG VO LOW is a unique character used by a distinct population for the same value as the UNIFIED TAI LETTER VO LOW.

 UNIFIED TAI LETTER TAI DAM THO LOW (xx49) vs.  UNIFIED TAI LETTER THO LOW (xx15)

The UNIFIED TAI LETTER THO LOW has the same body as the UNIFIED TAI LETTER THO HIGH—the distinction between them is in the direction of the tail. But the UNIFIED TAI LETTER TAI DAM THO LOW has a different body. In my legibility survey, 0 of 3 Tai Dam readers recognized the UNIFIED TAI LETTER THO LOW; one of them confused it for the UNIFIED TAI LETTER THO HIGH, indicating the form of the body may be more important than the direction of the tail. Thus, the UNIFIED TAI LETTER TAI DAM THO LOW contrasts with UNIFIED TAI LETTER THO LOW as a distinct character used by the traditional Tai Dam script for the same value.

 UNIFIED TAI LETTER TAI DON KO LOW (xx4A)

See UNIFIED TAI LETTER KO LOW (xx01), above.

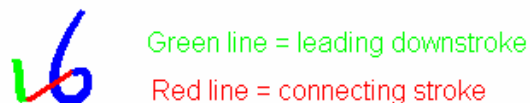
𑜏𑜤 UNIFIED TAI LETTER TAI DON NGO HIGH (xx4B) and 𑜏 UNIFIED TAI LETTER TAI DON DO HIGH (xx4D) vs. 𑜏 UNIFIED TAI LETTER NO HIGH and 𑜏 UNIFIED TAI LETTER NO LOW (compare also 𑜏 UNIFIED TAI LETTER NGO LOW)

Although the two Tai Dón letters bear a resemblance to the two letters from the Unified alphabet, I believe that are in fact different characters. Note that there is a clear contrast between the UNIFIED TAI LETTER TAI DON NGO HIGH and the UNIFIED TAI LETTER NO HIGH in some Tai Dón dialects. The principal difference seems to be that where the UNIFIED TAI LETTER NO HIGH has a small loop or point for the head in the lower right corner of the character, the UNIFIED TAI LETTER TAI DON NGO HIGH has a larger loop in that position. The same contrast seems to exist between the UNIFIED TAI LETTER TAI DON DO HIGH and the UNIFIED TAI LETTER NO LOW. My analysis is that the UNIFIED TAI LETTER TAI DON NGO HIGH is in fact a ligature of UNIFIED TAI LETTER HO HIGH + UNIFIED TAI LETTER NGO LOW (𑜏 + 𑜏) with a connecting stroke between them.



The addition of an “h” in front of a consonant is a common device in Thai languages that is used for changing a high tone consonant into a low tone consonant.

I believe that the UNIFIED TAI LETTER TAI DON DO HIGH is derived from the UNIFIED TAI LETTER NGO LOW in a similar fashion, but with the first part of the letter being a single down stroke in contrast to an ‘h’:



Thus these letters are distinct characters from the UNIFIED TAI LETTER NO HIGH and UNIFIED TAI LETTER NO LOW.

𑜏 UNIFIED TAI LETTER TAI DON SO LOW (xx4C)

UNIFIED TAI LETTER TAI DON SO LOW contrasts with 𑜏 UNIFIED TAI LETTER ALTERNATE FO LOW (xx35). The distinguishing feature is the double head in the former. They are different symbols used by the same community (Tai Don community) for different values.

UNIFIED TAI LETTER TAI DON SO LOW contrasts with 𑜏 UNIFIED TAI LETTER SO LOW (xx0D). They are different symbols used by different communities for the same value.

UNIFIED TAI LETTER TAI DON FO LOW (xx4E)

Among the Tai Dón characters listed by Nhim and Donaldson are the five listed in the first column of the following table.

Tai Don graph	Tai Don usage	Tai Dam graph	Tai Dam usage	Thai Song graph	Thai Song usage	Proposed Unicode name
	low /m/		low /p/			UNIFIED TAI LETTER PO LOW
	low /p ^h /		low /f/ in /fia4/			UNIFIED TAI LETTER PHO LOW
	high /s/		low /f/ ⁶		low /f/	UNIFIED TAI LETTER ALTERNATE FO LOW
	low /p/ ⁷		low /f/			UNIFIED TAI LETTER FO LOW
	low /f/					UNIFIED TAI LETTER TAI DON FO LOW

Because some dialects of Tai Dón use the UNIFIED TAI LETTER FO LOW for a low /p/ and the UNIFIED TAI LETTER ALTERNATE FO LOW for a high /s/, the UNIFIED TAI LETTER TAI DON FO LOW is required.

UNIFIED TAI LETTER TAI DON MO HIGH (xx4F)

Tai Dón uses UNIFIED TAI LETTER PO LOW as a low /m/. The corresponding high /m/ is formed by adding an “h” to that character, yielding the UNIFIED TAI LETTER TAI DON MO HIGH.

The UNIFIED TAI LETTER TAI DON MO HIGH contrasts with UNIFIED TAI LETTER MO HIGH in that the former is a ligature of “h” plus UNIFIED TAI LETTER PO LOW, while the latter is a ligature of “h” plus UNIFIED TAI LETTER MO LOW.

This form should be encoded even though it is a ligature to avoid ambiguity and for consistency with the other the other “h plus” ligatures. See “**h plus low class consonant**” ligatures above.

⁶ Tai Dam usage is limited to old samples from Nghia-lô and Van-bu (Finot 1917)

⁷ In the dialect represented by Nhim & Donaldson 1970. In some Tai Dón dialects, the usage is the same as the Tai Dam.

UNIFIED TAI LETTER THAI SONG KHO HIGH (xx50)

This character contrasts with UNIFIED TAI LETTER KHO HIGH (𑜀, xx02). It is a unique glyph used by a distinct population for the same value as the other.

Spacing vowels and tone marks vs. combining vowels and tone marks

The Unified Alphabet uses spacing characters for all vowels and tones, while the traditional scripts use combining marks for many vowels and tones. Because of the different behavior of the characters from the Unified Alphabet and the traditional scripts, both must be encoded.

Where the traditional script uses a spacing character, one character can serve both the Unified Alphabet and the traditional way of writing.

Vowel + final consonant ligatures

There are a number of ligatures formed from vowels and final consonants in the Tai scripts. These include:

- UNIFIED TAI VOWEL UHW – 𑜁 (xx62) = /-əw/
- UNIFIED TAI VOWEL AY – 𑜂 (xx63) = /-aj/
- UNIFIED TAI VOWEL AN – 𑜃 (xx64) = /-an/
- UNIFIED TAI VOWEL AM – 𑜄 (xx65) = /-am/
- UNIFIED TAI VOWEL TAI DON AT – 𑜅 (xx75) = /-at/
- UNIFIED TAI VOWEL LOW TONE AA – 𑜆 (xx76) = /-a:#/ tone 3

The case for encoding the first four is quite clear. They already exist in legacy fonts, and they should be encoded for backwards compatibility. In addition, if one attempted to deal with these as presentation forms, the parser would have to deal with a possible tone mark between the components of the ligature, the parser would have to analyze whether the consonant was the last consonant of the current syllable or the first consonant of the next syllable, and the rendering engine would have to reorder the glyphs in the first two cases.

With regards to the UNIFIED TAI VOWEL TAI DON AT, I don't know of any font which currently includes this character. The problem of a parser having to identify the components of the ligature when a tone mark intervenes between them is problematic. Because of collocation restrictions, Tai Dón words that end in a stopped consonant do not require a tone mark, but it is possible an author would write one any way. The parser would still have to analyze whether the letter 'd' was a final /t/ of the current syllable or the initial /d/ of the next syllable. I think it is better to avoid those potential problems and encode the ligature.

The last of these characters, UNIFIED TAI VOWEL LOW TONE AA, is the most unusual, in that it involves a tone rather than a final consonant. Of the forms shown here, this one could most easily be dealt with as a presentation form. But again, there may be some danger in trying to treat it as a presentation form. Not all authors mark tone, and if the tone is missing, a parser would not be able to identify the need for the ligature. So once again, I would encode the ligature.

Tai Daeng vowels

Tai Daeng has a significantly different inventory of vowels from that used by the other three languages. This is because Tai Daeng has a full set of long and short vowels, as well as some significant differences in form.

UNIFIED TAI VOWEL TAI DAENG A (xx6A)



This symbol is used by Aam and Aanu for marking certain short vowels, as shown in the table below. I have not found it in any other Tai Daeng text (or any text from the other Tai languages). It is most likely borrowed from Lao. In that Lao is closely related to the Tai languages and scripts, and the proximity of the Tai Daeng to Lao, it seems likely that some Tai Daeng communities will want to continue using this character for marking vowel length. I recommend encoding it.

IPA	long form	short form
iə	+ .	+ . ~
e	++ .	++ . ~
ɛ ~
a	.l	. ~ or .
ua	.u	.u ~
o	l.	l. ~ or .
ɔ	.a	.a ~

Use of  to mark vowel length in Tai Daeng

UNIFIED TAI VOWEL TAI DAENG II (xx6B)

The UNIFIED TAI VOWEL TAI DAENG II is required for writing the long vowel /ii/ in Tai Daeng.

A key question is whether the UNIFIED TAI VOWEL COMBINING I (, xx54) corresponds to the traditional form of the Tai Daeng short /i/ () or the long /i:/ shown above. Although in the example for the UNIFIED TAI VOWEL COMBINING I shown here, taken from the Tai Heritage font, the right side of the character has a slight curve, in most writing samples both sides are straight, like the Tai Daeng short /i/. In my legibility survey, the Tai Dam readers identified the short /i/ form as an /i/, but did not identify the

long /i:/ form. Therefore, in my judgment, UNIFIED TAI VOWEL COMBINING I corresponds to the Tai Daeng short /i/, and the Tai Daeng long /i:/ is the one that needs to be added to the code set.

☉ UNIFIED TAI VOWEL TAI DAENG UE (xx6C)

☉ UNIFIED TAI VOWEL TAI DAENG UUE (xx6D)

Compare the UNIFIED TAI VOWEL TAI DAENG UE with the UNIFIED TAI VOWEL COMBINING UE (☉, xx56). While I have considered these symbols to be distinct characters, there is a possibility they should be unified.

The UNIFIED TAI VOWEL TAI DAENG UUE is required for writing the Tai Daeng long /iiə/.

☉ UNIFIED TAI VOWEL TAI DAENG U (xx6E)

☉ UNIFIED TAI VOWEL TAI DAENG UU (xx6F)

These two symbols are used in Tai Daeng, and the first is used in Thai Song. Compare them with the UNIFIED TAI VOWEL COMBINING U (☉ – xx58).

In my intelligibility survey, Reader Three identified the UNIFIED TAI VOWEL TAI DAENG U as a /u/ in both the Tai Daeng and Thai Song samples. Reader One identified the UNIFIED TAI VOWEL TAI DAENG U as a /u/ in the Thai Song sample but not in the Tai Daeng sample. Reader Two did not complete this part of the survey.

On the other hand, this is a very limited data sample. I believe that most Tai Dam readers would find the use of the UNIFIED TAI VOWEL TAI DAENG U for /u/ to be unusual, and that Tai Daeng readers would find the use of the UNIFIED TAI VOWEL COMBINING U for short /u/ unusual. It might be possible to have these two as alternate forms of the same character, but I think it is better to disunify them.

☉ UNIFIED TAI VOWEL TAI DAENG EE (xx70)

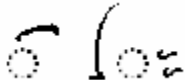
Compare the traditional forms of the Tai Daeng long /ee/, Tai Daeng long /iiə/, and Tai Dam /iə/, shown here:

Tai Daeng /ee/	☉ ☉
Tai Daeng /iiə/	☉ ☉ ☉
Tai Dam /iə/	☉

The Tai Daeng long /iiə/ and Tai Dam /iə/ are represented by the UNIFIED TAI VOWEL COMBINING IA (xx5E). The UNIFIED TAI VOWEL TAI DAENG EE, then, contrasts with the UNIFIED TAI VOWEL COMBINING IA as a distinct character used by a distinct population for a distinct value.

◌ UNIFIED TAI VOWEL TAI DAENG SHORT O (xx71)

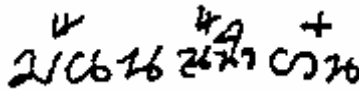
Aam and Aanu list two forms for the Tai Daeng short /o/,

Two forms for Tai
Daeng short /o/ 


but I have not found the second one used in any other document. The first form is used in the untitled Tai Daeng manuscript, and in Lafont's sample from southeast Houaphan. As compared to UNIFIED TAI VOWEL O (xx5C), it is a unique character used to write a distinct sound.

◌ UNIFIED TAI VOWEL TAI DAENG UUA (xx72)


To show contrast with UNIFIED TAI LETTER NO LOW --  (xx17), consider this phrase from page 11 of Aam and Aanu:




The first word, /mu:ən6/, contains both UNIFIED TAI VOWEL TAI DAENG UUA (second character) and UNIFIED TAI LETTER NO LOW (third character).

Contrast must also be established with UNIFIED TAI VOWEL UA – ◌ (xx61). The fact that the Tai Daeng form represents a long vowel is not in itself enough to establish contrast. The Tai Dam and Tai Dón do not have length contrast on the vowel /uə/. Thus, the UNIFIED TAI VOWEL UA has no length associated with it—it could be equivalent to either the long /u:ə/ or the short /uə/ in Tai Daeng. However, Tai Dam readers do not recognize the UNIFIED TAI VOWEL UA and the UNIFIED TAI VOWEL TAI DAENG UUA as equivalent. In my legibility survey, one reader thought the TAI DAENG UUA was a low /n/. A second did not recognize it at all. This indicates that the Tai Daeng form should be disunified from that used by the other languages.

◌ UNIFIED TAI VOWEL TAI DAENG SHORT UH (xx73)

This character is used for the short vowel /ə/ in Tai Daeng. It contrasts with the UNIFIED TAI VOWEL UHW – ◌ (xx62), which the Tai Daeng use for the long /ə:/.

◌ UNIFIED TAI VOWEL TAI DON A (xx74)

The straight UNIFIED TAI VOWEL COMBINING A – ◌ (xx51) is used by Tai Dam for /ə/ in open syllables and for /a/ in closed syllables. Some Tai Dón dialects use this

character as the vowel /ə/. They use the UNIFIED TAI VOWEL TAI DON A for the vowel /a/, creating a clear contrast between these two characters.

Also note that the style of the /a/ in Tai Daeng and Thai Song is closer to the Tai Dón version than the Tai Dam, although the line usually has a curve rather than a sharp angle, and they do not have any contrast with the straight line style.

☉ ✎ UNIFIED TAI VOWEL TAI DON AT (xx75)

See “**Vowel + final consonant ligatures**”, above

☉ ✎ UNIFIED TAI VOWEL LOW TONE AA (xx76)

See “**Vowel + final consonant ligatures**”, above

Additions for Vietnamese loan words

☉ UNIFIED TAI LETTER GO HIGH

☉ UNIFIED TAI LETTER GO LOW

☉ UNIFIED TAI LETTER RO HIGH

☉ UNIFIED TAI LETTER RO LOW


☉ UNIFIED TAI VOWEL RAISED A

These five characters not native to the Tai languages, but are wanted for writing Vietnamese loan words. In general, speakers of a language will adopt foreign words to their own phonology, and do not need extra symbols to write loan words. But much depends on the degree of assimilation of the loan word, the degree of contact between the societies, and whether the people speak the language from which the word was borrowed. In light of the close contact between Tai and Vietnamese speakers, the addition of these symbols is not unreasonable.

I am unsure of the best name for ☉ listed here as UNIFIED TAI VOWEL RAISED A. There is no native character in Tai that corresponds to it. It is equivalent to the Quốc Ngữ (Vietnamese) “â”, which is phonetically a /ə/ (i.e. a central vowel higher than /a/ but lower than /ə/).

Characters Not Encoded

The addition of several other vowels has been suggested in the past for purpose of writing Vietnamese loan words. In recent correspondence, Dr. Ngô Trung Viet has suggested that these are no longer considered necessary.

 vowel sign UYEE

 vowel sign UY

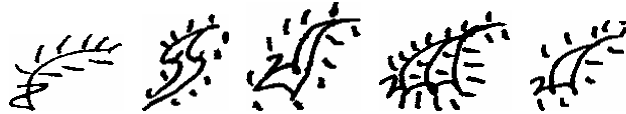
  vowel sign UEE; Quốc Ngữ “oe”

 Quốc Ngữ “uê”

The Tai languages also have a variety of ideographs, whose useage I have not been able to fully discern. The most common of these is the "ho hoi," which takes a variety of forms in script samples from different regions. The following symbol, for example, is part of the Chu Thai font, which was distributed with the proceedings of the *Workshop on the Preservation and Digitization of Tai Scripts*.



Lafont (1962) says that “ho hoi” means “love eternal”, and that the symbol is used in love letters. A second explanation comes from my personal correspondence with Dr. Ngo Trung Viet, who has told me “ho hoi” is only used at the beginning of a text. But on the other hand, it or symbols similar to it appear in many forms throughout the untitled Tai Daeng manuscript. Some examples, from pages 7, 11, and 15 of this text, are shown here:



The tendency of the Tai people to include one of these symbols with samples of their script indicates it is important to them. But it is not clear to me whether it carries lexical or grammatical meaning, and whether it is an orthographic symbol or a formatting device. I think eventually it should be encoded, but we need more information about it to firmly establish that.

I have also encountered some ideographic symbols which carry religious meaning. I have not examined them or included them in this study.

Bibliography

- ___ . “Các Mẫu Tử Thái Ở Miền Tây Bắc Việt Nam.” Internet:
<http://www.evertype.com/standards/tai/viet-thai-samples.pdf>
- ___ . *Conflict and Relocation*. <http://www.seasite.niu.edu/tai/TaiDam/article/a4.htm>.
- ___ . 1999. *Giới Thiệu Chương Trình Thái Học Việt Nam*. (การรู้หนังสือชนกลุ่ม
เผ่าต่าง ๆ ในภาคเหนือ; Introduction for Vietnam Programme of Thai Studies). Hanoi National
University, Hanoi.
- ___ . *Sip Hoc Chau Thai*. (Journal de l'Association A.R.T.F.) Orly, France. Cited by
<http://www.evertype.com/standards/tai/viet-thai-samples.pdf>.
- ___ . Song Petburi font. Internet: <http://www.seasite.niu.edu/tai/TaiDam/index.htm>.
- ___ . Untitled, undated manuscript in Tai Daeng.
- ອິນ and ອານຸ (Aam and Aanu). 1974. ການສຶກສາການອ່ານຂຽນພາສາໄຕດາງ (“be:p
hiə:n4 a:n2 lə? khiə:n1 nəŋ1 si:l taj4 deŋ1”, “Learning to Read and Write Tai Daeng”).
Vientiane.
- Baccam Don, Baccam Faluang, Baccam Hung, Dorothy Fippinger. 1989. *Tai Dam – English,
English – Tai Dam Vocabulary Book*. Summer Institute of Linguistics.
- Cam Trong. 2005. “Thai Scripts in Vietnam,” in *Workshop on the Preservation and Digitization
of Tai Scripts*. Hanoi, Vietnam.
- Donaldson, Jean and Jerold A. Edmondson. 1997. “A Preliminary Examination of Tai Tac,” in
Comparative Kadai, The Tai Branch, edited by Jerold A. Edmondson and David B. Solnit.
The Summer Institute of Linguistics and The University of Texas at Arlington. pp 235-266.
- Everson, Michael. Proposal for the Universal Character Set. Internet:
<http://www.evertype.com/standards/tai/viet-thai-comparisons.pdf>
- Ferlus, Michel. 1988. “Langues et écritures en asie du sud-est,” The 21st International
Conference on Sino-Tibetan Languages and Linguistics, University of Lund, Sweden.
- Ferlus, Michel. 1999. “Les dialectes et les écritures des Tai (Thai) du Nghệ An (Vietnam),”
Treizièmes journées de linguistique d'Asie orientale. Centre de recherches linguistiques sur
l'Asie orientale (EHESS-CNRS), 105 bd raspail, 75006 Paris.
- Ferlus, Michel. 2003 “L'intérêt linguistique du Húng Hóa Ký Lược de Phạm Thậ Dật,” Dix-
septièmes de Linguistiques d'Asie Orientale, Centre de Recherches Linguistiques sur l'Asie
Orientale (EHESS-CNRS).
- Finot, Louis, 1917. “Recherhes sur la literature laotienne,” *Bulletin de l'Ecole Française
d'Extrême-Orient*. 17(5).
- Fippinger, Jay and Dorothy. 1970. “Black Tai Phonemes with reference to White Tai.”
Anthropological Linguistics 12.3: 83-97.
- Lafont, Pierre-Bernard, 1962. “Les écritures 'Tay du Laos,” *Bulletin de l'Ecole Français
d'Extrême-Orient*.

- Gedney, William J. 1989. "A Comparative Sketch of White, Black and Red Tai." *Selected Papers on Comparative Tai Studies*, Michigan Papers on South and Southeast Asia no. 29, Center for South and Southeast Asia Studies, The University of Michigan.
- Lo Văn Mười (𑜋𑜰𑜫 𑜇𑜡𑜫 𑜏𑜢𑜤𑜰𑜫). 1966. Ép Sứ 'Táy Piên Peng. (𑜏𑜢𑜤𑜰𑜫 𑜇𑜡𑜫 𑜏𑜢𑜤𑜰𑜫 𑜇𑜡𑜫 𑜏𑜢𑜤𑜰𑜫), Learning the Revised Tai Alphabet.)
- Martini, Francois. 1954. "Romanisation des parlers 'Tay du Nord Vietnam." *Bulletin de l'Ecole Française d'extrême-orient*.
- Minot, Lieutenant. 1933. "Dictionnaire Français - Thay Blanc." Mưỡong Té.
- Minot, Georges. 1940. "Dictionnaire Tày Blanc-Français." *Bulletin de l'Ecole d'Extrême-Orient*, t. XL.
- Ngo Trung Viet. 2005. "ICT for Thai Ethnic Culture and Education Multilingual Projects," in *Workshop on the Preservation and Digitization of Tai Scripts*. Hanoi, Vietnam.
- Nhìm, Điều Chính and Jean Donaldson. 1970. *Pấp San Khhâm Pák Tãy-Keo-Eng* (Ngũ-Vụng Thái-Việt-Anh, Tai-Vietnamese-English Vocabulary). Saigon.
- Phan Anh Dung and Ngo Trung Viet. 2005. "Technical Design, Software for Inputting and Displaying Vietnam Thai Scripts," in *Workshop on the Preservation and Digitization of Tai Scripts*. Hanoi, Vietnam.
- Robert, R. 1941. *Notes sur les Tay Dèng de Lang Chánh (Thanh-Hoá, Annam)*. Institut Indochinois pour l'Etude de L'Homme, mémoire n' 1. Hanoi: Imprimerie d'Extrême-Orient.
- SIL. Tai Heritage font. Published by SIL International. Internet:
http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=SILTD_home
- Whitehouse, Ruth. 1975. *Phonemic Write-up, Lao Song Language*. Unpublished paper.