

Corrections to the Indic Syllabic Category for the Tai Tham Script

Proposer: Richard Wordingham

Date: 30 April 2017

Table of Contents

Introduction.....	1
Discussion.....	2
Alleged Final Consonants.....	2
Medial Consonants.....	5
Miscellaneous Marks.....	6
Proposal.....	9
Evidence.....	10
References.....	12
Appendix.....	12
Prescribed Syllable Structures Proscribed by the USE.....	12

Introduction

The Indic Syllabic Category property suffers the defect that the categories have never been properly defined. While some categories appear to be obvious, others are not.

The nearest widely known equivalent to a definition is the acceptance rules for Microsoft's Universal Shaping Engine (USE), which would not tolerate Burmese or the older Khmer writing system, and would only tolerate Thai because the spacing vowels are categorised as letters (gc=Lo). Examples are given in the appendix. Thus the USE can only be used as a guide to the semantics of the property. As the USE is the most notable user of the Indic syllabic Category property, I have noted the effect of these changes upon it.

I reported (Reference 1) a number of errors in the Tai Tham Indic syllabic categories of characters the Tai Tham script for the review of the Unicode 8.0.0 character properties. Consideration of these reported errors was deferred until Unicode 9.0.0, and then further deferred to Unicode 10.0.0 (Reference 2). As these have not yet been addressed (as of `IndicSyllabicCategory-10.0.0d2.txt`, current on 29 April 2017), I am now formally proposing their immediate correction for Unicode Version 10.0.0.

The errors strongly urged for correction all impact upon the rendering of the Universal Shaping Engine, increasing the temptation for a font simply to work round it by removing all dotted circles, thereby defeating the purpose of the USE inserting them in the first place.

As at least one version of Adobe Acrobat Reader does not render the Tai Tham text in the text of this proposal properly, a copy of all the in-line Tai Tham Text is included in the Appendix.

Discussion

Alleged Final Consonants

No justification was presented for most of the assignments of syllabic categories; it appears that they were assigned on the basis that the categories were 'obvious'. In the case of the Tai Tham script, there is the complication that there are several doublets of subjoined consonants, one encoded using the coeng model and the other encoded using the Tibetan model, and that three of the characters were added during ISO balloting, *after* the last proposal has been submitted. The result is that a whole swathe of consonant marks were initially categorised as Consonant_Final:

U+1A57 TAI THAM CONSONANT SIGN LA TANG LAI
U+1A58 TAI THAM SIGN MAI KANG LAI
U+1A59 TAI THAM CONSONANT SIGN FINAL NGA
U+1A5A TAI THAM CONSONANT SIGN LOW PA
U+1A5B TAI THAM CONSONANT SIGN HIGH RATHA OR LOW PA
U+1A5C TAI THAM CONSONANT SIGN MA
U+1A5D TAI THAM CONSONANT SIGN BA
U+1A5E TAI THAM CONSONANT SIGN SA

Of these, only SIGN FINAL NGA is an indubitable final consonant.

Note that the data file `IndicSyllableCategory.txt` helpfully explains that the category Consonant_Final is for “special final forms which do not take vowels”.

U+1A58 Sign Mai Kang Lai

SIGN MAI KANG LAI appears to be in the process of becoming a final consonant, though it is quite conceivable that this change will not happen in the former Lao cultural region of North-East Thailand and Laos. As the two characters have the same glyph in this region, the effect of such a change would be likely to be the replacement of SIGN MAI KANG LAI by SIGN FINAL NGA. Traditionally, it acts like a repha, and is displayed on the following consonant. For its traditional rendering, its categorisation would better be Consonant_Preceding_Repha.

I am aware of only one unbroken word where this might make any difference to USE, namely Northern Thai . At present, dotted circles are inserted thus  because it does not allow Consonant_Final within a cluster and, despite the precedent of the Khmer script and, historically, the Lao script, it does not allow subjoined consonants after vowels. If the categorisation were changed to Consonant_Preceding_Repha, USE would object because the 'repha' was not syllable initial. On the other hand, the change would cause problems with the Tai Khün line-breaking of  (modern Tai Khün style  <U+1A48 TAI THAM LETTER HIGH SA, U+1A58, U+1A25 TAI THAM LETTER LOW KHA> 'Buddhist community of monks', where the line-break

may occur after SIGN MAI KANG LAI, as shown in Figure 1. I do not have any examples of line-breaking before this character.

U+1A57 Consonant Sign La Tang Lai

Note that U+1A57 is an abbreviation or a ligature. It is derived by the addition of a stroke to the subscript form <U+1A60 TAI THAM SIGN SAKOT, U+1A43 TAI THAM LETTER LA>, which would appear to be a subscript form of U+1A26 TAI THAM LETTER NGA in origin. Abbreviations of the word /tangla:i/ 'all' using U+1A57 normally include at least <U+1A57, U+1A63 TAI THAM VOWEL SIGN AA>, so U+1A57 is not Consonant_Final. An example, apparently spelt  <U+1A26 TAI THAM LETTER NGA, U+1A57, U+1A76 TAI THAM SIGN TONE-2, U+1A63 TAI THAM VOWEL SIGN AA>, is given in Table 16 at <http://www.seasite.niu.edu/tai/TaiLue/graphic%20blends.htm>. I am not completely sure I have read the stroke above properly; it may be U+1A62 TAI THAM VOWEL SIGN MAI SAT, and thus be closer to the Northern Thai form , which uses plain <U+1A60, U+1A43> rather than U+1A57.

The effect on the misrendering of  by USE would be to move the dotted circle from after U+1A57 to before it. In the new position, it would be because USE does not allow subjoined consonants after vowels, a fault which must be corrected regardless if USE is to support the rendering of Tai Tham monosyllables.

U+1A5A Consonant Sign Low Pa

U+1A5A is reportedly restricted to a single word,  /kap p^haʔ/ 'pregnant', a Tai Khün borrowing of the Pali word *gabbha* with the primary meaning of 'womb'. However, should this Tai Khün word ever receive Pali inflexions, it will acquire a final written vowel. The encoding of the word is <1A23;TAI THAM LETTER LOW KA, 1A3D;TAI THAM LETTER LOW PHA, 1A5A;TAI THAM CONSONANT SIGN LOW PA>. Thus 'wombs' could not unreasonably be rendered as  /kap p^ha:/ *gabbhā*; but then U+1A5A would be followed by the vowel U+1A63 TAI THAM VOWEL SIGN AA, contrary to the definition of Consonant_Final, even though U+1A5A is indeed the special form of a consonant used to represent a final consonant! (There would be no problem with the more natural spellings  and  of *gabbhā*.) The coding and phonetic behaviour of U+1A5A is like U+17CC KHMER SIGN ROBAT in a Khmer dialect that has not lost final /r/, so the most appropriate category is Consonant_Succeeding_Repha.

The effect on USE of this correction ought to be to remove the dotted circle from the misrendering of the hypothetical word as . At present, USE treats 'Consonant_Succeeding_Repha' the same as a final consonant, for the consonants classified thus in scripts supported by USE are actually final consonants. The cleanest example of a 'Consonant_Succeeding_Repha' is U+17CC KHMER SIGN ROBAT, and that occurs before vowels, e.g. in  'camphor' <U+1780 KHMER LETTER KA, U+1794 KHMER LETTER BA, U+17CC, U+17BC KHMER VOWEL SIGN UU, U+179A KHMER LETTER RO> from Sanskrit *karpūra*.

U+1A5B Consonant Sign High Ratha or Low Pa

The word နိဗ္ဗာန <U+1A36 TAI THAM LETTER NA, 1A65 TAI THAM VOWEL SIGN I, 1A3B TAI THAM LETTER LOW PA, 1A5B, 1A64 TAI THAM VOWEL SIGN TALL AA, 1A36> /nippa:na/ 'nirvana' immediately demonstrates that U+1A5B is not Consonant_Final. Instead it is a subjoined consonant, rendered differently from <U+1A60, U+1A3B TAI THAM LETTER LOW PA> or <U+1A60, U+1A2E TAI THAM LETTER HIGH RATHA>.

The effect on USE of this correction would be to remove the dotted circle from the misrendering နိဗ္ဗာန.

U+1A5C Consonant Sign Ma

U+1A5C occurs in Pali proper names ending *-mmo* <U+1A3E TAI THAM LETTER MA, U+1A5C, U+1A6E TAI THAM VOWEL SIGN E, U+1A63 TAI THAM VOWEL SIGN AA> in the Tai Khün writing system, such as မာမ္မိဗ္ဗာနာ (Figure 2), so is clearly not Consonant_Final.

The effect on USE of this correction would be to remove the dotted circles from the misrendering မာမ္မိဗ္ဗာနာ.

U+1A5D Consonant Sign Ba

U+1A5D occurs in Northern Thai principally in one word, whose pronunciation is roughly /kəbɔː/. U+1A5D is not Consonant_Final in its phonetic effect. The word is a compound word (or perhaps just a visual compound), formed by chaining two syllables and striking out the duplicated characters. I have a text in which the constituents are to be encoded ကံ <U+1A23 TAI THAM LETTER LOW KA, U+1A74 TAI THAM SIGN MAI KANG> and ဃံ <U+1A37 TAI THAM LETTER BA, U+1A74, U+1A75 TAI THAM SIGN TONE-1>, so the chained word may reasonably be encoded ကံဃံ <U+1A23, U+1A74, U+1A5D, U+1A75> or ကံဃံ <U+1A23, U+1A5D, U+1A74, U+1A75>.

The effect on USE of this correction would be to remove the dotted circles from the misrendering of the latter as ကံဃံ. (The rendering of the former is also assailed by other problems in the application of USE to the Tai Tham script.)

U+1A5E Consonant Sign Sa

While all my examples of U+1A5E are word final, it seems to differ from <U+1A60, U+1A48 TAI THAM LETTER HIGH SA> on the basis of the room available for it. Both forms are used as a word final consonant. The only Pali consonant cluster ending in <s> is <ss>, and that is written using U+1A54 TAI THAM LETTER GREAT SA, so a non-final <s> will be rare. (I can find the Sanskrit-based combination /kʂ/ written with U+1A47 TAI THAM LETTER HIGH SSA due to the application of RUKI.) However, I feel it would be rash to presume that every example of U+1A5E will be a final consonant.

Medial Consonants

Distinction between Consonant_Medial and Consonant_Subjoined

It is not clear to me whether the choice of categorisation between Consonant_Medial and Consonant_Subjoined is intended to correspond to different behaviour within the syllable structure, and whether any such difference is to be relied upon.

Historically, there are two categories of Consonant_Medial in the Unicode standard. The first is in scripts where consonant stacking or otherwise combining is productive, and multiple forms are encountered. The commoner form, most typically seen in the writing of the onset of phonetic syllables, is encoded as a medial consonant. This is the case in the Gurmukhi, Myanmar, Tai Tham, Javanese, Zanabazar Square and Masaram Gondi scripts.

It follows from general phonological tendencies that these consonants are mostly liquids, but other resonants and even /h/ (in its most general sense) also occur as the original sounds of the consonants.

The second category is in scripts that lack a virama or other stacker, and for which formally adopting the Tibetan model would suggest that many characters had been overlooked. This is the case in the Lao, Cham and Ahom scripts.

An analysis of characters categorised as Consonant_Subjoined is interesting. The Tibetan and Marchen characters are simply a result of the scripts using the Tibetan model.

There is no apparent reason not to have classified the Limbu characters as Consonant_Medial rather than Consonant_Subjoined. (The Limbu final consonants are classified as Consonant_Final.) The Unicode Standard actually discusses the Sundanese and Lepcha characters of this category under heading of 'Medials'.

Consonant_Subjoined U+A9BD JAVANESE CONSONANT SIGN KERET appears to be a vowel sign; I presume it is declared a consonant because it is an alternative to the subscripted vowel letter U+A989 JAVANESE LETTER PA CERER, which is classed as a Consonant on the basis of its formal behaviour – it can be subscripted by means of U+A9C0 JAVANESE PANGKON and it can also taking the length mark/vowel U+A9B4 JAVANESE VOWEL SIGN TARUNG.

This leaves the Phags-pa 'subjoined consonants' as the only sporadic group which might be properly Consonant_Subjoined rather than Consonant_Medial!

Fighting Spoofing by Grammar Rules

One of the potential problems of locally non-linear writing systems is that different sequences of code points may have identical renderings. Sometimes that can be solved by canonical equivalence, but at others it cannot be. A particular problem arises with the subscript consonant forms that are written to the left (often a subscript form for RA) or to the right (often a subscript form of YA) of the base consonant; it may not be clear where it occurs in the sequence of such consonants. The distinctiveness may vary between closely related scripts - <U+103B MYANMAR CONSONANT SIGN MEDIAL YA, U+103D MYANMAR CONSONANT SIGN MEDIAL WA> is indistinguishable form <U+103D, U+103B>, but <U+1A60 TAI THAM SIGN SAKOT, U+1A40 TAI THAM LETTER HIGH YA, U+1A60, U+1A45 TAI THAM LETTER WA> is regularly rendered differently to <U+1A60, U+1A45, U+1A60, U+1A40>.

Now, it may be possible for a rendering engine to check that subscript consonants are encoded in a plausible order, and visibly disrupt the rendering if they are not. A very rough guide is that medial consonants tend to follow rather than to precede other types of subscript consonant. This heuristic largely works because Consonant_Medial is a category commonly appropriate to the consonants that tend to terminate onset clusters. However, I ask of the UTC:

Is the category of Consonant_Medial to be constrained by this heuristic?

If the category Constant_Medial merely identifies the *usual* use of a character, then this whole section on medial consonants is irrelevant to the categorisation of Tai Tham consonant signs.

Why Tai Tham Has No Medial Consonants

Although U+1A55 TAI THAM CONSONANT SIGN MEDIAL RA and U+1A56 TAI THAM CONSONANT SIGN MEDIAL LA are named as medial consonants, too much should not be read into such a description. Both are, very occasionally, immediately preceded by vowels. The usual example for U+1A55 is  <U+1A2F TAI THAM LETTER DA, U+1A6A TAI THAM VOWEL SIGN UU, U+1A55 TAI THAM CONSONANT SIGN MEDIAL RA, U+1A63 TAI THAM VOWEL SIGN AA> 'hearken to me'.

Examples for U+1A56 are shown in Figures 5 and 6; note the pronunciations shown. Reference 5 has another two examples –  <...U+1A3B TAI THAM LETTER LOW PA, U+1A6B TAI THAM VOWEL SIGN O, U+1A56> ('woollen garment', on p27, and spelt  on p32 of Reference 6) and  <... U+1A49 TAI THAM LETTER HIGH HA, U+1A6B, U+1A55 ...> ('great army', on p511). Note that the reader of these words will only realise that the sequence is <U+1A6B, U+1A55> /on/ and not <U+1A55, U+1A6B> /(l)o/ when he sees that under the latter interpretation the obligatory, written final consonant of the *phonetic* syllable is not supplied by the following *orthographic* syllable. When the last vowel is U+1A65 TAI THAM VOWEL SIGN I, it is only familiarity with the language that makes him realise that the last syllable is /k^hin/ and not /k^hi?/.

Both U+1A55 and U+1A56 are regularly followed by <U+1A60 TAI THAM SIGN SAKOT, U+1A40 TAI THAM LETTER HIGH YA> and <U+1A60, U+1A45 TAI THAM LETTER WA>, e.g.  'to embrace',  'to compare',  'banana' and  'nut'. While the latter two sequences most commonly represent vowels, the strictly consonantal cluster <U+1A49 TAI THAM LETTER HIGH HA, U+1A56, U+1A60, U+1A45> starts a few words beginning with the cluster /lw/, e.g.  'iron'. This is a behaviour the Universal Shaping Engine of Microsoft currently disallows for medial consonants.

I therefore recommend that, if a categorisation as Consonant_Medial is to be interpreted as restricting the legitimate usage of a character, U+1A55 and U+1A56 be recategorised from Consonant_Medial to Consonant_Subjoined.

Miscellaneous Marks

A number of marks require an appreciation of history to properly categorise:

U+1A61 TAI THAM VOWEL SIGN A

U+1A74 TAI THAM SIGN MAI KANG

U+1A7A TAI THAM SIGN RA HAAM

U+1A7C TAI THAM SIGN KHUEN-LUE KARAN

U+1A7F TAI THAM COMBINING CRYPTOGRAMMIC DOT

U+1A61 Vowel Sign A

Historically, U+1A61 TAI THAM VOWEL SIGN A is a visarga, like the closely related U+1038 MYANMAR SIGN VISARGA. Its primary effect is to represent a final glottal stop, but it may also be used to shorten the vowel of the syllable. When the vowel of an 'open' syllable is /a/, it has the effect of confirming its presence, whence its naming as a vowel.

The effect on USE of making a more historical would be to fix the misrendering of $\text{ᨾᨿ} = \langle \text{U+1A4B TAI THAM LETTER A, U+1A70 TAI THAM VOWEL SIGN OO, U+1A6C TAI THAM VOWEL SIGN OA BELOW, U+1A76 TAI THAM SIGN TONE-2, U+1A61 TAI THAM VOWEL SIGN A} \rangle$ as ᨾᨿᨶ . However, this fault in USE is a manifestation of its general prohibition on tone marks before following vowels (i.e. those on the right in Tai Tham); far more misrenderings would simply be cured by USE removing this faulty prohibition.

U+1A74 Sign Mai Kang

U+1A74 TAI THAM SIGN MAI KANG is of course the anusvara, though as in Lao it also serves as a vowel in the south (Thailand and Laos). Changing its category from Vowel_Dependent to Bindu would have several incidental benefits for rendering using USE:

- (a) The sequence $\langle \text{U+1A63 TAI THAM VOWEL SIGN AA, U+1A74} \rangle$ would not need any special treatment to enable it to render without dotted circles. Instead of /kam/ 'word' being misrendered as ᨾᨿᨶ , it would be rendered properly as ᨾᨿᨶ . Conversely, the misspelling $\langle \text{U+1A74, U+1A63} \rangle$ would be penalised with a dotted circle.
- (b) The Northern Thai spelling of /ɲɔː/ 'to lift' $\langle \text{1A3F;TAI THAM LETTER LOW YA, 1A6C;TAI THAM VOWEL SIGN OA BELOW, 1A74;TAI THAM SIGN MAI KANG} \rangle$ (see Reference 3 Section 5 Example 8 for the encoding of the compound vowel symbol) would be rendered properly as ᨾᨿᨶ , rather than being misrendered as ᨾᨿᨶ .

U+1A7A Sign Ra Haam and U+1A7C Sign Khuen-Lue Karan

U+1A7A TAI THAM SIGN RA HAAM is currently categorised as Syllable_Modifier, which appears to be a rubbish bin category. It has three functions:

- (a) It is a superscript syllable-final version of U+1A41 TAI THAM LETTER RA.

- (b) It silences a vowel. Usually this is an implicit vowel, but not always. If the silenced vowel follows a cluster of 2 (or more) consonants, then by the rules of Tai phonology all but the first of the consonants are silenced.
- (c) As a development of the above, it is occasionally used to silence a consonant.

Function (a) has become rare. If categorised for its commonest rôle, it is a Pure_Killer.

Changing its category to 'Pure_Killer' has several advantages for rendering with USE, which treats a Pure_Killer the same as Vowel_Dependent:

- (a) Tone marks follow SIGN RA HAAM visually, and changing its category to Vowel_Dependent would enable  (shown twice in Figure 4) to be rendered properly, instead of being misrendered as . However, that would still leave the unconfirmed issue of tone marks following U+1A59 TAI THAM CONSONANT SIGN FINAL NGA. A more recent analogue is the spelling of  (to give the Khün rendering) shown in Figure 3, in which the tone mark follows the superscript consonant, except that the superscript WA is represented by U+1A74 TAI THAM SIGN MAI KANG.

However, the current behaviour of USE may simply be wrong in this respect. New Tai Lue tone marks may follow final consonants, which are classified as Consonant_Final. (That U+19B0 NEW TAI LUE VOWEL SIGN VOWEL SHORTENER may also follow final consonants, as in words such as , is just evidence that U+19B0 should be classified as 'Visarga'.)

- (b) Once USE allows vowels before subscript consonants, it will for free be able to handle the combinations where SIGN RA HAAM visibly has the effect of silencing the first consonant of a cluster rather than the second, as in  /sɣ:p/ 'to serve';  would be pronounced /sɣ:n/.

U+1A7C TAI THAM SIGN KHUEN-LUE KARAN shares the second two functions of U+1A7A, and it therefore makes sense to also change it from 'Syllable_Modifier' to 'Pure_Killer'.

U+1A7B Sign Mai Sam

U+1A7B is principally a repetition mark, being a modification of  U+1A92 TAI THAM THAM DIGIT TWO, indicating the repetition of a word. As extensions of this role, it can also do at least the following:

- (1) Indicate a repeated (not geminate) consonant
- (2) Indicate an omitted implicit vowel (one omits an implicit vowel by replacing it with U+1A60)
- (3) Indicate an epenthetic vowel (extension of Rôle 2).

In Rôle (1), it serves as a subjoined consonant. In Rôles (2) and (3), it serves as a dependent vowel. For a shaper that does not constrain appearance, such as the Universal Shaping Engine, the best categorisation is probably 'Consonant_Subjoined'; it is certainly not 'Syllable_Modifier', which constrains it to be at the end of

an orthographic syllable. The proposal quite clearly states that it occurs as part of the syllable, not just at the end of it.

U+1A7F Combining Cryptogrammic Dot

I submit that U+1A7F should have its category corrected from Syllable_Modifier to Nukta. Its rôle is to change the value of letters to another letter, albeit one already in the alphabet. I believe this change will have no deleterious effects.

Proposal

1. I strongly urge that the following recategorisations be made:

Codepoint	Name	Change Category	
		From	To
1A57	TAI THAM CONSONANT SIGN LA TANG LAI	Consonant_Final	Consonant_Subjoined
1A5B	TAI THAM CONSONANT SIGN HIGH RATHA OR LOW PA	Consonant_Final	Consonant_Subjoined
1A5C	TAI THAM CONSONANT SIGN MA	Consonant_Final	Consonant_Subjoined
1A5D	TAI THAM TAI THAM CONSONANT SIGN BA	Consonant_Final	Consonant_Subjoined

2. I also request that the following changes be made:

Codepoint	Name	Change Category	
		From	To
1A55	TAI THAM CONSONANT SIGN MEDIAL RA	Consonant_Medial	Consonant_Subjoined
1A56	TAI THAM CONSONANT SIGN MEDIAL LA	Consonant_Medial	Consonant_Subjoined
1A5E	TAI THAM CONSONANT SIGN SA	Consonant_Final	Consonant_Subjoined
1A74	TAI THAM SIGN MAI KANG	Vowel_Dependent	Bindu
1A7A	TAI THAM SIGN RA HAAM	Syllable_Modifier	Pure_Killer
1A7C	TAI THAM SIGN KHUEN-LUE KARAN	Syllable_Modifier	Pure_Killer
1A7F	TAI THAM COMBINING CRYPTOGRAMMIC DOT	Syllable_Modifier	Nukta

3. The following changes should be considered:

Codepoint	Name	Change Category	
		From	To
1A58	TAI THAM SIGN MAI KANG LAI	Consonant_Final	Consonant_Preceding- _Repha
1A5A	TAI THAM CONSONANT SIGN LOW PA	Consonant_Final	Consonant_Succeeding- _Repha
1A61	TAI THAM VOWEL SIGN A	Vowel_Dependent	Visarga
1A7B	TAI THAM SIGN MAI SAM	Syllable_Modifier	Consonant_Subjoined

Evidence

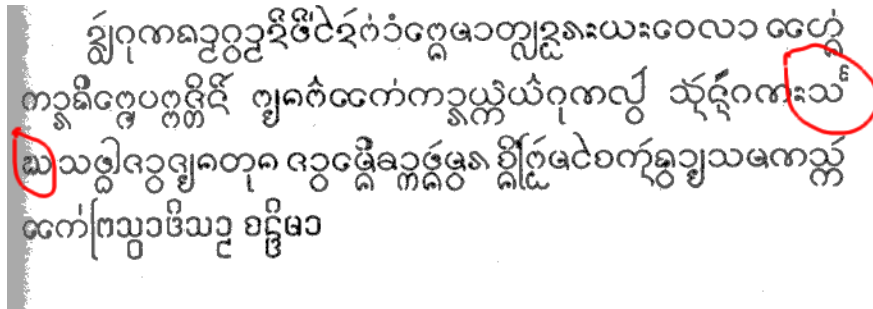


Figure 1 Tai Khün Line-Breaking of ၁၁၁ (Reference 4 p205)

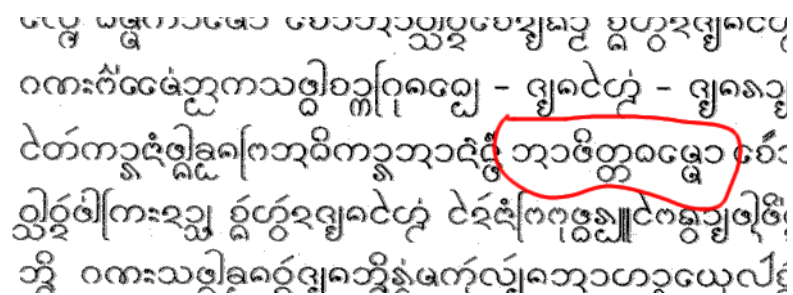


Figure 2 Tai Khün Pali Monk's Name Adittadhammo (Reference 4 p264)

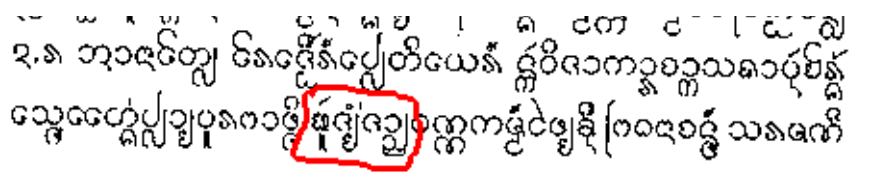


Figure 3 Example of ၁၁၁ (Reference 4 p184)

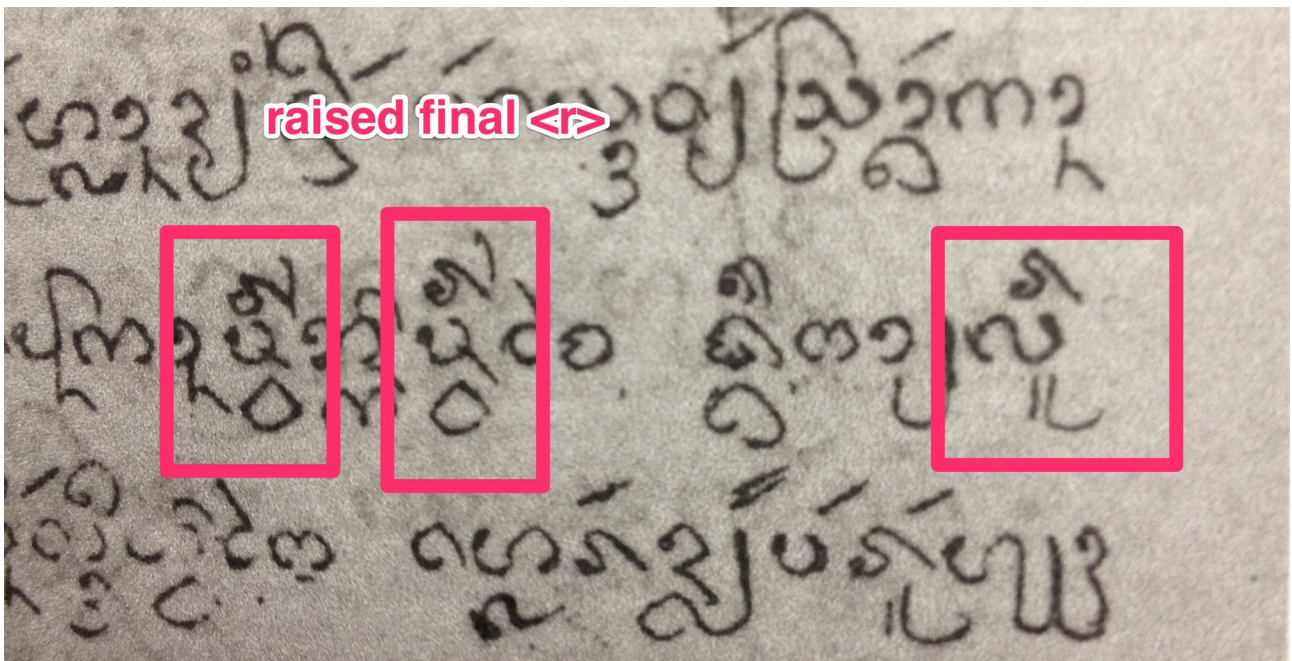


Figure 4 Tone Mark Visually Following RA HAAM in Tai Khün (c. 1949 (furnished by Patrick Chew)

เสามงคล	เสามงคล / เสา-ม่ง-ค่น / saw5-mop1-kon1
เม่อผู้	น. เสาเอก- เสาเล่มแรกที่ยกกลงหลุมในการสร้างบ้าน house pillar: the first one put up when builing the house
	888/ พ่อแม่เจ้าเรือนเค้าตันหว่างเสามงคลกับอีกเสานางนอนกลาง /เจีย/181
เสาอินทิล	เสาอินทิล / เสา-อิน-ตะ-ขึ้น / saw5-ʔin1-ta1-khin5
เม่อผู้	น. เสาหลักเมือง (อินทร์- พระอินทร์, ซิล- เสา) Inthakhin, city pillar
	พันเสาอินทิล ที่ใหม่ 330 เบี้ย (พัน)หลักเสมาเมืองใหม่ 110 เงิน /มังราย/1/4

Figure 5 Words with MEDIAL LA as final consonant (Reference 5 p691)

เดือนผลคุณ	เดือนผลคุณ / เดือน-ผน-ละ-ค่น / d+an2-phon5-laʔ4-khun1
เม่อผู้	น. เดือน 6 เหนือ the sixth month
	ปีรวายลีเดือนผลคุณไต่ว่าเดือน 6
	/จ.1/จ.ภ.4/(พระธาตุศรีจอมทอง/104)112

Figure 6 Double-Acting MEDIAL LA (Reference 5 p259)

References

1. Wordingham J.R., in *Accumulated Feedback on PRI #297*, <http://www.unicode.org/review/pri297/feedback.html> (submitted February 2015)
2. Unicode Consortium, *Open Action Items*, Action 143-A54, <http://www.unicode.org/L2/L-SD2.htm> (April 2017)

(This is a live document, updated regularly.)
3. Michael Everson, Martin Hosken, and Peter Constable, *Revised proposal for encoding the Lanna script in the BMP of the UCS*, ISO/IEC JTC1/SC2/WG2, L2/07-007R, <http://www.unicode.org/L2/L2007/07007r-n3207r-lanna.pdf> (March 2007)
4. *Chiang Tung: Its Way of Life* ຄຊະມຸດຕູນຈຸງຕູນ [Khemarattha Nakon Cheng Tung] เขมรรัตนนครดียงตุง [Khemarat Nakhon Chiang Tung] (in Thai, Tai Khün, French and English) (Chiang Mai: Wat Tha Kradas วัดท่ากระดาศ) (1998)
5. Arunrat Wichiankhiao et al. อรุณรัตน์ วิเชียรเขียว, กาอตุญฺทึว นัฏ จขเฎี= กิ๋มิ โก๊ กุ๊ ๕๔๒๖๗๖ พจนานุกรมศัพท์ล้านนาเฉพาะคำที่ปรากฏในใบลาน *The Northern Thai Dictionary of Palm-Leaf Manuscripts*. ISBN 974-7067-77-2 (1996)
6. Rungrueangsi, Udom (2004) [1991]. *Lanna-Thai Dictionary, Princess Mother Version* พจนานุกรมล้านนา ~ ไทย ฉบับแม่ฟ้าหลวง กาอตุญฺทึว นัฏ ~ ๕๒๒ จขญจจเฎีวภูฏ [Photchananukrom Lanna ~ Thai, Chabap Maefa Luang] (in Thai) (Revision 1 ed.). Chiang Mai: Rongphim Ming Mueang (โรงพิมพ์เมือง). ISBN 974-8359-03-4.

Appendix

Prescribed Syllable Structures Proscribed by the USE

These examples are taken from scripts not supported by the USE.

Language	Word	Meaning	Encoding	Violation of USE
Burmese	ကျွန်	slave	1000 103B 103D 1014 103A	Order of Consonant_Medial – right before bottom!
Burmese	ယောက်ျား	man, male; husband	101A 1031 102C 1000 103A 103B 102C 1038	Pure_Killer before Consonant_Medial. In this case, ASAT is being used as a pure killer; U+1000 is <i>implicitly</i> repeated between U+103A and U+103B.

Language	Word	Meaning	Encoding	Violation of USE
Khmer	ហើរ	already	17A0 17BE 17D2 1799	Invisible_Stacker between Vowel_Dependent and Consonant
Thai	บ้าน	house	0E1A 0E49 0E32 0E19	Tone_Mark before Vowel_Dependent, except that the constraint does not apply because U+0E32 has gc=Lo.

In-Line Tai Tham Text

Section

Tai Tham Examples

U+1A58 Sign Mai Kang Lai

ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ

U+1A57 Consonant Sign La Tang Lai

ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ

U+1A5A Consonant Sign Low Pa

ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ

U+1A5B Consonant Sign High Ratha or Low Pa

ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ

U+1A5C Consonant Sign Ma

ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ

U+1A5D Consonant Sign Ba

ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ

Why Tai Tham Has No Medial Consonants

ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ

U+1A61 Vowel Sign A

ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ

U+1A74 Sign Mai Kang

ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ

U+1A7A Sign Ra Haam and U+1A7C Sign Khuen-Lue Karan

ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ ᩈᩢ᩠ᨦ