

Proposal for encoding New Tai Lue script in Unicode/ISO-IEC 10646

Doc Type: draft

Title: Proposal for encoding New Tai Lue script in Unicode/ISO-IEC 10646

Source: Peter Constable, SIL International

Status:

THIS IS A PRELIMINARY DRAFT OF A PROPOSAL, SUBMITTED FOR INFORMATION PURPOSES ONLY. NO REQUEST IS BEING MADE BY THE SUBMITTER FOR ACTION ON THIS PROPOSAL UNTIL FURTHER RESEARCH HAS BEEN CONDUCTED.

Action:

Date: 1999-08-13

A. Administrative

1. Title

Proposal for encoding New Tai Lue script in Unicode/ISO-IEC 10646

2. Requester's name

Peter Constable, SIL International

3. Requester type

Expert contribution

4. Submission date

1999-08-13

5. Requester's reference

6. Completion

This is a PRELIMINARY DRAFT OF A complete proposal.

B. Technical—General

1a. New script? Name?

This is a new script. New Tai Lue.

1b. Addition of characters to existing block? Name?

No.

2. Number of characters

71

3. Proposed category

Category A

4. Proposed level of implementation and rationale

Level 2 (contains combining characters)

5a. Character names included in proposal?

Yes.

5b. Character names in accordance with guidelines?

Yes.

5c. Character shapes reviewable?

Yes (see below).

6. Who will provide computerized font?

SIL can provide a font.

7a. Are references provided?

7b. Are published examples of used of proposed characters attached?

(can be provided)

8. Does the proposal address other aspects of character data processing?

Yes (see below).

C. Technical-Justification

1. Has this proposal for addition of character(s) been submitted before?

No.

2a. Has contact been made to members of the user community?

Yes.

2b. References

3. Information on the user community

This script is used by speakers of Tai Lue in Xishuangbanna Dai Autonomous Region of Yunnan Province, PRC. According to the SIL Ethnologue, there are up to 770,000 in China, with an estimated 303,000 speakers in Myanmar, Thailand, Laos and Vietnam. This script is in use by speakers in China. The community in China also use a traditional script (Lanna script with minor variations), from which this script was adapted.

4. The context of use for the proposed characters.

Commonly used for printing. (Need to determine whether it is also commonly used for writing.)

5. Are the proposed characters in current use by the user community?

Yes, in China.

6a. Should characters be entirely in BMP?

Yes.

6b. Rationale

Contemporary use.

7. Should the proposed characters be kept in a contiguous range?

Yes.

8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

9. Can any of the proposed characters be considered to be similar to an existing character?

No. These characters are most closely related to Lanna script (not yet included in standard), and arguments could perhaps be made for unification of these two scripts. Due to significant differences between these scripts, however, it is my view that they should be encoded separately.

10a. Does the proposal include use of combining characters and/or composite sequences?

Yes.

10b. Is a list of sequences and corresponding glyph images provided?

Yes (see below).

11. Does the proposal contain characters with any special properties, such as control functions?

No.

D. Administrative

E. Proposal

Introduction

New Tai Lue script was adapted from a variant of Lanna that had been in use for writing the Tai Lue language for several centuries. In the adaptation process, however, significant simplifications were made with the result that there is far less in the way of complex script behaviour than is typical of Indic scripts.

For those familiar with closely related scripts of Southeast Asia, Lanna and Myanmar script in particular, a quick review of the representative glyphs will immediately reveal many character shapes in common with those scripts, but many that are distinct. The character shapes that underwent the greatest transformation in the adaptation from Lanna script are those that correspond to combining vowel diacritics in Lanna script.

Structural characteristics of New Tai Lue script and implications for this encoding proposal

New Tai Lue script is a left-to-right, alphabetic script with some structural similarities to other scripts of Indic derivation, though it is far less complex than most Indic scripts.

The character inventory of the script consists of 39 consonant letters, 17 vowel letters, 2 tone letters, and one modifying diacritic. With the exception of this diacritic, all letters, including tones are written directly on the baseline.

The single diacritic is a combining, subscript form of ‘o’ `xx1F LETTER LOW WA`, and is used in writing labialised consonant clusters. In this proposal, a single character is allocated for this diacritic, `xx3B MODIFIER W`. For example,



(Note: it is possible that some users perceive these labialised clusters as graphemic units, as with the high /h/ + sonorant combinations described below. If so, a case might be made that each of the labialised consonant clusters should be given separate character allocations and the combining modifier dispensed with. It is my current thinking that the solution proposed here is the better alternative.)

In the writing systems based on many related scripts of Southeast Asia (e.g. Thai, Lao, Northern Thai), letters for writing most obstruent consonants come in pairs, often referred to as “high” and “low”, each having distinct tone properties. This is also true of New Tai Lue script. Thus, for example, ‘`๓`’ `xx04 HIGH KHA` and ‘`๔`’ `xx07 LOW KHA`, are used to write the same phoneme but in syllables of differing tone.

In these other writing systems, it is also typical that letters for writing sonorant consonants do not come in pairs, but that the “high” counterpart of a sonorant consonant is written as a digraph combining that sonorant consonant letter with the letter for high /h/. Whereas in Thai and Lao, both consonants are written on the baseline, in Lanna script the sonorant consonant is written as a diacritic below the HIGH HA. In some cases, the shape of this diacritic is the closely similar to the base consonant letter, but in other cases it is quite different.

New Tai Lue script is exactly like Lanna script in this respect, although the high /h/ + diacritic sonorant combinations are customarily viewed as a single, composite grapheme. Thus, discussion of New Tai Lue script do not normally speak of high /h/ + diacritic sonorant combinations but rather of high sonorant / low sonorant letter

pairs. This proposal follows this customary view of sonorants in New Tai Lue script. So, for example, ‘’ xx05 HIGH NGA and ‘’ xx08 LOW NGA.

In Lanna script, most or all consonants can be written as diacritic subscripts below baseline characters; this is most typical done with syllable-final consonants. In New Tai Lue script, this has changed such that syllable-final consonants are written on the baseline, but with contextual-variant shapes. So, for example, /k/ is nominally written ‘’, but syllable-final it is written ‘’. Due to the syllable restrictions of Tai Lue phonology and orthographic constraints adopted in the revision of Tai Lue writing, there are only 7 consonant letters that are written syllable-final:

non-final							
final							

In this proposal, final forms of these consonants are encoded by means of a single combining character, xx3C MODIFIER SIGN FINAL. For example,

 +  → 

The New Tai Lue digits are very similar in form to those of Lanna script. The alternate digits used in Lanna Burmese texts have not been retained in New Tai Lue script, though there is another variant form for the digit corresponding to ‘1’. This alternate form has been included in this proposal, though it may be preferable to unify these.

While scripts of South Asia often encoded in “logical” (phonological) order, Thai and Lao are encoded in visual order. Since all vowel characters have been given equal structural status in New Tai Lue script as letters written on the baseline, it is most appropriate that they all be encoded in visual order. The combining modifiers xx3B MODIFIER W and xx3C MODIFIER SIGN FINAL, of course, follow the character they modify.

Names and ordering

Names for consonant and vowel letters are based on a typical Romanisation of the Tai Lue names for these characters. The order of characters follows the sort order most commonly used for this script.

Unicode character properties

Spacing letters: category “Lo”, bidi category “L” (strong left-to-right)

Combining modifiers: category: “Mn”

Numbers: category “Nd”, bidi category “L”, numeric value per names

It might reasonably be argued that xx3C MODIFIER SIGN FINAL is in some way similar to the virama of South Asian scripts. On that basis, it might be deemed appropriate to assign this character to Canonical Combining Class 9. It is unclear to me whether xx3B MODIFIER W should be assigned to Class 0, Class 220, or to a fixed position class.

Font implementations

It is advised that syllable-final consonants and labialised consonant clusters be rendered using contextually selected, composite presentation glyphs rather than over-striking diacritic glyphs.

Because of the shapes of certain glyphs, kerning is needed to provide optimal spacing of glyphs. For applications that do not support kerning, the type designer must choose whether to have certain pairs appear too-widely spaced, or to allow certain pairs to crowd and collide. The latter option in combination with a thin space may provide users an adequate compromise.

Other implementation issues

No standards for encoding of New Tai Lue script exist. Accordingly, backwards compatibility is not a concern.

I am not aware at the present time of any standards with regard to input methods.

New Tai Lue script is like Thai in that word breaks are not always overtly indicated with spaces. It may be necessary for some implementations to devise algorithms for detecting word boundaries. Since no standards for encoding this script have been established to date, it would be possible to consider incorporating such algorithms into input methods and to insert ZERO WIDTH SPACES where necessary. Not all implementers would necessarily want to make the needed investment in developing input methods, however.

Further research must be conducted to determine if searches should always equate syllable-final consonants with non-final consonants, equate them optionally, or never. Likewise for xx3B MODIFIER W and xx1F LOW W.

There may be certain expectations in sort orders with regard to the ordering of labialised consonants and non-labialised consonants. In this regard, it may be deemed necessary to treat consonant + MODIFIER W sequences as a unit for sorting purposes. Also, it is quite possible that reordering of preceding vowels may be necessary, as for Thai. Further research in this regard is required.

Bibliography

DRAFT

TABLE XXX – ROW XX: NEW TAI LUE

	xx0	xx1	xx2	xx3	xx4
0		ᨡ	ᨢ	ᨣ	ᨤ
1	ᨦ	ᨧ	ᨨ	ᨩ	ᨪ
2	ᨫ	ᨬ	ᨭ	ᨮ	ᨯ
3	ᨰ	ᨱ	ᨲ	ᨳ	ᨴ
4	ᨵ	ᨶ	ᨷ	ᨸ	ᨹ
5	ᨺ	ᨻ	ᨼ	ᨽ	ᨿ
6	ᩀ	ᩁ	ᩂ	ᩃ	ᩄ
7	ᩅ	ᩆ	ᩇ	ᩈ	ᩉ
8	ᩊ	ᩋ	ᩌ	ᩍ	ᩎ
9	ᩐ	ᩑ	ᩒ	ᩓ	ᩔ
A	ᩕ	ᩖ	ᩗ	ᩘ	ᩙ
B	ᩛ	ᩜ	ᩝ	ᩞ	
C	᩟	᩠	ᩡ	ᩢ	
D	ᩣ	ᩤ	ᩥ		
E	ᩦ	ᩧ	ᩨ		
F	ᩩ	ᩪ	ᩫ		

dec	hex	Name
	00	(This position shall note be used.)
	01	NEW TAI LUE LETTER HIGH QA
	02	NEW TAI LUE LETTER LOW QA
	03	NEW TAI LUE LETTER HIGH KA
	04	NEW TAI LUE LETTER HIGH KHA
	05	NEW TAI LUE LETTER HIGH NGA
	06	NEW TAI LUE LETTER LOW KA
	07	NEW TAI LUE LETTER LOW KHA
	08	NEW TAI LUE LETTER LOW NGA
	09	NEW TAI LUE LETTER HIGH CA
	0A	NEW TAI LUE LETTER HIGH SA
	0B	NEW TAI LUE LETTER HIGH YA
	0C	NEW TAI LUE LETTER LOW CA
	0D	NEW TAI LUE LETTER LOW SA
	0E	NEW TAI LUE LETTER LOW YA
	0F	NEW TAI LUE LETTER HIGH TA
	10	NEW TAI LUE LETTER HIGH THA
	11	NEW TAI LUE LETTER HIGH NA
	12	NEW TAI LUE LETTER LOW TA
	13	NEW TAI LUE LETTER LOW THA
	14	NEW TAI LUE LETTER LOW NA
	15	NEW TAI LUE LETTER HIGH PA
	16	NEW TAI LUE LETTER HIGH PHA
	17	NEW TAI LUE LETTER HIGH MA
	18	NEW TAI LUE LETTER LOW PA
	19	NEW TAI LUE LETTER LOW PHA
	1A	NEW TAI LUE LETTER LOW MA
	1B	NEW TAI LUE LETTER HIGH FA
	1C	NEW TAI LUE LETTER HIGH WA
	1D	NEW TAI LUE LETTER HIGH LA
	1E	NEW TAI LUE LETTER LOW FA
	1F	NEW TAI LUE LETTER LOW WA
	20	NEW TAI LUE LETTER LOW LA
	21	NEW TAI LUE LETTER HIGH HA
	22	NEW TAI LUE LETTER HIGH DA
	23	NEW TAI LUE LETTER HIGH BA
	24	NEW TAI LUE LETTER LOW HA
	25	NEW TAI LUE LETTER LOW DA
	26	NEW TAI LUE LETTER LOW BA
	27	NEW TAI LUE LETTER LAE

dec	hex	Name
	28	NEW TAI LUE VOWEL LETTER A
	29	NEW TAI LUE VOWEL LETTER AA
	2A	NEW TAI LUE VOWEL LETTER AAY
	2B	NEW TAI LUE VOWEL LETTER AW
	2C	NEW TAI LUE VOWEL LETTER AWY
	2D	NEW TAI LUE VOWEL LETTER II
	2E	NEW TAI LUE VOWEL LETTER OEY
	2F	NEW TAI LUE VOWEL LETTER UE
	30	NEW TAI LUE VOWEL LETTER UEY
	31	NEW TAI LUE VOWEL LETTER U
	32	NEW TAI LUE VOWEL LETTER UU
	33	NEW TAI LUE VOWEL LETTER UUY
	34	NEW TAI LUE VOWEL LETTER OY
	35	NEW TAI LUE VOWEL LETTER E
	36	NEW TAI LUE VOWEL LETTER EE
	37	NEW TAI LUE VOWEL LETTER O
	38	NEW TAI LUE VOWEL LETTER AY
	39	NEW TAI LUE TONE LETTER ONE
	3A	NEW TAI LUE TONE LETTER TWO
	3B	NEW TAI LUE MODIFIER W
	3C	NEW TAI LUE MODIFIER SIGN FINAL
	3D	(This position shall note be used.)
	3E	(This position shall note be used.)
	3F	(This position shall note be used.)
	40	NUMBER 0 NEW TAI LUE
	41	NEW TAI LUE DIGIT ONE
	42	NEW TAI LUE DIGIT TWO
	43	NEW TAI LUE DIGIT THREE
	44	NEW TAI LUE DIGIT FOUR
	45	NEW TAI LUE DIGIT FIVE
	46	NEW TAI LUE DIGIT SIX
	47	NEW TAI LUE DIGIT SEVEN
	48	NEW TAI LUE DIGIT EIGHT
	49	NEW TAI LUE DIGIT NINE
	4A	NEW TAI LUE DIGIT ALTERNATE ONE
	4B	(This position shall note be used.)
	4C	(This position shall note be used.)
	4D	(This position shall note be used.)
	4E	(This position shall note be used.)
	4F	(This position shall note be used.)

Names for xx3B and xx3C are tentative.