Universal Multiple-Octet Coded Character Set International Organization for Standardization Organisation internationale de normalisation Международная организация по стандартизации

**Doc Type:** Working Group Document

Title: Revised proposal for encoding the Tai Le script in the BMP of the UCS

Source: Ireland

**Status:** National Body contribution

Action: For adoption by JTC1/SC2/WG2

Date: 2001-10-06

Document N966 from China was introduced in 1994 and contained 58 characters. The repertoire of Tai Le characters was discussed (by experts from China, Ireland, the UK, and the US) and modified (to contain 35 characters) at the Athens meeting of WG2, and the results of this were published in N2292. This document contains the tables and character names for the script, and contains the proposal summary form. China has recently published N2371, which contains the same repertoire but gives the tone markers in different positions. We prefer the positions here, but consider that the Tai Le script repertoire and names are mature and ready for encoding *apart from verification of ordering requirements*. The names of the tone letters in N2371 differ from those given here, but the ones given here are in accordance with WG2 decisions taken at San Jose regarding such names.

The Tai Le script has a history of 700-800 years, during which period several orthographic conventions were used. The modern form of the script, which is proposed for encoding here, was developed in the years following 1954; it rationalized the older system and added a systematic representation of tones with the use of combining diacritics. In 1988, the new system was again revised, and spacing tone marks were introduced to replace the combining diacritics. This proposal treats both orthographies.

Tai Nüa, Dehong Dai, Tai Mau, Tai Kong, and Chinese Shan are other names by which the Tai Le language is known. This encoding prefers *Tai Le*, which is a transliteration of the indigenous designation, myn  $\mathfrak{All}$  [c [tai² lə⁶] (in older orthography myn  $\mathfrak{All}$ ). Tai Le orthography is simple and straightforward: initial consonants precede vowels, vowels precede final consonants, and tone marks, if any, follow the entire syllable. There is a one-to-one correspondence between the tone mark letters now used and existing nonspacing diacritics in the UCS, used prior to 1988. In both orthographies the tone mark is the last character in a syllable-string. Typographically, when one of the combining diacritics follows a tall letter, it is represented as a spacing mark. It is recommended to let the font do this, and not to employ sp or NBSP in the encoded text to force this behaviour, or to use spacing diacritics with tall letters.

7	ka	។	ka	7B	ki	ИB	ki
٩л	$ka^2$	៉ា	$ka^2$	чЮл	$ki^2$	чÐ	$ki^2$
чe	ka³	Ϋ́	$ka^3$	иlle	ki³	<b>∽l</b> ]^	$ki^3$
70	$ka^4$	৸	$ka^4$	чBа	ki <sup>4</sup>	<b>∽l</b> }′	ki <sup>4</sup>
70	$ka^5$	Ä	$ka^5$	ଧ୍ୟା	ki <sup>5</sup>	۸ <b>.</b>	$ki^5$
7C	$ka^6$	Í	ka <sup>6</sup>	чBС	$ki^6$	ηÐ	ki6

ulginö		د "المه	لد نو	انزا	P.		भानं औन
	H	18.	784	26,	Ð.	表	
Mo.	N	M.	M	M.	N.	M.	יוליה מינך יוליה מיניים
	Apo	18:	180	780	Mi	<b>78</b> 6	نوائد
	No	Mo	Mi	180	No	200	
							"n <sub>y</sub> "
My	29/2	190	39%	Mi	ناوار ا	Aft.	امر سرمام بدال
and the second second second							you wing
dy 10		M.,	3-yì .	تېلا	$\mathcal{C}$		شرك تراه
20			Mi			<b>\</b>	العام الله بزلا
Service report			I		-Do	5 - 1 - 1 - 1 - 2 - 1	

Sample of Tai Le in old orthography, 1988.

**Digits.** In China, European digits (U+0030..U+0039) are mainly used, though Myanmar digits, (U+1040..U+1049) are also used with slight glyph variants.

၀၁၂၃၄၅၆၇၈၉ Myanmar digits

000 J 2 5 9 6 7 A B Myanmar digits as used by Tai Le in China Apparently a script-specific set of numerals are used by Tai Le writers in Myanmar; these are not known in China, and are not proposed here pending further information.

8826977610Script-specific Tai Le digits

1997	МЛУ		6	လိူဆ်ဂူ၊;	လိုခ6်ရွှၵ်း	7081
ကျေးတိတ် SUN	16	တင်းဂၢခင်း TUE	MED රන්	မေတ်း THURS	သု <sub></sub> ်း FRI	మర్ SAT
SUN	J WOW	IOE		8		
				ຄາ.ພຣິ: <b>1</b>	ຄາຽ, ໝໍ 2	လမ်းသ် 3
	17					
	1					
၈၂.ပစ်း	ໄດ້ຄຸນ: 12	იან: <b>ყ</b> ნ; 13	oi:a 14	ಕ್ಟ್ 15	გგ:ა <b>ა</b> ი 16	၈တ်းမှတ် <b>17</b>
1?	1h 3 gene 19	11 005.305 20	<b>ο</b> 21	ინ,ინ 22	? ონ:ინ; <b>23</b>	ი ისან: <b>24</b>
٨		9	Z 200 200	8 ×	8	10
OOLOGY)		puter ung Aung	(⊗a	lpha co	lor Sepa	ration

Calendar from Myanmar showing script-specific Tai Le digits. The year is given as 2091 (= 1997).

**Punctuation.** Both CJK punctuation and western punctuation are used. Typographically, digits are about the same height and depth as the tall characters  $\{ \}$  and  $\{ \}$ ; in some fonts, the baseline for punctuation is the depth of those characters.

# **Unicode Character Properties**

Spacing letters, category "Lu" (uppercase), bidi category "L" (strong left to right) xx00-xx1D, xx20-xx24

# Bibliography

Coulmas, Florian. 1996. *The Blackwell encyclopedia of writing systems*. Oxford and Cambridge: Blackwell. (s.v. *Dehong writing*, pp. 118-119.) ISBN 0-631-19446-0

าบอยไอทุบอ: มตรปิก าปายาปิอกาปุทอ. มีในงมีหงยในงาบอาบุบปหกาปโอ. ( $Tsa\ va^4\ m\acute{a}^3\ h\acute{o}\ va^3$ : la ta $^6\ m\acute{e}^2\ s\acute{a}^3$  seh va $^2\ xo\ \eta a^3$ . Yi na $^5\ l\acute{a}\ na^5\ mi\ na^5\ su^4\ su^4\ p\acute{a}\ na^2\ se^3$ .) 1997. ISBN 7-5367-1455-6.

# A. Administrative

#### 1. Title

Revised proposal for encoding the Tai Le (Dehong Dai) script in the BMP of the UCS

# 2. Requester's name

Ireland

# 3. Requester type

National Body contribution

# 4. Submission date

2001-03-19

#### 5. Requester's reference

# 6a. Completion

This is a complete proposal.

# 6b. More information to be provided?

No

# **B.** Technical -- General

# 1a. New script? Name?

Yes. Tai Le.

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

No.

# 2. Number of characters

35

# 3. Proposed category

Category A

# 4. Proposed level of implementation and rationale

Level 1 for the new orthography; Level 3 for the older orthography (but the characters used in that orthography are already encoded and are not part of this proposal).

# 5a. Character names included in proposal?

Yes

# 5b. Character names in accordance with guidelines?

Yes

# 5c. Character shapes reviewable?

Yes

# 6a. Who will provide computerized font?

Michael Everson, Everson Typography

# 6b. Font currently available?

Yes

#### 6c. Font format?

TrueType

# 7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?

Yes. See bibliography.

# 7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?

Yes. See below.

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

Yes.

# C. Technical -- Justification

# 1. Contact with the user community?

Yes.

# 2. Information on the user community?

There are approximately 250,000 Tai Le users in China, 120,000 users in Laos, and 72,400 in Myanmar (figures from the SIL *Ethnologue*).

# 3a. The context of use for the proposed characters?

To write the Tai Le language.

3b. Reference

4a. Proposed characters in current use?

Yes

4b. Where?

China, Laos, and Myanmar.

5a. Characters should be encoded entirely in BMP?

Yes.

5b. Rationale

Contemporary use and in keeping with the Roadmap.

6. Should characters be kept in a continuous range?

Yes.

7a. Can the characters be considered a presentation form of an existing character or character sequence?

No.

7b. Where?

7c. Reference

8a. Can any of the characters be considered to be similar (in appearance or function) to an existing character?

No.

8b. Where?

8c. Reference

9a. Combining characters or use of composite sequences included?

No.

9b. List of composite sequences and their corresponding glyph images provided?

No.

10. Characters with any special properties such as control function, etc. included?

No

# D. SC2/WG2 Administrative

To be completed by SC2/WG2

- 1. Relevant SC 2/WG 2 document numbers:
- 2. Status (list of meeting number and corresponding action or disposition)
- 3. Additional contact to user communities, liaison organizations etc.
- 4. Assigned category and assigned priority/time frame

**Other Comments** 

**TABLE XXX - Row x0: TAI LE** 

ADLL		INOW A	). IAI L
	xx0	xx1	xx2
0	4	ষ	Л
1	শ্ব	ข	e
2	n	И	а
3	11	1	v
4	น	Ð	С
5	u	Π	
6	m	Ŋ	
7	m	L	
8	Ŋ	L	
9	U	IJ	
Α	w	Ø	
В	A	[]	
С	ŋ	Γ	
D	u	Ţ	
E	ŋ		
F	ŋ		

G = 00 P = 00

# **TABLE XXX - Row x0: TAI LE**

hex	Name	hex	Name
80 81 82 83 84 85 86 87 88 88 88 88 88 88 89 99 99 99 99 99 99	TAI LE LETTER KA TAI LE LETTER XA TAI LE LETTER YA TAI LE LETTER SA TAI LE LETTER YA TAI LE LETTER YA TAI LE LETTER TA TAI LE LETTER TA TAI LE LETTER LA TAI LE LETTER PA TAI LE LETTER PA TAI LE LETTER PA TAI LE LETTER PA TAI LE LETTER MA TAI LE LETTER HA TAI LE LETTER HA TAI LE LETTER NA TAI LE LETTER SHA TAI LE LETTER SHA TAI LE LETTER TSHA TAI LE LETTER TSH TAI LE LETTER TSH TAI LE LETTER TSH TAI LE LETTER E TAI LE LETTER E TAI LE LETTER E TAI LE LETTER E TAI LE LETTER TOO TAI LE LETTER TO TAI LE LETTER O TAI LE LETTER O TAI LE LETTER O TAI LE LETTER O TAI LE LETTER TONE-3 TAI LE LETTER AU TAI LE LETTER TONE-3 TAI LE LETTER TONE-5 TAI LE LETTER TONE-6 (This position shall not be used)		