

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, **ᩉ᩠ᩅᩢᩣᨾᩯ** [tai² lə⁶] (in older orthography **ᩉ᩠ᩅᩢᩣᨾᩯ**). 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.

ᩉ	ka	ᩉ	ka	ᩉ᩠ᩅ	ki	ᩉ᩠ᩅ	ki
ᩉ᩠ᩅ	ka ²	ᩉ᩠ᩅ	ka ²	ᩉ᩠ᩅ᩠ᩅ	ki ²	ᩉ᩠ᩅ᩠ᩅ	ki ²
ᩉ᩠ᩅ	ka ³	ᩉ᩠ᩅ	ka ³	ᩉ᩠ᩅ᩠ᩅ	ki ³	ᩉ᩠ᩅ᩠ᩅ	ki ³
ᩉ᩠ᩅ	ka ⁴	ᩉ᩠ᩅ	ka ⁴	ᩉ᩠ᩅ᩠ᩅ	ki ⁴	ᩉ᩠ᩅ᩠ᩅ	ki ⁴
ᩉ᩠ᩅ	ka ⁵	ᩉ᩠ᩅ	ka ⁵	ᩉ᩠ᩅ᩠ᩅ	ki ⁵	ᩉ᩠ᩅ᩠ᩅ	ki ⁵
ᩉ᩠ᩅ	ka ⁶	ᩉ᩠ᩅ	ka ⁶	ᩉ᩠ᩅ᩠ᩅ	ki ⁶	ᩉ᩠ᩅ᩠ᩅ	ki ⁶

ထိုအခါ၌ အသံလဲ၍ "ဝ", "ဂါဝ" ဟု ခေါ်တွင်သည်။
 ထိုအခါ၌ အသံလဲ၍ "ဝ", "ဂါဝ" ဟု ခေါ်တွင်သည်။

အက္ခရာ	အသံလဲ၍ ခေါ်တွင်သည်						အသံလဲ၍ ခေါ်တွင်သည်
ဝ	ဝ	ဝ	ဝ	ဝ	ဝ	ဝ	ဝ ဝ ဝ ဝ
	ဝ	ဝ	ဝ	ဝ	ဝ	ဝ	
	ဝ	ဝ	ဝ	ဝ	ဝ	ဝ	
	ဝ	ဝ	ဝ	ဝ	ဝ	ဝ	

ထိုအခါ၌ အသံလဲ၍ "ဝ", "ဂါဝ" ဟု ခေါ်တွင်သည်။

အက္ခရာ	အသံလဲ၍ ခေါ်တွင်သည်						အသံလဲ၍ ခေါ်တွင်သည်
ဝ	ဝ	ဝ	ဝ	ဝ	ဝ	ဝ	ဝ ဝ ဝ ဝ

ထိုအခါ၌ အသံလဲ၍ "ဝ", "ဂါဝ" ဟု ခေါ်တွင်သည်။

အက္ခရာ	အသံလဲ၍ ခေါ်တွင်သည်						အသံလဲ၍ ခေါ်တွင်သည်
ဝ			ဝ				ဝ
ဝ					ဝ		"

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.

၀ ၁ ၂ ၃ ၄ ၅ ၆ ၇ ၈ ၉ Myanmar digits

၀ ၁ ၂ ၃ ၄ ၅ ၆ ၇ ၈ ၉ 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.

၀ ၁ ၂ ၃ ၄ ၅ ၆ ၇ ၈ ၉ Script-specific Tai Le digits

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

	xx0	xx1	xx2
0	ᄁ	ᄂ	ᄃ
1	ᄄ	ᄅ	ᄆ
2	ᄇ	ᄈ	ᄉ
3	ᄊ	ᄋ	ᄌ
4	ᄍ	ᄎ	ᄏ
5	ᄐ	ᄑ	
6	ᄒ	ᄓ	
7	ᄔ	ᄕ	
8	ᄖ	ᄗ	
9	ᄘ	ᄙ	
A	ᄚ	ᄛ	
B	ᄜ	ᄝ	
C	ᄞ	ᄟ	
D	ᄠ	ᄡ	
E	ᄢ		
F	ᄣ		

G = 00
P = 00

