

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

Doc Type: Working Group Document
Title: Proposal to add minority characters to Myanmar script
Source: Martin Hosken
Action: For consideration by JTC1/SC2/WG2
Date: 2011-04-19

This prerelease version contains the main proposal but should not be considered final since further input is being sought from the user communities. It is expected that representative glyph shapes may change and other examples added.

Introduction

This proposal is to add 17 extra characters to the Myanmar script for the Tai Laing and Shwe Palaung languages.

The Tai Laing are a language group of about 100,000 speakers living along the Irrawaddy River in Myanmar. The writing system is part of their history that has not completely died out and there is interest in reviving it. While the script is not taught formally in schools, it is taught during school breaks. The Shwe Palaung have about 200,000 speakers, primarily in Shan State, Myanmar. 20% of Shwe Palaung are literate in Shwe Palaung with ongoing literacy development happening indigenously. The orthography has been in development since the 1930s with the latest revision occurring in the 1980s when the proposed characters were added.

Charts: The proposed characters are to be added to two existing Myanmar Extended blocks. This fills out Myanmar Extended-A and adds to Myanmar Extended-B.

U+AA70

	AA7
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
A	
B	
C	◌̄
D	◌̄-
E	ချ
F	ဆချ

Myanmar Extended-A

AA7C	◌̄	MYANMAR SIGN TAI LAING TONE-2
AA7D	◌̄-	MYANMAR SIGN TAI LAING TONE-5
AA7E	ချ	MYANMAR LETTER SHWE PALAUNG CHA
AA7F	ဆချ	MYANMAR LETTER SHWE PALAUNG SHA

U+A9E0

	A9E	A9F
0		၀
1		၁
2		၂
3		၃
4		၄
5		၅
6		၆
7	၇	၈
8	၉	၉
9	၀	၀
A		
B		
C		
D		
E		
F		

Myanmar Extended-B

Consonants

A9E7	၀	MYANMAR LETTER TAI LAING SHA
A9E8	၁	MYANMAR LETTER TAI LAING NYA
A9E9	၂	MYANMAR LETTER TAI LAING FA

Digits

A9F0	၀	MYANMAR TAI LAING DIGIT ZERO
A9F1	၁	MYANMAR TAI LAING DIGIT ONE
A9F2	၂	MYANMAR TAI LAING DIGIT TWO
A9F3	၃	MYANMAR TAI LAING DIGIT THREE
A9F4	၄	MYANMAR TAI LAING DIGIT FOUR
A9F5	၅	MYANMAR TAI LAING DIGIT FIVE
A9F6	၆	MYANMAR TAI LAING DIGIT SIX
A9F7	၇	MYANMAR TAI LAING DIGIT SEVEN
A9F8	၈	MYANMAR TAI LAING DIGIT EIGHT
A9F9	၉	MYANMAR TAI LAING DIGIT NINE

Rationale

The various subgroups of characters will be considered separately, in encoding order.

Tai Laing Tone Marks. Tai Laing has 5 tone marks. Of these, 3 are already encoded and this proposal adds the remaining two.

ဟ်းင်ဟွက်ဟွက်ဟ်းင်

—	—	—	—	—	—
ဟ်းင်ကပ်	ဟ်းင်တ့်	ဟ်းင်လှိုင်	လှမ်လှ	လှမ်တ့်	ဟ်းင်ဟိုင်-

As can be seen in the example above, the tone marks will position before an interacting vowel in the same diacritic space. The stored sequence, though, is still with the tone mark stored finally. With reference to UTN#11 Diacritic Storage Order, the tone marks are added to the Visarga slot class.

AA7C;MYANMAR SIGN TAI LAING TONE-2;Mn;0;NSM;;;;N;;;;
 AA7D;MYANMER SIGN TAI LAING TONE-5;Mn;0;NSM;;;;N;;;;

Another example of the use of these characters:

ကံဟံတိ-ထံ
 ပီ-မံဟံလှ-
 တိ-ထံတၢ-ဟိုင်း
 ဟော်ဟံထံလှး
 -မံဟိုင်ပီ-ဟိုင်း။

Shwe Palaung Consonants: The two proposed characters for Shwe Palaung have a visual representation that is very close to an existing sequence:

ချ	AA7E	ချ	1001 103B 103E
ဆျ	AA7F	ဆျ	1006 103B 103E

The difficulty comes if we introduce a medial wa into the sequence:

ချ	AA7E 103D	ချ	1001 103B 103D 103E
ဆျ	AA7F 103D	ဆျ	1006 103B 103D 103E

The medial is added into the middle of the sequence representing the consonant. This is a problem for data entry. For sorting, while the example dictionary here sorts them as though the components are medials, being distinct consonants, these characters should have their own consonantal position. The encoding of these characters follows in the tradition of encoding such characters atomically. Other examples are: U+106F (c.f. U+101F U+103E), U+1070 (c.f. U+1003 U+103E), U+107E (c.f. U+107D U+103E).

AA7E;MYANMAR LETTER SHWE PALAUNG CHA;Lo;0;L;;;;N;;;;
 AA7F;MYANMAR LETTER SHWE PALAUNG SHA;Lo;0;L;;;;N;;;;

It is also worth noting that the rendering behaviour with medial wa may be different to the way the sequence is rendered. Some styles prefer them to be the same, others for them to be different.

ဆွဲ၊
ကတံမ်ဂိုန်ခဲဆွဲ၊
ကတံမ်ဆိုရ်ဆွဲ၊
ဆွဲဝ်
ကညောမ်အိတ်ဆွဲဝ်
ဒေပါင်ကတဲ
ရွာ
ရွာ
ချူရွာချူချီ
ရွာ၊
ဲရွာ၊
(ကိုင်ကွာ၊ကွာ၊ဆနာ၊
ပါင်ဆွဲ၊ ဝါင်

ဆဲညို၊
ကိုရ်ဆို
ဆိုညို၊ညို၊
ဆျ
ကတံမ်ဆျ
ဆျ
ဆျ
ကိုရ်ဆျ
ပို၊ပိုင်ကောဒေကိုရ်ဆျ
မည်ပိုင်၊ကိုရ်ဆျဘိန်ပ်
လဲဆျ
ဆျာ
ဘိုပ်ဆျာ
ဆျာ၊
မိအဆျာ၊အော
ဆျ။
ဘိုပ်ဆျ။တကိုင်ဒေ
(ဘိုပ်ပုတကြောင်ဒေ၊
(ဘိုပ်ညိုတကတို၊ဒေ
ဆံင်
ဆံင်
ဆံပနသ်

For IDN purposes, the sequence and the unit should be considered confusable.

Tai Laing Consonants: The proposed characters are listed as part of the alphabet for Tai Laing.

မေ-ကု-လိက်တံးလိင်-ဟမ်ပကုပ်တုင်-

က ka	ခ ka	င ga
လ sa	ယ su	ဟ na
တ ta	ထ ta	န na
ပ pa	ဖ pa	ဖ fa
မ ma	ယ ja	လ ra
လ la	ဝ wa	ဟ ha
	ဟ za	

The proposed characters have been circled with all the other characters already supported in the UCS.

A9E7;MYANMAR LETTER TAI LAING SHA;Lo;0;L;;;N;;;;;
 A9E8;MYANMAR LETTER TAI LAING NYA;Lo;0;L;;;N;;;;;
 A9E9;MYANMAR LETTER TAI LAING FA;Lo;0;L;;;N;;;;;

Tai Laing Digits: Tai Laing has its own set of digits, which are proposed here:

လမ်းတို့ဟာဂ်မာ်

လမ်းဖွင့်	တံးလိုင်-	တံး တူဟိုင်းဟိုင်း	ဟာဂ်းကလိက်
ပျိုင်	၀	၀	၀
လိုင်	၁	၁	၁
ဟွိုင်	၂	၂	၂
ဟာမ်	၃	၃	၃
ဟီ	၄	၄	၄
ဟု	၅	၅	၅
ဟူတ်	၆	၆	၆
လိတ်	၇	၇	၇
ပျိတ်	၈	၈	၈
ကွပ်	၉	၉	၉
ဟိပ်	၁၀	၁၀	၁၀
ဟိပ်ဟာတ်	၁၁	၁၁	၁၁
ဟာဂ်းလိုင်-	၂၀	၂၀	၂၀

The columns are: character name in Tai Laing, Tai Laing digit, Shan digit (from U+1090-U+1099), Arabic digit.

A9F0;MYANMAR TAI LAING DIGIT ZERO;Nd;0;L;;0;0;0;N;;;;;
 A9F1;MYANMAR TAI LAING DIGIT ONE;Nd;0;L;;1;1;1;N;;;;;
 A9F2;MYANMAR TAI LAING DIGIT TWO;Nd;0;L;;2;2;2;N;;;;;
 A9F3;MYANMAR TAI LAING DIGIT THREE;Nd;0;L;;3;3;3;N;;;;;
 A9F4;MYANMAR TAI LAING DIGIT FOUR;Nd;0;L;;4;4;4;N;;;;;
 A9F5;MYANMAR TAI LAING DIGIT FIVE;Nd;0;L;;5;5;5;N;;;;;
 A9F6;MYANMAR TAI LAING DIGIT SIX;Nd;0;L;;6;6;6;N;;;;;
 A9F7;MYANMAR TAI LAING DIGIT SEVEN;Nd;0;L;;7;7;7;N;;;;;
 A9F8;MYANMAR TAI LAING DIGIT EIGHT;Nd;0;L;;8;8;8;N;;;;;
 A9F9;MYANMAR TAI LAING DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;

Sort Order

The default sort order is integrated into the existing default sorting for the Myanmar script. Given that all Myanmar based languages require complex sort tailoring, the precise values here can be somewhat arbitrary. The sort order information given here also includes the other characters from the Myanmar Extended-B block as specified in N3906 (L2/10-345).

```
&1003 MYANMAR LETTER GHA < A9E0 MYANMAR LETTER SHAN GHA
&1006 MYANMAR LETTER CHA < A9E1 MYANMAR LETTER SHAN CHA
&AA62 MYANMAR LETTER KHAMTI CHA < AA7E MYANMAR LETTER SHWE PALAUNG CHA
&105B MYANMAR LETTER MON JHA < A9E2 MYANMAR LETTER SHAN JHA
&1061 MYANMAR LETTER SGAW KAREN SHA < AA7F MYANMAR LETTER SHWE PALAUNG SHA < A9E7
MYANMAR LETTER TAI LAING SHA
&AA65 MYANMAR LETTER KHAMTI NYA < A9E8 MYANMAR LETTER TAI LAING NYA
&106E MYANMAR LETTER EASTERN PWO KAREN NNA < A9E3 MYANMAR LETTER SHAN NNA
&108E MYANMAR LETTER RUMAI PALAUNG FA < A9E9 MYANMAR LETTER TAI LAING FA
&1018 MYANMAR LETTER BHA < A9E4 MYANMAR LETTER SHAN BHA
&1068 MYANMAR VOWEL SIGN WESTERN PWO KAREN UE < A9E5 MYANMAR SIGN SHAN SAW
&108D MYANMAR SIGN SHAN COUNCIL EMPHATIC TONE << AA7C MYANMAR SIGN TAI LAING TONE-2
&1089 MYANMAR SIGN SHAN TONE-5 < AA7D MYANMAR SIGN TAI LAING TONE-5
&AA70 MYANMAR MODIFIER LETTER KHAMTI REDUPLICATION < A9E6 MYANMAR MODIFIER LETTER
SHAN REDUPLICATION
```

MYANMAR TAI LAING DIGIT characters are sorted following their corresponding MYANMAR SHAN DIGIT characters.

Bibliography

Hosken, Martin “Representing Myanmar in Unicode” (Unicode Technical Note 11, version 3).

ထန်ငဝ် ငိုန်လုံတိုးလင်-

SonNgaw Phawnla TaiLaing

O Thuwa Palaung-Burmese Dictionary (Namhfan, 2003)

Acknowledgements

Thanks go to Payap University Linguistics Institute, Chiang Mai, Thailand, under whose auspices this work is done.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹.

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.
 See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	<i>Lao Extensions</i>
2. Requester's name:	<i>Martin Hosken</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>
4. Submission date:	<i>21/04/11</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<input checked="" type="checkbox"/>
(or) More information will be provided later:	<input type="checkbox"/>

B. Technical – General

1. Choose one of the following:	
a. This proposal is for a new script (set of characters):	
Proposed name of script:	
b. The proposal is for addition of character(s) to an existing block:	<input checked="" type="checkbox"/>
Name of the existing block:	<i>Myanmar Extended-A, Myanmar Extended-B</i>
2. Number of characters in proposal:	<i>17</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary <input checked="" type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>
D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>
4. Is a repertoire including character names provided?	<i>yes</i>
a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?	<i>yes</i>
b. Are the character shapes attached in a legible form suitable for review?	<i>yes</i>
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?	<i>SIL</i>
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:	
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<i>yes</i>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<i>yes</i>
7. Special encoding issues:	
Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?	<i>yes</i>

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N3102-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	<i>no</i>
If YES explain	
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?	<i>yes</i>
If YES, with whom?	<i>local experts</i>
If YES, available relevant documents:	<i>see bibliography</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	<i>yes</i>
Reference:	<i>this document</i>
4. The context of use for the proposed characters (type of use; common or rare)	<i>common</i>
Reference:	
5. Are the proposed characters in current use by the user community?	<i>yes</i>
If YES, where? Reference:	<i>see bibliography</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	<i>yes</i>
If YES, is a rationale provided?	<i>yes</i>
If YES, reference:	<i>addition to existing BMP blocks</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<i>no</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	<i>yes</i>
If YES, is a rationale for its inclusion provided?	<i>yes</i>
If YES, reference:	<i>this document</i>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	<i>no</i>
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?	<i>yes</i>
If YES, is a rationale for its inclusion provided?	<i>yes</i>
If YES, reference:	<i>this document</i>
11. Does the proposal include use of combining characters and/or use of composite sequences?	<i>yes</i>
If YES, is a rationale for such use provided?	<i>yes</i>
If YES, reference:	<i>this document</i>
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	<i>no</i>
If YES, reference:	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	<i>no</i>
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility character(s)?	<i>no</i>
If YES, is the equivalent corresponding unified ideographic character(s) identified?	
If YES, reference:	