L2/12-012
(replaces L2/11-130)

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:	**2011-05-23**

Introduction

This proposal is to add 28 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.

	AA7
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
A	
B	
C	◌̄
D	◌–
E	ဩ
F	ဪ

AA7C	◌̄	MYANMAR SIGN TAI LAING TONE-2
AA7D	◌–	MYANMAR SIGN TAI LAING TONE-5
AA7E	ဩ	MYANMAR LETTER SHWE PALAUNG CHA
AA7F	ဪ	MYANMAR LETTER SHWE PALAUNG SHA

Myanmar Extended-B

	A9E	A9F
0		၀
1		ၡ
2		ၢ
3		ၣ
4		ၤ
5		ၥ
6		ၦ
7	ၧ	ၧ
8	ၨ	ၨ
9	ၩ	ၩ
A	ၪ	ၪ
B	ၫ	ၫ
C	ၬ	ၬ
D	ၭ	ၭ
E	ၮ	ၮ
F	ၯ	

Consonants

A9E7	ၧ	MYANMAR LETTER TAI LAING NYA
A9E8	ၨ	MYANMAR LETTER TAI LAING FA
A9E9	ၩ	MYANMAR LETTER TAI LAING GA
A9EA	ၪ	MYANMAR LETTER TAI LAING GHA
A9EB	ၫ	MYANMAR LETTER TAI LAING JA
A9EC	ၬ	MYANMAR LETTER TAI LAING JHA
A9ED	ၭ	MYANMAR LETTER TAI LAING DDA
A9EE	ၮ	MYANMAR LETTER TAI LAING DDHA
A9EF	ၯ	MYANMAR LETTER TAI LAING NNA

Digits

A9F0	၀	MYANMAR TAI LAING DIGIT ZERO
A9F1	ၡ	MYANMAR TAI LAING DIGIT ONE
A9F2	ၢ	MYANMAR TAI LAING DIGIT TWO
A9F3	ၣ	MYANMAR TAI LAING DIGIT THREE
A9F4	ၤ	MYANMAR TAI LAING DIGIT FOUR
A9F5	ၥ	MYANMAR TAI LAING DIGIT FIVE
A9F6	ၦ	MYANMAR TAI LAING DIGIT SIX
A9F7	ၧ	MYANMAR TAI LAING DIGIT SEVEN
A9F8	ၨ	MYANMAR TAI LAING DIGIT EIGHT
A9F9	ၩ	MYANMAR TAI LAING DIGIT NINE

Consonants

A9FA	ၪ	MYANMAR LETTER TAI LAING LLA
A9FB	ၫ	MYANMAR LETTER TAI LAING DA
A9FC	ၬ	MYANMAR LETTER TAI LAING DHA
A9FD	ၭ	MYANMAR LETTER TAI LAING BA
A9FE	ၮ	MYANMAR LETTER TAI LAING BHA

3

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.
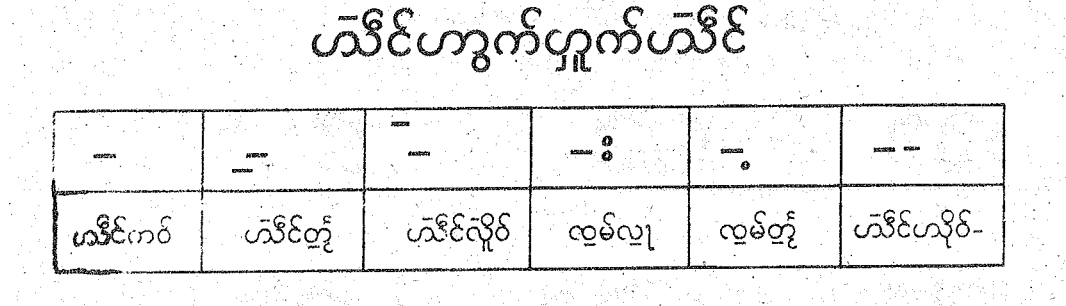

Figure 1. Tai Laing tone marks

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;;;;;
```
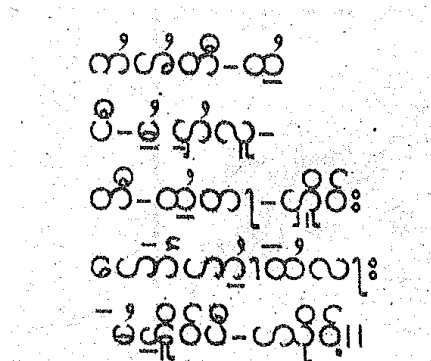
Another example of the use of these characters:

Figure 2. Use of Tai Laing tone marks in conjuction with other diacritics.

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

4

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.

Figure 3. Shwe Palaung consonant sorting.

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.

The proposed characters have been circled in figure 4, with all the other characters already supported in the UCS. Notice that the labelling of the sa and s^{h}a characters is wrong when compared with the Pali based shiksha. In addition the shape of the s^{h}a belies its underlying encoding of U+AA6C MYANMAR LETTER KHAMTI SA, as can be seen in figure 5.

5

မေ်-ကၡ-လီက်ၲတံးလ်ၢင်-ဟ�won်ပ်ကၠဝ်တုၢ်-

Figure 4. Tai Laing modern alphabet

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

Tai Laing Pali Consonants: As with most Myanmar script based writing systems, Tai Laing adds character support for the Pali language.

Figure 5. Shiksha showing Devanagari, Burmese, Tai Laing and Roman scripts.

Notice that the character for pha is the same as U+A9E4 MYANMAR LETTER SHAN BHA and that likewise the code for bha is based on that shape.

```
A9E9;MYANMAR LETTER TAI LAING GA;Lo;0;L;;;;;N;;;;;
A9EA;MYANMAR LETTER TAI LAING GHA;Lo;0;L;;;;;N;;;;;
A9EB;MYANMAR LETTER TAI LAING JA;Lo;0;L;;;;;N;;;;;
A9EC;MYANMAR LETTER TAI LAING JHA;Lo;0;L;;;;;N;;;;;
A9ED;MYANMAR LETTER TAI LAING DDA;Lo;0;L;;;;;N;;;;;
A9EE;MYANMAR LETTER TAI LAING DDHA;Lo;0;L;;;;;N;;;;;
A9EF;MYANMAR LETTER TAI LAING NNA;Lo;0;L;;;;;N;;;;;

A9FA;MYANMAR LETTER TAI LAING LLA;Lo;0;L;;;;;N;;;;;
A9FB;MYANMAR LETTER TAI LAING DA;Lo;0;L;;;;;N;;;;;
A9FC;MYANMAR LETTER TAI LAING DHA;Lo;0;L;;;;;N;;;;;
A9FD;MYANMAR LETTER TAI LAING BA;Lo;0;L;;;;;N;;;;;
A9FE;MYANMAR LETTER TAI LAING BHA;Lo;0;L;;;;;N;;;;;
```

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

လၢ:တၠုၞၠ်	တၠ:လၠၠင်-	တၠ: ကူတႂုၠင်:ဟႂုၠင်:	ဟၠၠၠ်:ကလၠၠၠၠ်
ပၠ̊င်	0	0	0
လၠၠင်	၉	l	1
ဟၠၠႂၠင်	၄	?	2
ဟၠၠၠ̄ၠ်	၈	b	3
ဟၠၠ̊	၃	Λ	4
ၮၠၠ	6	d	5
ၮၠၠၠ်	၉	q	6
လ̄ၠၠ်	၇	Z	7
ပၠ̊ၠ်	၆	X	8
ကၠ̄ၠ်	၈	X	9
ဟၠ̊ၠ်	၉0	10	10
ဟၠ̊ၠ်ဟၠ̊ၠ်	၉၉	11	11
ဟၠၠၠ်:လၠၠင်-	၈0	၇0	20

Figure 6. Tai Laing Digits. The colums 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 < A9EA MYANMAR LETTER TAI LAING 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
&AA64 MYANMAR LETTER KHAMTI JHA < A9EC MYANMAR LETTER TAI LAING JHA
&1061 MYANMAR LETTER SGAW KAREN SHA < AA7F MYANMAR LETTER SHWE PALAUNG SHA
&AA65 MYANMAR LETTER KHAMTI NYA < A9E7 MYANMAR LETTER TAI LAING NYA
&106E MYANMAR LETTER EASTERN PWO KAREN NNA < A9E3 MYANMAR LETTER SHAN NNA < A9EE MYANMAR
      LETTER TAI LAING NNA
&108E MYANMAR LETTER RUMAI PALAUNG FA < A9E8 MYANMAR LETTER TAI LAING FA
&1018 MYANMAR LETTER BHA < A9E4 MYANMAR LETTER SHAN BHA < A9FE MYANMAR LETTER TAI LAING 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
&AA60 MYANMAR LETTER KHAMTI GA < A9E9 MYANMAR LETTER TAI LAING GA
&105B MYANMAR LETTER MON JA < A9EB MYANMAR LETTER TAI LAING JA
&AA68 MYANMAR LETTER KHAMTI DDA < A9ED MYANMAR LETTER TAI LAING DDA
&AA69 MYANMAR LETTER KHAMTI DDHA < A9EE MYANMAR LETTER TAI LAING DDHA
&100F MYANMAR LETTER NNA < A9EF MYANMAR LETTER TAI LAING NNA
&1020 MYANMAR LETTER LLA < A9FA MYANMAR LETTER TAI LAING LLA
&107B MYANMAR LETTER SHAN DA < A9FB MYANMAR LETTER TAI LAING DA
&1013 MYANMAR LETTER DHA < A9FC MYANMAR LETTER TAI LAING DHA
&107F MYANMAR LETTER SHAN BA < A9FD MYANMAR LETTER TAI LAING BA
```

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).

ထၵ်ႄ္ငဝ်   ၐ္ုိၵ်လ္ုႄတႃးလ္ိင်-

SonNgaw   Phawnla TaiLaing

O Thuwa   Palaung-Burmese Dictionary (Namhfan, 2003)

Acknowledgements

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
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:**	*Lao Extensions*
2. Requester's name:	*Martin Hosken*
3. Requester type (Member body/Liaison/Individual contribution):	*Individual contribution*
4. Submission date:	*21/04/11*
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	*X*
(or) More information will be provided later:	

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: *X*
       Name of the existing block: *Myanmar Extended-A, Myanmar Extended-B*
2. Number of characters in proposal: *28*
3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary	*X*	B.1-Specialized (small collection)	
C-Major extinct		B.2-Specialized (large collection)	
		D-Attested extinct	
		E-Minor extinct	
F-Archaic Hieroglyphic or Ideographic		G-Obscure or questionable usage symbols	

4. Is a repertoire including character names provided?   *yes*
   a. If YES, are the names in accordance with the "character naming guidelines"
       in Annex L of P&P document?   *yes*
   b. Are the character shapes attached in a legible form suitable for review?   *yes*
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for
   publishing the standard?   *SIL*
   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?   *yes*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources)
   of proposed characters attached?   *yes*
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)?   *yes*

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? *no*
      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.)? *yes*
            If YES, with whom? *local experts*
            If YES, available relevant documents: *see bibliography*

3. Information on the user community for the proposed characters (for example:
      size, demographics, information technology use, or publishing use) is included? *yes*
      Reference: *this document*

4. The context of use for the proposed characters (type of use; common or rare) *common*
      Reference:

5. Are the proposed characters in current use by the user community? *yes*
      If YES, where? Reference: *see bibliography*

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
      in the BMP? *yes*
            If YES, is a rationale provided? *yes*
            If YES, reference: *addition to existing BMP blocks*

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? *no*

8. Can any of the proposed characters be considered a presentation form of an existing
      character or character sequence? *yes*
            If YES, is a rationale for its inclusion provided? *yes*
            If YES, reference: *this document*

9. Can any of the proposed characters be encoded using a composed character sequence of either
      existing characters or other proposed characters? *no*
            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? *yes*
            If YES, is a rationale for its inclusion provided? *yes*
            If YES, reference: *this document*

11. Does the proposal include use of combining characters and/or use of composite sequences? *yes*
      If YES, is a rationale for such use provided? *yes*
            If YES, reference: *this document*
      Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? *no*
            If YES, reference:

12. Does the proposal contain characters with any special properties such as
       control function or similar semantics? *no*
            If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility character(s)? *no*
      If YES, is the equivalent corresponding unified ideographic character(s) identified?
            If YES, reference: