ISO/IEC JTC1/SC2/WG2 N3423R L2/08-181 2008-04-29

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 Khamti Shan Characters to the Myanmar Blocks

Source: Martin Hosken

Action: For consideration by JTC1/SC2/WG2

References:

Date: 2008-04-29

Introduction: Khamti Shan is spoken by 50,000 people in Myanmar and India. The language has a long literary tradition which was lost as people forgot their script. In the last 5 years the script has undergone a resurgence in a new form and after a pilot literacy programme is to be taught throughout the region. The new script has been accepted by all the Khamti and looks to have a strong future ahead of it. Aiton and Phake are smaller language communities of around 2,000 each.

Khamti Shan bears a superficial resemblence to the Myanmar script using many of the same shapes, but adds stylistic dots to the glyphs. The question arises whether this script should be unified with the Myanmar script or be given its own block. This proposal unifies these writing systems with the Myanmar script for the following reasons:

- It is acknowledged that the dots are purely stylistic and can vary in size considerably.
- The various writing systems will place dots in different places on what are acknowledged to be the same character. For example compare the Aiton and Phake forms for ca and pha.
- The level of overlap between characters in Khamti Shan script and those already encoded in the Myanmar block, assuming that the dots are stylistic, is 80%.
- The community requested unification with Myanmar script. This arises from a realisation that the dots are stylistic and may change over time even perhaps disappearing all together.

Therefore, this proposal is to add the extra characters needed to support modern Khamti Shan to the Myanmar block. Khamti, Aiton and Phake are based on Shan and as such follow the Myanmar model of encoding including diacritic vowels, medial consonants and even stacking consonants.

Khamti Shan

The Khamti Shan writing system proposed for encoding is the revised Khamti Shan in contrast to the historic writing system which is all but lost and of academic interest. Another revision was made in India and is documented by Stephen Morey. But this Indian writing system did not gain popularity and the wider Khamti community has met and decided to use the revisions described here.

Consonants: The consonants in Khamti Shan are listed below along with their unified values where unification is appropriate. The glyphs shown are Khamti style to be compared to the Burmese style in the existing charts. The proposed charts will use Burmese style glyphs, in contrast.

က	ຄ	c	ဢ	ಏ	ၯ	တ	∞	ଶ	ပ
k	kh	ng	c	s	nj	t	th	n	p
U+1000	U+1075	U+1004	U+AA60	U+AA61	U+AA62	U+1010	U+1011	U+AA63	U+1015

က	೪	ယ	N	ಣ	0	႒	ဢ	က	€
ph	m	у	r	1	w	h	?	x	z
U+1078	U+1019	U+101A	U+AA64	U+101C	U+101D	U+AA65	U+1022	U+AA66	U+AA67

O O	ย	Ø	න
d	b	f	th
U+107B	U+107F	U+AA68	U+1080

U+AA62 (MYANMAR LETTER KHAMTI NYA), while behaving the same as the corresponding Shan character (U+107A MYANMAR LETTER SHAN NYA) has a significantly different glyph shape with all the key features of cup direction and leg direction, inverted.

Vowels: Like the consonants, there are various vowel characters used, often in combination and all of them may be unified with other characters in the Myanmar block. Diacritics are shown in conjunction with wa (U+101D MYANMAR LETTER WA)

÷	1	1	°°	8 O	ε O	00			ေ
-ai	-aa	-aa	-i	-ii	-au	-aai	-u	-uu	-ae
U+108A	U+1062	U+1083	U+102D	U+102E	U+1085	U+1032	U+102F	U+1030	U+1031

ε	000	ò
-e	-aue	-am
U+1084	U+1082 U+103A	U+1038

Tones: There are 6 tone marks in Khamti Shan. Although these glyphs are filled in, they are the same characters as the Shan tones, for the most part, with two additions.

0.	٥.	0'	0,	0;	0:
tone 1	tone 2	tone 3	tone 4	tone 5	tone 6
U+109A	U+1089	U+109B	U+1087	U+1088	U+1038

Tone 1 is not the same character as U+108B (MYANMAR SIGN SHAN COUNCIL TONE-2). Khamti Shan is a closely related language to Shan and the use of the raised dot is considered different and not a stylistic variant of U+108B, just as U+108B is not a stylistic variant of U+1089 (MYANMAR SIGN SHAN TONE-5).

Digits: Khamti Shan uses the Shan digits from the range U+1090 .. U+109A.

Miscellaneous: Various other characters are used.

٦		00	o	I	II	ಆ
-у	-r	-w	reduplication	comma pause	period pause	exclamation
U+103E	U+103C	U+103D	U+AA69	U+104A	U+104B	U+109F

U+AA69 (MYANMAR SIGN KHAMTI REDUP) is functionally equivalent to U+0E46 (THAI CHARACTER MAIMAYOK) and is spacing.

Subjoined characters: Khamti Shan does not use subjoined characters.

Line Breaking and segmentation: Khamti Shan follows Shan and therefore Burmese models for line breaking and also for cluster and syllable segmentation.

Sorting: Khamti Shan follows a Burmese model for sorting with final consonants taking precedence over vowels.

Aiton and Phake

The Aiton and Phake scripts are very closely related and can be analysed together. Top row Aiton style, second row Phake.

m	S	c	wo	w	సు	တ	∞	23	O
က	S	c	m	w	సు	စာ	∞	n	O
k	kh	ng	c	S	ny	t	th	n	p
U+1000	U+1075	U+1004	U+AA60	U+AA61	U+107A	U+1010	U+1011	U+AA63	U+1015

ന	&	ω	<u></u>	ಎ	Ð	જ	n
က	8	ω	ණ	ಎ	Ð	<i>s</i> >	က
ph	m	у	r	1	w	h	?
U+1078	U+1019	U+101A	U+AA6A	U+101C	U+101D	U+AA65	U+1022

U+AA6A (MYANMAR LETTER AITON RA) while, theoretically this could be encoded as U+101C (MYANMAR LETTER LA) U+103C (MYANMAR CONSONANT SIGN MEDIAL RA) this would not reflect the unitary nature of this character.

U+107A (Myanmar Letter shan Nya) is used rather than following the Khamti U+AA62 (Myanmar Letter Khamti Nya) since the character shape follows Shan rather than Khamti.

Medials: Medials are taken from Mon.

<u>დ</u>	ಟ್	ભ
6	್ತ	ભ
kr-	kw-	ky-
U+103C	U+105E	U+103B

Subjoined Consonants: Aiton and Phake have a system of subjoining consonants to chain syllables in a polysyllabic word. This system follows that of Burmese and is encoded in the same way with U+1039 (MYANMAR SIGN VIRAMA) followed by the code of the consonant being subjoined. The following characters may take a subjoined form which is the same shape as the base character but smaller, when following U+1039 (MYANMAR SIGN VIRAMA): U+1000, U+AA60, U+1010, U+1011, U+1015, U+101A, U+101C. No other subjoined characters are known in Aiton and Phake.

Vowels: Vowels follow Shan for the most part. The examples here are with a ka (U+1000 MYANMAR LETTER KA) initial

M	ယႏ	ಣಿ	u	ന	ന്നേ	ကုစ်
സ	ಟ :	ಣಿ	53	ന	ണ്	ကုစ်
-aa	-a	-ii	-uu	-e	-aw	-0
U+1083	U+109C	U+102E	U+1030	U+1031	U+1031 U+1083	U+102F U+101D U+103A

ભ ိ 6	<u> గ</u> ోక	ભૄર્દ	നൂട്	ကိုင်
ကိုစ်	ශි දි	ભૄ	നൂര്	ကိုင်
-ue	-i-	-u-	-aw-	-ue-
U+102D U+102F U+101D U+103A	U+102D	U+102F	U+103D	U+102D U+102F

Dipthongs: Aiton and Phake have a number of dipthongs

					<u> </u>		
ಣೆ	ಣಿ	ಣಿ	<i>ત્</i>	ന്റ	જે:	ર્ભ	ರ್ನ
ಣೆ	ರ್	ಚ	ಭ	ಞ	rq;	ಗ್ಗ	දිදී
-am	-em	-ai	-awi	-oi	-ui	-au	-aue
U+1036	U+103A U+1036	U+109D		U+103D U+1031	U+102D U+102F U+109A	U+103D U+103A	U+105E U+103A

Ligatures: Aiton and Phake have an interesting set of ligature symbols that follow the same principles used for U+109E (MYANMAR SYMBOL SHAN ONE) and U+109F (MYANMAR SYMBOL SHAN EXCLAMATION). Thus they are symbols that constitute a word in their own right and do not take diacritics.

Á	Ħ	?
Å,	ત્ર	کاں
'EXCL'	'one'	'two'
U+AA6B	U+AA6C	U+AA6D

Tones: Traditionally tones are not marked in Aiton and Phake, although short -a can be used as a type of tone marker. All proposed patterns for adding tone marking to Aiton and Phake would be within the scope of tone marks covered by Shan and Khamti Shan.

Samples

နော့ မိ'ကူ' ရွင် နားဝ် မန်းယုံ, အူဝ်; ပီ န်းငံး မူးကျူး? တေု့ ီ ပီကပ် ့ဆ်ုံးယဝ်, လူင်၍း။ ကဝ်;မဲ' အုက် ့ဆုံ နာန်းပူဝ် ၅ဝင်း ပန်းနာမူမှာ ှား'ဝဲ ့နဲ ့ယင်း အူ။ နဲ ့နော့မဝ်ကူး နာန်းကဝ်းနာုံ ထွမ်နားမ်းလူင် ၅ာုံ ရုဝ်'နီး; နီ ရိုန်းနားင်မျူးအိုက်နော့ နာဝ်ကျေ့ နှင့်မားမ်ာဆူး နာမ်အိုက်များယဝ် ့။ ဆုက် ့နာုံကဝ်နာုံ ထုရှိုင်းကာင်း ကာက် ့၅ာ'ကွတ်။ လူင်ကဝ်းကျေ, မိုဝ်းငှားပုန်း ထိုက် ့မဝ်နွန်မား နာကေု့ လုက် ့မိုဝ်းနာ ့ကူဝ်မီလိုက် တဲး ၅ဝင်းနွန်းနာဝင်းတာ။ ကဝ်းနာုံ အနာ့်ကို အိုးဘာ။

ပၢင်;မွ ත်လိက် ့မဲ'ကျေ, ကျမွ ත်ကီး စီး့ဘူ။ လုက် ့ပီး තဲ ့ကူဝ်မီ တဲး အမ်းတီး ဗာဝ်းကူးကူဝ် ့လိက် ့တဲး ၅ာဝ်း ၅ာိပူဝ်းဆာဝ်း ုတဲ ့ ကူးတီး မဲ'ပာင်းမွ ရှိလိက် ့တဲးကူဝ်

 $\frac{8}{8}; \cos \frac{1}{6}; \cos \frac{1}{6} = \cos \frac{1}{6}; \cos \frac{1}{6$

မီလ: ကဆမ်း လွင်း နာမ်မို ကမ်းထမ်း လွင်းကုမ် ရှိုက် ကွမ် တုန်း ကျိုင်းမ်ား ကိုင်း ကို လွင်း နာမ်မို ကောင်း ကျွန်း ကျွန်း ကို ကောင်း ကျွန်း ကျွန်း ကို ကောင်း ကောင်း ကျွန်း ကျွန်း

၂၂. လွင်း မိုက်ဗျာရုပ်

- 1. ပျာရာတီ.ပုတ္ ေစေရမ႑ိဳး.။ = ပုုရာတီ.ပုုတ္၊ ဝေလျမျာဳ ကုုရဲ့နာမ်း လွင်; တာင်းကျိုင်း ကူမူဝံလး မွ်ရာ တူဝံ; တိုုပ်တတ်. ဆွ်းတူဝ်းပိုုရ်း ကုုယ္လ်, ။ ရီးမီး ရီးမျာတီ. စိုးလွင်း မြိုတ်တုု ကိုက်ထုု လွင်း ကူမူ, ရာ ရှိုဝ်,မိုုဝ် မတ်တွုု ရှုက်.ရှု မျိုရ်မျိုင်း ကုုယ္လ်, ။
- 2. **ကျတိုက်ကးတျာ၊ ဝေလျမျာဂီ.။** = ကျတိုက်ကးတျာ၊ ဝေလျမျာဂီ. ကျာဂ်, ကမ်း ရိုဝ် ရိုင်းရီး ခွင်ပိုက်; က်, ကျာဂင်း တွေ့နှင့် မွ်ဂ်ားမီ မွ်ဝျးတွ်းယွ်, ။
- 3. ကျမေမှု,မိအ်အျအာျလး ဝေလျမျာရီ.။ = လွင်းကဲ, ကဆမ်း လွင်း ကျတင်ကျား မဲ လျတီ. ၊ မွ်ထိတ် မေး မမ်လုက်. ပို့ရ်း ကျယွ်, ။
- 4. မူ မျာဝုတာ ဝေ မာ မာ ကီ.။ = ဧ ကီ, က ဆမ်း မျာတီ. ရုံးပိုင်းရာမ်းတာကို ထွင်းတွ်းမွ်ကီး စာမ်းထိုုတ်. စာမီးထာင်း စာမီးထုုင် စာမီးထုုင် စာမီးထုုင် စာမီးတုုင် စာမီးကိုုမ်း ရုံးမိုုင်မှုုင်းမဲ့ စာမီ စိုင်းမျိုုင် ကြုုးပီ၏တွင် ပူင်းပော်း ထွန်းစုံး ဝဲ, ကာုယွ်, ။ ရုံးလွင်း ဝုဏီကမ် မူမာုုဝုုတာ စာ,မိ မီတွင်း လွင်းပိုုင်းစ်, ယင်းယွ်, ။
- (၂) ပီးမှုကားဝျသား စားမ်းရာ ပိုက် ကျင်ပိုက် မွ်တွ် အီး ကျင်ပိုက် သိုက် (၂)

39

Bibliography

Morey, S.D., 2005, The Tai languages of Assam - a grammar and texts, (Canberra: Pacific Linguistics). 413p. : ill., tables. Accompanying CD

Saonoai Mannhan, 2008 Five Sacred Words

Ven. Vimalatissa Bhukkhu, Therevada Buddhist Devotion (Lohit, Arunchai Pradesh: Chongkham Raj

Vihar Society)

U+1090

109 0 1 2 3 4 5 6 7 8 9 Α \bigcirc ° B ీ <u>ိ</u> D ੇ E F

Myanmar

Tone Marks

109A $_{\odot^{\circ}}$ MYANMAR SIGN KHAMTI TONE-1

109B ° MYANMAR SIGN KHAMTI TONE-3

Aiton Vowel Signs

109C S MYANMAR VOWEL SIGN AITON A

109D d MYANMAR VOWEL SIGN AITON AI

U+AA60

AA6 AA7 0 \mathfrak{M} W 2 ၯ 3 3 4 \mathbf{N} 5 ၅ 6 ന **S** 8 M 9 Α B ป D 9

F

Myanmar Extended-A

Khamti Consonants

AA60 MYANMAR LETTER KHAMTI CA AA61 MYANMAR LETTER KHAMTI SA MYANMAR LETTER KHAMTI NYA **AA62 AA63** MYANMAR LETTER KHAMTI NA AA64 MYANMAR LETTER KHAMTI RA MYANMAR LETTER KHAMTI HA AA65 **AA66** MYANMAR LETTER KHAMTI XA MYANMAR LETTER KHAMTI ZA **AA67** MYANMAR LETTER KHAMTI FA **AA68**

Khamti Extensions

AA69 6 MYANMAR SIGN KHAMTI REDUPLICATION

Aiton Extensions

AA6A MYANMAR LETTER AITON RA

AA6B MYANMAR SYMBOL AITON
EXCLAMATION

AA6C MYANMAR SYMBOL AITON ONE

AA6D MYANMAR SYMBOL AITON TWO

Character Properties

```
109A; MYANMAR SIGN KHAMTI TONE-1; Mc; 0; L;;;;; N;;;;
109B; MYANMAR SIGN KHAMTI TONE-3; Mc; 0; L;;;;; N;;;;;
109C; MYANMAR VOWEL SIGN AITON A; Mc; 0; L; ;; ;; N; ;; ;;
109D; MYANMAR VOWEL SIGN AITON AI; Mn; 0; L;;;;; N;;;;;
AA60; MYANMAR LETTER KHAMTI CA; Lo; 0; L;;;;; N;;;;;
AA61; MYANMAR LETTER KHAMTI SA; Lo; 0; L;;;;; N;;;;;
AA62; MYANMAR LETTER KHAMTI NYA; Lo; 0; L;;;;; N;;;;;
AA63; MYANMAR LETTER KHAMTI NA; Lo; 0; L;;;;; N;;;;
AA64; MYANMAR LETTER KHAMTI RA; Lo; 0; L;;;;; N;;;;;
AA65; MYANMAR LETTER KHAMTI HA; Lo; 0; L;;;;; N;;;;;
AA66; MYANMAR LETTER KHAMTI XA; Lo; 0; L;;;;; N;;;;
AA67; MYANMAR LETTER KHAMTI ZA; Lo; 0; L;;;;; N;;;;;
AA68; MYANMAR LETTER KHAMTI FA; Lo; 0; L;;;;; N;;;;;
AA69; MYANMAR SYMBOL KHAMTI REDUPLICATION; Mn; 0; L; ;; ;; N; ;; ;
AA6A; MYANMAR LETTER AITON RA; Lo; 0; L;;;;; N;;;;;
AA6B; MYANMAR SYMBOL AITON EXCLAMATION; So; 0; L;;;;;;N;;;;;
AA6C; MYANMAR SYMBOL AITON ONE; So; 0; L;;;;; N;;;;;
AA6D; MYANMAR SYMBOL AITON TWO; So; 0; L;;;;; N;;;;
```

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.

 $\label{lem:please} \begin{tabular}{ll} Please read Principles and Procedures Document (P \& P) from $$\underline{$http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html}$$_ for guidelines and details before filling this form. \\ \end{tabular}$

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: Myanmar Khamti Shan Extensions							
2. Requester's name: Martin Hosken							
3. Requester type (Member body/Liaison/Individual contribution): Individual contribution							
4. Submission date:							
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, Myanmar Extended-A							
2. Number of characters in proposal:	18						
3. Proposed category (select one from below - see section 2.2 of P&P document):							
A-Contemporary X B.1-Specialized (small collection) B.2-Specialized (large collection)							
C-Major extinct 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?							
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools							
used: http://scripts.sil.org/Padauk							
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)?	no						
8. Additional Information:							
Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Scrip							
in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of su							
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.	101						

^{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? Stephen Morey; Khamti Shan Literacy Committee					
If YES, available relevant documents:					
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:					
6. After giving due considerations to the principles in the P&P document must the proposed characters be entire	ely				
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?	no				
If YES, is a rationale for its inclusion provided?					
If YES, reference:					
9. Can any of the proposed characters be encoded using a composed character sequence of either					
existing characters or other proposed characters?	yes				
If YES, is a rationale for its inclusion provided?					
If YES, reference: this document					
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?					
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:					