

## Notes on Khamti

Compiled by Ben Mitchell with contributions from Norbert Lindenberg, Deborah Anderson, Stephen Morey, Patrick Chew and Jayasāro Bhikku.

### Background

The Khamti script<sup>1</sup> uses characters that are encoded as Burmese, Shan and Aiton, as well as Khamti. This presents two challenges. Firstly, certain characters take different forms in different texts. For example, Khamti letter Ga may take the Burmese form ဂ or the Shan form ၵ; Jha may take the Shan form ၶ or the Khamti form ၷ; and Ra may take the Burmese form ရ, the Aiton form ၸ or the “Khamti form” ၹ. As a result, Khamti text in one style is encoded differently from Khamti text in another style, so interchange is not possible (searching for ရ will not return ၸ or ၹ.) Having more than one possible encoding for glyph variations means it is unclear which is the ‘standard’ character to use for Khamti.

Secondly, the preferred style of Khamti uses dots<sup>2</sup> in most letters — for example Burmese Ka က becomes ၵ or ၶ<sup>3</sup> in Khamti — which means their visual representation needs to differ from their Burmese, Shan and Aiton counterparts. The current implementation uses a variation selector to substitute the dotted forms, but this mechanism currently applies to only a subset of Burmese, Shan and Khamti characters, and not all those needed in Khamti language.

We request the following:

**1** Allow dotted forms for the characters not currently covered by the variation selector mechanism:

**1a** Add new Standardized Variants (and include the dotted forms in the nameslist) for:

- ▶ uni1077 Shan Ga ၵ [see pages 6–7];
- ▶ uniA9E1 Shan Cha ၶ [see pages 6–7];
- ▶ uni107C Shan Na ၷ and/or uniA9E3 Shan Nna ၸ [see pages 6–7 and note 13 on page 18];
- ▶ uni101B Myanmar Ra ရ [see pages 3–5];
- ▶ uniAA6D Khamti Ha ၹ [see pages 6–7];
- ▶ Khamti Shan logograms uniAA74 Oay ၺ, uniAA75 Qn ၻ and uniAA76 Hn ၼ [see [p7](#) in original proposal L2/08-276].

**1b** Allow use of variation selectors with the following, or establish how dotted forms can be implemented for:

- ▶ Vowel sign Aiton -ai ၽ 109D [see page 10];
- ▶ Myanmar sign anusvara ၾ 1036;
- ▶ Myanmar sign asat ၿ 103A [see page 15];
- ▶ Myanmar sign Khamti tone 1 ၿ 109A;
- ▶ Myanmar sign Shan tone 5 (used in Khamti as tone 2) ၿ 1089;
- ▶ Myanmar sign Khamti tone 3 ၿ 109B;
- ▶ Myanmar sign Shan tone 2 (used in Khamti as tone 4) ၿ 1087;

---

1 'Khamti script' in this document is used a shorthand for the script used by the Khamti, Aiton, Phake and Khamyang. Minor letterform variations can be identified, but the principles discussed here apply to all four language groups, whose script would locally be called “to lik tai”.

2 Stephen Morey explains “Black dot style...was much more salient in earlier manuscripts, but the 'black dots' were added to the script in 1997 at the direct request of community members. At the time I made a 'dot free' form at the request of the late Aimya Khang Gohain, who felt that printing would be more comfortable without the dots. But this hasn't been accepted by the community.”

3 Closer examination of manuscripts and discussion with community members should clarify the preference.

- ▶ Myanmar sign Shan tone 3 (used in Khamti as tone 5) ◌̃; 1088
- ▶ Myanmar sign visarga (used in Khamti as tone 6) ◌̃; 1038
- ▶ Myanmar sign Shan tone 6 (used in Khamti as tone 8) ◌̃; 108A;

**1c** Request mention of dotted forms as Standardized Variants in the text of the “Khamti Shan” section of 16.3 of the Unicode Standard.

**2** Assign the alias “Myanmar Letter Khamti LLA” to uniAA6E ၵ. The current character name “Myanmar Letter Khamti HHA” is incorrect. [See pages 3, 6, 7]

**3** Encode Khamti numerals. Three sources [see pages 11-12] attest to the existence of distinct Khamti numerals. This example telephone number compares them with Burmese, Shan and Tai Laing counterparts. We are seeking further attestations and comments from users.

Number	011-23978046
<b>Khamti</b>	၀၁၁-၂၃၉၇၈၀၄၆
Burmese	၀၁၁-၂၃၉၇၈၀၄၆
Shan	၀၁၁-၂၃၉၇၈၀၄၆
Tai Laing	၀၁၁-၂၃၉၇၈၀၄၆

If these are proposed and encoded<sup>4</sup>, the text of the Unicode Standard 16.3 will need updating (“Khamti Shan uses the Shan digits from the range U+1090...U+109A”)

### Recommendations sought

The following questions may be outside the purview of the Unicode committees, or it may be useful to add text to the Unicode Standard / UTN11 to address them. Either way, they are noted here as a reference for font makers, and should be considered open questions while we liaise with users.

**4** The consonant Ra and the -ai vowel each have three possible encodings. If Khamti Unicode permits all three versions of each, texts will diverge in their encoding. If the aim is standardised encoding, how could this be achieved? One possible solution could be compatibility mappings in the Unicode data.

**4a** The vowel sign -ai has been encoded at three codepoints: uni1032 ◌̃ for Standard Burmese, uni1086 ◌̃ for Shan and uni109D ◌̃ for Aiton [see page 10]. All three glyphs can be found in different Khamti-Aiton-Phake texts, but the preferred, dotted form is ◌̃. This is visually closest to uni109D ◌̃, but UTN11 suggests a dotted form of uni1032 ◌̃<sup>5</sup>. See page 10.

**4b** UTN11 lists uni101B ၵ for Khamti consonant Ra, and uniAA73 ၵ as the historic form; this conflicts with Khamti proposal 08-276, where uniAA7A ၵ is the historic form, and uniAA73 ၵ is the revised form. We suggest ၵ is used especially for the parts of manuscripts written in Aiton language while dotted ၵ is found in the parts in Pali or Burmese language. In Khamti language, digital texts mainly prefer ၵ, while handwritten texts may even use letter La (uni101C ၵ) in place of Ra. The ၵ form encoded as “Myanmar Letter Khamti Ra” (uniAA73) is not familiar to any of the collaborators on this document, but could potentially be an alternative form of La. We are investigating further.

**5** Syllable reduplication is encoded using uniAA70 ◌̂ “Myanmar modifier letter Khamti reduplication”. In most digital fonts, and in some manuscripts, this symbol is visible in text, while in other handwritten sources, syllable repetition is indicated by doubling the syllable’s vowel. UTN11 suggests that uniAA70 should have this doubling effect on two diacritics, but we discuss others, and how users are actually typing doubled vowels on page 13. See pages 13-16.

---

<sup>4</sup> It will need to be decided whether to present dotted or undotted forms in the code charts; if undotted, the variation selector will need to be applied to obtain dotted forms.

<sup>5</sup> as mentioned in 1b above, this dotted form cannot be obtained with the variation selector mechanism.

**Khamti attestations**

ထူး;မေး;လိက်.တံး:နံး:တီး;

က	ခ	ဂ	ဂ	င
Ka	Kha	Ga	Gha	Ṃa
ဏ	ဆ	ဆ	ဇ	ဗ
Ca	Cha	Ja	Jha	Ṃa
တ	ထ	တ	ထ	ဆ
Ta	Ṭha	Ḍa	Ḍha	Ṃa
တ	ထ	တ	ထ	န
Ta	Tha	Da	Dha	Na
ပ	ဖ	ပ	ဖ	မ
Pa	Pha	Ba	Bha	Ma
ယ	ရ	လ	ဝ	သ
Ya	Ra	La	Va	Sa
	ဟ	လ	ာ	
	Ha	Ḷa	A	

Three Khamti consonant charts showing Ra ရ in the standard construction, with and without a dot. Either way is acceptable and the variation selector mechanism would allow users the choice. Lla လ is incorrectly encoded as uniAA6E Hha. Font makers may also note the cursive variant form of Nya ဗ with the tail connecting to the final bowl.

ပိန်လိက်.တံး:နံး:တီး;

မေး;လိက်.

က	ခ	ဂ	ဂ	င	ဏ	ဆ	ဆ	ဇ	ဗ	တ	ထ	တ	ထ	ဆ	တ
တ	ထ	န	ပ	ဖ	ပ	ဖ	မ	ယ	ရ	လ	ဝ	သ	ဟ	လ	ာ

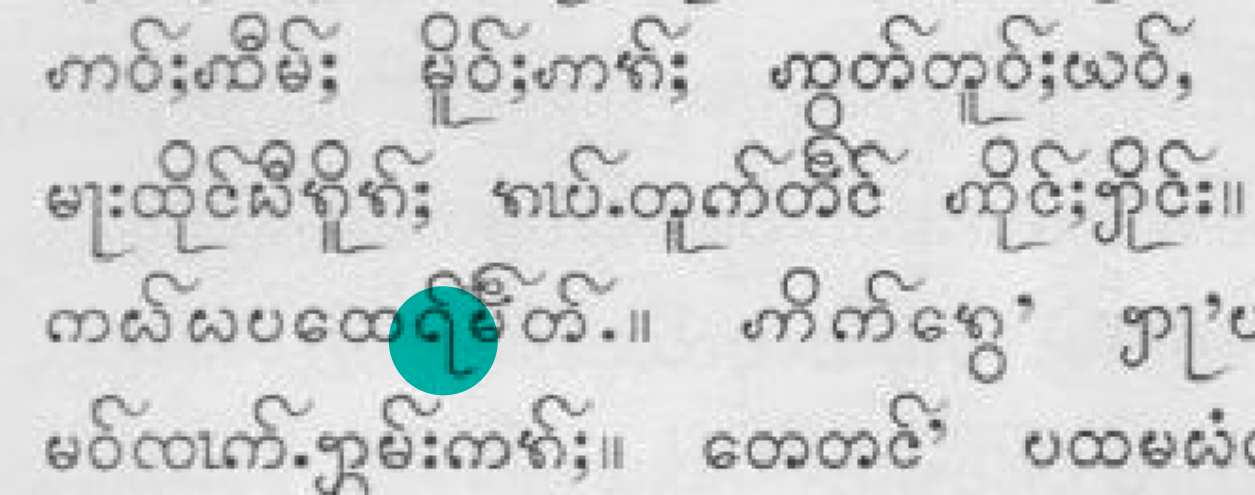
မေး;လိက်.

ပိန်လိက်.တံး:နံး:တီး;

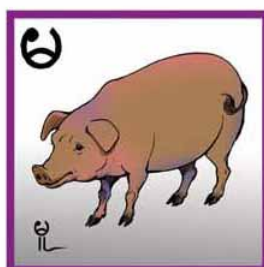
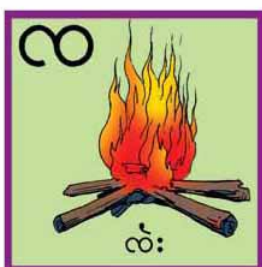
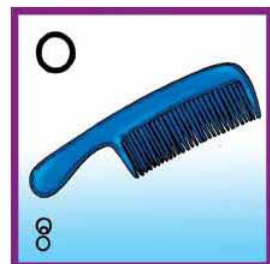
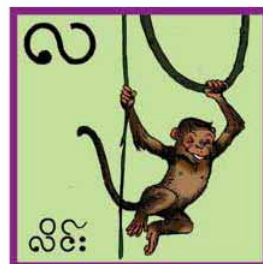
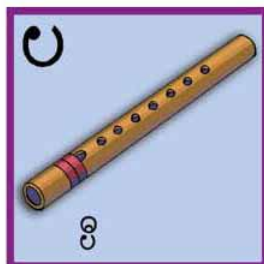
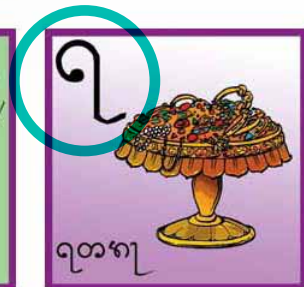
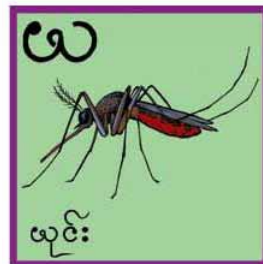
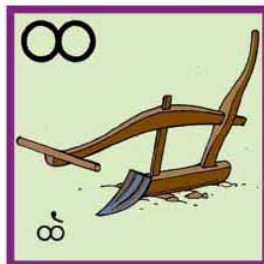
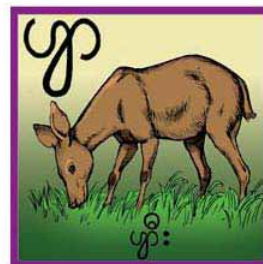
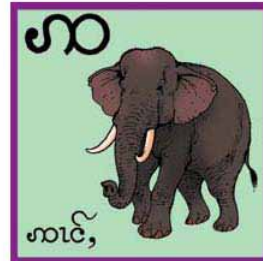
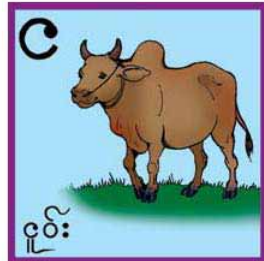
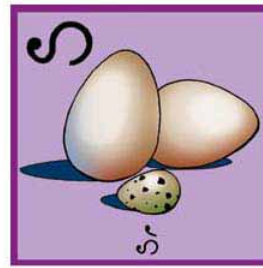
က	ခ	ဂ	ဂ	င	ဏ	ဆ	ဆ	ဇ	ဗ	တ	ထ	တ	ထ	ဆ	တ	ထ
တ	ထ	န	ပ	ဖ	ပ	ဖ	မ	ယ	ရ	လ	ဝ	သ	ဟ	လ	ာ	

13

<sup>1</sup> ရှိုဝ်းဆတ်ဝိင်းကန်တိကုတ်.မ်' ပရော့လိတ်.ရ်,ခဝါ မြုရ်,ခဝါယင်းကျ။ ခဝါယျ, ဗျာဗျာ  
စိကုကန်; ဣဝ်,ပုဗ္ဗင်၊ လုကိကုန်းမိုင်းကုရော့ရင်း၊ မျးစိန် (လူ'တံးကုဝ်,ကဝ်  
ဆတ်'ကျော့ဂျတ်.ကန်တိပ) စကျ မေးလုခဝ်စမ်။ <sup>2</sup> မိုဝ်;ခဝ် ဝ်;ဆတ်'ကြူးင်း ဣတ်.တွင်,င်း  
ယွန်းပူးနီကုရ်, ခွန်ဆတ်'ကြူးယျ, “မေးလုစကျဗျာဗျာခဝ်မ်' ဆျာ့ဂျင်းဣတ်ကမူးကဝ်; ကွဲး



Two Khamti typefaces with dotted Ra in place of the standard Burmese looped Ra ရ.



Khamti consonants with dotted Ra.

## The 33 Consonants

တုၣ်းမေ; လိက်.တဲး 33 တုၣ်း;

တုၣ်းမေ;လိက်.တဲး 33 တုၣ်း; ဖဲ.ကအမ်း တုၣ်းမေ;လိက်.ကပ် 18 တုၣ်း; လွမ်.ကိက်. 15 တုၣ်း;  
 သီၣ်ဗက် လုက်.လိက်. ဗျုလိ ကျုယပ်.၊

က K	ခ Kh	ဂ G	ဂ Gh	င ṅ
ဇ C	ဇ Ch	ည J	ည Jh	ဗျာ ṅ
တ T	ထ Th	ဒ D	ဒ Dh	န N
တ T	ထ Th	ဒ D	ဒ Dh	န N
ပ P	ဖ Ph	ဗ B	ဖ Bh	မ M
ယ Y	ရ R	လ L	ဝ W	
သ S	ဟ H	လ L	က A	

Khamti consonant table showing dotted forms of Khamti Gha (uni1077 Shan Ga), Cha (uniA9E1 Shan Cha), Nna (uni107C Shan Na or uniA9E3 Shan Nna) and Ha (uniAA6D Khamti Ha). In yellow, uniAA6E is annotated as Lla (not Hha) using a double-looped stylistic variant.

Vagga ဝဂ္ဂ	Ṭhāna- karaṇa ဌာနကရဏ	Sithila သိထိလ		Dhanita ဓနိတ		Nāsiga နာသိဂ
Ka Vagga	Velar ကဏ္ဍဇ	k [k]	kha [kʰ]	g [g]	gh [gʰ]	ṅ [ŋ]
		က	ခ	ဂ	ခ	င
		ဂ	ခ	ဂ	ခ	င
Ca Vagga	Palatal တာလုဇ	c [tʃ]	ch [tʃʰ]	j [tʃ]	jh [tʃʰ]	ñ [ɲ]
		ဆ	သ	ဆ	ဆ	ဗ
		ဆ	သ	ဆ	သ	ဗ
Ṭa Vagga	Lingual မုမုဇ	t [t]	th [tʰ]	ḍ [d]	ḍh [dʰ]	ṇ [ɳ]
		တ	ထ	ဓ	ဓ	ဏ
		တ	ထ	ဓ	ဓ	ဏ
Ta Vagga	Dental ဒန္တဇ	t [t]	th [tʰ]	d [d]	dh [dʰ]	n [n]
		တ	ထ	တ	ထ	န
		တ	ထ	တ	ထ	ဏ
Pa Vagga	Labial သြဋ္ဌဇ	p [p]	ph [pʰ]	b [b]	bh [bʰ]	m [m]
		ပ	ဖ	ပ	ဖ	မ
		ပ	ဖ	ပ	ဖ	မ
Three Sound		Middle-အလယ်		High-မြင့်သော		Low-အနိမ့်

Vigga ဝဂ္ဂ	Semi-vowel Open or rattled				Fricative					Lingual
Avagga အဝဂ္ဂ	y	r	l	v	śa	ṣa	s	h	ṅ	!
	[j]	[r]	[l]	[v]			[s]	[h]	[ŋ]	[!]
	ယ	ရ	လ	ဝ	ရှ	ဗ	ဆ	ဟ	င	ဏ
Three Sounds	Middle-အလယ်				High-မြင့်သော					Low-နိမ့်စွာ
	a	ā	i	ī	u	ū	e	o		ṁ
	[a]	[aː]	[i]	[iː]	[u]	[uː]	[eː]	[oː]		[ṁ]
Sara သရ	ဃ	၂	ဝ	စ	၂	၂	ဇ	၂		•
	ဃ	၂	ဝ	စ	၂	၂	ဇ	၂		•
	ဃ	၂	ဝ	စ	၂	၂	ဇ	၂		•

Chart showing Khamti characters with their Shan counterparts. Gha, Cha, Nna and Ha are dotted again, and Lla is presented as IPA [!].

မောင်ကွက်တူဝ်းမေးလိက်. (hh)လုက်.

		စပင်းမောင်(၁)စပင်း						
		မောင်ကူက်	မောင်စျောင်း	မောင်စန်း	မောင်စက်	မောင်ဗူးစန်း		
တီးကွက်မောင်(၁)တီး	ကွင်းကွင်း	က	ဂ	ဏ	ခ	င	ဗ	ဘ
	ကန်'ရှိုဝါက်	ဂ	င	ဆ	ဇ	ဗ	ဃ	
	ပါးလိက်,	တ	ထ	ဓ	ဇ	ဆ	ရ	လ
	ဒိတ်'	တ	ထ	ဓ	ဇ	န	လ	မ
	ပိင်းညှပ်	ပ	လ	စ	ဇ	မ		ဝ

Khamti consonant table in the undotted style, showing a short-tailed stylistic variant of 101B Ra.

က	ဂ	င
ဏ	ဆ	ဗ
တ	ထ	ဆ
ပ	လ	မ
ဃ	ဗ	ဃ
ဝ	ဗ	ဏ

Khamti consonants in an older style showing variant forms for Na and Ra. The two-bowl form of AA6B န Na is further shown on the following page.

တြား တွတ် ဟပ် (ဆင်ရှင် ထံး လှိုင်)

ဘဲခေး တာယး ကကဝး တေး ဟာလှာ တေး  
ဟံမာ ဟံပုတ္တယး

စံလေး ပြိုင် ခင် ကေး တာ ဟေး ပွတ် ဟပ် ဟင်  
တုလ် ဟိလ် ဟံး ဝဲး ပဲ ဟင် ဝှိလ် ဟျင် တီ ခြတ် ဟပ်  
ဟင် ဝဲး ပဲ ဟင် ဟိုဝ်း ဂွင် တွတ် ထံး ဟာ ဟာ ဟာ ဟာ  
ဟင် ဂှာ ဂှာ ဂှိလ် ဟိ ဟိ ကာ ကပ် တံး ဟင် ဟာ ဟိ ခြိတ် ဟာ ဟင်  
ဟွတ် တွတ် ပာဝ် ဟင် ဟိလ် ဟျင် ဟေး တိုလ် ဟိ ဟိ  
ဟင် ဟင် တွတ် တုလ် ဟိုလ် ဂှိလ် တွတ် ဂှာ ဟင် ဟိး က  
တ် ဟာ ဟေး တေ ခြတ် ဟိ က် ဟိး ဂှာ ပက် ဂှာ ဟင် တာ ကာ  
ခင် ကက် ဂှာ ဟင် ကာ ဟင် တေ တင် ပထမး ဟင် ဟာ ဟာ  
ဟာ မေး ထံး တလ် လှိုင် ဟာ ခင် ကွင် တြား ဟင် ဟိလ် ကပ် မှာ  
လိး ဟိုလ် ကပ် တံး ဟုပ် ပာ ဟင် လိ ထာ ဟာ ဟင် ဟင်  
ဟာ ဟာ ဟာ ဟာ တိုလ် ဟေး ပွတ် ဟပ် ဟတ် ဟာ မေး တေ ကျ  
လ် ဟျင် ဟိ ဟင် ဟင် ဟာ ဟင် ဟိ ဟင် ပက် ဟင် ချိး ဟျင်  
လှက် ဟပ် ဟင် မေး ဟိုဝ်း ထံး တွတ် ဟေး ဟင် ဟာ ဟာ ဟာ  
ဟွတ် တွတ် လိ က် ကာ ဟင် တံး ဟာ တင် လှိုင်

Pali text handwritten in Khamti script, with consonant Na ၫ AA6B in the two-bowl variant form.

## Khamti vowel -ai

UTN11 encodes the Khamti -ai vowel using the Burmese character uni1032 ဲ.



Khamti text in the undotted style, with vowel -ai highlighted.

At present this character is not covered by the variation selector mechanism, so the normal dotted form ဲ needs to be implemented using a different approach, likely involving language-tagged text triggering an OpenType <locl> feature substitution.



Khamti text in the dotted style, showing the dot may sit on top of the outward stroke, or (more commonly in handwriting) beneath it.

In place of the Burmese character uni1032, Khamti texts often use Shan character uni1086 ဝံ, or Aiton character uni109D ဝံ. Since these are merely variant forms of the same vowel, having a choice of three possible characters means the same text can be encoded in three different ways. See also "Notes on Arakanese and Mon", p6, in which the Arakanese -ai vowel takes the form currently encoded as Shan / Aiton.



Aiton (left) or Shan (right) character may be used for the -ai vowel in Khamti.

**Khamti numerals**

**NOVEL CORONAVIRUS (COVID-19)**

**Protect yourself and others!**

**Do's** Follow these Do's and Don'ts

**Do's**

- Practice frequent hand washing.** Wash hands with soap and water or use alcohol based hand rub. Wash hands even if they are visibly clean
- Cover your nose and mouth with handkerchief/tissue while sneezing and coughing**
- Throw used tissues into closed bins immediately after use**
- See a doctor if you feel unwell (fever, difficult breathing and cough). While visiting doctor wear a mask/cloth to cover your mouth and nose**
- Avoid participating in large gatherings**

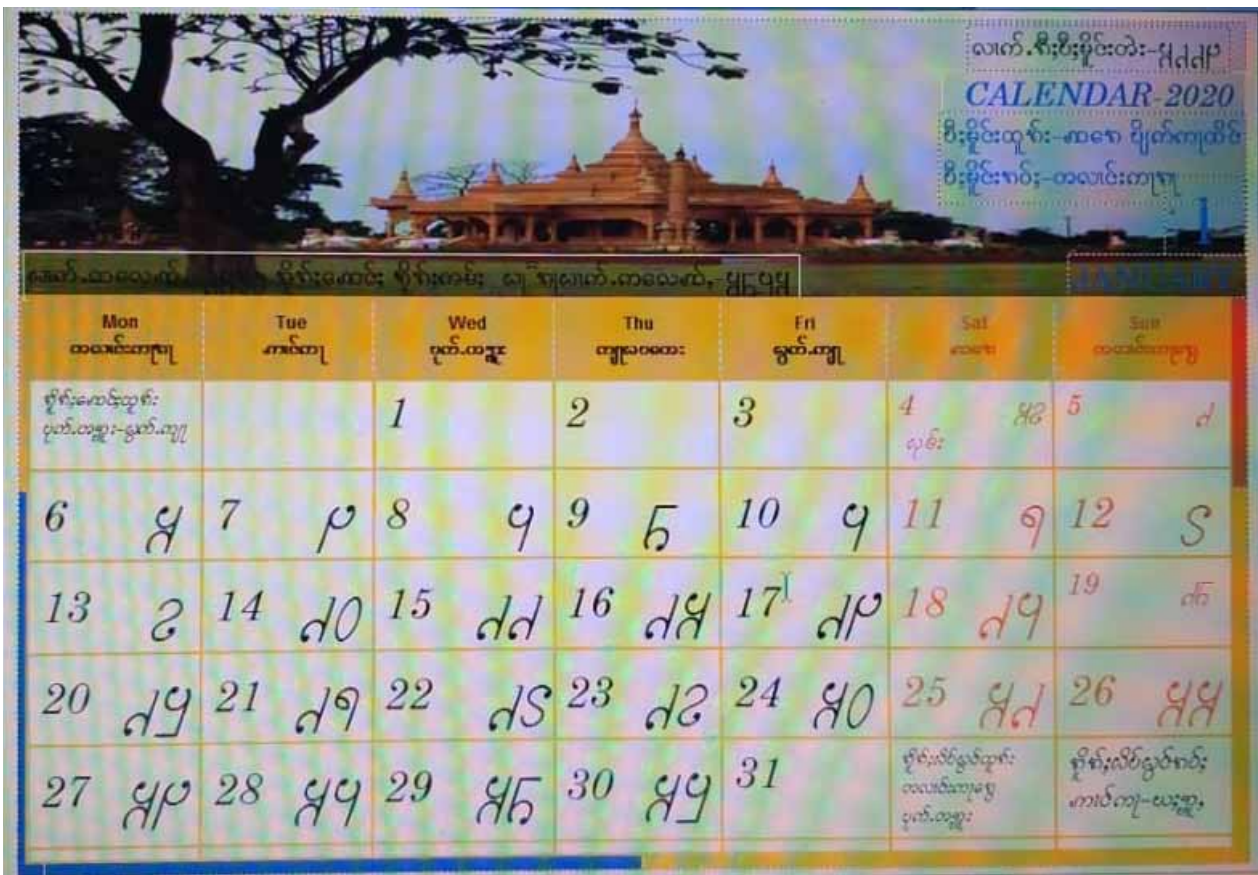
**Don'ts**

- Have a close contact with anyone, if you're experiencing cough and fever**
- Touch your eyes, nose and mouth**
- Spit in public**

**Together we can fight Coronavirus**

**For further information :**  
 Call at Ministry of Health, Govt. of India's 24X7 control room number  
 +91-11-2397 8046  
 Email at [neov2019@gmail.com](mailto:neov2019@gmail.com)

Khamti health poster attesting distinct Khamti numerals, as yet unencoded. Shan, Burmese or Hindu-Arabic numerals may also be used with Khamti text.



Khamti calendar showing distinct numerals (note they do not correspond with the Hindu-Arabic numerals).

(ထူဝ်ဂးစိုး တီး ခြင်း)

၁	ရွ	=	ရွ	ရွရွ	=	ဃိပ် ကိတ်
၂	ဗ	=	ဗ	ရွဗ	=	ဃိပ် ဗ
၃	ဒ	=	ဃမ်	ရွဒ	=	ဃိပ် ဃမ်
၄	င	=	ဃိ	ရွင	=	ဃိပ် ဃိ
၅	စ	=	ခါ	ရွစ	=	ဃိပ် ခါ
၆	ဗ	=	ခါတ်	ရွဗ	=	ဃိပ် ခါတ်
၇	ဇ	=	ကိတ်	ရွဇ	=	ဃိပ် ကိတ်
၈	န	=	ပျတ်	ရွန	=	ဃိပ် ပျတ်
၉	ဂ	=	ကပ်	ရွဂ	=	ဃိပ် ကပ်
၁၀	ရွ၀	=	ဃိပ်	ဗ၀	=	ဃိုး ရွ

Alternate forms of Khamti numerals, similar to Tai Laing numerals (yellow)

## Syllable reduplication

Syllable repetition can be indicated visually in two ways. The syllable's diacritic or vowel character can be doubled (approach 1), or the reduplication sign <sup>◌̆</sup> AA70<sup>1</sup> can appear directly after the syllable (approach 2). Though the reduplication sign appears in digital texts (e.g p16), for many combinations we suggest that a visible reduplication sign is not acceptable. In this understanding, the appearance of the reduplication sign with certain vowels is an indicator that the font doesn't support the required doubled vowel glyphs. We are liaising with user communities to confirm this.

UTN11 notes "the reduplication character ligates with<sup>2</sup> two diacritics": the -ai vowel ဲ 1032 and the asat ် 103A, and it shows ဲ̆ 1032 AA70 and ်̆ 103A AA70 (Preferred dotted forms would be ဲ̇ and ်̇.) Additionally, the -aa vowel ာ 1083, the -ii vowel ိ 102E and the anusvara ံ 1036 can be doubled. Visually these would appear as follows:

Approach 1	Transcription	Representation <sup>3</sup>			Unicode
	mai mai	မိ	မိ	မိ	1019 FE00 1032 [FE00] AA70
	kan kan	ကန်	ကန်	ကန်	1000 FE00 AA6B FE00 103A [FE00] AA70
	la la	လာ	လာ	လာ	101C FE00 1083 AA70
	miimii	မိ	မိ	မိ	1019 FE00 102E AA70
	mam mam	မံ	မံ	မံ	1019 FE00 1036 AA70

Note [FE00] are not yet supported so the correct representations are unobtainable

In many digital fonts, this doubling is not available, so the reduplication character itself appears in text.

Approach 2	Transcription	Representation			Unicode
	mai mai	မိ̆	မိ̆	မိ̆	1019 FE00 1032 [FE00] AA70
	kan kan	ကန်̆	ကန်̆	ကန်̆	1000 FE00 AA6B FE00 103A [FE00] AA70
	la la	လာ̆	လာ̆	လာ̆	101C FE00 1083 AA70
	miimii	မိ̆	မိ̆	မိ̆	1019 FE00 102E AA70
	mam mam	မံ̆	မံ̆	မံ̆	1019 FE00 1036 AA70

We have highlighted in red the representations we consider unacceptable or unlikely (even though they appear in digital texts in the following pages); these all include an above-base mark. For others<sup>4</sup> including the third line, we may now have a conflict, since approach 1 and approach 2 are encoded the same way, and there is no mechanism to pick one instead of the other. Currently users are typing vowel characters (including above-base marks) twice to show doubling; this allows doubled signs to be stored differently than a visible reduplication sign, but seems to contradict the current encoding scheme.

Finally, syllables may be duplicated more than once [pp14–15], how are these to be encoded?

1 Should this character be interpreted as a doubled asat sign (uni103A ်̆)?

2 'Ligates with' is understood to mean that the reduplication character disappears and the preceding diacritic is doubled.

3 Left column is a custom build of Noto Serif Myanmar, centre column is Kellen Parker van Dam's Ghin Kao font, right column is Stephen Morey's Aiton Unicode font.

4 Other reduplicated syllables found in digital texts include ကိ̆ (kiki), ကု̆ (kuku), တူ̆ (tuutuu) and ကေ̆ (keke).

Handwritten Burmese script on aged paper. A yellow circle highlights a specific character or syllable in the second line.

Handwritten Burmese script on a white background. Multiple yellow circles highlight specific characters or syllables throughout the text, illustrating doubled and tripled/doubled syllables.

Handwritten texts showing doubled -ai vowel (above) and tripled and doubled syllables (below).

Handwritten text in a cursive script, likely Burmese. Several characters are highlighted with yellow circles, indicating specific linguistic features or errors.

Handwritten text in a cursive script, likely Burmese. Several characters are highlighted with yellow circles, indicating specific linguistic features or errors.

Attestations showing tripled and quadrupled reduplications of (dotted) uni103A ် asat and uni1036 ် anusvara.

ကကိဝ်းလွင်;မာင်ကလၢ

ကူဝ်; နဝ်'မာင်ကိဃျ်း

လွင်;မာင်ကလၢတၢ;စီ; မာမ်သိပ်ပိတ်ပိုင်;ရဲ,ဂဝ်မဲ' ပူဝ်းကမ်;ကိင်;  
ကိးယၢ, နမဲ' စကေး: လၢနမူရဲ,ဂဝ်မဲ' ပေ,မိ နဲ' နေး:ကိမ်ချမ်မၢကူ။  
လံးပိုင်;ရဲ,မိ ကိ ကူရဲ: တင်းလွင်မဲ' မာင်ကလၢတၢ;စီ;မာမ်သိပ်ပိတ်ပိုင်;ရဲ,  
ကကမ်: မာင်ကလၢကရဲ;ဂိတ်:မိတ်.ရမိ နဝ်'မာင်ကိဃျ်း မွဲ:မတ်.ဝဲ,တၢ။

တြးတြးကူဝ်: မာင်ကလၢရဲ,မုတ်ယဝ်,။

မိုဝ်းလွဲ: လးရး: ရဲ,ကွဲ။ နွဲ,ကွမ်ဂုရဲ'မဲ'ကေး,  
လွဲ: ကုင်;ကၢတိုဝ်း။ ရဲ;မဲ'ထိုင်ကိ: ကွမ်စေ့  
ကွမ်စေ့မိ ကၢတိုဝ်း။ ရဲ,မဝ်ကွဲ;ကိးယၢ, မွင်ဂိုဝ်  
တွဲ;တၢင်းမဲ' မွင်;ရမ်မိုဝ်။ ကရဲ;ကွးပုဝ်ဂွေကဝ်'  
မွဲရဲ,ထွမ်စီ; တိုဝ်း။ ရး: ။

တင်းမိုဝ်း: ကိပ်. လိပ်. ရဲ;ကွဲ'ကၢဝ်,။  
ရဲ,မဲ'ထွဲ'ရဲ,မာဝ် မရဲ:ယၢ, ကူဝ်; ..... ပိတ်;ဝဲ:  
မး:ကျး? တေ. ပိကပ်.ကွဲ;ယဝ်, လွင်ဃျ်း။ ကဝ်;မဲ'  
ကုတ်.ကၢ ကရဲ;ပုဝ်ကွဲ: ပရဲ;ကမူပျ်ကွဲ'ဝဲ,ရဲ,ယင်း:  
ကူ။ ရဲ,မေ,မဝ်ကး: ကရဲ;ကဝ်;ရဲ' ထွမ်ဂၢမ်းလွင်ကွဲ'  
ကွဲ'စီ; မိ ရဲ;ဂၢင်မး:ကိတ်ရဲ,ဂဝ်ကေး, မွင်မာမ်ကူ:

Attestations showing reduplication character itself appearing in text, rather than vowels appearing twice.

## Appendix

The table below lists the characters used in Khamti, showing their default form and dotted forms where applicable. Shaded rows indicate letters already covered by the variation selector mechanism detailed in Unicode document 15-320 [\[7\]](#). Of the others, some do not require dotted versions and the rest (indicated by ∅) are discussed below and should be allowed with the variation selector FE00.

Default letterform	Unicode	Unicode name	Used as Khamti letter	Khamti letterform	Khamti Unicode
က	1000	Myanmar letter Ka	Ka	က	1000 FE00
ခ	1075	Myanmar letter Shan Ka <sup>2</sup>	Kha	ခ	1075 FE00
က <sup>1</sup>	AA60	Myanmar letter Khamti Ga	Ga	က	AA60 FE00
ဂ	1002	Myanmar letter Ga <sup>2</sup>	Gha	ဂ / ဂ	1002 FE00 <sup>11</sup>
ခ	1077	Myanmar letter Shan Ga <sup>2</sup>	Gha	ခ <sup>5</sup>	∅ <sup>10</sup>
င	1004	Myanmar letter Nga	Nga	င	1004 FE00
ဆ <sup>1</sup>	AA61	Myanmar letter Khamti Ca	Ca	ဆ	AA61 FE00
ဆ <sup>1</sup>	AA62	Myanmar letter Khamti Cha	Cha	ဆ	AA62 FE00
ဆ	A9E1	Myanmar letter Shan Cha	Cha	ဆ <sup>5</sup>	∅ <sup>10</sup>
ဆ <sup>1</sup>	AA63	Myanmar letter Khamti Ja	Ja	ဆ	AA63 FE00
ဆ <sup>1</sup>	AA64	Myanmar letter Khamti Jha	Jha	ဆ / ဆ	AA64 FE00 <sup>11</sup>
ဆ	A9E2	Myanmar letter Shan Jha	Jha	ဆ / ဆ <sup>5</sup>	A9E2 FE00 / A9E2
ဌ <sup>1, 4</sup>	AA65	Myanmar letter Khamti Nya	Nya	ဌ	AA65 FE00
ဋ <sup>4</sup>	107A	Myanmar letter Shan Nya	Nya	ဋ	107A FE00
တ <sup>1</sup>	AA66	Myanmar letter Khamti Tta	Tta	တ	AA66 FE00
ထ	AA67	Myanmar letter Khamti Ttha	Ttha	ထ	AA67 (no dots needed)
ထ	AA68	Myanmar letter Khamti Dda	Dda	ထ	AA68 (no dots needed)
ထ	AA69	Myanmar letter Khamti Ddha	Ddha	ထ	AA69 (no dots needed)
ဆ	107C	Myanmar letter Shan Na <sup>2, 13</sup>	Nna	ဆ / ဆ	107C / ∅ <sup>10</sup>
တ	1010	Myanmar letter Ta	Ta	တ	1010 FE00
ထ	1011	Myanmar letter Tha	Tha	ထ <sup>6</sup>	1011 FE00
ဆ	1080	Myanmar letter Shan Tha <sup>12</sup>	Tha	ဆ	1080 FE00
တ	107B	Myanmar letter Shan Da	Da	တ	107B (no dots needed)
ထ	AA6A	Myanmar letter Khamti Dha	Dha	ထ	AA6A (no dots needed)
န <sup>1</sup>	AA6B	Myanmar letter Khamti Na	Na	န	AA6B FE00
ပ	1015	Myanmar letter Pa	Pa	ပ	1015 FE00
ထ	1078	Myanmar letter Shan Ca <sup>2</sup>	Pha	ထ	1078 FE00
ပ	107F	Myanmar letter Shan Ba	Ba	ပ	107F (no dots needed)
ဆ	1079	Myanmar letter Shan Za <sup>2</sup>	Bha	ဆ	1079 (no dots needed)
မ	1019	Myanmar letter Ma	Ma	မ	1019 FE00
ယ	101A	Myanmar letter Ya	Ya	ယ	101A FE00
ဇ <sup>9</sup>	AA7A	Myanmar letter Aiton Ra	Ra	ဇ ဇ <sup>6</sup>	AA7A FE00
န <sup>1, 9</sup>	AA73	Myanmar letter Khamti Ra	Ra	န	AA73
ရ <sup>9</sup>	101B	Myanmar letter Ra	Ra	ရ	∅ <sup>10</sup>
လ	101C	Myanmar letter La	La	လ	101C FE00
ဝ	101D	Myanmar letter Wa	Wa	ဝ <sup>6</sup>	101D FE00
ဆ <sup>1</sup>	AA6C	Myanmar letter Khamti Sa	Sa	ဆ	AA6C FE00

Default letterform	Unicode	Unicode name	Used as Khamti letter	Khamti letterform	Khamti Unicode
ှ	AA6D	Myanmar letter Khamti Ha	Ha	ှ/ှ	AA6D / 0 <sup>10</sup>
လ	AA6E	Myanmar letter Khamti Hha <sup>3</sup>	Lla	လ/လ	AA6E / 0 <sup>11</sup>
၎ <sup>1</sup>	AA6F	Myanmar letter Khamti Fa	Fa	၎	AA6F FE00
ာ	1022	Myanmar letter Shan A	A	ာ	1022 FE00
ေ	1031	Myanmar vowel sign E	-e	ေ	1031 FE00
်	103A	Myanmar sign asat	asat	်	0 <sup>10</sup>
ဲ <sup>1, 7</sup>	109D	Myanmar vowel sign Aiton Ai	-ai	ဲ	0 <sup>10</sup>
ံ	1036	Myanmar sign anusvara		ံ	0 <sup>10</sup>
့ <sup>1</sup>	109A	Myanmar sign Khamti tone 1	tone 1	့	0 <sup>10</sup>
့	1089	Myanmar sign Shan tone 5	tone 2	့	0 <sup>10</sup>
့ <sup>1</sup>	109B	Myanmar sign Khamti tone 3	tone 3	့	0 <sup>10</sup>
့	1087	Myanmar sign Shan tone 2	tone 4	့	0 <sup>10</sup>
့	1088	Myanmar sign Shan tone 3	tone 5	့	0 <sup>10</sup>
့	1038	Myanmar sign visarga	tone 6	့	0 <sup>10</sup>
့	108A	Myanmar sign Shan tone 6	tone 8	့	0 <sup>10</sup>
့ <sup>1</sup>	AA74	Myanmar logogram Khamti Oay		့	0 <sup>10</sup>
့ <sup>1</sup>	AA75	Myanmar logogram Khamti Qn		့	0 <sup>10</sup>
့ <sup>1</sup>	AA76	Myanmar logogram Khamti Hm		့	0 <sup>10</sup>

**1** These codepoints were added solely for Khamti-Aiton-Phake but have been given default glyphs that do not look like Khamti-Aiton-Phake.

**2** In Khamti, these represent characters different to their defaults.

**3** AA6E လ Lla is incorrectly encoded as Khamti Hha.

**4** For letter Nya, Khamti uses AA65 ဟ while Aiton and Phake use the Shan letter 107A ရာ.

**5** Khamti may use Shan Ga and Jha rather than the Burmese and Khamti codepoints.

**6** The styling of wraparound Ra varies, this can be decided by font makers.

**7** UTN11 indicates ဲ uni1032 for this vowel character in Khamti. As noted on p2, ံ uni1086 or ဲ uni109D represent the same character and are also used.

**9** UTN11 lists ရ 101B for Khamti consonant Ra, and လ AA73 as the historic form; this conflicts with Khamti proposal 08-276, where ရှ AA7A is the historic form, and လ AA73 is the revised form. The dotted version of ရ 101B, is not currently covered by the variation selector mechanism.

**10** Dotted versions of these are not currently enabled with the variation selector mechanism. Letters Na 107C, and Ha AA6D may appear either with or without dots. The others always have the dots, unless the entire Khamti range takes an undotted style. Unicode document 15-320 [\[1\]](#) states the tone marks should always be filled in Khamti. Anusvara (uni1036) and the logograms (AA74, AA75 and AA76) are presented with filled dots in proposal 08-276 [\[2\]](#) but without dots in the code chart. Note also that variation selectors are not allowed for non-spacing marks, so an alternative solution will be needed for these.

**11** The choice of a dot or a loop is a stylistic one. Fontmakers can choose which forms to use.

**12** This letter is not used Khamti language, and its inclusion in the variation selector mechanism is likely the result of confusion of Burmese Sa သ / Shan Tha သ (pronounced /θa/ and romanised 'tha') with Tha ထ (/tʰa/, romanised 'hta').

**13** In UTN11, the Khamti letter Nna ဓ is represented using uni107C ဓ Shan Na, rather than uniA9E3 ဓ Shan Nna. 107C has the correct form in an undotted Khamti style, but is not letter Nna, while A9E3 is closer to the correct form in a dotted style, and is the equivalent letter.