

Shan-Related Issues with N3080: Preliminary proposal for encoding Karen, Shan, and Kayah Characters

Submission by: Dr Richard Wordingham (richard.wordingham@ntlworld.com)
Mr Patrick Chew, M.A. (Linguistics)
Date: 7 May 2006

We would like the UTC to consider the following eleven Shan-related issues when discussing agenda item C.13.6: Preliminary proposal for encoding Karen, Shan, and Kayah characters. Figure references are to N3080.

Contents

1.	Naming of Consonants.....	1
2.	Doubtful Distinctions.....	2
3.	MYANMAR LETTER SHAN 'THA'.....	2
4.	MYANMAR SIGN SHAN COUNCIL EMPHATIC TONE.....	2
5.	Hsipaw Letters, SHAN BA and SHAN DDDA.....	2
6.	Stacking Behaviour.....	4
7.	Annotations for Shan Tones.....	4
8.	SHAN CONTEXTUAL VOWEL VARIANTS.....	4
8.1	Analysis.....	4
8.2	Vowels for /e/.....	5
8.3	Vowels for /ɛ/.....	5
8.4	Vowels for /a/.....	5
8.5	Conclusion.....	6
9.	SHAN Vowel AY.....	6
10.	SHAN Vowel /ɔ/.....	6
11.	Underdot Virama.....	6

Issues 1, 3, 5 and 8 relate to the naming of characters

Issues 2, 3, 4, 8 and 11 relate to the unification of characters.

Issues 4 and 6 are requests for information needed for the rendering of text.

Issues 7, 10 and 11 relate to the annotation of characters names.

Issues 8 and 9 relate to shaping.

1. Naming of Consonants

Cannot the phrase 'SHAN PALI' be simplified to 'SHAN' in all cases? In particular, it seems inappropriate in 'LETTER SHAN PALI GA', for this is used in Shan for words of foreign origin – see N3080 Figure 1.

2. Doubtful Distinctions

Two distinctions between Burmese and Shan consonants seem to depend on the 'plain-text monofont requirement', that a single non-language sensitive font be able to render languages encoded in the Myanmar font as each language's speakers expect. If this requirement is not valid, then the case for making them separate characters is also invalid:

MYANMAR LETTER KHA v. MYANMAR LETTER SHAN KHA
The significant difference is the slight kink in the Shan glyph.

MYANMAR LETTER NNYA v. MYANMAR LETTER SHAN NYA

3. MYANMAR LETTER SHAN 'THA'

The name 'THA' has been proposed because one of the uses of this letter ၵ is to represent the sound of ၵ MYANMAR LETTER SA, namely [θ]. This name suggests that Shan uses it instead of MYANMAR LETTER THA, which is grossly misleading. A better name would be 'MYANMAR LETTER SHAN THHA'.

From Figure 1 Row 2 Character 2 it may be seen that this character is also used in Pali (in the Shan tradition) as the CHA character. This suggests the alternative name MYANMAR LETTER SHAN CHA.

Again, this is a language-specific variant of a Burmese character, namely ၵ MYANMAR LETTER CHA, so should they be unified?

4. MYANMAR SIGN SHAN COUNCIL EMPHATIC TONE

More information is required on how MYANMAR SIGN SHAN COUNCIL EMPHATIC TONE is to be rendered:

- (1) In CVC syllables, is it the final consonant that is underlined?
- (2) Does the underline normally only extend to the last cell of a consonant, or is that an artefact of the simple font or typewriter used in the examples?
- (3) How does it interact with subscript vowels and consonants?

If its behaviour is the same, should it be distinguished from U+0331 COMBINING MACRON BELOW? The name suggests that it is actually based on the latter.

5. Hsipaw Letters, SHAN BA and SHAN DDDA

No evidence is provided for the 'SHAN HSIPAW' letters – the purported evidence demonstrates the existence of two hitherto unproposed consonants:

MYANMAR LETTER SHAN BA (same glyph as given for MYANMAR LETTER SHAN HSIPAW RA) and

MYANMAR LETTER SHAN DDDA (same glyph as given for MYANMAR LETTER SHAN HSIPAW FA).

Figure 1 shows 𑄠 SHAN BA twice, in Row 5 Column 3 (labelled by 𑄡 MYANMAR LETTER BA and 𑄢 THAI CHARACTER PHO PHAN) and in Row 8 Column 3 (labelled by 𑄡 MYANMAR LETTER BA and 𑄣 THAI CHARACTER BO BAI MAI). (Rows 1 to 7 show correspondences for Pali; Row 8 seems to show correspondences for Thai sounds that are not native to standard Shan.) Figure 1 shows 𑄤 SHAN DDDA in Row 8 Column 1 (labelled by 𑄥 MYANMAR LETTER DA and 𑄦 THAI CHARACTER DO DEK.)

Figure 2 is less easy to interpret, for the sample is labelled only in Shan. The first section is labelled 'mae3 seng ngao3' meaning 'letters of basic sounds' and lists the Shan letters for the sounds /k/, /kʰ/, /ŋ/, /s/, /sʰ/, /ɲ/, /t/, /tʰ/, /n/, /p/, /pʰ/, /f/, /m/, /j/, /r/, /l/, /w/, /h/, /ʔ/. /f/ is actually limited to eastern Shan dialects and SHAN NYA is still found in native words (rare) as an initial, though the majority of dictionary entries seem to be Burmic, Pali, or "regional differences". The sound /r/ is, however, actually of foreign origin in Shan. The second section is labelled 'mae3 seng kaem' meaning 'letters for borrowed sounds'. It then lists the four letters SHAN (PALI) GA, SHAN DDDA, SHAN BA and SHAN TH(H)A. Now, what would make sense as borrowed sounds needing letters? I suggest the most important sounds would be /g/, /d/, /b/ and /θ/, all needed for Burmese.

SHAN DDDA is a Shan invention, possibly under the inspiration of the Thai split of Brahmic TA into THAI CHARACTER DO DEK and THAI CHARACTER TO TAO. Visually, there is a graphic and mostly auditory correspondence amongst the Shan letters:

𑄡 GA : 𑄢 KA 𑄤 DDDA : 𑄥 TA
 𑄠 BA : 𑄡 PA 𑄦 TH(H)A : 𑄣 SA

which would explain the order of these four letters – it is the Brahmic order of KA, TA, PA, SA.

I will now explain the name 'DDDA'. In mainland South East Asia, there is a rough correspondence between voicing and implosion – the modern Tai voiced stops were generally formerly pre-glottalised, and in loanwords corresponded to Khmer and Mon implosive stops. For Brahmic scripts, the conventions are:

B/G/D	plain voiced
DD	retroflex
BB/GG/DDD	Implosive (c.f. the Sindhi implosive stops and the proposed MYANMAR CHARACTER MON BBA and MYANMAR CHARACTER MON BBE)

We may also recall the extended pinyin use of 'bb' and 'dd' for voiced consonants.

6. Stacking Behaviour

The stacking behaviour of the Shan additions has not been described. Although three Burmese medials are used, I see no other evidence that the additions stack. Indeed, the Pali examples (Figure 1) suggest that they do not.

7. Annotations for Shan Tones

MYANMAR SIGN VISARGA should be annotated to record that it is, inter alia, one representation of Shan Tone 4. MYANMAR SIGN VISARGA and MYANMAR SIGN SHAN COUNCIL TONE-4 should cross-reference one another.

Cushing calls tone 5 the emphatic tone. This should be recorded in the annotations because of potential confusion with MYANMAR SIGN SHAN COUNCIL EMPHATIC TONE.

8. SHAN CONTEXTUAL VOWEL VARIANTS

8.1 Analysis

A quick review of the evidence in N3080 shows that within the Shan language, four simple vowels have contrasting medial (i.e. CVC) and final (i.e. CV) forms – AA /a/, E /e/, AE /ɛ/, and I /i/. (The U vowels do not function as contextual variants – within closed syllables at least, U represents /u/ and UU represents /o/.) However, the /i/ vowels contrast in Pali and occasionally in Shan, so there is no possibility of unifying the two I forms.

For the three remaining pairs, the following encodings would yield appropriate glyphs:

Medial /e/	MYANMAR VOWEL SIGN SHAN EE ABOVE (proposed)
Final /e/	MYANMAR VOWEL SIGN E (established)
Medial /ɛ/	MYANMAR VOWEL SIGN SHAN E ABOVE (proposed)
Medial /a/	MYANMAR VOWEL SIGN SGAW KAREN EU (proposed in N3043)
Final /a/	MYANMAR VOWEL SIGN SHAN AA (proposed)

Specimen forms from Figure 2, using မ MA as the final consonant, are:

Vowel	Final Vowel	Medial Vowel
/a/	မာ	မာ်
/ɛ/	မေ	မေ်
/e/	မေ	မေ်

Note that MYANMAR VOWEL SIGN SHAN FINAL Y makes a syllable a CVC context, so in the fourth section of Figure 2 we have the example form မာ်ယံ.

8.2 Vowels for /e/

Unification with the Myanmar script makes it confusing to unify these forms which are contextual variants in the Shan language. The medial form appears to be identical with a symbol used in the Mon language. What is the function of this symbol in the Mon language? Is it a vowel?

It is confusing to have different vowel names for the two forms. I suggest the name of the new character be changed to MYANMAR VOWEL SIGN SHAN E ABOVE.

8.3 Vowels for /ɛ/

MYANMAR VOWEL SIGN SHAN E ABOVE is proposed for the medial form representing the sound /ɛ/. There is also a final form, dubbed 'VOWEL SIGN SHAN E' in N3080 Figure 2. If these two forms are to be unified, their behaviour needs to be documented. I believe that this alternating behaviour should only be allowed if MYANMAR VOWEL SIGN SHAN AA also has such behaviour. It is not certain that this alternating behaviour can be implemented – see 'Shan Vowel AY' below.

I believe the name should be changed – 'AE' or 'EH' would be better.

8.4 Vowels for /a/

There is a medial form identical to that of the proposed VOWEL SIGN SGAW KAREN EU. It is labelled VOWEL SIGN SHAN A in Figure 2, and is the vowel of the second word in the rows of words with closed syllables in Figure 2. Note that medial form is used when the syllable has MYANMAR VOWEL SIGN SHAN FINAL Y.

Should this vowel have contextual shaping?

For unification:

- 1) No known contrast between the two forms in the Shan *language*.

Against unification:

- 1) Medial and final E are better not unified; unifying the AA and unifying the AE but not unifying the E would cause confusion.
- 2) One 'Shan-Thai reader' uses the medial AA for the final vowels in Pali.
- 3) Shan vowel AY may prevent identification of the context.

If unification is rejected, then a decision must be made as to whether there should be two visually and behaviourally identical vowel signs VOWEL SIGN SGAW KAREN EU and, say, VOWEL SIGN SHAN MEDIAL AA.

8.5 Conclusion

If contextual shaping is rejected for these vowels, six characters will be required, which I suggest should be named:

Medial /e/	MYANMAR VOWEL SIGN SHAN E ABOVE (proposed, but with EE not E in the name)
Final /e/	MYANMAR VOWEL SIGN E (established)
Medial /ɛ/	MYANMAR VOWEL SIGN SHAN AE ABOVE (proposed, but with E not AE in the name)
Final /ɛ/	MYANMAR VOWEL SIGN SHAN AE (not yet proposed)
Medial /a/	MYANMAR VOWEL SIGN SGAW KAREN EU (proposed in N3043) <i>or</i> MYANMAR VOWEL SIGN SHAN MEDIAL AA (not yet proposed)
Final /a/	MYANMAR VOWEL SIGN SHAN AA (proposed)

9. SHAN Vowel AY

This compound symbols represents /aw/.

This vowel symbol was not proposed, though it is shown in Figure 1 between the illustrations of MYANMAR VOWEL SIGN SHAN FINAL Y and, in one vowel combination, MYANMAR CONSONANT SIGN SHAN MEDIAL WA and VOWEL SIGN SHAN AA. It occurs in such common words as nay⁴ 'in', may² 'new' and hay³ 'give'. Perhaps it was not proposed because it was seen as composed of Shan medial WA and asat.

The use of asat for this vowel complicates syllable-breaking and thus:

- 1) simple checks that Shan is spelt with the correct form of I, E, AE/EH or AA., or contextual shaping if that is to be employed.
- 2) spell-checking word look-ups.
- 3) line-breaking.

The following issues need to be resolved:

- (a) Component ordering
- (b) Interaction with Shan medial WA representing /w/. (There may not be any - /ay/ does not occur in many words - Thai orthography retains only 20 instances.)

Shan Vowel OA (low) / Burmese Medial WA

10. SHAN Vowel /ɔ/

A statement needs to be made somewhere that it is the Burmese Medial WA, not the stacked WA, that is used as a vowel in Shan. This vowel sign is frequently triangular.

11. Underdot Virama

Figure 1 shows the use of an underdot virama, fully comparable to THAI CHARACTER PHINTHU. If this is to be unified with MYANMAR SIGN DOT BELOW, there should be a statement to that effect.