

## Two Issues Relating to N3044: Proposal to encode Mon and S'gaw Karen characters

Submission by: Dr Richard Wordingham  
 Contact e-mail: [richard.wordingham@ntlworld.com](mailto:richard.wordingham@ntlworld.com)  
 Date: 5 May 2006

I would like the UTC to consider the following two issues when discussing agenda item C.4: Proposal to encode Mon and S'gaw Karen characters.

### Mon JHA v. Standard Myanmar JHA

Should the proposed character MYANMAR LETTER MON JHA (\*U+105B) actually be distinct from MYANMAR LETTER JHA (U+1008)?

Michael Everson has already answered,

"Because it is a language-specific variant which differs significantly from the "default" letter. This can be seen easily in the large alphabet charts. We have regularly disunified non-lookalikes, and this is one of those. Language specific fonts do not meet the "plain-text monofont" requirement for the encoding of the Myanmar-script languages of Union of Myanmar."

Is this sufficient reason? They are no more different than TUS 4.0 reference glyphs for U+0067 (LATIN SMALL LETTER G) and U+0261 (LATIN SMALL LETTER SCRIPT G). In normal English practice they are treated as one and the same letter - young children's school books use the latter, and as a young boy I can remember asking what the former was because I did not realise it was the same letter.

Returning to the Burmese/Mon forms, corresponding glyphs are used non-contrastively in the sister Lanna script. I have looked through several Lanna (mostly Thai) fonts, and none provides both glyphs. The differences between the glyphs in the Northern Thai fonts are remarkable – other characters do not differ so much. The Lao Tham glyph shows an intermediate form. For those not acquainted with the Lanna script, these glyphs are:

Lannaworld (N. Thai)	CR-Father (N. Thai)	PhrasongLao (Lao Tham)
		

While these are not from the Myanmar script, the scripts have much in common. Whether it is recent cultural influence or shared inheritance from the Mon culture I do not know.

## Encoding of Mon NGA

Mon texts show three different forms of the letter NGA – a form like the Burmese base form, a subscript form (described in the proposal as the 'diacritic'), and what can seem to be the Burmese base form with the 'diacritic' attached – the full Mon form.

The rules determining the choice of the Burmese base form versus the full Mon form (the representative glyph implicitly proposed for MYANMAR LETTER MON NGA) have not been presented, and Michael Everson has stated, 'This is a matter of spelling... We aren't doing contextual shaping with this aspect of the script'. On the other hand, I have also been told by someone else that the difference 'is not a matter of spelling', though possibly that just meant that it was a matter of taste.

If the difference is not a matter of spelling, then the lay-out rules need to be documented. If they cannot be documented, then the matter does become one of spelling.

According to the proposal, two encoding methods were considered:

- A. Encode the diacritic (parallel to the signs for medial consonants), and represent the full Mon form as the Burmese base form (U+1004) plus the diacritic.
- B. Encode the full Mon form as a letter (\*U+105E) and encode the diacritic as a conjoined form (U+1039 \*U+105E).

The proposal rejected method A because the full Mon form sometimes includes an extra loop. As conjunct forms will not in general need to be encoded separately, this does not seem a cogent reason to reject method A.

Method A seems a more economical method, and I would like the committee to consider its adoption instead. Under this scheme, both in Burmese and in Mon there will be two NGA forms – the base form, and the subscript form. The subscript forms will be different, normally depending on the language of the text.