

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document**Title: Further discussion of Myanmar medials****Source: Michael Everson****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2006-03-20**

Under the current encoding model for Myanmar, the VIRAMA is used to create subjoined characters and the special medial forms:

Sequence	Rendering	Unsupported rendering
က ka + ့ virama + က ka	ကက kka	
က ka + ့ virama + ဝ ga	ကဝ kga	
က ka + ့ virama + န na	ကန kna	
က ka + ့ virama + မ ma	ကမ kma	
က ka + ့ virama + ယ ya	ကျ kya	ကယ kya
က ka + ့ virama + ရ ra	ကရ kra	ကရ kra
က ka + ့ virama + လ la	ကလ kla	
က ka + ့ virama + ဝ wa	ကဝ kwa	ကဝ kwa
က ka + ့ virama + သ sa	ကသ ksa	
က ka + ့ virama + ဟ ha	ကဟ kha	ကဟ kha

The “unsupported” renderings here are required in Sanskrit and Pali. Some minority languages require very different support, which also cannot be achieved with the current encoding model.

က ka + ့ virama + ဝ ga	ကဝ kga (S’gaw Karen)
က ka + ့ virama + န na	ကန kna (Mon)
က ka + ့ virama + မ ma	ကမ kma (Mon)
က ka + ့ virama + ယ ya	ကယ kya (S’gaw Karen)
က ka + ့ virama + လ la	ကလ kla (Mon)
က ka + ့ virama + လ la	ကျ kl (S’gaw Karen)
က ka + ့ virama + ဝ wa	ကဝ kwa (Shan)
က ka + ့ virama + ဝ wa	ကဝ kwa (also Shan)

Each of these base characters has a different behaviour in Burmese, so the required behaviour in the minority languages cannot be achieved in plain text without recourse to special language-tagging or special language-specific fonts. It is not reasonable to expect minority languages using the Myanmar script to be *required* to use such methods to represent their languages, and not to be able to exchange plain text reliably among the Myanmar-script-using community. Consider a national list of personal

names in their natural orthographies. Should the database be expected to tag languages and switch fonts? It could certainly not be a plain-text database. If medials are not encoded uniquely in the Myanmar script, then we have the following problems:

Sequence	Rendering	Unsupported rendering
က ka + ့ virama + ဝ ya	ကျ kya	ကွ kya
က ka + ့ virama + ?? ya	ကွ kya (S'gaw Karen)	
က ka + ့ virama + ရ ra	ကြ kra	ကွ kra
က ka + ့ virama + လ la	ကွ kla	
က ka + ့ virama + ?? la	ကျ kla (S'gaw Karen)	
က ka + ့ virama + ?? la	ကွ kla (Mon)	
က ka + ့ virama + ဝ wa	ကွ kwa	ကွ kwa
က ka + ့ virama + ?? wa	ကွ kwa (Shan)	
က ka + ့ virama + ဟ ha	ကျ kha	ကွ kha
က ka + ့ virama + ?? ga	ကျ kga (S'gaw Karen)	

There aren't any source consonants for five of the minority medials. There is no way to resolve this in plain text with the current encoding model, unless pseudo-coding with variation selectors or more hacking with zero-width characters were employed. There is no advantage to that.

The proposed changes to the encoding model proposed in N3043 solves the problem simply and easily. It does not add any new architectural feature to the UCS, because medial consonants have already been encoded for other scripts (Limbu and Lepcha), and because subjoining is exemplified already by the Khmer COENG model. It *does* change the encoding model for the Myanmar script—solving many implementation problems and supporting minority languages. The minority medials are proposed in N3044.

က ka + ◻ virama + က ka	ကွ kka	
က ka + ◻ virama + ဝ ga	ကွ kga	
က ka + ◻ virama + န na	ကွ kna	
က ka + ◻ virama + မ ma	ကွ kma	
က ka + ◻ virama + ဝ ya	ကွ kya	
က ka + ◻ virama + ရ ra	ကွ kra	
က ka + ◻ virama + လ la	ကွ kla	
က ka + ◻ virama + ဝ wa	ကွ kwa	
က ka + ◻ virama + သ sa	ကွ ksa	
က ka + ◻ virama + ဟ ha	ကွ kha	
က ka + ◐ medial Mon na	ကွ kna (Mon)	
က ka + ◑ medial Mon ma	ကွ kma (Mon)	
က ka + ◒ medial ya	ကျ kya	also ကျ kla (S'gaw Karen)
က ka + ◓ medial S. Karen ya	ကွ kya (S'gaw Karen)	also ကွ kla (Mon)
က ka + ◔ medial ra	ကြ kra	
က ka + ◕ medial wa	ကွ kwa	
က ka + ◖ medial Shan wa	ကွ kwa (Shan)	
က ka + ◗ medial ha	ကျ kha	also ကွ kga (S'gaw Karen)