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

Doc Type:	**Working Group Document**
Title:	**Requirements leading to the characters added to the Myanmar script**
Source:	**Michael Everson**
Status:	**Individual contribution**
Action:	**For consideration by JTC1/SC2/WG2**
Date:	**2006-05-17**

This document outlines, for the convenience of reviewers, the critical requirements for the support of Burmese script that the current encoding fails to satisfy. The seven characters proposed in N3043 lead to a new model which satisfies those requirements; their addition was approved by WG2 for inclusion in the PDAM 3.2 ballot.

1. Requirement for plain-text. The Myanmar script is used by national, state, and local administrations. Over 100 languages are spoken in Myanmar; at least eight of these (Burmese, Mon, S'gaw Karen, Eastern and Western Pwo Karen, Geba Karen, Kayah, and Shan) use the Myanmar script; the Ethnologue estimates these languages to comprise 91% of the population of the Union of Myanmar. The requirement for plain-text monofont support for text (including personal names) in these languages has been stated by authorities in Myanmar. Language-tagging and language-specific fonts are not acceptable solutions.

2. Requirement for full conjunct behaviour. In the current encoding, most of the consonants in the Myanmar script as used for Burmese subjoin when preceded by VIRAMA; four of them (YA, RA, WA, HA) do not.

Current sequence	Supported	Unsupported
က *ka* + ◌ *virama* + ယ *ya*	ကျ *kya*	ကၠ *kya*
က *ka* + ◌ *virama* + ရ *ra*	ကြ *kra*	ကꩡ *kra*
က *ka* + ◌ *virama* + ဝ *wa*	ကွ *kwa*	ကꩠ *kwa*
က *ka* + ◌ *virama* + ဟ *ha*	ကှ *kha*	ကꩦ *kha*

The full-form subjoined renderings in the left-hand column are sometimes required in Pali and Sanskrit contexts; the proposed characters add support for all of them:

New sequence	Supported
က *ka* + ◌ *virama* + ယ *ya*	ကꩡ *kya*
က *ka* + ◌ *virama* + ရ *ra*	ကꩡ *kra*
က *ka* + ◌ *virama* + ဝ *wa*	ကꩠ *kwa* (really a fallback since *medial wa* is preferred)
က *ka* + ◌ *virama* + ဟ *ha*	ကꩦ *kha*
က *ka* + ◌ *medial ya*	ကျ *kya*
က *ka* + ◌ *medial ra*	ကြ *kra*
က *ka* + ◌ *medial wa*	ကွ *kwa*
က *ka* + ◌ *medial ha*	ကှ *kha*

3. Requirement for Burmese/Karen rendering. In the current encoding, the *medial ya* shape is bound to the letter ယ *ya*, but in S'gaw Karen it must be bound to letter လ *la* and S'gaw Karen uses a different shape for *-ya*.

Current sequence	Burmese supported	S'gaw Karen unsupported
က *ka* + ◌ *virama* + ယ *ya*	ကျ *kya*	ကၞ *kya*
က *ka* + ◌ *virama* + လ *la*	က္လ *kla*	ကျ *kla*

The proposed characters support both:

New sequence	Burmese supported	S'gaw Karen supported
က *ka* + ◌ *virama* + ယ *ya*	ကျ *kya*	
က *ka* + ◌ *virama* + လ *la*	ကျ *kla*	
က *ka* + ◌ *medial ya*	ကျ *kya*	ကျ *kla*
က *ka* + ◌ *Mon medial la*		ကၞ *kya*

4. Requirement for handling AA and LONG AA. In the current encoding, the vowel sign ◌ာ/◌ါ *ā* shapes contextually depending on the previous consonant. This practice works for modern Burmese, but Mon, which can also be said to shape contexually, does not agree entirely as to which letters should be followed by the tall form. Moreover, in S'gaw Karen, the tall form is used with all consonants, and the short form, with ◌် *asat*, is used *not* as a vowel sign, but as a tone mark. Note that Burmese has an optional Burmese spelling in the word ပွား/ပွါး *pwā:* 'proliferate' cannot be supported by the current encoding.

Current sequence	Supported	Unsupported
ဟ *ha* + ◌ာ *ā*	ဟာ *hā*	ဟါ *hā* (S'gaw Karen)
ပ *pa* + ◌ာ *ā*	ပါ *pā*	
ပ *pa* + ◌ *vir.* + ဝ *wa* + ◌ာ *ā* + း	ပွား *pwā:* (Burmese)	ပွါး *pwā:* (Burmese; Mon)
ဖ *pha* + ◌ာ *ā*	ဖာ *phā*	ဖါ *phā* (Mon)

The proposed character supports both:

New sequence	Supported
ဟ *ha* + ◌ာ *ā*	ဟာ *hā*
ဟ *ha* + ◌ါ *ā*	ဟါ *hā* (S'gaw Karen)
ပ *pa* + ◌ာ *ā* + ◌် *asat*	ပာ် *pa²* (S'gaw Karen)
ပ *pa* + ◌ါ *ā*	ပါ *pā*
ပ *pa* + ◌ *medial wa* + ◌ာ *ā* + း	ပွား *pwā:* (Burmese)
ပ *pa* + ◌ *medial wa* + ◌ါ *ā* + း	ပွါး *pwā:* (Burmese; Mon)
ဖ *pha* + ◌ာ *ā*	ဖာ *phā*
ဖ *pha* + ◌ါ *ā*	ဖါ *phā* (Mon)

5. Requirement for handling GREAT SA. In the current encoding, stacked *sa*-over-*sa* is unsupported, though this is needed in some Buddhist texts. It is best to consider GREAT SA as a new letter, like Latin *v* + *v* = *w*.

Current sequence	Supported	Unsupported
သ *sa* + ◌ *virama* + သ *sa*	ဿ *ssa* ("Great Sa")	ဿ *ssa*

The full-form subjoined renderings in the left-hand column are sometimes required in Pali and Sanskrit contexts; the proposed character supports both types of rendering, and virama stacks *sa* just as any other consonant:

New sequence	Supported
သ *sa* + ◌ *virama* + သ *sa*	ဿ *ssa*
ဿ *ssa*	ဿ *ssa*

Note that GREAT SA will behave like any other consonant (though this form does not occur):

ဿ *ssa* + ◌ *virama* + ဿ *ssa*	ဿ္ဿ *ssssa*

6. Requirement for representing KILLER without excess complication. In the current encoding, the high-frequency character KILLER is rendered by following VIRAMA with the ZERO WIDTH NON-JOINER.

Current sequence	Supported
က *ka* + ◌ *vir.* + က *ka*	က္က *kka*
က *ka* + ◌ *vir.* + ZWNJ + က *ka*	က်က *kka*
ဇ *ja* + ေ◌ *e* + ◌ာ *ā* + *vir.* + ZWNJ	ေဇာ် *jau* 'Zaw'

The simplified encoding treats VIRAMA analogously to the Khmer COENG; it has no visible shape. The KILLER *asat* is a simple diacritical mark. This is advantageous because the KILLER is used on vowel signs to indicate tones, and does not function only as a vowel killer as VIRAMA does in Devanagari and other scripts. The use of ZERO WIDTH NON-JOINER is avoided.

New sequence	Supported
က *ka* + ◌ *virama* + က *ka*	က္က *kka*
က *ka* + ◌် *asat* + က *ka*	က်က *kka*
ဇ *ja* + ေ◌ *e* + ◌ာ *ā* + ◌် *asat*	ေဇာ် *jau* 'Zaw'

7. Requirement for representing all forms of *kinzi*. In the current encoding, the selection of ◌ *kinzi* as opposed to a subjoined letter is ambiguous in certain clusters. Note in particular that the ZWNJ in "*ṅraṅ:*" below breaks the syllable before the tonemark ":".

Current sequence	Correct	Incorrect
င *ṅa* + ◌ *vir.* + ဝ *wa* + ေ◌ *e*	ေင်ွ *ṅwe* 'silver'	ေင်ွ
င *ṅa* + ◌ *vir.* + ရ *ra* + င *ṅa* + ◌ *vir.* + ZWNJ + ː	ြင်ႏ: *ṅraṅ:* 'refuse'	ရ်င်း

The proposed characters support both forms predictably. They also support both spelling forms of the word for 'Singapore', as well as the ◌ *repha* in Sanskrit words:

New sequence	Supported
င *ṅa* + ◌ *medial-wa* + ေ◌ *e*	ေင်ွ *ṅwe* 'silver'
င *ṅa* + ◌် *asat* + ◌ *vir.* + ဝ *wa* + ေ◌ *e*	ေင်ွ
င *ṅa* + ◌ *medial-ra* + င *ṅa* + ◌် *asat* + ː	ြင်ႏ: *ṅraṅ:* 'refuse'
င *ṅa* + ◌် *asat* + ◌ *vir.* + ရ *ra* + င *ṅa* + ◌် *asat* + ː	ရ်င်း
စ *ca* + င *ṅa* + ◌် *asat* + က *ka* + ◌ာ *ā* + ပ *pa* + ◌ *ū*	စင်္ကာပူ *caṅkāpū* 'Singapore'
စ *ca* + င *ṅa* + ◌် *asat* + ◌ *vir.* + က *ka* + ◌ာ *ā* + ပ *pa* + ◌ *ū*	စင်္ကာပူ *caṅkāpū* 'Singapore'
န *na* + ◌ိ *i* + ရ *ra* + ◌် *asat* + ဝ *wa* + န *na*	နိရ္ဝန *nirwana* 'nirvaṇa'
န *na* + ◌ိ *i* + ရ *ra* + ◌် *asat* + ◌ *vir.* + ဝ *wa* + န *na*	နိရ်ဝန *nirwana* 'nirvaṇa'