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

Doc Type: **Working Group Document**
Title: **Proposal to encode seven additional Myanmar characters in the UCS**
Source: **Ireland (NSAI), United Kingdon (BSI), Myanmar Language Commission, Myanmar Unicode and Natural Language Processing Research Center, Myanmar Computer Federation**
Status: **Member Body contribution**
Replaces: **N2827, N1883R**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Date: **2006-02-28**

Request. This document asks for a number of disunifications which simplify Myanmar script processing and rendering. The disunifications are necessary for the encoding of minority languages which use the Myanmar script, and solve in addition a number of long-standing problems which have prevented the successful implementation of the Myanmar script for the first official language of the Union of Myanmar. *For reasons of urgency of implementation, we ask that the characters requested here be added to the current balloting FPDAM3 of ISO/IEC 10646*. The disunifications requested are:

- A new *U+102B MYANMAR VOWEL SIGN TALL AA is disunified from a contextual variant of U+102C MYANMAR VOWEL SIGN AA.
- A new *U+103A MYANMAR SIGN ASAT is disunified from the existing sequence U+1039 MYANMAR SIGN VIRAMA U+200C ZERO WIDTH NON-JOINER. The glyph for VIRAMA is changed.
- A new *U+103B MYANMAR CONSONANT SIGN MEDIAL YA is disunified from the existing sequence U+1039 MYANMAR SIGN VIRAMA U+101A MYANMAR LETTER YA.
- A new *U+103C MYANMAR CONSONANT SIGN MEDIAL RA is disunified from the existing sequence U+1039 MYANMAR SIGN VIRAMA U+101B MYANMAR LETTER RA.
- A new *U+103D MYANMAR CONSONANT SIGN MEDIAL WA is disunified from the existing sequence U+1039 MYANMAR SIGN VIRAMA U+101D MYANMAR LETTER WA.
- A new *U+103E MYANMAR CONSONANT SIGN MEDIAL HA is disunified from the existing sequence U+1039 MYANMAR SIGN VIRAMA U+101F MYANMAR LETTER HA.
- A new *U+103F MYANMAR LETTER GREAT SA is disunified from the existing sequence U+101E MYANMAR LETTER SA U+1039 MYANMAR SIGN VIRAMA U+101E MYANMAR LETTER SA.
- The glyph for U+104E MYANMAR SYMBOL AFOREMENTIONED is changed.

(The asterisk is used to show characters which are not yet encoded.)

If this proposal is adopted, the following characters will exist:

102B MYANMAR VOWEL SIGN TALL AA

1039 MYANMAR SIGN VIRAMA [glyph change and note change]
• shape shown is arbitrary, not rendered

်	103A	MYANMAR SIGN ASAT = killer, atha	
ျ	103B	MYANMAR CONSONANT SIGN MEDIAL YA • used for medial la in S'gaw Karen	
ြ	103C	MYANMAR CONSONANT SIGN MEDIAL RA	
ွ	103D	MYANMAR CONSONANT SIGN MEDIAL WA	
ှ	103E	MYANMAR CONSONANT SIGN MEDIAL HA	
ဿ	103F	MYANMAR LETTER GREAT SA	
၎	104E	MYANMAR SYMBOL AFOREMENTIONED	[glyph change]

with the following properties:

```
102B;MYANMAR VOWEL SIGN TALL AA;Mc;0;L;;;;;N;;;;;
1039;MYANMAR SIGN VIRAMA;Mn;9;NSM;;;;;N;;;;;
103A;MYANMAR SIGN ASAT;Mn;0;NSM;;;;;N;;;;;
103B;MYANMAR CONSONANT SIGN MEDIAL YA;Mc;0;L;;;;;N;;;;;
103C;MYANMAR CONSONANT SIGN MEDIAL RA;Mc;0;L;;;;;N;;;;;
103D;MYANMAR CONSONANT SIGN MEDIAL WA;Mn;0;NSM;;;;;N;;;;;
103E;MYANMAR CONSONANT SIGN MEDIAL HA;Mn;0;NSM;;;;;N;;;;;
103F;MYANMAR LETTER GREAT SA;Lo;0;L;;;;;N;;;;;
104E;MYANMAR SYMBOL AFOREMENTIONED;Po;0;L;;;;;N;;;;;
```

Rationale for medial disunification

The initial rationale for this request came from a recognition that support for S'gaw Karen requires the disunification of ျ MEDIAL YA from ယ YA because S'gaw Karen uses this shape for its ျ MEDIAL LA and an entirely different shape for its ၟ MEDIAL YA. This disunification allows greater simplicity in the rendering of Myanmar subjoined consonants: ယ YA, ရ RA, ဝ WA and ဟ HA which do, in older texts and in minority texts, occur in full subjoined forms. Compare S'gaw Karen ကျ *kla* and ကၟ *kya* with Burmese ကြ *kla* and ကျ *kya* and with Mon ကၟ *kla* and ကျ *kya*. Note that this is not simply a matter of "spelling". If the model is not changed, and S'gaw Karen were to use ယ YA to represent ျ to be read *-la*, there is still no letter which could yield ၟ S'gaw Karen *-ya* (Mon *-la*); လ LA doesn't work, because it has a normal subscript form ္လ in Burmese and Pali. (A new ၟ MYANMAR LETTER SGAW KAREN MEDIAL YA will be proposed in a subsequent document dealing with S'gaw Karen, Mon, and other minority languages.)

The current sequences (VIRAMA + YA/RA/WA/HA) remain valid sequences but for different renderings. Those renderings do not occur in modern Burmese, but they do occur in older orthography, in Pali and Sanskrit. Encoding the explicit medials allows for simple representation of both kinds of orthography.

Sequence	**Current rendering**	**Proposed rendering**
က *ka* + ် *virama* + ယ *ya*	ကျ *kya*	ကြ *kya*
က *ka* + ် *virama* + ရ *ra*	ကြ *kra*	ကၠ *kra*
က *ka* + ် *virama* + ဝ *wa*	ကွ *kwa*	ကွ *kwa*
က *ka* + ် *virama* + ဟ *ha*	ကှ *kha*	ကှ *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*

Note that *kwa* with MEDIAL WA may take a teardrop or triangular WA shape, which is never the case with true subjoined WA (which is rare, though it occurs in Sanskrit).

Rationale for ASAT ("killer") disunification

A big advantage of simplifying the use of ◌် VIRAMA is that the model becomes similar to the familiar encoding model for Khmer and Kharoshthi. The killer ◌် ASAT, which occurs with very high frequency in all of the languages of Myanmar, may now simply be rendered as a combining diacritical mark. The ongoing difficulties of implementation involving ZERO WIDTH NON-JOINER are now moot. Thus processes which remove ZWJ and ZWNJ now can handle Myanmar script without problem. (See Figure 3.)

Rationale for TALL AA disunification

Since S'gaw Karen only has one form of the the AA vowel, namely ◌ါ, the previous contextual variation of the AA between short and tall form is no longer universally applicable for the Myanmar script. Disunification of these two forms is therefore necessary and the introduction of VOWEL SIGN TALL AA has been proposed. Speakers of Burmese are taught to make the size distinction when they learn to write, and keyboards contain two separate keys for the short and tall forms, so no disadvantage will be had by Burmese-speaking users, who already make a distinction in practice if not in the current UCS encoding. It is important to note that in sorting and searching, TALL AA and AA should be considered equivalent. Compare S'gaw Karen ကါ *kā* and ဝါ *wā* with Burmese ကာ *kā* and ဝါ *wā*. In sorting, both ကာ and ကါ should be equivalent at the first level, with ကာ preceding ကါ to break ties.

Rationale for GREAT SA disunification

A consequence of the simplified encoding model is that SA VIRAMA SA will stack: သ္သ. The special conjunct form called GREAT SA is best rendered with the use of a single character ဿ, because GREAT SA is not just a ligature of two SAs: consider Pali သဿ *sassa* 'corn, crop'; this could also be written သ္သ or သသ်သ (though by convention it is not). Note that in sorting and searching, GREAT SA and SA VIRAMA SA should be considered equivalent; in sorting, they are equivalent at the first level, with ဿ preceding သ္သ to break ties.

Advantages for representation of *kinzi*

Another advantage of the simplifications to the encoding model is the rationalisation of the approach to encoding *kinzi*, which can now be considered a simple unambiguous sequence: NG + ASAT + VIRAMA. This is simply a rendering rule, where the NGA and the ASAT are drawn in reduced size and where the subjoined character retains its full size and position relative to the baseline. Thus NG + ASAT + VIRAMA + KA produces င် *ṅka*, while the unattested syllable form *င်္က *ṅka* must be produced by NGA + ASAT + ZWJ + VIRAMA + KA. The simplified encoding model allows the distinction between modern သင်္ချာ, older သင်္ခာရ, and even unattested *သင်္ခ်ာရ *saṅkhāra*. The superscript Sanskrit form of *repha* can also be met with this same model: နိဗ္ဗန် *nirvana* (Pali နိဗ္ဗန *nibbana*) is encoded NA + I + RA + ASAT + VIRAMA + WA + NA; when *repha* is not preferred, ASAT is deleted for နိဗ္ဗန NA + I + RA + VIRAMA + WA + NA.

An example of the problems with the previous model can be found with the word *ṅraṅ:* 'to refuse'. The current UCS encoding model requires this to be represented NGA + VIRAMA + RA | NGA + VIRAMA + ZWNJ + VISARGA. Instead of rendering correctly as ငြင်း, it renders incorrectly as *ရ်င်း. The proposed model will allow the user to choose either form (even the incorrect one) simply and predictably: NGA + MEDIAL-RA | NGA + ASAT + VISARGA, and NGA + ASAT + VIRAMA + RA | NGA + ASAT + VISARGA.

Rationale for change of glyph for SYMBOL AFOREMENTIONED

The base symbol ၎ used to represent the abbreviation လည်းကောင်း *lañ:koṅ:* has been drawn with a following ်း in the code charts: ၎်း. But the abbreviation can also be written with *kinzi*: ၌. The first of these should be represented explicitly with SYMBOL AFOREMENTIONED + NGA + ASAT + VISARGA, the second, *kinzi* form, should be represented as NGA + ASAT + VIRAMA + SYMBOL AFOREMENTIONED.

Impact on current implementations

All of the implementors who took part in the workshop—including local implementors in Myanmar as well as foreign implementors, including members of SIL—are in unanimous agreement that the changes requested in this proposal meet their requirements, solve their problems, and are the way forward. Impact on existing implementations of Unicode 4.1 text can be considered effectively nil. All implementations within Myanmar are still experimental and we are aware of only *one* complete Unicode 4.1 compliant implementation outside of Myanmar—and that implementor (Martin Hosken of SIL) has stated that he would be only too happy to change his software to support these changes and make any transcoding changes required.

Rationale for fast-tracking request

It is worth revisiting the previous WG2 decision in N1883R (1998-09-24):

> "The Myanmar delegation agreed after some discussion that [KA + VIRAMA + YA —> ကျ *kya*] was a possible and consistent representation, and that it would avoid the problem of possible alternative representations of the same text (the problem which would arise if other subjoined consonants were to be encoded when using the virama model). They agreed to engage in testing of Myanmar text representation with virama and without a separate encoding for glide consonants. In the meantime, they agreed that 10646 need not include the glide consonant forms, as long as sufficient structural gaps were left so that if implementation experience proves that they must be encoded, they could be added later without significant disruption of the core ordering of the chart. This requirement is met by the proposed disposition of comments, and this position was unanimously assented to by the Ad Hoc Committee."

Since 1998 it has been shown that the model without the medials does *not* work adequately. While all current implementations are at research level, some are ready for production and delaying a change could result in considerable text to transcode. *Participants in the Workshop on Myanmar Language Processing (Yangon, 13–15 February 2006) are unanimous in their desire to move forward in implementation with the new characters and code positions as listed in this proposal.*

Acknowledgements

This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley).

Figure 1. From left to right, some supporters of this proposal:
U Tun Tint (Myanmar Language Commission), U San Lwin (Director General, Myanmar Language Commission), U Thein Oo (President, Myanmar Computer Federation), Dr Kyaw Thein (Vice-President, Myanmar Computer Federation), U Aung Myint (Deputy Minister, Civil Service Selection & Training Board, Union of Myanmar), U Soe (Principal, Post and Telecommunication Training Center), Michael Everson (Evertype), Martin Hosken (Payap University), Dr Myint Myint Than (Director, Myanmar Computer Federation).

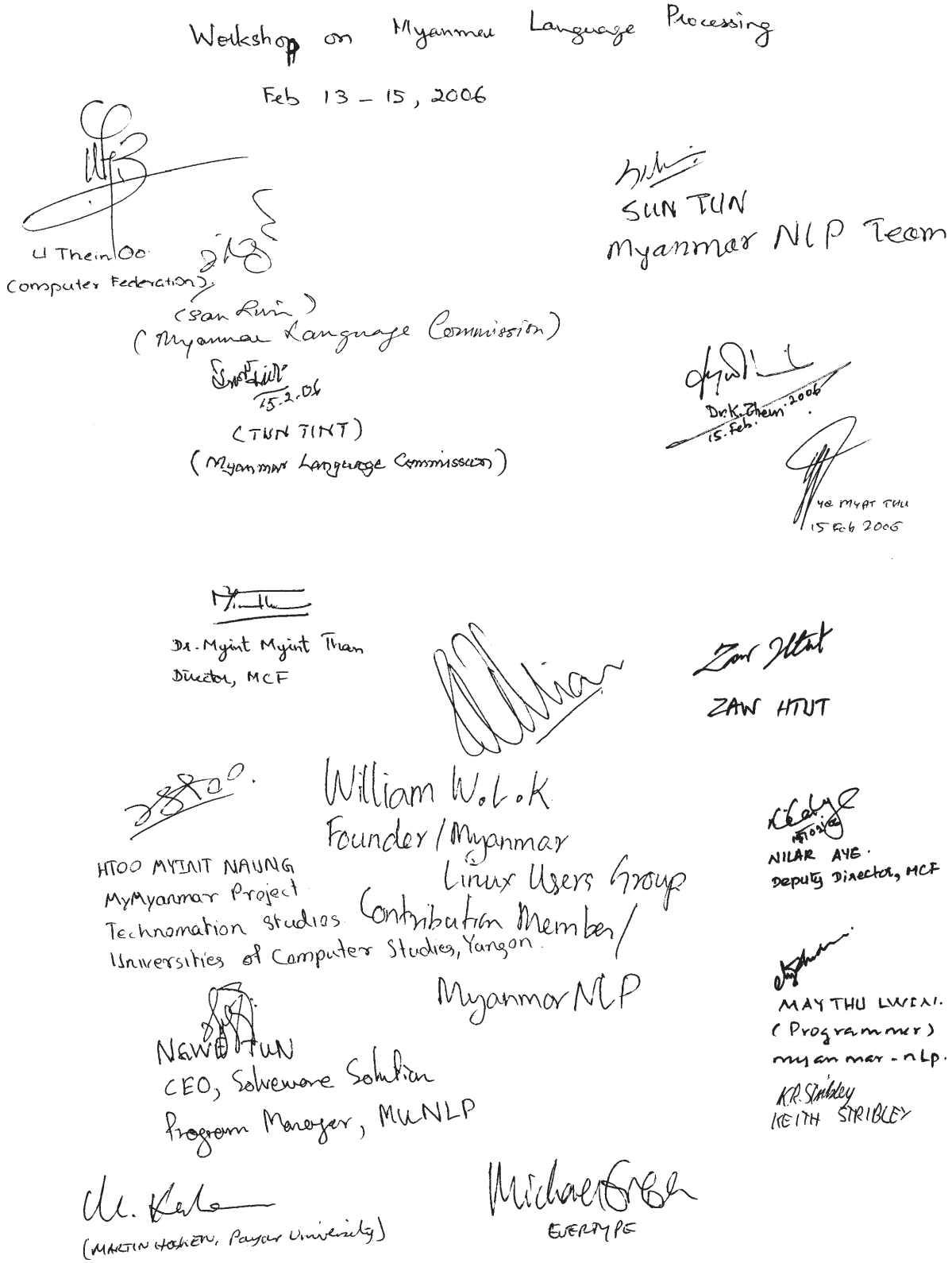


Figure 2. Signatures in support of this proposal. Signatories include a number of implementors, as well as representatives of the Myanmar Language Commission and the Myanmar Computer Federation

	Issue	Proposal A: UTN 11	Proposal B: NLP
1	File size – kinzi is rare	This may take more space, given kinzi's frequency compared to the medials used with nga, however, this factor is negligible on modern computers.	Perhaps slightly reduced file sized, but negligible, especially if compression is used. This argument has already been ignored for the case U+1004 U+1039 U+200C
2	Collation rules	In glibc locale 10 extra rules required – but this is negligible compared to the total number of rules required for Myanmar.[1]	No change to normal rules.
3	Unicode 4 compliance Note: UTN11 is only a technical note, it is not part of the standard.	Kinzi – always compliant. YRWH medials – deviate in special cases as specified in UTN 11.	Kinzi – deviates in special cases. YRWH medials – always compliant. UTN 11 – non-compliant.
4	Searching for Kinzi	Search for U+1004 U+1039 ![U+200C] where ! indicates a logical NOT This is a simpler than case B because U+200D will always be inserted before a medial if it follows U+1004 when used as a base. In practice you would normally search for a complete word, so this search will be rare.	Search for U+1004 U+1039 ![U+200C U+101A U+101B U+101C U+101D] where ! indicates a logical NOT and [...] matches any one of the characters inside the bracket. This is more complicated than A because of the medial forms. The multiple NOT is difficult for a user to specify.
5	Searching for Medial	e.g. medial U+101F ![U+1004] U+1039 U+101F This correctly finds the medial, but is slightly more complicated. Since Kinzi on YRWH is very rare the number of false positives without specifying ![U+1004] is very small (c.f. issue 1).	e.g. medial U+101F U+1039 U+101F This is slightly easier.
6	Old Myanmar – with the example given: ꩜ There are 2 possibilities: a) ဃ is a medial with ⌀ as the base consonant b) ⌀ is attached to another base consonant (equivalent to -ိဃ) with ဃ as the main consonant for a second syllable. This distinction is crucial for collation. It would be helpful to see some examples in real words.	Medial case: dealt with using U+1004 U+200D U+1039 U+101C Stacked case: Unspecified. Since this would normally give Kinzi when stacked, the NLP suggestion seems a sensible extension in this case.	Medial case: Unspecified, but since it does occur in medial form in ꩜ (drop) this must be representable somehow.[2] It must be distinguished from the normal stacked case for collation. Stacked case: U+1004 U+1039 U+101C
7	Rendering complexity – crude estimate of number of context specific rules required. This comparison is too simplistic: it is implementation dependent.[3]	Kinzi – one rule YRWH – two rules each Total: 9 rules Note: if character classes are used then this is only 3 rules.	Kinzi – two rules YRWH – one rule Total: 3 rules Fewer rules if character classes not supported.
8	Input Method complexity – assuming Kinzi is typed after the consonant it sits on top of. This relative complexity is implementation dependent.[4]	Context length 2 required for determining YRWH medial. i.e. after the sequence U+1004 [U+1031]? insert U+200D U+1039 [U+101A U+101B U+101D U+101F] before U+1031 where ? means the character occurs 0 or 1 times. (Note: the medials already need a much longer context than to support reordering of multiple medials)	Context length 2 required for determining which Kinzi to use. After [U+101A U+101B U+101D U+101F] [U+1031]? substitute U+1004 U+1039 U+200D before the sequence. (Note the substitution inserts 2 characters back, not 1, but this complexity is already necessary for Myanmar Input Methods).
9	Implied syllable breaking – this is the argument used in UTN11.	ZWJ signifies that the medial is linked to the base consonant within the syllable.	ZWJ appears (incorrectly) to join Kinzi with the consonant of the next syllable.

1 Arguably these rules should be present even for proposal B so that standalone medials could be sorted.

2 See example in UTN 11.

3 Implementation example using Graphite for rendering Padauk:
Proposal A: 200D just prevents the Kinzi rule from matching, no extra rule is required with 200D. The "Take Kinzi" character class has 4 extra characters compared to proposal B.
Proposal B: One extra rule is required, so it could be argued that A is marginally easier to implement.

4 For the Keyman Input Method (which uses character classes) two rules are needed to implement both proposals so there is no difference in complexity. You can switch between the two proposals by commenting out 2 lines and uncommenting the other 2.

Figure 3. Example of the kinds of problems which *kinzi* caused implementors endeavouring to follow the current UCS encoding model for the Myanmar script, with examples of the different attempts they used to formulate solutions. The implementors in case (A) are Martin Hosken and Maung Tun Tun Lwin; in case (B) they are the Myanmar Language Commission and Natural Language Processing Research Lab. Every bit of this complexity is moot under the new simplified encoding model used in this proposal.

Figure 4. MYANMAR SYMBOL AFOREMENTIONED shown with two spellings, ၎င်း and ၌.

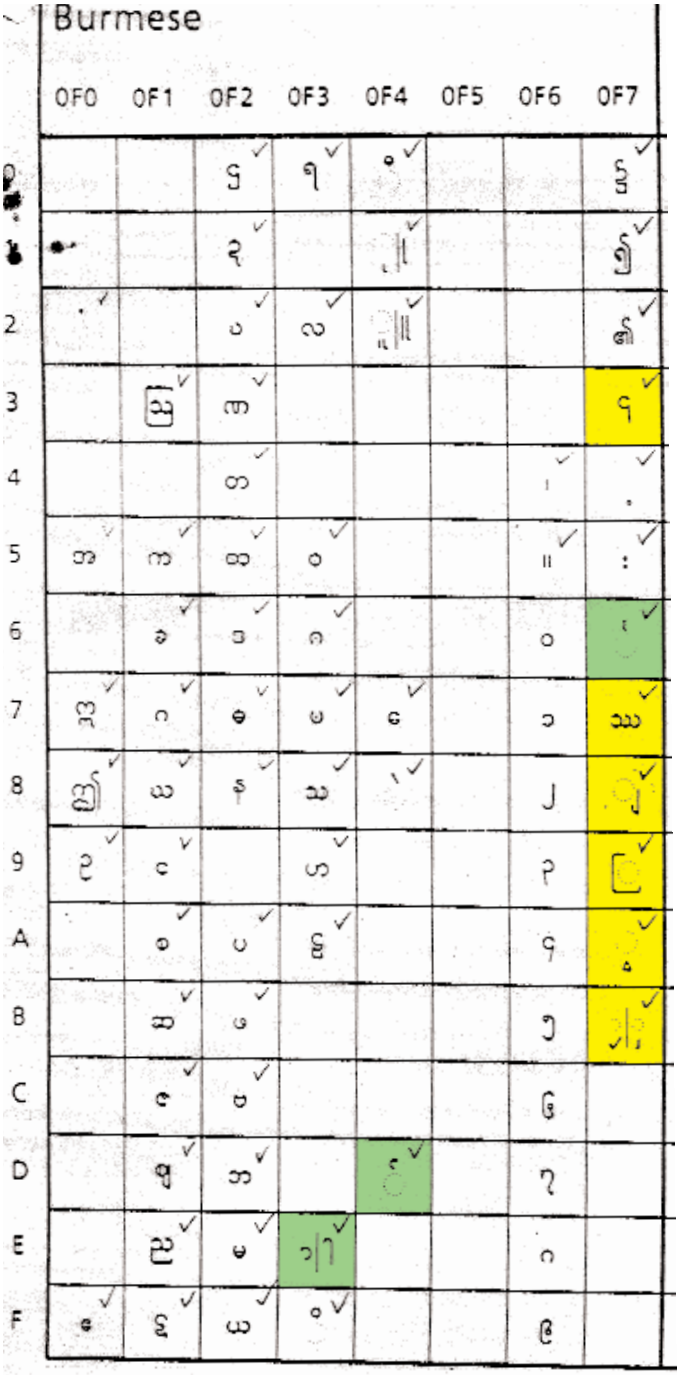
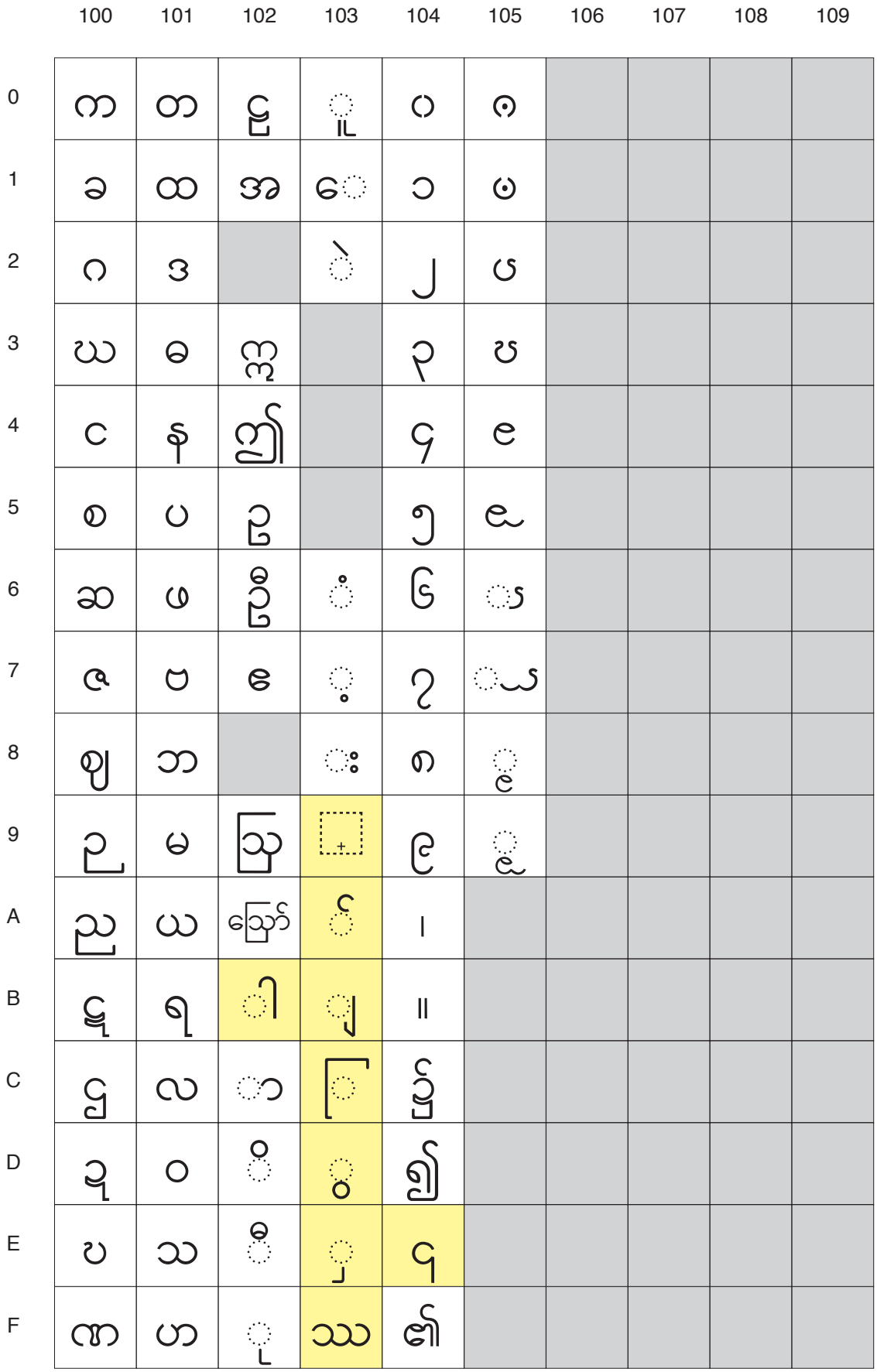


Figure 5. A draft pre-Unicode 1.0 chart for the Myanmar script, showing the base glyph for *0F73 SYMBOL AFOREMENTIONED, and encoding *0F77 GREAT SA, *0F78 MEDIAL YA, *0F79 MEDIAL RA, *0F7A MEDIAL WA, and *0F7C MEDIAL HA. This pre-Unicode 1.0 proposal used contextual rendering for AA and LONG AA (*0F3E), and did not distinguish between VIRAMA and ASAT (*0F4D). *Kinzi* was represented by a combining character in this pre-Unicode 1.0 proposal, which would have been less advantageous than the representation given in the present proposal. (The identification of the document isn't entirely certain. Burmese is given from *0F00-0F7F, followed by Khmer from 0F80. After the Burmese table is a list of names, at the end of which is the note: "Burmese letter names (very approximate etymological transliterations & notes by Lloyd Anderson)".)

We present this table here only to show that the encoding model we propose is not new. N1883R shows that there was clear consensus from representatives of the Ireland, Myanmar, the UK, and the US to adopt the current model. Although that model has proved inadequate, we are confident that the present proposal will solve the problems and enable Myanmar script processing with UCS encoding.

TABLE XX - Row 10: MYANMAR



G = 00
P = 00

TABLE XX - Row 10: MYANMAR

hex	Name
00	MYANMAR LETTER KA
01	MYANMAR LETTER KHA
02	MYANMAR LETTER GA
03	MYANMAR LETTER GHA
04	MYANMAR LETTER NGA
05	MYANMAR LETTER CA
06	MYANMAR LETTER CHA
07	MYANMAR LETTER JA
08	MYANMAR LETTER JHA
09	MYANMAR LETTER NYA
0A	MYANMAR LETTER NNYA
0B	MYANMAR LETTER TTA
0C	MYANMAR LETTER TTHA
0D	MYANMAR LETTER DDA
0E	MYANMAR LETTER DDHA
0F	MYANMAR LETTER NNA
10	MYANMAR LETTER TA
11	MYANMAR LETTER THA
12	MYANMAR LETTER DA
13	MYANMAR LETTER DHA
14	MYANMAR LETTER NA
15	MYANMAR LETTER PA
16	MYANMAR LETTER PHA
17	MYANMAR LETTER BA
18	MYANMAR LETTER BHA
19	MYANMAR LETTER MA
1A	MYANMAR LETTER YA
1B	MYANMAR LETTER RA
1C	MYANMAR LETTER LA
1D	MYANMAR LETTER WA
1E	MYANMAR LETTER SA
1F	MYANMAR LETTER HA
20	MYANMAR LETTER LLA
21	MYANMAR LETTER A
22	(This position shall not be used)
23	MYANMAR LETTER I
24	MYANMAR LETTER II
25	MYANMAR LETTER U
26	MYANMAR LETTER UU
27	MYANMAR LETTER E
28	(This position shall not be used)
29	MYANMAR LETTER O
2A	MYANMAR LETTER AU
2B	MYANMAR VOWEL SIGN TALL AA
2C	MYANMAR VOWEL SIGN AA
2D	MYANMAR VOWEL SIGN I
2E	MYANMAR VOWEL SIGN II
2F	MYANMAR VOWEL SIGN U
30	MYANMAR VOWEL SIGN UU
31	MYANMAR VOWEL SIGN E
32	MYANMAR VOWEL SIGN AI
33	(This position shall not be used)
34	(This position shall not be used)
35	(This position shall not be used)
36	MYANMAR SIGN ANUSVARA
37	MYANMAR SIGN DOT BELOW
38	MYANMAR SIGN VISARGA
39	MYANMAR SIGN VIRAMA
3A	MYANMAR SIGN ASAT
3B	MYANMAR CONSONANT SIGN MEDIAL YA
3C	MYANMAR CONSONANT SIGN MEDIAL RA
3D	MYANMAR CONSONANT SIGN MEDIAL WA
3E	MYANMAR CONSONANT SIGN MEDIAL HA
3F	MYANMAR LETTER GREAT SA
40	MYANMAR DIGIT ZERO
41	MYANMAR DIGIT ONE
42	MYANMAR DIGIT TWO
43	MYANMAR DIGIT THREE
44	MYANMAR DIGIT FOUR
45	MYANMAR DIGIT FIVE
46	MYANMAR DIGIT SIX
47	MYANMAR DIGIT SEVEN
48	MYANMAR DIGIT EIGHT
49	MYANMAR DIGIT NINE
4A	MYANMAR SIGN LITTLE SECTION
4B	MYANMAR SIGN SECTION
4C	MYANMAR SYMBOL LOCATIVE
4D	MYANMAR SYMBOL COMPLETED
4E	MYANMAR SYMBOL AFOREMENTIONED
4F	MYANMAR SYMBOL GENITIVE
50	MYANMAR LETTER SHA
51	MYANMAR LETTER SSA
52	MYANMAR LETTER VOCALIC R
53	MYANMAR LETTER VOCALIC RR
54	MYANMAR LETTER VOCALIC L
55	MYANMAR LETTER VOCALIC LL
56	MYANMAR VOWEL SIGN VOCALIC R
57	MYANMAR VOWEL SIGN VOCALIC RR
58	MYANMAR VOWEL SIGN VOCALIC L

hex	Name
59	MYANMAR VOWEL SIGN VOCALIC LL
5A	(This position shall not be used)
5B	(This position shall not be used)
5C	(This position shall not be used)
5D	(This position shall not be used)
5E	(This position shall not be used)
5F	(This position shall not be used)
60	(This position shall not be used)
61	(This position shall not be used)
62	(This position shall not be used)
63	(This position shall not be used)
64	(This position shall not be used)
65	(This position shall not be used)
66	(This position shall not be used)
67	(This position shall not be used)
68	(This position shall not be used)
69	(This position shall not be used)
6A	(This position shall not be used)
6B	(This position shall not be used)
6C	(This position shall not be used)
6D	(This position shall not be used)
6E	(This position shall not be used)
6F	(This position shall not be used)
70	(This position shall not be used)
71	(This position shall not be used)
72	(This position shall not be used)
73	(This position shall not be used)
74	(This position shall not be used)
75	(This position shall not be used)
76	(This position shall not be used)
77	(This position shall not be used)
78	(This position shall not be used)
79	(This position shall not be used)
7A	(This position shall not be used)
7B	(This position shall not be used)
7C	(This position shall not be used)
7D	(This position shall not be used)
7E	(This position shall not be used)
7F	(This position shall not be used)
80	(This position shall not be used)
81	(This position shall not be used)
82	(This position shall not be used)
83	(This position shall not be used)
84	(This position shall not be used)
85	(This position shall not be used)
86	(This position shall not be used)
87	(This position shall not be used)
88	(This position shall not be used)
89	(This position shall not be used)
8A	(This position shall not be used)
8B	(This position shall not be used)
8C	(This position shall not be used)
8D	(This position shall not be used)
8E	(This position shall not be used)
8F	(This position shall not be used)
90	(This position shall not be used)
91	(This position shall not be used)
92	(This position shall not be used)
93	(This position shall not be used)
94	(This position shall not be used)
95	(This position shall not be used)
96	(This position shall not be used)
97	(This position shall not be used)
98	(This position shall not be used)
99	(This position shall not be used)
9A	(This position shall not be used)
9B	(This position shall not be used)
9C	(This position shall not be used)
9D	(This position shall not be used)
9E	(This position shall not be used)
9F	(This position shall not be used)

Group 00 **Plane 00** **Row 10**

A. Administrative

1. Title

Proposal for encoding seven additional Myanmar characters in the UCS.

2. Requester's name

Ireland (NSAI), United Kingdom (BSI), Myanmar Language Commission, Myanmar Unicode and Natural Language Processing Research Center, Myanmar Computer Federation

3. Requester type (Member body/Liaison/Individual contribution)

Member body contribution.

4. Submission date

2006-02-28

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

No.

Proposed name of script

1b. The proposal is for addition of character(s) to an existing block

Yes.

1c. Name of the existing block

Myanmar.

2. Number of characters in proposal

7

3. Proposed category (see section II, Character Categories)

Category A.

4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)

Level 2

4b. Is a rationale provided for the choice?

Yes.

4c. If YES, reference

Brahmic Level 2 implementation.

5a. Is a repertoire including character names provided?

Yes.

5b. If YES, are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?

Yes.

5c. Are the character shapes attached in a legible form suitable for review?

Yes.

6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

No.

7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

No. The characters to be represented are not new to SC2.

8. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

Yes.

9. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script.

See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

Yes. See N2827, N1883R, and Unicode 1.0, which are all relevant documents.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

Yes.

2b. If YES, with whom?

San Lwin (Director General, Myanmar Language Commission), **Tun Tint** (Myanmar Language Commission), **Thein Oo** (President, Myanmar Computer Federation), **Kyaw Thein** (Vice-President, Myanmar Computer Federation), **Myint Myint Than** (Director, Myanmar Computer Federation), **Zaw Htut** (Myanmar Computer Professional Association, Myanmar's NET), **Htoo Myint Naung** (MyMyanmar Project, Technomation Studios, Universities of Computer Studies Yangon), **Myint Thu** (MyMyanmar Project, Myanmar Heritage Publications), **Ngwe Tun** (Mon—Myanmar Computer Professional Association, Solveware Solution, Myanmar Info-Tech), **Maung Maung Thant** (Myanmar Computer Professional Association), **Jai Pah Bung Mein** (Shan—SSi Technologies), **Saw Hare Sei** (S'gaw Karen—Ayeyawady Data Centre), **Saw Baldwin Khaing Oo** (S'gaw Karen—Ayeyawady Data Centre), **Nant Silver Tun** (Western Pwo Karen—Pwo Kayin Conference), **William Wai Lin Kyaw** (Myanmar Computer Professional Association, Myanmar Linux Users Group), **Ye Myat Thu** (Alpha Mandalay, Alpha Info-Tech), **Martin Hosken** (Payap University), **Keith Stribley** (Thanlwinsoft).

2c. If YES, available relevant documents

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

People in Myanmar.

4a. The context of use for the proposed characters (type of use; common or rare)

Common.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

In Myanmar.

6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Contemporary use and accordance with the Roadmap.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

N/A.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

Yes.

8b. If YES, is a rationale for its inclusion provided?

This proposal requests disunifications and a change in the sequences currently specified for Myanmar because those sequences do not work.

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

Yes.

11b. If YES, is a rationale for such use provided?

Yes.

11c. If YES, reference

Brahmic vowel and consonant signs.

12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

12b. If YES, reference

13a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

13b. If YES, describe in detail (include attachment if necessary)

14a. Does the proposal contain any Ideographic compatibility character(s)?

No.

14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?