Universal Multiple-Octet Coded Character Set

Doc Type:	Working Group Document
Title:	**Myanmar (Burmese) Ad-Hoc Meeting Report**
Source:	WG2
Status:	Working Document

References: WG2 N1729 (Khmer and Burmese Ad-Hoc Meeting Report, Redmond)
WG2 N 1792 (Comment on Proposal for Burmese Script, WG2 N 1729, UK)
L2/89-265 (Response to Burmese Proposal N1729, Expert Contribution, Zaw Htut)
WG2 N 1815 (Myanmar Character Code in ISO 10646, MITSC)
WG2 N1826 (Amendment 26 PDAM text, = N 1780)
WG2 N 1836 (Summary of Voting)
Revised Proposal for Encoding Myanmar Script in ISO 10646 (distributed at WG2, MITSC)

Meeting Dates: 1998-09-21 through 1998-09-24
Attendees: Aung Maw (Myanmar)
John Clews (SESAME,UK)
Michael Everson (Ireland)
Khin Maung Lwin (Myanmar)
Kyaw Thein (Myanmar, Chairman, MITSC)
John Okell (invited expert, SOAS)
Hugh McGregor Ross (UK)
Thaung Tin (Myanmar)
Thein Htut (Myanmar)
Ken Whistler (USA)

Synopsis

This document constitutes the meeting report of the ad hoc committee on Myanmar, which met on several occasions during the WG2 meeting, starting on 1998-09-21. The meeting report consists of three parts:

1. A summary of the consensus reached among the participants, together with a short outline of the remaining points in contention and how the committee resolved them.
2. A proposed disposition of comments on the Burmese PDAM (WG2 N1826, Amendment 26).
3. Detailed instructions to the editor on how to modify the PDAM text of WG2 N1826 in order to prepare FPDAM text, based on the proposed resolution of comments and the committee consensus on issues not directly raised in the national body comments.

Summary of Consensus Reached by Myanmar Ad-Hoc Committee

The Ad-Hoc Committee met on several occasions and discussed various technical issues related to the PDAM for Amendment 26 (N 1826), the comments of the national bodies (N 1836), and various other expert contributions, including official Myanmar contributions. The various expert contributions commented either on the Burmese (= Myanmar) encoding in N 1826 or responded to the working

document (WG2 N 1729) and its antecedents (e.g. WG2 N 881, X3L2/96-061, WG2 N 1523, L2/98-049) that were the basis for the PDAM text.

The committee reached consensus on a number of important technical issues regarding the Myanmar encoding and the PDAM text. Minor issues related directly to the national body comments are simply included in the proposed disposition of comments below. In this section, the major issues and the consensus opinion regarding them are summarized, so that the principles which were used in the disposition of comments are clear to all.

Repertoire of Vowels

There were fundamental differences between various country comments. Some requested the encoding of more vowel characters (both independent and dependent), omitted in the PDAM. Others requested the removal of one or more vowel characters from the PDAM. These differences of opinion result from different analyses and approaches that one might take regarding the Myanmar system of vowels. Many vowel symbols are clearly composed of smaller units which are also available in the encoding (vide 1732 in the PDAM text = 1730 + 172B), and the difference boils down to whether one encodes all of the structurally significant vowels (and then whether to include archaic vowels for Pali/Sanskrit) or only the subset of those vowels which can be used to build those displayed with composite forms. In the end, the committee chose a conservative compromise. The composite forms from the PDAM (1722, 1728, 1732, 1733) should be removed; however, the positions for them in the chart are left uncoded, so that in the conceivable event that implementation experience should in the future show that having the composite forms also encoded as characters is required, an encoding option would be available that would not result in fundamentally disordering the expected sequence of Myanmar vowels in the code chart. (This conservative approach towards the complex Brahmi-based scripts has served WG2 well in the past, as for Tibetan, when further study revealed the actual need for several characters for which the structural "gap" in the encoding chart had been prudently left, just in case.) Positions for two vowels requested in country comments that were not in the PDAM text itself were also left uncoded. Two vowels which are themselves technically composite (1726 and 172A) remain encoded instead of being removed, as they were determined to be treated as basic, unanalyzed units. This approach received the unanimous assent of the Ad Hoc Committee.

Confirmation of the Virama Model

The PDAM text assumes a virama-based model of encoding, where the textual representation of a consonant with a subjoined consonant consists of the first consonant followed by the virama, followed by the second consonant. In this model, subjoined consonants (which are quite frequent in Myanmar text) are not represented by separate character codes, but rather simply by the base consonant characters. The virama model is fundamentally different from the subjoined consonant model advocated in the Japanese national body comments. The text model advocated in the Myanmar revised proposal from MCITS (N1815, etc.) at first also seemed to be different; however, the ad hoc committee determined that the use of the LINK character in N1815 is functionally equivalent to that of virama in the PDAM text. With this understanding, the Myanmar delegation agreed that use of the virama was desirable and matched their own implementation model. Therefore, the virama-based model of the PDAM is to be retained. This conclusion received the unanimous assent of the Ad Hoc Committee.

Unification of Virama and Killer

National body comments pointed out that it is not necessary, and indeed is confusing, to maintain a coding distinction between virama (1737) and killer (1738). In Myanmar text encoded as implied by the PDAM text, the virama would be used to form subjoined consonants and the killer would be used to show the visible character that marks the drop of a final vowel. However, the strong structural analogy with other Brahmi-based scripts encoded in 10646 suggested that the latter function (traditionally known as "killer" in Myanmar) should instead use the sequence of the virama and ZERO WIDTH NON-JOINER (ZWNJ). The committee agreed with this analysis. The implication is that 1738 should be removed from the repertoire. With input from the Myanmar delegation, the committee determined that the unified character should

retain the name VIRAMA. The term "killer" can be reserved to refer to those visible manifestations of the virama character at the ends of syllables. Implementation issues regarding how input of virama for subjoining versus virama for final vowel killing could be arranged to match current keyboard practice in Myanmar were answered to the satisfaction of the Myanmar delegation. The decision to unify the virama and killer and retain the name VIRAMA received the unanimous assent of the Ad Hoc Committee.

Subjoined Glide Characters

Myanmar text uses four common forms of glide consonants known as YA PIN, YA YIT, WA HSWE, and HA HTOU. Structurally, these are the subjoined versions of the consonants YA, RA, WA, and HA, respectively. However, they are taught and conceived of as basic letters of the Myanmar alphabet, are typically keyed separately on Myanmar keyboards, and have significant implications for sorting of Myanmar data. (Documents L2/98-165 and N1815 refer to these subjoined glide characters and others resulting from combinations of them as "vowel modifiers", because these glides modify the vocalic nucleus of a Myanmar syllable.) The PDAM text does not contain these four characters, as they can be unambiguously represented in the virama model: KA + VIRAMA + YA ➜ *kya* (displayed as KA/YA PIN). The Myanmar delegation agreed after some discussion that this was a possible and consistent representation, and that it would avoid the problem of possible alternative representations of the same text (the same 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.

Kinzi

The kinzi (1739) was determined to be a conjunct form of 1704. Since no other conjunct forms are encoded (a policy which accords with the 10646 encoding of other Brahmi-based scripts), 1739 should not be encoded. This position received the unanimous assent of the Ad Hoc Committee.

Glyph Corrections

172B in the PDAM text does not show the typical form of the vowel sign AA. It is to be corrected to use the form shown for the right half of 1732. The glyph shown for 174F is actually a glyph fragment for a complexly shaped but common character; it is to be corrected to show the entire, usual form of this character. This position received the unanimous assent of the Ad Hoc Committee.

Font

Although not part of any official comment on the ballot, the extensive and thorough evaluation provided by the expert, Zaw Htut, suggested many ways in which the font used in the tables should be systematically improved to match current best Myanmar practice. The Myanmar delegation concurred and provided a suitable font with these improvements which may be used for the preparation of the FPDAM text and the standard itself. The other members of the Ad Hoc Committee appreciated this input and agreed that the new font should be used.

Sorting

The Ad Hoc Committee discussed sorting of Myanmar data, since the key principles of the encoding in the proposal from MCITS (N 1815) involve table layout decisions, encoding of vowel-tone combinations, and encoding of dummy consonant and vowel placeholder characters to facilitate auto-sorting of Myanmar text. The committee concluded that in the context of the encoding of Myanmar for 10646, where Myanmar data may be mixed with other data, it is desirable to distinguish issues of encoding per se from the detailed handling required for generating correct sort orders. This is a long-standing position of SC2 regarding

character encoding. Therefore, none of the special features of the N 1815 proposal designed specifically for Myanmar sorting are to be included in list of characters for the FPDAM. This position was unanimously assented to by the Ad Hoc Committee.

Terminology: Myanmar versus Burmese

At the expressed request of the Myanmar delegation, the Ad Hoc Committee determined that wherever possible the FPDAM text should follow the convention of referring to the script in question as the "Myanmar script" rather than the "Burmese script". The implications of this extend to the title for the FPDAM, the new block name, collection names, and the character names. This was a strongly held opinion of the Myanmar delegation, and the other members of the Ad Hoc Committee believe that that opinion should weigh heavily, despite the long English-based tradition of referring to the script as "Burmese."

Basic versus Extended Myanmar

The Myanmar delegation and other experts and national bodies had conflicting opinions about the encoding of additional extended Myanmar letters intended for Pali/Sanskrit text The Myanmar delegates pointed out that these are never used for Myanmar proper, and are not included among the standard set of characters for general use and education in Myanmar. Other experts pointed out that these are required for scholarly work on Pali texts, and are only the first of a number of additional Myanmar script characters that will be required for the representation of other minority languages spoken in Myanmar and neighboring countries. The Myanmar delegation agreed that those characters could stay in the encoding, as long as some clear mechanism was provided to distinguish between Basic Myanmar characters and extensions provided for minority languages and for scholarly work on archaic texts. After feedback from the plenary session of WG2, by the editor of 10646, and by the president of AFII, the Ad Hoc Committee determined that the most feasible compromise on this issue was to retain a single block name "MYANMAR" for the Basic Myanmar characters and all extensions for languages other than Myanmar; this would be in accord with established WG2 practice for many other scripts. However, to maintain the required distinction, two collections should be defined in the FPDAM: one for BASIC MYANMAR and one for EXTENDED MYANMAR. Both would be open collections, as there is the potential for the addition of new characters in the future to either collection. While this does not have the desired degree of visual segregation in the code charts preferred by the Myanmar delegation, it does provide unambiguous means for distinguishing the two sets, both for procurement and other legal needs, and internally by applications. This position is the best compromise that the Ad Hoc Committee could arrive at to accomodate both the needs of the Myanmar delegation and the requirements of WG2.

Name Changes

The Myanmar delegation pointed out that the names of 174A and 174B had been interchanged. The Ad Hoc Committee unanimously determined that this was simply an error and should be corrected. Other proposed name changes seem to reflect local name conventions for some characters versus the phonologically based character names typical for Brahmi-derived scripts in 10646. As there was a difference on opinion on these other name changes, it is probably safest to simply leave the other character names as they are (except, of course, for replacing "BURMESE" with "MYANMAR".

Encoding Position for Myanmar Script

The PDAM text encoded the Burmese (= Myanmar) script at positions 1700..177F in the BMP. However, expert feedback regarding the number of characters likely to be required in the future for Myanmar minority languages (Kayah, Karen, Chin, Mon, and Shan) has enabled us to estimate the required number to be somewhere in the vicinity of 50 ~ 60 additional extended characters. This will easily overflow the current block allocation. Unfortunately, since Khmer is encoded at 1780..17FF, any such overflow would require the creation of a separate, discontiguous block for a small number of Myanmar extension characters (as WG2 recently had to do for Bopomofo). Avoidance of such discontiguous blocks is clearly desirable if there are no countervailing factors. It was the unanimous opinion of the Ad Hoc Committee that it would be better to relocate the entire Myanmar block to the positions 1000..109F that has since been vacated in

the Roadmap by the encoding of Mongolian elsewhere. The additional two columns available at that position in the BMP would enable all Myanmar extensions to be encoded in one continguous block.

Proposed Disposition of Comments on PDAM 14

The Ad Hoc Committee has prepared a proposed draft for the Disposition of Comments on PDAM 14, for use by the editor in producing the final Disposition of Comments. This draft is very succinct. Technical comments are labeled "T1", "T2", ... and editorial comments are labeled "E1.", "E2"... Where the rationale for a proposed disposition of a comment is covered by the extended discussion of the consensus positions reached by the Ad Hoc Committee, that rationale is not recapitulated. The Final Disposition of Comments can presumably simply point to the consensus discussion for details.

Ireland

T1: Accepted. The BURMESE SIGN KINZI will be deleted.
T2: Rejected. The parenthetical comments following the character names will be deleted.
T3: Rejected. The Pali/Sanskrit characters will stay as they are, encoded in a separate, extended range.
T4: Rejected. See consensus discussion (Repertoire of Vowels)
T5: Accepted. VIRAMA and KILLER will be unified.

Israel

Noted. No action required. The consent of the official Myanmar delegation should confirm the interpretation of the Israeli vote as "YES".

Japan (N)

T1: Rejected. See consensus discussion (Confirmation of Virama Model, also Subjoined Glide Characters and Kinzi)
T2: Accepted. These two characters will be deleted.
T3: Rejected. These two items are judged to be words, not letters of the alphabet by the Myanmar delegation.
T4: Rejected. The display presentation of allographs is a font display issue, not a character coding issue.
T5: Accepted in principle. The Japanese comment is conditional. Since the graphically discontinuous character 1732 is to be deleted, it follows that there is no need to encode the requested additional forms for logical consistency.
T6: Accepted. VIRAMA and KILLER will be unified.
T7: Rejected. These vowel signs may be font variants or extensions only appropriate for archaic usage. The Japanese national body is invited to submit more complete contextual information about the usage of these vowel signs which might lead to a future proposal to add them to the Myanmar script if there is sufficient evidence that they are distinguished as characters.
T8: Accepted. The glyph will be corrected as noted.
T9: Rejected. See consensus discussion (Basic versus Extended Myanmar).
It is hoped that the Japanese national body comments have been sufficiently accomodated that they will consider changing their vote on the PDAM to "YES".

Sweden

Noted. No action required. The consent of the official Myanmar delegation should confirm the interpretation of the Swedish vote as "YES".

UK (N)

T1: Rejected. See consensus discussion (Repertoire of Vowels).

T2: Accepted.
It is hoped that the UK national body comments have been sufficiently accomodated that they will consider changing their vote on the PDAM to "YES".

USA

T1: Accepted. The BURMESE SIGN KINZI will be deleted.
E1: Accepted. The usual (short) glyph will be used for this character.
T2: Accepted in principle. VIRAMA and KILLER will be unified.
T3: Accepted. The parenthetical comments following the names will be deleted.
E2: Noted. A font will be provided by Myanmar and Ireland.

Instructions to the Editor for Preparation of FPDAM text for Amendment 26

1. Notify SC2 secretariat that the title of Project 1.02.18.01.26 should be amended to "... AMENDMENT 26: Myanmar" instead of "...AMENDMENT 26: Burmese".
2. On page 1 of the PDAM text change all instances of "Burmese" to "Myanmar" (2) and all instances of "BURMESE" to "MYANMAR" (2). Change citations of the range "1700-177F" to "1000-109F" (2) and citations of the row "17" to row "10" (2). Change the words "after the entry OGHAM" to "after the entry BASIC TIBETAN".
3. Throughout, change all citations of Table "240" to the appropriate table number,-if moving the MYANMAR block to 1000-109F affects the table numbering.
4. On page 2 of the PDAM text, replace the code chart with the appropriately formatted version of the code chart appended to this document. The appropriately formatted version will be supplied by Ireland.
5. In the headers for the table, update the table number (if necessary), the row to "10", and the block name to "MYANMAR".
6. On page 3 of the PDAM text, replace the names list with the appropriately formatted version of the names list appended to this document. The appropriately formatted version will be supplied by Ireland.
7. On page 4 of the PDAM text, replace the existing text with that shown below.

Page 700, Annex A
In the list of collection numbers and names, insert two new lines after 84 SINHALA (see Amd. 21) as follows:

87	BASIC MYANMAR	1000-104F, 200C, 200D
90	EXTENDED MYANMAR	1050-109F

Page 703, Annex B
In B.1, List of all combining characters, after the entry 0FB9 TIBETAN SUBJOINED LETTER KSSA, insert new entries, as shown below

102B   MYANMAR VOWEL SIGN AA

...

1059   MYANMAR VOWEL SIGN VOCALIC LL
[[complete list to be supplied to the editor in plain text via email]]

Page 716, Annext E
After the entry 2665 MUTED SHARP SIGN insert new entries, as shown below

1048   MYANMAR DIGIT EIGHT

...

1057   MYANMAR VOWEL SIGN VOCALIC RR
[[complete list to be supplied to the editor in plain text via email]]

Page 754, Annex N
Delete the entry
-       Burmese

Table 240 - Row 10: Myanmar

	100	101	102	103	104	105	106	107
0	0	16	32	48	64	80		
1	1	17	33	49	65	81		
2	2	18		50	66	82		
3	3	19	35		67	83		
4	4	20	36		68	84		
5	5	21	37		69	85		
6	6	22	38	54	70	86		
7	7	23	39	55	71	87		
8	8	24		56	72	88		
9	9	25	41	57	73	99		
A	10	26	42		74			
B	11	27			75			
C	12	28	44		76			
D	13	29	45		77			
E	14	30	46		78			
F	15	31	47		79			

G= 00
P= 00

2

Table 240 - Row 10: Myanmar

HEX	NAME	HEX	NAME
1000	MYANMAR LETTER KA	1044	MYANMAR DIGIT FOUR
1001	MYANMAR LETTER KHA	1045	MYANMAR DIGIT FIVE
1002	MYANMAR LETTER GA	1046	MYANMAR DIGIT SIX
1003	MYANMAR LETTER GHA	1047	MYANMAR DIGIT SEVEN
1004	MYANMAR LETTER NGA	1048	MYANMAR DIGIT EIGHT
1005	MYANMAR LETTER CA	1049	MYANMAR DIGIT NINE
1006	MYANMAR LETTER CHA	104A	MYANMAR SIGN LITTLE SECTION
1007	MYANMAR LETTER JA	104B	MYANMAR SIGN SECTION
1008	MYANMAR LETTER JHA	104C	MYANMAR SYMBOL LOCATIVE
1009	MYANMAR LETTER NYA	104D	MYANMAR SYMBOL COMPLETED
100A	MYANMAR LETTER NNYA	104E	MYANMAR SYMBOL AFOREMENTIONED
100B	MYANMAR LETTER TTA	104F	MYANMAR SYMBOL GENITIVE
100C	MYANMAR LETTER TTHA	1050	MYANMAR LETTER SHA
100D	MYANMAR LETTER DDA	1051	MYANMAR LETTER SSA
100E	MYANMAR LETTER DDHA	1052	MYANMAR LETTER VOCALIC R
100F	MYANMAR LETTER NNA	1053	MYANMAR LETTER VOCALIC RR
1010	MYANMAR LETTER TA	1054	MYANMAR LETTER VOCALIC L
1011	MYANMAR LETTER THA	1055	MYANMAR LETTER VOCALIC LL
1012	MYANMAR LETTER DA	1056	MYANMAR VOWEL SIGN VOCALIC R
1013	MYANMAR LETTER DHA	1057	MYANMAR VOWEL SIGN VOCALIC RR
1014	MYANMAR LETTER NA	1058	MYANMAR VOWEL SIGN VOCALIC L
1015	MYANMAR LETTER PA	1059	MYANMAR VOWEL SIGN VOCALIC LL
1016	MYANMAR LETTER PHA	105A	(This position shall not be used)
1017	MYANMAR LETTER BA	105B	(This position shall not be used)
1018	MYANMAR LETTER BHA	105C	(This position shall not be used)
1019	MYANMAR LETTER MA	105D	(This position shall not be used)
101A	MYANMAR LETTER YA	105E	(This position shall not be used)
101B	MYANMAR LETTER RA	105F	(This position shall not be used)
101C	MYANMAR LETTER LA	1060	(This position shall not be used)
101D	MYANMAR LETTER WA	1061	(This position shall not be used)
101E	MYANMAR LETTER SA	1062	(This position shall not be used)
101F	MYANMAR LETTER HA	1063	(This position shall not be used)
1020	MYANMAR LETTER LLA	1064	(This position shall not be used)
1021	MYANMAR LETTER A	1065	(This position shall not be used)
1022	(This position shall not be used)	1066	(This position shall not be used)
1023	MYANMAR LETTER I	1067	(This position shall not be used)
1024	MYANMAR LETTER II	1068	(This position shall not be used)
1025	MYANMAR LETTER U	1069	(This position shall not be used)
1026	MYANMAR LETTER UU	106A	(This position shall not be used)
1027	MYANMAR LETTER E	106B	(This position shall not be used)
1028	(This position shall not be used)	106C	(This position shall not be used)
1029	MYANMAR LETTER O	106D	(This position shall not be used)
102A	MYANMAR LETTER AU	106E	(This position shall not be used)
102B	(This position shall not be used)	106F	(This position shall not be used)
102C	MYANMAR VOWEL SIGN AA	1070	(This position shall not be used)
102D	MYANMAR VOWEL SIGN I	1071	(This position shall not be used)
102E	MYANMAR VOWEL SIGN II	1072	(This position shall not be used)
102F	MYANMAR VOWEL SIGN U	1073	(This position shall not be used)
1030	MYANMAR VOWEL SIGN UU	1074	(This position shall not be used)
1031	MYANMAR VOWEL SIGN E	1075	(This position shall not be used)
1032	MYANMAR VOWEL SIGN AI	1076	(This position shall not be used)
1033	(This position shall not be used)	1077	(This position shall not be used)
1034	(This position shall not be used)	1078	(This position shall not be used)
1035	(This position shall not be used)	1079	(This position shall not be used)
1036	MYANMAR SIGN ANUSVARA	107A	(This position shall not be used)
1037	MYANMAR SIGN DOT BELOW	107B	(This position shall not be used)
1038	MYANMAR SIGN VISARGA	107C	(This position shall not be used)
1039	MYANMAR SIGN VIRAMA	107D	(This position shall not be used)
103A	(This position shall not be used)	107E	(This position shall not be used)
103B	(This position shall not be used)	107F	(This position shall not be used)
103C	(This position shall not be used)		
103D	(This position shall not be used)		
103E	(This position shall not be used)		
103F	(This position shall not be used)		
1040	MYANMAR DIGIT ZERO		
1041	MYANMAR DIGIT ONE		
1042	MYANMAR DIGIT TWO		
1043	MYANMAR DIGIT THREE		