

Revised Proposal for Encoding Burmese in Unicode

Lee Collins
Apple Computer
12/10/97

- [Overview](#)
- [Difference from Michael Everson's Proposal](#)
- [Character Code Assignments and Names](#)
- [Code Chart](#)

Overview

This document proposes revising Michael Everson's *Proposal for encoding the Burmese script in ISO 10646* <<http://www.indigo.ie/egt/standards/my/mv.html>>. The revision is based on comments from experts at the Myanmar Language Commission concerning both the original Unicode proposal for encoding Burmese in Unicode Technical Report #1 and Everson's proposal. Character names and transliterations used in Everson's document are retained.

Contact information for the Myanmar Language Commission is

U San Lwin
Director General
Myanmar Language Commission
No. 27, 6 1/2 Mile, Pyi Road
Yangon
Fax + 95 1 549369

Differences from Everson's proposal

- 1) This proposal is not based on the Unicode layout used for the Indian scripts. It assigns codepoints in standard Burmese alphabetic order, beginning with BURMESE LETTER KA.
- 2) This proposal assumes that the character BURMESE VOWEL SIGN E is stored before its consonant. The actual requirement of the Myanmar Language Commission is that input and display be correct (before the consonant). If implementations permit this, then storage order may be phonetic (after the consonant). This needs to be resolved by Unicode.
- 3) This proposal removes 8 characters from Everson's proposal because they are considered to be composites of other encoded characters:

BURMESE LETTER AA = LETTER A + VOWEL SIGN AA
BURMESE LETTER UU = LETTER U + VOWEL SIGN II
BURMESE LETTER AU = VOWEL SIGN E + LETTER O + VOWEL SIGN AA + SIGN KILLER
BURMESE LETTER AI = LETTER A + LETTER AI
BURMESE LETTER UI = LETTER A + LETTER I + LETTER U
BURMESE VOWEL SIGN O = VOWEL SIGN E + VOWEL SIGN AA
BURMESE VOWEL SIGN AU = VOWEL SIGN E + VOWEL SIGN AA + SIGN KILLER
BURMESE VOWEL SIGN UI = LETTER I + LETTER U

- 2) This proposal adds the 4 subscript letters because they are considered core letters of the alphabet:

U+14B3 BURMESE SUBSCRIPT LETTER YA (ya pang., ya pin)
U+14B4 BURMESE SUBSCRIPT LETTER RA (ra rac, ya yi)
U+14B5 BURMESE SUBSCRIPT LETTER VA (wa chwe, wa hswe)
U+14B6 BURMESE SUBSCRIPT LETTER HA (ha thui:, ha htou)

The remaining subscripts and ligatures are to be formed with the BURMESE VIRAMA.

- 4) This proposal assigns the Burmese script letters used for Sanskrit and minority languages spoken in Myanmar such as

Mon, Karen, and Shan to a separate range beginning at U+14D0. The Myanmar Language Commission is still trying to determine the full set of such characters. The 10 Sanskrit letters encoded in Everson's proposal are assigned the code points U+14D0 - U+14D9. The code points U+14DA- U+14FF are set aside for the minority languages pending future assignment.

Character Code Assignments and Names

U+1480 BURMESE LETTER KA (ka. krii:, ka ji)
 U+1481 BURMESE LETTER KHA (kha. khwe:, hka gwei)
 U+1482 BURMESE LETTER GA (ga. ngay, ga nge)
 U+1483 BURMESE LETTER GHA (gha. khrii:, ga ji)
 U+1484 BURMESE LETTER NGA (nga.)
 U+1485 BURMESE LETTER CA (ca. lum:, hsa loun)
 U+1486 BURMESE LETTER CHA (cha. lim, hsa lein)
 U+1487 BURMESE LETTER JA (ja. khwai:, za gwe)
 U+1488 BURMESE LETTER JHA (jha. myang-chwai:, za myinzwe)
 U+1489 BURMESE LETTER NYA (nya. kale:, nya galei)
 U+148A BURMESE LETTER NNYA (nnya. krii:, nya ji)
 U+148B BURMESE LETTER TTA (tta. samlyang:khyit, ta talinjei)
 U+148C BURMESE LETTER TTHA (ttha. wam:bhai:, hta wunbe)
 U+148D BURMESE LETTER DDA (dda. rang-kok, da yingau)
 U+148E BURMESE LETTER DDHA (ddha. re-mhut, da yeihmou)
 U+148F BURMESE LETTER NNA (nna. krii:, na ji)
 U+1490 BURMESE LETTER TA (ta. wam:puu, ta wunbu)
 U+1491 BURMESE LETTER THA (tha. chang-thuu:, hta hsindu)
 U+1492 BURMESE LETTER DA (da. twe:, da dwei)
 U+1493 BURMESE LETTER DHA (dha. okkhyuik, da auhcai)
 U+1494 BURMESE LETTER NA (na. ngay, na nge)
 U+1495 BURMESE LETTER PA (pa. cok, pa zau)
 U+1496 BURMESE LETTER PHA (pha. uuthep, hpa outhou)
 U+1497 BURMESE LETTER BA (ba. thakkhyuik, ba lahcai)
 U+1498 BURMESE LETTER BHA (bha. kun:, ba goun)
 U+1499 BURMESE LETTER MA (ma.)
 U+149A BURMESE LETTER YA (ya. paklak, ya pale)
 U+149B BURMESE LETTER RA (ra. kok, ya gau)
 U+149C BURMESE LETTER LA (la.)
 U+149D BURMESE LETTER WA (wa.)
 U+149E BURMESE LETTER SA (sa., tha)
 U+149F BURMESE LETTER HA (ha.)
 U+14A0 BURMESE LETTER LLA (lla. krii:, la ji)
 U+14A1 BURMESE LETTER A (a.)
 U+14A2 BURMESE LETTER I (Paa-lii. atkharaa i., Pali ehkaya i)
 U+14A3 BURMESE LETTER II (atkharaa i., ehkaya i)
 U+14A4 BURMESE LETTER U (atkharaa u., ehkaya u)
 U+14A5 BURMESE LETTER E
 U+14A6 BURMESE LETTER O (o.)
 U+14A7 BURMESE VOWEL SIGN AA (re: khyā., yei hca)
 U+14A8 BURMESE VOWEL SIGN I (lum:krii: tang, lounji tin)
 U+14A9 BURMESE VOWEL SIGN II (lum:krii: tang chan khat, lounji tin hsan hka)
 U+14AA BURMESE VOWEL SIGN U (takhyong: ngang, tahcaun ngin)
 U+14AB BURMESE VOWEL SIGN UU (nhac-khyong: ngang, hnacaun ngin)
 U+14AC BURMESE VOWEL SIGN E (sawe: thui:, thawei htou)
 U+14AD BURMESE VOWEL SIGN AI (nok pac, nay pyi)
 U+14AE BURMESE SIGN ANUSVARA (se:se tang, theidhei tin)
 U+14AF BURMESE SIGN DOT BELOW (ok-ka. mrac, auka myi)
 U+14B0 BURMESE SIGN VISARGA (rhe.ka. pok, hyeiga pai)
 U+14B1 BURMESE SIGN KILLER (asat, atha)
 U+14B2 BURMESE SIGN VIRAMA
 U+14B3 BURMESE SUBSCRIPT LETTER YA (ya pang., ya pin)
 U+14B4 BURMESE SUBSCRIPT LETTER RA (ra rac, ya yi)
 U+14B5 BURMESE SUBSCRIPT LETTER VA (wa chwe, wa hswe)
 U+14B6 BURMESE SUBSCRIPT LETTER HA (ha thui:, ha htou)
 U+14B7 (This position shall not be used)
 U+14B8 (This position shall not be used)
 U+14B9 (This position shall not be used)
 U+14BA (This position shall not be used)
 U+14BB (This position shall not be used)
 U+14BC (This position shall not be used)
 U+14BD (This position shall not be used)

U+14BE (This position shall not be used)
 U+14BF (This position shall not be used)
 U+14C0 BURMESE DIGIT ZERO
 U+14C1 BURMESE DIGIT ONE
 U+14C2 BURMESE DIGIT TWO
 U+14C3 BURMESE DIGIT THREE
 U+14C4 BURMESE DIGIT FOUR
 U+14C5 BURMESE DIGIT FIVE
 U+14C6 BURMESE DIGIT SIX
 U+14C7 BURMESE DIGIT SEVEN
 U+14C8 BURMESE DIGIT EIGHT
 U+14C9 BURMESE DIGIT NINE
 U+14CA BURMESE SIGN SECTION (pudma., pouma)
 U+14CB BURMESE SIGN LITTLE SECTION (pudse:, pouthel)
 U+14CC BURMESE SYMBOL COMPLETED (atkharaa rwe:, ehkaya ywei)
 U+14CD BURMESE SYMBOL GENITIVE (atkharaa i., ehkaya i)
 U+14CE BURMESE SYMBOL LOCATIVE (atkharaa nhuik, ehkaya hnai)
 U+14CF BURMESE SYMBOL 4NG (atkharaa leng:, ehkaya lagaun)
 U+14D0 BURMESE LETTER SHA (sha.)
 U+14D1 BURMESE LETTER SSA (ssa.)
 U+14D2 BURMESE LETTER VOCALIC R
 U+14D3 BURMESE LETTER VOCALIC RR
 U+14D4 BURMESE LETTER VOCALIC L
 U+14D5 BURMESE LETTER VOCALIC LL
 U+14D6 BURMESE VOWEL SIGN VOCALIC R
 U+14D7 BURMESE VOWEL SIGN VOCALIC RR
 U+14D8 BURMESE VOWEL SIGN VOCALIC L
 U+14D9 BURMESE VOWEL SIGN VOCALIC LL
 U+14DA (This position shall not be used)
 U+14DB (This position shall not be used)
 U+14DC (This position shall not be used)
 U+14DD (This position shall not be used)
 U+14DE (This position shall not be used)
 U+14DF (This position shall not be used)
 U+14E0 (This position shall not be used)
 U+14E1 (This position shall not be used)
 U+14E2 (This position shall not be used)
 U+14E3 (This position shall not be used)
 U+14E4 (This position shall not be used)
 U+14E5 (This position shall not be used)
 U+14E6 (This position shall not be used)
 U+14E7 (This position shall not be used)
 U+14E8 (This position shall not be used)
 U+14E9 (This position shall not be used)
 U+14EA (This position shall not be used)
 U+14EB (This position shall not be used)
 U+14EC (This position shall not be used)
 U+14ED (This position shall not be used)
 U+14EE (This position shall not be used)
 U+14EF (This position shall not be used)
 U+14F0 (This position shall not be used)
 U+14F1 (This position shall not be used)
 U+14F2 (This position shall not be used)
 U+14F3 (This position shall not be used)
 U+14F4 (This position shall not be used)
 U+14F5 (This position shall not be used)
 U+14F6 (This position shall not be used)
 U+14F7 (This position shall not be used)
 U+14F8 (This position shall not be used)
 U+14F9 (This position shall not be used)
 U+14FA (This position shall not be used)
 U+14FB (This position shall not be used)
 U+14FC (This position shall not be used)
 U+14FD (This position shall not be used)
 U+14FE (This position shall not be used)
 U+14FF (This position shall not be used)

Code chart

Note that this code chart, unlike the code charts found in *The Unicode Standard, Version 2.0*, does not include the dotted circle for combining characters.

	148	149	14A	14B	14C	14D	14E	14F
0	က 1480	တ 1490	ဇ 14A0	း 14B0	ဝ 14C0	ဂ 14D0		
1	ခ 1481	ဆ 1491	အ 14A1	င 14B1	ဘ 14C1	ဖ 14D1		
2	ဂ 1482	ဒ 1492	ဃ 14A2	င 14B2	၂ 14C2	ဇ 14D2		
3	ဃ 1483	စ 1493	ဤ 14A3	၂ 14B3	၃ 14C3	ဗ 14D3		
4	င 1484	န 1494	၃ 14A4	[14B4	၄ 14C4	ဇ 14D4		
5	စ 1485	ပ 1495	ဇ 14A5	၁ 14B5	၅ 14C5	ဇ 14D5		
6	ဆ 1486	ဖ 1496	ဩ 14A6	✓ 14B6	၆ 14C6	၄ 14D6		
7	ဇ 1487	ဗ 1497	၇ 14A7		၇ 14C7	၄ 14D7		
8	ဈ 1488	ဘ 1498	ဝ 14A8		ဂ 14C8	ဇ 14D8		
9	ည 1489	မ 1499	စ 14A9		ဇ 14C9	ဇ 14D9		
A	၃ 148A	ယ 149A	၂ 14AA		၂ 14CA			
B	င 148B	၇ 149B	၂ 14AB		၂ 14CB			
C	င 148C	လ 149C	၆ 14AC		၅ 14CC			
D	၃ 148D	ဝ 149D	၂ 14AD		၅ 14CD			

E	၃ 148E	၃ 149E	• 14AE		၃ 14CE			
F	၃ 148F	၃ 149F	• 14AF		၃ 14CF			