ISO
INTERNATIONAL ORGANIZATION FOR STANDARDIZATION

ISO/IEC JTC1/SC2/WG2
Multi-Octet Coded Character Set

Title: Response to Michael Everson

Source: Ngo Thanh Nhan
Status: Expert Contribution
Action: For consideration by JTC1/SC2/WG2
Date: 1999-02-26

The "Virama model" proposed for encoding Cham causes basic text processing problem. I will present here crucial evidence for the "graphic model" like Thai and Lao.

1. The main justification for extending the meaning of virama in Brahmic type script, esp. Cham is in point 1 of N1960:

    "The virama model uses the virama (Cham character U+xx3F) to 'kill' the inherent −a vowel of a consonant, and often causes a following consonant to change in some way to join with the first consonant."

    and

    "Consider   ka + VIRAMA + ya = kya
                  ka + VIRAMA + ra = kra
                  ka + VIRAMA + la = kla
                  ka + VIRAMA + va = kva"

2. The virama model also handles final consonant graphic extension, with the help of a ZERO-WIDTH NON-JOINER (ZWNJ) character (where a period represents the syllable ending mark). For examples

    "Consider   ka + VIRAMA + ZWNJ = k.
                  ka + VIRAMA + ZWNJ = t." etc.
    then

                  "ka + VIRAMA + ZWNJ + ra = k.ra"

To keep ra, la, ya, va from becoming semivowels.

This contribution comments on the virama model.

Consider the sequences kRa and –kra (where R is a semivowel, and r is a consonant). The virama model represents these two strings by

     i.   ka + VIRAMA = k. (Cham rule, not shown in N1960)
     ii.   ka + VIRAMA + ra = k.ra (Cham rule, not shown in N1960)
     iii.   ka + VIRAMA + ra = kRa (according to point 1 in N1960)
     iv.   ka + VIRAMA + ZWNJ + ra = k.ra (according to point 3 in N1960)

Obviously, either i. or iii. is incorrect; either ii. or iv. is incorrect. I will argue that extending the meaning of the virama by the virama model ends up creating another character ZWNJ, thus complicate the Cham encoding in many ways:

a. The semivowels r, l, w, y in Cham are different graphically and psychologically from the syllables ra, la, va, ya. Simply because the character-based European writing systems "consider" them as the same, we cannot consider them the same in syllable-based writing systems.
b. Extending the meaning of the VIRAMA complicates the encoding as well as keyboard entry. Although we do not speak about keyboard entry issue here, how will the character VIRAMA and ZWNJ automatically come into being in the representation ?
c. In old scripts and other fonts, Cham syllables in texts are written continuously without a SPACE. We can say that the VIRAMA serves as a syllable ending marker in place of SPACE. Thus, extending the meaning of VIRAMA causes serious problems in text representation.

Therefore, we propose an additional 4 code points for the semivowels in Cham.

Since the virama model has the ability graphically to prepose R semivowel, we will propose an extension to the graphic model to represent kRa as ka+R in that order—thus solving sorting inconsistency in the graphic model used for Lao and Thai.