# Comment on L2/12-125: Encoding Model for Kawi Script

Naga Ganesan (naa.ganesan@gmail.com)

I read Mr. Anshuman Pandey's interesting research work on Kavi script. Java island (*yaavakam* in Tamil) seems to be named by a tree called "yaa" tree in Sangam Tamil literature: yaa = Shorea robusta, saal tree known to withstand severe droughts, and during that period elephants get their water from the barks of "yaa" tree (many poems in classical Tamil literature). More on yaa tree, and seems to be the source for the name of Yaavakam, modern Java island in Indonesia:
http://groups.google.com/group/santhavasantham/msg/a140b8d12aa70d1f

Kavi script has a structure very similar to Tamil script. As in Tamil script, Kawi lacks the aspirated consonants, and has the PuLLi (Viraama) of Tamil as its Viraama. Interestingly, Ancient (= tol in Tamil, vRddha) grammarian belonging to Kapi (= brown) gotra – called Tolkaappiyan wrote the grammar for Tamil 2000 years ago (He seems to be a Jaina and scholars have seen Mahabhaashya's influences in his Tamil grammar). Will Kawi script be referring to the ancient Tamil grammarian's gotra, Kapi (brown)?

(1) In the available corpus of Kawi script, how are Sanskrit loan words written? Do they also lack aspirated consonants?

(2) Have you looked at the encoding model of Thaana script of Maldives? That may better suit Kawi script encoding: that is, encode "pure" consonants (K, G, C, J, T, D, ... without inherent vowel /a/) as basic code points, and the vowel modifiers, for A, I & U vowels. Thaana script outwardly looks Arabic glyphs, but its structure is basically Tolkaappiyam grammar.
http://en.wikipedia.org/wiki/Thaana
http://www.omniglot.com/writing/thaana.htm
http://homepage.ntlworld.com/stone-catend/Tha02.pdf
If you can work with Bruce D. Cain, who did his PhD under James Gair, Cornell University, it will be very helpful for encoding Kawi script which was carried to Java by Tamils - for pointers, the island name, Kawi script's puLLi virama & lack of aspirated consonants are there. For the PuLLi virama of Kawi script, compare it against "sukun" puLLi of Thaana script.

Ideally, people say, Telugu, Kannada, Grantha, should have a "subjoined" model like Tibetan now in Unicode. And, Tamil should have had "pure" consonants as basic code points plus vowel modifiers including one for removing puLLi to create KA, CA, TA, ... etc., [1] Anyways, Telugu, Hindi, Grantha all now have Virama concept first codified in Tamil grammar 2000 years ago as the core design feature in unicode.

Thanks for considering Thaana with that of Kawi script.

N. Ganesan

[1] Mey-Uyir encoding model for Tamil: (old post from Jan. 2008):
http://nganesan.blogspot.com/2008/01/blog-post.html
(a) In the MeyUyir encoding, only vowels and consonants (மெய்யெழுத்துக்கள்) of Tamil are encoded atomically.

(b) All uyirmey letters are generated by font intelligence: that is க் followed by உ will automatically produce கு. Nowadays, in Open Type fonts this is trivial, no rendering engine etc., So, Tamil grammar is followed: க் + உ = கு.

(c) In few and rare instances where தமிழ்இனம், mey and uyir letters have to be shown separated, use zwnj (zero-width non-joiner).

(d) க்ஷ is left as non-conjunct by default. E.g., பக்ஷி (bird, name of a Muslim male). Conjunct க்ஷ is created only when needed using zwj (zero width joiner).

Pure Consonant and Vowels encoding model (my MeyUyir model): If Unicode is newly designed today, my choice is this model for Tamil (and, it will not be optimal for Hindi/Sanskrit). Now, unfortunately, current Indic Unicode is optimal for Hindi, and the overhead on Tamil is high.

[2] Where Did the Maldives People Come From? Dravidian substratum in Maldives:
http://www.iias.nl/iiasn/iiasn5/insouasi/maloney.html