Review of a Proposal placed before Unicode Technical Committee
entitled 'Proposal to add Tamil Letter SHA' (L2/03-273)
Reviewed by
Peri Bhaskararao

1. Purpose of the Proposal: The proposal under review is for encoding the character SHA for Tamil. It should be pointed out that Devanagari script as used for Sanskrit language possesses 'akṣaras' for four fricatives. In conventional romanization (for linguistic and philological purposes), the four fricatives are transcribed as: {š} {ṣ} {s} {h}. A transcriptional and transliteratory variant of {š} is {ś}. In a very old romanization method it was also transcribed as {ç}. All these four letters occur in the native vocabulary of all the Indo-Aryan languages (which are derivatives of Sanskrit). However, there has been some merger in the pronunciation of the first three in different combinations in some of the modern Indo-Aryan languages. However, they are maintained in the writing systems of these languages.

The situation of Dravidian languages is somewhat different. {š} and {ṣ} occur only in words originally borrowed from Sanskrit and later from other Indo-Aryan languages as well as from English. The other three Dravidian languages, viz., Malayalam, Telugu and Kannada have internalized the letters for all these four. ISCII included letters for {ṣ} {s} {h} but left out the letter for {š} in Tamil. This resulted in a major deficiency in the code – for instance, there is no way of representing the backing string of a very important 'akṣara' in the language viz., {šrī}. Probably this is taken care of by fontsets by includig some glyph somewhere outside the U+0B80 to U+BFF range. This position needs to be corrected.

The proposal under review correctly points out that the letter {š} is an essential letter in writing and reading Tamil documents and illustrates the point by showing several documents to that effect. Though this letter occurs in words of Indo-Aryan origin, its correct usage is important for the integrity of the document. It cannot be substituted by any nearby letter such as {s} for maintaining the integrity of the document and for correct pronunciation.

2. Combinations of {š}: The proposal discusses 'Data Processing and 'Ligature' issues of this character. In general this character behaves like {ṣ} in its combination with vowel-mātrās as well clustering with other consonants. As mentioned in §1 above, the 'akṣara' {šrī} requires a separate glyph. There is also a separate glyph for {šri} as in the word {āšrita} (e.g., in the phrase {āšrita-rakṣaka-šrī-rāma}). Both {šrī} and {šri} will be treated as Grapheme-compounds and the compound-glyphs will be derived by rule.

3. Recommendation: The proposed character for Tamil {š} be encoded and be placed at U+0BB6 as requested by the proposers.

Prof. *Peri Bhaskararao,* Ph.D.
ILCAA, Tokyo University of Foreign Studies
Asahi-cho 3-11-1, Fuchu-shi, TOKYO 183-8534, Japan
Phones:+81-42-330-5607 and +81-42-330-5609
Fax: +81-42-330-5607
Email: bhaskar@aa.tufs.ac.jp