Follow-up #2 to Extended Tamil proposal
Shriramana Sharma, jamadagni-at-gmail-dot-com, India
2010-Oct-16

§1. Attestations for some more variants of Extended Tamil
In my proposal document L2/10-256R I have clearly shown that there are very many variants of Extended Tamil, and that these variants cannot be represented by the existing Unicode prescription of using superscript digits. If only appropriate characters for the vocalic vowels, aspirated/voiced consonants etc are provided at the encoding level, all these variants can be taken care of at the font level. There is no need to encode each of these variants separately. Some more variants of Extended Tamil which further go to prove that the existing prescription of using superscript digits is insufficient are provided now.

§2.1. Using subscript digits
Many contemporary publications in Extended Tamil use subscript digits instead of superscript ones. To illustrate this we provide samples from publications of Gita Press, Gorakhpur, which is a very respected publisher of Hindu religious books in all the major Indian languages/scripts. (See http://www.gitapress.org.)

From p 14 of Sundara Kandam, Tr: T S Kodandaraman, 2007, Gita Press, ISBN: 81-293-0892-4, we provide the transcription table for the notation used in that book:
Note the use of the horizontal stroke below for JHA and Vocalic R. And that is apart from subscript digits being used for all other consonants.

Here is another sample showing subscript digits from p 642 of another publication of the same Gita Press, Gorakhpur: viz Gita Tattva Vivechani (Tamil), Tr: Swaminatha Atreya, 2004; ISBN: 81-293-0058-3:

It is quite obvious that the existing Unicode prescription of using the superscript digits at U+00B2, U+00B3 and U+2074 will not cater to this form of Extended Tamil. It is also unlikely that suggesting the use of U+2082, U+2083 and U+2084 would be appropriate, as that is an obvious step in the wrong direction - catering at the encoded representation level to stylistic variations with zero semantic difference. Further, seeing as this Gita book has had 417,500 copies printed as of 2010 as per the publication information given in the latest edition of the book, this variant of Extended Tamil can certainly not be ignored.

§2.2. Importing Grantha characters (another variant)
Quite apart from the usage of digits, it has already been pointed out that Grantha written forms may be imported in a ‘liberal’ variant of Extended Tamil. An edition of the Vishnu Sahasranama Stotra published in 2008 by one Shri Prasanna Venkatachalapati Perumal Charitable Trust, Gunaseelam, Tiruchi shows yet another ‘liberal’ variant. (See next page.)

In fact, this book claims uniqueness among all books using Extended Tamil in that it does not limit itself to superscript 2, 3 and 4 which are applied only to consonants, but also caters to other minute variations (by the usage of Grantha written forms). This is mentioned by the publishers in their introduction on p 3 which also clearly indicates (see
the highlighted portions) that the publishers consider this book to be a book printed in the Tamil script and not any ‘Linear Grantha’ or other script form.

Usage sample, p 17:

Introduction, p 3:

If any organization claims to truly represent the Tamil community, they should realize that these publishers are also native Tamilians and part of the Tamil community and have extended the script in this way for writing Sanskrit. Imposing a narrow view of what is Tamil and what is not on the community at large claiming to ensure the ‘purity’ of the script will only suffocate it from growth and productive expansion.
The samples provided below from the transcription table on pp 10-13 of the same book also shows that the publishers do not consider the script they have used as the Grantha script or any variant thereof (such as ‘Linear’ or ‘Extended’ Grantha). They have placed their notation system in a column called Tamil Ravi (after the name of the font they developed for this Extended Tamil variant) quite apart from the Grantha column.

This table shows that in Tamil Ravi, superscripts 2, 3 and 4 are used for all the class consonants, but for the vocalic vowels the Grantha written forms have been imported. For the anusvara, an asterisk has been placed on the letter MA. The chandrabindu as seen in other scripts used to write Sanskrit like Devanagari and Grantha is also used with Tamil characters as seen in the transcription table (and hence it will also have to be included in the Extended Tamil encoding as proposed in L2/ 10-256R).
§2.3. Using superscript V and subscript I

Yet another publication shows even more innovation in extending the Tamil script. For voiced consonants, it places a superscript Latin letter V after the consonant. (V for voiced!) For aspirates, it places an I-like vertical stroke in subscript. For voiced aspirates, it uses both superscript V and subscript I. SHA is represented by CA with an anudatta-like stroke below. This publication first released in 1967 and reprinted many times since by one T S Parthasarathy, West CIT Nagar, Madras (Chennai) is a collection of almost 700 devotional songs by the Hindu saint Thyagaraja in the Telugu and Sanskrit languages (mostly Telugu). Throughout this book those songs have been printed using this variant of Extended Tamil.

The transcription table given before the preface of this book is reproduced here:
I must admit here that in this variant of Extended Tamil, the diacritics are observed to be placed after vowel signs placed to the right of the consonant and not between the consonant and vowel signs. However, the fact remains that in most forms of Extended Tamil (including the Gita book mentioned previously running to almost 420,000 copies) the diacritics are placed between the consonant and any vowel signs placed to the right. We have also remarked (in L2/10-085 p 11) that the diacritic should rightfully semantically gravitate to the glyph that it qualifies. Thus in the interests of standardization, one would prefer that even in this ‘V-I system’ of Extended Tamil, the diacritic(s) is/are placed immediately after the consonant. Smart fonts based on an Extended Tamil encoding can nevertheless achieve even the rendering shown here by appropriate substitution tables.
§2. On the anunasika sign/chandrabindu for Extended Tamil

On pp 9-11 of my Extended Tamil proposal I had discussed (among other things) the character that should be used in Extended Tamil corresponding to the chandrabindu of other Indic scripts. The chandrabindu as such is seen to be used in Extended Tamil also. Some users have however employed other written forms (such as MA + VIRAMA + SUPER-3) as well. The crux of the matter is that if the chandrabindu as such is to be consistently used for Extended Tamil also, then the character should take GC=Mn and should be named TAMIL (EXTENDED) SIGN CANDRABINDU in line with the other Indic scripts. If however other variants are to be entertained which use spacing written forms (such as MA + VIRAMA + SUPER-3 as mentioned above) then it may be required for the character to take GC=Mc, and it would also not be entirely appropriate for the character to be named CANDRABINDU.

In my proposal I had advocated, in the interests of entertaining other variants, that this character should indeed take GC=Mc and be named TAMIL SIGN ANUNASIKA (where the semantics-based term ANUNASIKA is more generic than the glyph-based term CANDRABINDU). Further reflections on this issue, along with discussions with other native users who use Extended Tamil, now suggests that the above may not be the best way to go.

First, only the use of the chandrabindu is attested in existing printed publications. (See pp 3-4 of this very document for attestation samples.) The written forms involving MA + VIRAMA are to a large extent theoretical and merely suggested to be employed in e-text for want of a TAMIL CANDRABINDU character. I have confirmed this from other Extended Tamil users such as the author of http://tamilcc.org/thoorihai/Manual.pdf (where MA + VIRAMA is used). Such forms involving MA + VIRAMA are not attested to by printings. Thus it would be appropriate to keep to the chandrabindu, especially seeing as this would complete the set of chandrabindu-s for all major Indic scripts. (Chandrabindu-s for the other major South Indian scripts are proposed by L2/10-392.) Thus this character should indeed take GC=Mn and be named TAMIL SIGN CANDRABINDU or TAMIL EXTENDED SIGN CANDRABINDU.

When I am instructed to do so by the UTC, I will submit a finalized Extended Tamil proposal complete with proposal summary form, code chart with appropriate glyphs and appropriate Unicode character properties listing.

§3. Conclusion

Extended Tamil is an attested and real writing form. While it has many variants, the underlying semantic content and set of characters is the same throughout all these
variants, comprising a complete complement of Brahmic vowels, consonants and other signs as found in other Indic scripts. Scholars and publishers all over India (from Gorakhpur in Uttar Pradesh to Chennai and Tiruchi in Tamil Nadu) have recognized this writing form as a natural extension of the Tamil script. It is not possible to represent this writing form (especially with all its variants) by the characters existing in Unicode. Therefore it is justified to encode new characters for this purpose. When such characters are encoded, it will be possible to achieve all the different variants by mere alteration of fonts.

As for text search, since in Sanskrit one searches “by phonological content and not by orthography” (as already said in L2/09-372 pp 8 and 34) and since there is a general consistency of principle in applying the diacritics, whatever written form they may take, there would be no problem, as users will readily recognize the unity between the variants. As for collation, it is obviously the Sanskrit collation order as shown in L2/09-372 pp 48-49.

The use of Grantha written forms is limited to only one form of Extended Tamil, that which we have called Extended Tamil Liberal. The majority of Extended Tamil printings (as exemplified by the books running to almost 420,000 copies) are however of the ‘conservative’ variant and do not use Grantha written forms but instead diacritics in the form of numbers or letters or other marks such as strokes applied to the regular Tamil written forms. Even when Grantha written forms are used in Extended Tamil Liberal, the overall script structure and grammar is that of Tamil and not Grantha. It is hence inappropriate to characterize this Extended Tamil writing as either Grantha, Linear Grantha, Extended Grantha or any other kind of Grantha.

It is thus entirely justified to encode the characters required for Extended Tamil with script=tamil and with the word TAMIL in their character names (with or without an additional adjective EXTENDED). Any objections to this need not be entertained as they are based not on logic or any solid technical ground but only on meaningless anti-Sanskrit or at least artificial Tamil purist attitudes.

Further, before concluding I should note that while I have repeatedly referred to Sanskrit throughout this document as it defines the major usage case of Extended Tamil, what I have said should also apply to Extended Tamil as used for the representation of other languages such as Saurashtra, Hindi, Marathi, Telugu and Kannada which are also represented by Extended Tamil. Extended Tamil is a real and living written form and hence should definitely be uniquely represented in Unicode.