

## Follow-up to Extended Tamil proposal L2/10-256R

Shriramana Sharma, jamadagni-at-gmail-dot-com, India

2010-Sep-30

This is a follow-up to my Extended Tamil proposal L2/10-256R. It reflects some further thought I have been giving to the matter of how Extended Tamil and related script-forms should be represented at the encoding level. It also describes use of Extended Tamil for contexts I had not considered earlier. No part of this document alters any of the Extended Tamil characters or their names or properties as proposed by L2/10-256R, however. The only intention is to clarify the details of the implementation and usage of Extended Tamil.

### §1. Encoding model of Extended Tamil and related script-forms

#### §1.1. Tamil script for Tamil language

Just to cover the entire spectrum, I first note that characters from the Tamil block are used to denote the Tamil language (obviously). The point is that Extended Tamil characters which are intended for the proposed Tamil Extended block are *not* used for this:

ஐந்து கரத்தனை யானை முகத்தனை  
இந்தின் இளம்பிறை போலும் எயிற்றனை  
நந்தி மகன்தனை ஞானக் கொழுந்தினைப்  
புந்தியில் வைத்தடி போற்றுகின் றேனே.

(The above verse is the first verse from the Tirumantiram, a text on the Shaiva Vedanta religion attributed to one sage Tirumūlar.)

#### §1.2. Grantha script for Sanskrit language

At the other end of the spectrum, the Grantha script – to be precise, characters from the Grantha block – are used to denote Sanskrit as in this verse from the Bhagavad Gītā (18-66):

ஸவ்யுயஜ்ஞாந் வாரிகுஜ்ஞ  
ஜாஹ் னகம் ஸரணம் வ்ரஜ |  
கூஹம் சூவா ஸவ்யுவாபெஹ்யூ  
ஜொகூயிஷ்யூஜி ஜா ஸூஹ: ||

### §1.3. Extended Tamil script (Liberal variant) for Sanskrit language

The same verse in Extended Tamil, using characters from the Tamil and Tamil Extended blocks and a font that displays those characters in the orthographic style we have called in our proposal L2/10-256R as ET-L or Extended Tamil Liberal:

ஸர்வயர்மாந் பரித்யஜ்ய  
மாம் ஏகம் ஸரணம் வ்ரஜ |  
அஹம் த்வா ஸர்வபாபேஹ்யோ  
மோக்ஷயிஷ்யாமி மா ஸூசஃ ||

The language is Sanskrit. Both Tamil-style and Grantha-style glyphs are seen. Only characters from the Tamil and Tamil Extended block are used. No characters from the Grantha block are used.

### §1.4. Extended Tamil script (Conservative variant) for Sanskrit language

The same verse in ET-C or Extended Tamil Conservative:

ஸர்வத<sup>4</sup>ர்மாந் பரித்யஜ்ய  
மாம் ஏகம்<sup>2</sup> ஸரணம்<sup>2</sup> வ்ரஜ |  
அஹம்<sup>2</sup> த்வா ஸர்வபாபேப்<sup>4</sup>யோ  
மோக்ஷயிஷ்யாமி மா ஸூசஃ ||

The language is Sanskrit. Only Tamil-style glyphs along with superscript digits (or other diacritics in other variants) are seen. Only characters from the Tamil and Tamil Extended block are used. No characters from the Grantha block are used.

### §1.5. Mixed Grantha and Tamil scripts for mixed Sanskrit and Tamil languages

A translation of the same verse above into Tamil interlaced with Sanskrit words, in which the Sanskrit portions – often parts of words and not even whole words – are represented in the Grantha script (marked in blue):

ஸவ்யுயஜுங்களையும் வரித்யஜ்யம் செய்து என்னொருவனையே  
அடைக்கலம் புகுவாய். நான் உன்னை  
ஸவ்யுபாவங்களிலிருந்தும் விடுவிப்பேன், கவலைப்படாதே.

This writing style, called Manipravalam, is marked by a pronounced increase in the usage of Sanskrit word-roots, especially Sanskrit-style compounds and Sanskrit verb-roots, which however mostly take Tamil-style word endings (as seen above) while sometimes Sanskrit-style word endings may also be seen. When the writer seeks to accurately represent the Sanskrit-style portions, s/he uses Grantha.

In such a writing style, the language is variously Sanskrit and Tamil. Sanskrit portions are composed entirely using characters from the Grantha block and Tamil portions entirely from the Tamil block. No characters from the Tamil Extended block are used. The Sanskrit/Grantha portions will exhibit Grantha orthographic rules while the Tamil/Tamil portions will exhibit Tamil ones. Thus this is a true mixture of two scripts.

#### *§1.6. Extended Tamil script (Conservative variant) for Sanskrit fragments in Tamil*

When significant portions of a Tamil passage are judged to be in Sanskrit and are desired to be accurately represented, the mixed Tamil-Grantha writing as described above is used. When there is no desire for such accuracy, especially when Sanskrit words are considered full imports into the Tamil language, such Sanskrit-derived words are written in the Tamil script without any extensions. Most of the time, Sanskrit-aware readers can infer the original Sanskrit word with its proper phonetic content from the context.

Occasionally, however, a writer will desire to mark a Sanskrit-derived Tamil word so that it is not confused with a similar native Tamil or another Sanskrit-derived Tamil word. For example, the word bhāvam (from Skt bhāva = state of mind) in Tamil would be written the same as pāvam (from Skt pāpa = sin, used to express sympathy). In a context where the writer seeks to convey to the reader it is the former word that is intended and not the latter, the writer will leave no room for doubt and use ET-C for the consonant, as follows:

பகவானிடம் உயர்ந்த ப<sup>4</sup>ாவம் கொண்டவர் இவர் ... பாவம்  
இப்படிக் கஷ்டப்படுகிறார்.

Whether an Extended Tamil character has here been used for a Sanskrit word or a Tamil one is moot, and in fact, it is irrelevant. Extended Tamil is intended to enable the Tamil script to distinguish more sounds that it is natively able to. Such a distinction was required here, and was hence accomplished using Extended Tamil. The next section will demonstrate that Extended Tamil is not limited to the writing of Sanskrit. Thus even if it were argued that it is a Tamil word here that Extended Tamil was used to denote, no fault can be found with that on principle.

### §1.7. Concluding words on encoded representation

The condensed summary of the above analysis can be stated as follows. **Characters from the Grantha block are only used in two very clearly defined contexts: when the writing is entirely Grantha/Sanskrit, and for the Grantha/Sanskrit parts of Manipravalam texts. For all other cases, only Tamil and Extended Tamil characters are to be used, even if Grantha-style glyphs are seen in some renderings thereof** and irrespective of the language represented. When one clearly understands this demarcation, all confusions will disappear. There is no “mixing” of scripts at the encoding level even if any seems to exist at the superficial glyphic level.

Even Manipravalam, which comes close to “mixing” scripts, is encoding-wise stable. Even when characters with script=grantha and script=tamil are “mixed” in the sense of being present in the same “word”, they are *never* mixed in rendering – i.e. combinations of base characters and combining marks with different script properties will *never* occur and do *not* need to be supported. (In fact, such combinations should *not* be supported in the interest of ensuring sane data input.) As to where exactly the script boundary occurs in a given Manipravalam text, i.e. where the sequence of script=grantha characters ends and that of script=tamil characters starts, it is up to the inputter to carefully compose the text such that there will be no problem when the text is later used. “Caveat scriptor”!

## §2. The use of Extended Tamil for other than Sanskrit

In my Extended Tamil proposal, I had predominantly considered the use of Extended Tamil for the writing of Sanskrit. In passing, I had also cast a glance at its use for writing Saurashtra assuring that Extended Tamil as proposed by me would also support this. I was content to consider only Sanskrit and Saurashtra because these are the only two languages *explicitly* mentioned in TUS chapter on Tamil as requiring the use of numeral diacritics.

However, I realized that Extended Tamil can be and is also used to write very many other languages. I provide samples for Hindi, Marathi, Telugu and Kannada. These are Hindu religious compositions in these respective languages by Goswami Tulasi Das, Sant Jnaneshvar, Tyagaraja Swami and Purandara Dasa, who are all recognized to have chiefly employed these particular languages in their extensive literary output. The source for these samples is a book *Śrī Bhāgavata Sammelana Sampradāya Bhajana Mālā* published in 2008 by Śrī Jñānānanda Sēvā Samājam, Mandaveli, Chennai which is a respected institution in these Hindu religious circles. The page numbers are given along with the samples.

Hindi, p 236

ப<sup>4</sup>ஜமன ராம் சரண தி<sup>3</sup>ன ராதி  
ரஸனா கஸபஜோ தும் ஹரிபத  
ஸுரத பிரதக்<sup>3</sup>யோம் அவஸாதி  
ஜாகே கஹத் த<sup>3</sup>ஹத் து<sup>3</sup>க்க த<sup>3</sup>ாருண்  
ஸுனோ த்ரய தாப பு<sup>3</sup>ஜாதி  
ராமசந்த்<sup>3</sup>ரகே நாம அமியரஸ .  
ஸோரஸ கோபி<sup>3</sup>ன க<sup>3</sup>ாதி  
துளஸ்த<sup>3</sup>ாஸ ப்ரபு<sup>4</sup> வினய கரத் ஹை  
ப்ரத<sup>2</sup>ம ஆரஜகீ பாதி

Marathi, p 245

ஸாரஸார நாம விடே<sup>2</sup>ாப<sup>3</sup>ா நாம துஜே<sup>2</sup> ஸார  
ம்ஹணோனீ ஸுலபாணீ ஜபதாஹேம் வாரம்வார  
ஆதி<sup>3</sup> மத்<sup>3</sup>ய அந்தீம் நிஜ பீ<sup>3</sup>ஜ ஒங்கார  
பதீத அக்ஞானி ஜட தரலே அபார  
பு<sup>4</sup>க்தி ஆணி முக்தி ஸுக<sup>2</sup> த<sup>3</sup>ாயக ஸாசார  
த்ருவ ப்ரஹ்லாத<sup>3</sup> அம்ப<sup>3</sup>ா<sup>3</sup>ஷேம்கேலா ஹா நிர்<sup>3</sup>த<sup>3</sup>ார  
தி<sup>3</sup>வஸோ தி<sup>3</sup>வஸீ வ்யர்த<sup>2</sup> ஜாதோஹா ஸம்ஸார  
ப<sup>3</sup>ய ரகு<sup>3</sup>மா தே<sup>3</sup>வீ வஸு விட்<sup>2</sup>ட<sup>2</sup>லேம் ஆத<sup>3</sup>ார

Telugu, p 230

ராகம் : சுத்தபங்காள  
ப : ராம ப<sup>4</sup>க்தி ஸாம்ராஜ்யம்  
ஏமானவுல கப்பே<sup>3</sup>னோ மனஸா  
அ : ஆமானவுல ஸந்த<sup>3</sup>ா<sup>3</sup>ஸனம்  
அத்யந்த ப்<sup>3</sup>ரஹ்மானந்த<sup>3</sup>மே  
ச : ஈலாக<sup>3</sup>னி விவரிம்பலேனு  
சாலா ஸ்வானுப<sup>4</sup>வ வேத்<sup>3</sup>யமே  
லீலா ஸ்ருஷ்ட ஜக<sup>3</sup>த்ரயமனே  
கோலாஹல த்யாக<sup>3</sup>ராஜ<sup>3</sup>னுது<sup>3</sup>கு

Kannada, p 203

ஜக<sup>3</sup>து<sup>3</sup>த்<sup>3</sup>த<sup>3</sup>ாரன ஆடி<sup>3</sup>ஸித்<sup>3</sup>ளெ யஸோதே<sup>3</sup>  
1 ஜக<sup>3</sup>து<sup>3</sup>த்<sup>3</sup>த<sup>3</sup>ாரன மக்<sup>3</sup>னெந்து<sup>3</sup> திளியுத  
ஸுகுணாந்த ரங்க<sup>3</sup>ன ஆடி<sup>3</sup>ஸித்<sup>3</sup>ளெ யஸோதே<sup>3</sup>  
2. நிக<sup>3</sup>மகெ ஸிலுகத்<sup>3</sup> அக்<sup>3</sup>ணித மஹிமன  
மகு<sup>3</sup>க்<sup>3</sup>ள மாணிக்யன ஆடி<sup>3</sup>ஸித்<sup>3</sup>ளெ யஸோதே<sup>3</sup>  
3. அணோரணீயன மஹதோ மஹீயன  
அப்ரமேயன ஆடி<sup>3</sup>ஸித்<sup>3</sup>ளெ யஸோதே<sup>3</sup>  
4. பரமபுருஷன பரவாஸு<sup>3</sup>தே<sup>3</sup>வன  
புரந்த<sup>3</sup>ர விட்<sup>2</sup>ட<sup>2</sup>லன ஆடி<sup>3</sup>ஸித்<sup>3</sup>ளெ யஸோதே<sup>3</sup>

These samples indicate that Extended Tamil is used for languages other than Sanskrit and Saurashtra as well, It is also to be noted that such usage for South Indian languages like Telugu and Kannada will mean that the contrast between short E/O and long EE/OO exists. (Such contrast is seen in the Kannada sample above, for example.) That does not alter the fact that Sanskrit and Saurashtra are the *major* usage cases for Extended Tamil, however.

-0-0-0-