

## 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

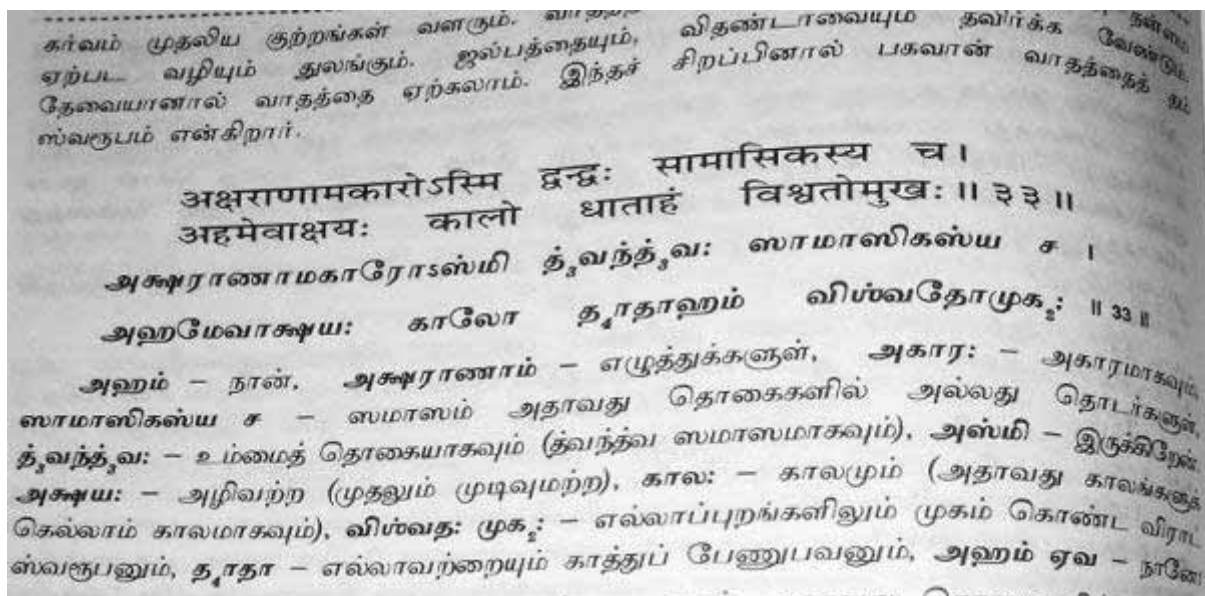
Many contemporary publications in Extended Tamil use subscript digits instead of superscript ones. To illustrate this we provide samples from publications of Gita Press, Gorakhpur, which is a very respected publisher of Hindu religious books in all the major Indian languages/scripts. (See <http://www.gitapress.org>.)

From p iv 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:

ஸ்ரீஹரி			
தேவநாகரி எழுத்துகள்			
வடமொழிச் சொற்களைப் பிழையின்றி உச்சரிக்க உதவியாக, ஆங்கில எழுத்துகளின் துணைகொண்டு 2, 3, 4 என்னும் குறியீடுகளுடன் வடமொழி எழுத்துகளுக்கு நிகரான தமிழ் ஒலி வடிவங்களும் தரப்பட்டுள்ளன.			
[ka] க க	[kha] க <sub>2</sub> க <sub>2</sub>	[ga] க <sub>3</sub> க <sub>3</sub>	[gha] க <sub>4</sub> க <sub>4</sub>
[ca] ச ச	[cha] ச <sub>2</sub> ச <sub>2</sub>	[ja] ஜ ஜ	[jha] ஜ <sub>2</sub> ஜ <sub>2</sub>
[ta] ட ட	[tha] ட <sub>2</sub> ட <sub>2</sub>	[da] ட <sub>3</sub> ட <sub>3</sub>	[dha] ட <sub>4</sub> ட <sub>4</sub>
[ta] த த	[tha] த <sub>2</sub> த <sub>2</sub>	[da] த <sub>3</sub> த <sub>3</sub>	[dha] த <sub>4</sub> த <sub>4</sub>
[pa] ப ப	[pha] ப <sub>2</sub> ப <sub>2</sub>	[ba] ப <sub>3</sub> ப <sub>3</sub>	[bha] ப <sub>4</sub> ப <sub>4</sub>
[sha] ஶ ஶ	[sha] ஷ ஷ	[sa] ஸ ஸ	[gna] ங்ஞ ங்ஞ
[ksha] ஶ்ஷ ஶ்ஷ	[ri] ற்ரு ற்ரு		
'ஸ' என்னும் எழுத்து பிசுரு, கசடு என்னும் சொற்களில் வரும் 'ச' வைப் போல் ஒலிக்கும்.			
'ரு' என்னும் எழுத்து எருது, விருது என்னும் சொற்களில் வரும் 'ரு' வைப் போல் ஒலிக்கும்.			
ஸம்ஸ்கிருத கலோகங்களில் சில இடங்களில் 'ஸ' அல்லது 'ஸ' என்னும் குறியீடுகளைக் காணலாம். இதற்கு 'அவக்ரஹம்' என்று பெயர். இந்த இடங்களில் முறையே 'அ', 'ஆ' என்னும் எழுத்துகள் தொக்கி நிற்கின்றன என்பது பொருள்.			

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:

யதஸ் ஸர்வாணி பூ<sup>4</sup>தானி ப<sup>4</sup>வந்த்யாதி<sup>3</sup> யுக<sup>3</sup>ா ஸ்<sup>3</sup>க்<sup>3</sup>மே |  
யஸ்மிம்ஸ்ச ப்ரளயய் யாந்தி புன ரேவ யுக<sup>3</sup> க்ஷயே ||

தஸ்ய லோக ப்ரத<sup>4</sup>ானஸ்ய ஜக<sup>3</sup>ந்நாத<sup>2</sup>ஸ்ய பூ<sup>4</sup>பதே |  
விஷ்ணோர் நாம ஸஹஸ்ரம் மே ஸ்யணு பாப ப<sup>4</sup>யா ஸபஹம் ||

யானி நாமானி கௌணானி விக<sup>2</sup>யாதானி மஹா ஸ்<sup>3</sup>த்மன: |  
ஜஷிபி<sup>4</sup>: பரி கீ<sup>3</sup>தானி தானி வக்ஷ்யாமி பூ<sup>4</sup>தயே ||

3. ஜஷி:, தே<sup>3</sup>வதா, ச<sup>2</sup>ந்த<sup>3</sup>:, பூர்வ ந்யாஸ:

ஜஷிர் நாம்நாம் ஸஹஸ்ரஸ்ய-வேத<sup>3</sup> வ்யாலோ மஹா முனி: |  
ச<sup>2</sup>ந்தோ ஸனுஷ்டுப் தத<sup>2</sup>ா தே<sup>3</sup>வோ ப<sup>4</sup>க்<sup>3</sup>வான் தே<sup>3</sup>வக்<sup>3</sup> ஸுத: ||

அம்யுதாம் ஸு<sup>3</sup>த<sup>3</sup>ப்<sup>4</sup>வோ பீ<sup>3</sup>ஜம் ஸக்திர் தே<sup>3</sup>வகி நந்த<sup>3</sup>ன: |  
த்ரி ஸாமா ஹு<sup>3</sup>த<sup>3</sup>யந் தஸ்ய ஸாந்த்யர்தே<sup>2</sup> வினியுஜ்யதே ||

விஷ்ணுஞ் ஜிஷ்ணும் மஹா விஷ்ணும் ப்ரப<sup>4</sup>விஷ்ணும் மஹேஸ்வரம் |  
அனேக ரூப தை<sup>3</sup>த்யாந்தந் நமாமி புருஷோத்தமம் ||

Introduction, p 3:

அறிமுகம்தான். ஸ்தோத்திரங்கள் மற்றும் வேத மந்திரங்களின் தமிழ்ப் பதிப்புகளில் இது ஒரு புதிய அத்தியாயம். வேத மந்திரங்களுக்கும் வடமொழி ஸ்தோத்திரங்களுக்கும் பல தமிழ்ப் புத்தகங்கள் உள்ளன. ஆனால் இப்பதிப்பு மற்ற தமிழ்ப் பதிப்புகளிலிருந்து முற்றிலும் மாறுபட்ட ஒன்று. ஏனெனில், இது நூறு சதவிகிதம் வடமொழிக்கு இணையான உச்சரிப்புடன் தமிழில் வெளிவரும் முதல் பதிப்பு எனலாம்.

இதுவரை வெளிவந்துள்ள பதிப்புகளில் வடமொழி உச்சரிப்புகளுக்காக க, ச, ட, த, ப என்ற எழுத்துகளுடன் சிறிய வடிவில் 2, 3, 4 எண்களும் 'ஸ்' என்னும் எழுத்தும் பயன் படுத்தப் பட்டுள்ளன. ஆனால் இவை மட்டுமின்றி வடமொழியில் இன்னும் பல நுணுக்கமான உச்சரிப்பு வேறுபாடுகள் உண்டு. இவை அனைத்தையும் முழுமையாகப்

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).

பிரிவு	ஆங்கிலம்	வட மொழி	கிரந்தம்	தமிழ் ரவி	உதாரணம் (உச்சரி)	பிரிவு	ஆங்கிலம்	வட மொழி	கிரந்தம்	தமிழ் ரவி	உதாரணம் (உச்சரி)
'க' வ ர் க் ம்	ka	क	क	க	கூல்	ஊஷம்	sha	श	श	சா	சாவி
	kha	ख	ख	க்	அகிலம்		Sha	ष	ष	ஷ	விஷ்ணு
	ga	ग	ग	க்	க்ணபதி		sa	स	स	ஸ	ஸரஸ்வதி
	gha	घ	घ	க்	ராக்வ		ha	ह	ह	ஹ	ஹரி
	-Na	ङ	ङ	ங	இங்கனம்		La	ळ	ळ	ள	வளம்
'ச' வ ர் க் ம்	cha	च	च	ச	சென்னை	'அ' ச் க ள்	aru	ठ	ठ	ஊ(ருஷி), அமு	
	Cha	छ	छ	ச்	இச்சர்		aru	ड	ड	பிதுன் (பித்ருன்)	
	ja	ज	ज	ஜ	ஜன்னல்		alu	ढ	ढ	மாலா களப்தம்	
	jha	झ	झ	ஜ்	ஜர்		nasal ya	ण	ण	ய	பரமய்ய: (பரம
-na	ञ	ञ	ஞ	மஜ்ஜன்	nasal la	त	त	ல	தைவதல் லோடு		
'ட' வ ர் க் ம்	Ta	ट	ट	ட	பட்டம்	வ	nasal va	थ	थ	வ்	விஸ்வல் விஷ்ணு
	Tha	ठ	ठ	ட்	வைகுண்டம்		##	{m}	०	ம்	ஹம்ஸ (ஹ்ம்)
	Da	ड	ड	ட்	கடல்						
	Dha	ढ	ढ	ட்	மூடம்தே						
Na	ण	ण	ண	க்ணபதி							
'த' வ ர் க் ம்	ta	त	त	த	தலை						
	tha	थ	थ	த்	பரதம்ம்						
	da	द	द	த்	வந்தனம்						
	dha	ध	ध	த்	தர்மம்						
	na	न	न	ந	நாடு						
'ப' வ ர் க் ம்	pa	प	प	ப	பால்						
	pha	फ	फ	ப்	பலம் (phalam)						
	ba	ब	ब	ப்	அன்பு						
	bha	भ	भ	ப்	பக்தி						
	ma	म	म	ம	மலர்						
அந் த ஸ் த் ம்	ya	य	य	ய	யானை						
	ra	र	र	ர	மரம்						
	la	ल	ल	ல	மலர்						
	va	व	व	வ	வயல்						

### §2.3. Using superscript V and subscript I

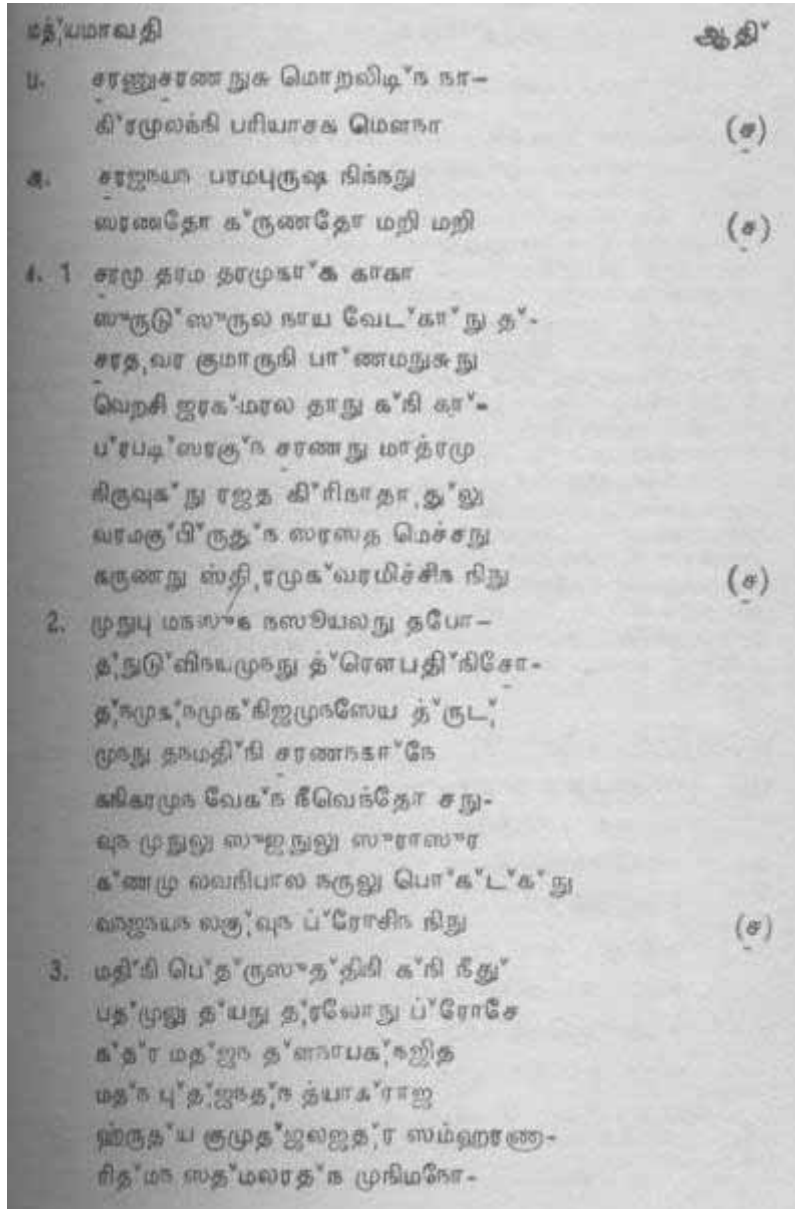
Yet another publication shows even more innovation in extending the Tamil script. For voiced consonants, it places a superscript Latin letter V after the consonant. (V for voiced!) For aspirates, it places an I-like vertical stroke in subscript. For voiced aspirates, it uses both superscript V and subscript 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:

உச்சரிப்புப் பட்டியல்					
க	க	க	த <sup>v</sup>	த	த
க <sup>v</sup>	ச	ச	க	ச	ச
க <sup>v</sup>	ர	ர	ப	ப	ப
க <sup>v</sup>	ல	ல	ப <sub>I</sub>	ப <sub>I</sub>	ப <sub>I</sub>
ங	ங	ங	ப <sup>v</sup>	ப <sup>v</sup>	ப <sup>v</sup>
ச	ச	ச	ம	ம	ம
ச <sup>v</sup>	ர	ர	ய	ய	ய
ச <sup>v</sup>	ல	ல	ச	ச	ச
ச <sup>v</sup>	ர	ர	ற	ற	ற
ச <sup>v</sup>	ல	ல	ல	ல	ல
ச <sup>v</sup>	ர	ர	வ	வ	வ
ச <sup>v</sup>	ல	ல	ச	ச	ச
ச <sup>v</sup>	ர	ர	ச <sup>v</sup>	ச <sup>v</sup>	ச <sup>v</sup>
ச <sup>v</sup>	ல	ல	ச <sup>v</sup>	ச <sup>v</sup>	ச <sup>v</sup>
ச <sup>v</sup>	ர	ர	ஸ	ஸ	ஸ
ச <sup>v</sup>	ல	ல	ஹ	ஹ	ஹ
ச <sup>v</sup>	ர	ர	ள	ள	ள
ச <sup>v</sup>	ல	ல	சு	சு	சு

நெலுக்கு, சம்ஸ்கிருதம் ஆகிய பாணிகளில் நகரம் ஒன்றே வாதலால் கசுநு 'ந' என்று தமிழ் எழுத்தினால் குறிப்பிடப் பெற்றுள்ளது. நெலுக்கு மொழியில் மாத்திரம் 'ர' 'ற' ஆகிய இரண்டு எழுத்துக்களும் வழக்கில் உள்ளன.

Here is an entire page (p 143) from the book showing this form of Extended Tamil:



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.

-0-0-0-