

Request to allocate a Tamil Supplementary block in the Unicode SMP Roadmap

Shriramana Sharma

Cuddalore, Tamil Nadu, India

jamadagni-at-gmail.com

Introduction

This is a request to allocate a Tamil Supplementary Block in the SMP Roadmap of Unicode. The Tamil script has been placed in the BMP in the block 0B80-0BFF. Apart from the basic language character repertoire of consonants and vowels, this block also contains the Tamil numerals – the digits 0 to 9 and the numbers 10, 100 and 1000. It also contains various symbols that were formerly used in Tamil day-to-day life to denote generic concepts like ‘as above’ (akin to “ditto”) and ‘number’, calendrical elements like year, month and day, and commercial concepts like debit, credit and the currency rupee.

Digits

0BE6	௦	TAMIL DIGIT ZERO
0BE7	௧	TAMIL DIGIT ONE
0BE8	௨	TAMIL DIGIT TWO
0BE9	௩	TAMIL DIGIT THREE
0BEA	௪	TAMIL DIGIT FOUR
0BEB	௫	TAMIL DIGIT FIVE
0BEC	௬	TAMIL DIGIT SIX
0BED	௭	TAMIL DIGIT SEVEN
0BEE	௮	TAMIL DIGIT EIGHT
0BEF	௯	TAMIL DIGIT NINE

Tamil numerics

0BF0	௧௦	TAMIL NUMBER TEN
0BF1	௧௦௦	TAMIL NUMBER ONE HUNDRED
0BF2	௧௦௦௦	TAMIL NUMBER ONE THOUSAND

Tamil symbols

0BF3	௩௦	TAMIL DAY SIGN
0BF4	௩௧	TAMIL MONTH SIGN
0BF5	௩௨	TAMIL YEAR SIGN
0BF6	௩௩	TAMIL DEBIT SIGN
0BF7	௩௪	TAMIL CREDIT SIGN
0BF8	௩௫	TAMIL AS ABOVE SIGN

Currency symbol

0BF9	௩௬	TAMIL RUPEE SIGN
------	----	------------------

Tamil symbol

0BFA	௩௭	TAMIL NUMBER SIGN
------	----	-------------------

Portion of the Unicode Standard 5.1 chart of the Tamil block showing the currently encoded repertoire of Tamil numerals and other Tamil symbols

However, apart from the symbols currently encoded in the Tamil block for simple numerals and the above-mentioned concepts, various sources testify that Tamils have used even more symbols for commerce, agriculture, weights and measures and mathematical fractions. Some samples follow:

Abbreviations & Contractions

ௐ	<i>om</i> (cf. below)
ௐ	= க்கு until
சீ	= சிரஞ்சீவி long-lived
ஸ்ரீ & ஸ்ரீ	= சிரீ <i>śrī</i>
பீ	= பிள்ளை <i>pillai</i> , a title of rank
ம-ந-ந-ஸ்ரீ	= மகாராஜ ராஜசிரீ <i>mahārājarājaśrī</i> , abbr. "M. R. Ry."
மு	= முதல் from the first
வு	= வகையரா belonging to
ஷ	= மேற்படி afore-mentioned
ஹ	= முதலியவை "etc."
– <i>Accounting</i>	
கூ	= கூக total
நு & நு	= நட்டார் number
நீ	= நிலுவை balance

ரூ & ரூ = ரூபாய் a *Rupee* = 16 *Annas*

– *Calendar*

வரு = வருஷம் year

பீ = மீசம் month

உ & உ = தேதி date

நா = நாள்து current

– *Weights & Measures*

பா = பாரம் a *Candy* (an “Indian ton”)

ம = குறுணி one *Marc* (measure of capacity)

ஹ = பதக்கு two *Marcals*

தம = ஐங்குறுணி five *Marcals*

வத = இருதாணி eight *Marcals*

கல = கலம் a *Kalam* = 12 *Marcals*

கா = காணி a *Cawney* (measure of land)

கச = கசம் a yard

குடி = குடி a square yard

Sign.	Name.	Power.
வத	முத்திரை is the three hundred and twentieth part of the Integer,	$\frac{1}{320}$
ந	அரைக்காணி	$\frac{1}{160} = \frac{2}{320}$
ரி	காணி	$\frac{1}{80} = \frac{4}{320}$
சு	அரைமா	$\frac{1}{40} = \frac{8}{320}$
சு	முக்காணி	$\frac{3}{80} = \frac{12}{320}$
ப	மா or ஒருமா	$\frac{1}{20} = \frac{16}{320}$
பரி	மாகாணி or வீசம்	$\frac{1}{16} = \frac{20}{320}$
பி	இரண்டிமா or இருமா	$\frac{1}{10} = \frac{32}{320}$
ஞ	அரைக்கால் or இரண்டிமாவரை	$\frac{1}{8} = \frac{40}{320}$
ந	மும்மா or மூன்றா	$\frac{3}{20} = \frac{48}{320}$
ச	நாலுமா	$\frac{1}{5} = \frac{64}{320}$
வ	கால்	$\frac{1}{4} = \frac{80}{320}$
உ	அரை	$\frac{1}{2} = \frac{160}{320}$
ஐ	முக்கால்	$\frac{5}{4} = \frac{240}{320}$

Symbols for various fractions

<http://kanimozhi.org.in/kanimozhi/?p=18> (retrieved 2009-Sep-21)

கருட - -	ஒருச்சுவடு - - - - -	1 Suvadu.
உளு - -	இருச்சுவடு - - - - -	2
நுளு - -	மூச்சுவடு - - - - -	3
சுளு - -	நாற்சுவடு - - - - -	4
ஹ - -	ஆழாக்கு - - - - -	5 Suvadu 1 Alhāccu.
ஹ - -	உழக்கு - - - - -	2
ஹு - -	உழக்காழாக்கு - - - - -	3
ஹரி - -	உரி - - - - -	4
ஹரி - -	உரியாழாக்கு - - - - -	5
ஹ - -	மூவுழக்கு - - - - -	6
ஹு - -	மூவுழக்காழாக்கு - - - - -	7
ஹ - -	நாழி - - - - -	8 Alhāccu 1 Nālhi.
ஹ - -	இருநாழி - - - - -	2
ஹ - -	முந்நாழி - - - - -	3
ஹ - -	நாநாழி - - - - -	4
ஹ - -	ஐந்நாழி - - - - -	5 Nālhi.
ஹ - -	அறுநாழி - - - - -	6
ஹ - -	எழுநாழி - - - - -	7
ஹ - -	குறுணி - - - - -	8 Nālhi 1 Curuni.
ஹ - -	பதக்கு - - - - -	2
ஹ - -	முக்குறுணி - - - - -	3
ஹ - -	தூணி - - - - -	4
ஹ - -	ஐங்குறுணி - - - - -	5
ஹ - -	தூணிப்பதக்கு - - - - -	6
ஹ - -	எழுகுறுணி - - - - -	7
ஹ - -	இருதூணி - - - - -	8
ஹ - -	இருதூணிக்குறுணி - - - - -	9
ஹ - -	இருதூணிப்பதக்கு - - - - -	10
ஹ - -	இருதூணிமுக்குறுணி - - - - -	11
ஹ - -	கலம் - - - - -	12 Curuni 1 Calam.

Symbols for measures, especially related to agriculture

<http://kanimozhi.org.in/kanimozhi/?p=18> (retrieved 2009-Sep-21)

Requirement for encoding in Unicode

It is thus evident that the ancient Tamils have been very productive in devising and using various symbols for all facets of life. Though most (but not all) of these symbols have gone out of current use, they are a part of Tamil heritage and hence the world's heritage. As Unicode has encoded many such historical scripts and symbols for their very historical value, it must also encode these symbols used by the Tamils.

Requirement of a separate block

The existing Tamil block in the BMP does have some space left over in the sixth, seventh and eighth columns, and in fact has some of these symbols already encoded there. However, it is evident even by seeing the few samples we have given here, that all the attested symbols cannot be encoded within that limited space. Therefore we request for a separate block to be allocated in the SMP Roadmap. Considering the number of these symbols, we request that the block span at least six columns. A full half-row block might be preferable seeing as further research may reveal even more valid characters.

Common repertoire of Tamil and Grantha

It should be noted here that many of these symbols are also used in Sanskrit texts written in Grantha. Especially those for the major fractions such as quarter, half, and three-quarters are used in Vedic texts to denote minute timing aspects of pronunciation.

Ancient Tamil culture and Tamil heritage combines the richness of both the Tamil and Sanskrit languages and the corresponding literatures. These symbols, being a part of that Tamil culture, are hence used equally in texts composed in both languages. Since Grantha is the script traditionally used for Sanskrit in the Tamil-speaking region, Grantha texts also use these symbols when needed.

The Tamil and Grantha scripts have been and continue to be mutually complementary in representing the two major languages of Tamil culture – Tamil and Sanskrit. They already share a large number of consonant and vowel characters. It is only to be expected that they also share the symbols for the numerals and other concepts. In my recent document “Comments on Mr Ganesan’s Grantha Proposal” submitted to the UTC via e-mail sent to Rick McGowan on 2009-Sep-11, I have clearly proved based on samples from printed Grantha texts that Grantha does not use a distinct repertoire of numerals than those used in Tamil. That Grantha also shares the symbols for the fractions is but an extension of that. Hence Grantha requires no separate symbols for numerals or fractions.

Choice of name

The new block may be named “Tamil Supplementary” as the word ‘Tamil’ has in current usage a broader cultural connotation than the script-specific word Grantha. It would also clarify the fact that the symbols to be encoded in this block are used in Tamil texts. Since it is already recognized that Grantha texts use the numerals encoded in the existing Tamil block, it goes without saying that symbols encoded in a Tamil Supplementary block will be accepted by the Grantha user community for usage in Grantha texts.

Background of this request and possibility of a future proposal

Recently (2009-Sep-19 02:30 UTC+0530) on the indic@unicode.org mailing list, our friend Sri Ramadoss M (ramadasan-at-amachu.net) raised this issue of the Tamil symbols for fractions not being encoded in Unicode:

Refer to Ramanraj's article at <http://kanimozhi.org.in> about the inability of Unicode to represent symbols of araikani, kaani, araima, mukkani, maa, 1/4, 1/2, 3/4 etc. used by Tamils. <http://kanimozhi.org.in/kanimozhi/?p=18> to which Shriramana Sharma (the author of this document) responded:

Why don't you make a well-researched proposal for the inclusion of these characters in the Tamil block? I'm sure the UTC will be glad to add these characters to enable Unicode better represent such characters. ... If you feel the existing Tamil block is insufficient for adding all these characters, you can ask for and get allotted in the SMP additional blocks [sic] for "Tamil Extended" just like Devanagari extended!

after which N Ganesan (naa.ganesan-at-gmail.com) added the following information:

There is an organization working on these issues, <http://infitt.org> Please join it & we can look into these points. ... As far as Tamil is concerned, you may know that INFITT has done many proposals now implemented - zero digit, U+0BB6 (sha), OM sign ... INFITT is aware of the archaic and unused [sic] Tamil fractions, and many symbols of Tamil. ... we can list all of them in an INFITT proposal, & get them encoded all in SMP.

Thus, it is to be reasonably expected that a proposal with further source material and attestations will be forthcoming in the future. Any and all Tamil-Grantha symbols hitherto attested and to be attested in the future and which are used to denote concepts beyond the sounds and words of the actual Tamil and Sanskrit languages may be placed in this block thereby freeing up the Tamil and Grantha blocks for language content alone.

-o-o-o-