

# Tamil Script LGR Proposal

Introduction, Current Analysis and Next Steps

Dr. Shanmugam Rajabhathar

NBGP F2F Meeting, Colombo  
13th December 2017



# Agenda

---

1

Introduction to Tamil  
Script

2

Classification of  
Characters

3

Repertoire Analysis

4

Within & Cross-  
Script Variants

5

WLE Rules

6

Current Status and  
Next Steps for  
Completion

# Introduction to Tamil Script

---

- ⊙ Population:
- ⊙ Geographical area:
- ⊙ Languages written in Tamil script:
- ⊙ .

# Classification of Characters: Consonants

Tamil consonants have been categorized as 3 groups according to their phonetic properties:  
Stops (vəllinəm) , Medial (idəyīnəm) & Nasal(mellinəm)

STOP	ஃ U+0B95	ஈ U+0B9A	ஊ U+0B9F	஋ U+0BA4	஌ U+0BAA	஍ U+0BB1
MEDIAL	ஞ U+0B99	஠ U+0B9E	஡ U+0BA3	஢ U+0BA8	ண U+0BAE	த U+0BA9
NASAL	஥ U+0BAF	஦ U+0BB0	஧ U+0BB2	ந U+0BB5	ன U+0BB4	ப U+0BB3

Table Above : Group classification of consonants; Table Below : IPA classification of Tamil consonants

	Bilabial	Lab-Dental	Dental	Alv	Post-Alv	Retroflex	Palatal	Velar	Uvu	Glottal
Plosive	p (ப) b (ப)		t (த) d (த)			ʈ (ட) ɖ (ட)		k (க) g (க)		
Nasal	m (ம)		n (ந)	ɳ (ண)		ɳ (ண)	ɲ (ஞ)	ŋ (ங)		
Tap/Flap				ɾ (ர)						
Trill				ɽ (ற)						
Fricative				s (ச)						h (ஃ)
Approx		ʋ (வ)				ɻ (ழ)	j (ய)			
Lat Approx				ɻ (ல)		ʎ (ள)				
Affricate							tʃ (ச) dʒ (ஜ)			

# Classification of Characters: Vowels

Separate symbols exist for all Vowels, which are pronounced independently either at the beginning or after a vowel sound. To indicate a Vowel sound other than the implicit one, a Vowel sign (Matra) is attached to the consonant. Since the consonant has a built in 'a', there are equivalent Matras for all vowels excepting the अ.

The correlation is shown as under:

Vowel	Corresponding vowel sign (Matra)
अ U+0B85	
आ U+0B86	ा U+0BBE
इ U+0B87	ी U+0BBF
ऋ U+0B88	ी U+0BC0
ॠ U+0B89	ी U+0BC1

Table : Vowels with corresponding Matras

Vowel	Corresponding vowel sign (Matra)
अ U+0B8A	ं U+0BC2
ए U+0B8E	े U+0BC6
ऐ U+0B8F	ै U+0BC7
ऋ U+0B90	ै U+0BC8
ॠ U+0B92	ै U+0BCA
ऌ U+0B93	ै U+0BCB
ॡ U+0B94	ै U+0BCC

# Repertoire Included - 1

Sr. No	Unicode Code Point	Glyph	Character Name	Unicode General Category (gc)	Indic Syllabic Category	Reference	Y/N
1.	0B83	ஃ	TAMIL SIGN VISARGA	Lo	Visarga	[1003]	
2.	0B85	அ	TAMIL LETTER A	Lo	Vowel	[1001]	
3.	0B86	ஆ	TAMIL LETTER AA	Lo	Vowel	[1001]	
4.	0B87	இ	TAMIL LETTER I	Lo	Vowel	[1001]	
5.	0B88	ஈ	TAMIL LETTER II	Lo	Vowel	[1001]	
6.	0B89	உ	TAMIL LETTER U	Lo	Vowel	[1001]	
7.	0B8A	ஊ	TAMIL LETTER UU	Lo	Vowel	[1001]	
8.	0B8E	எ	TAMIL LETTER E	Lo	Vowel	[1001]	
9.	0B8F	ஏ	TAMIL LETTER EE	Lo	Vowel	[1001]	
10.	0B90	ஐ	TAMIL LETTER AI	Lo	Vowel	[1001]	

# Repertoire Included - 2

Sr. No.	Unicode Code Point	Glyph	Character Name	Unicode General Category (gc)	Indic Syllabic Category	Reference	Y/N
11.	0B92	ஓ	TAMIL LETTER O	Lo	Vowel	[1001]	
12.	0B93	ஔ	TAMIL LETTER OO	Lo	Vowel	[1001]	
13.	0B94	ஔள	TAMIL LETTER AU	Lo	Vowel	[1001]	
14.	0B95	ஈ	TAMIL LETTER KA	Lo	Consonant	[1002]	
15.	0B99	ஈங்	TAMIL LETTER NGA	Lo	Consonant	[1002]	
16.	0B9A	ஈச	TAMIL LETTER CA	Lo	Consonant	[1002]	
17.	0B9C	ஈஜ	TAMIL LETTER JA	Lo	Consonant	[1002]	
18.	0B9E	ஈஞ	TAMIL LETTER NYA	Lo	Consonant	[1002]	
19.	0B9F	ஈட	TAMIL LETTER TTA	Lo	Consonant	[1002]	
20.	0BA3	ஈண	TAMIL LETTER NNA	Lo	Consonant	[1002]	

# Repertoire Included - 3

Sr. No	Unicode Code Point	Glyph	Character Name	Unicode General Category (gc)	Indic Syllabic Category	Reference	Y/N
21.	0BA4	த	TAMIL LETTER TA	Lo	Consonant	[1002]	
22.	0BA8	ந	TAMIL LETTER NA	Lo	Consonant	[1002]	
23.	0BA9	ண	TAMIL LETTER NNNA	Lo	Consonant	[1002]	
24.	0BAA	ப	TAMIL LETTER PA	Lo	Consonant	[1002]	
25.	0BAE	ம	TAMIL LETTER MA	Lo	Consonant	[1002]	
26.	0BAF	ய	TAMIL LETTER YA	Lo	Consonant	[1002]	
27.	0BB0	ர	TAMIL LETTER RA	Lo	Consonant	[1002]	
28.	0BB1	ற	TAMIL LETTER RRA	Lo	Consonant	[1002]	
29.	0BB2	ல	TAMIL LETTER LA	Lo	Consonant	[1002]	
30.	0BB3	ள	TAMIL LETTER LLA	Lo	Consonant	[1002]	



# Repertoire Included - 4

Sr. No.	Unicode Code Point	Glyph	Character Name	Unicode General Category (gc)	Indic Syllabic Category	Reference	Y/N
31.	0BB4	ழ	TAMIL LETTER LLLA	Lo	Consonant	[1002]	
32.	0BB5	வ	TAMIL LETTER VA	Lo	Consonant	[1002]	
33.	0BB6	ஸ	TAMIL LETTER SHA	Lo	Consonant	[1002]	
34.	0BB7	ஷ	TAMIL LETTER SSA	Lo	Consonant	[1002]	
35.	0BB8	ஸ	TAMIL LETTER SA	Lo	Consonant	[1002]	
36.	0BB9	ஹ	TAMIL LETTER HA	Lo	Consonant	[1002]	
37.	0BBE	ா	TAMIL VOWEL SIGN AA	Mc	Matra	[1002]	
38.	0BBF	ி	TAMIL VOWEL SIGN I	Mc	Matra	[1002]	
39.	0BC0	ீ	TAMIL VOWEL SIGN II	Mn	Matra	[1002]	
40.	0BC1	ு	TAMIL VOWEL SIGN U	Mc	Matra	[1002]	

# Repertoire Included - 5

Sr. No.	Unicode Code Point	Glyph	Character Name	Unicode General Category (gc)	Indic Syllabic Category	Reference	Y/N
41.	0BC2	ஊ	TAMIL VOWEL SIGN UU	Mc	Matra	[1002]	
42.	0BC6	஋	TAMIL VOWEL SIGN E	Mc	Matra	[1002]	
43.	0BC7	஌	TAMIL VOWEL SIGN EE	Mc	Matra	[1002]	
44.	0BC8	஍	TAMIL VOWEL SIGN AI	Mc	Matra	[1002]	
45.	0BCA	ஐ	TAMIL VOWEL SIGN O	Mc	Matra	[1002]	
46.	0BCB	ஓ	TAMIL VOWEL SIGN OO	Mc	Matra	[1002]	
47.	0BCC	ஔ	TAMIL VOWEL SIGN AU	Mc	Matra	[1002]	
48.	0BCD	஁	TAMIL SIGN VIRAMA	Mn	Matra	[1002]	

# Repertoire Excluded

- Code Points in MSR-2 but excluded because they are either not in common use or used for special purpose only (historic, religious, etc.):

Code Point	Glyph	Type	Notes
U+0BD7	௰	TAMIL AU LENGTH MARK	Not in modern usage. Excluded as per conservatism principle.

# Within Script Variants for Tamil - 1

List of code points which are variants

Group 1: visually similar characters in Tamil Script

Code Point 1 + Glyph 1	Code Point 2 + Glyph 2	Type (allocatable or blocked)
ெௌ U+0BCC	ெௌ U + 0BC6 U+ 0BB3	
ஔௌ U+0B94	ஔௌ U+0B92 U+0BB3	

# Within Script Variants for Gujarati - 2

- List of code points which are variants:  
Group 2: visually similar Tamil character combinations, due to presence dots and other characters

Code Point 1 + Glyph 1	Code Point 2 + Glyph 2	Type (allocatable or blocked)

# Cross Script Variants for Tamil and <script>

- List of code points which are variants because Tamil and <script> scripts are closely related to each other:

Code Point 1 + Glyph 1	Code Point 2 + Glyph 2	Type (blocked)

# Whole Label Evaluation Rules (in plain English) - 1

---

- ⊙ C → Consonant
- ⊙ M → Matra
- ⊙ V → Vowel
- ⊙ X → Visarga / Aytham
- ⊙ H → Halant / Virama / Pulli

# Whole Label Evaluation Rules (in plain English)

---

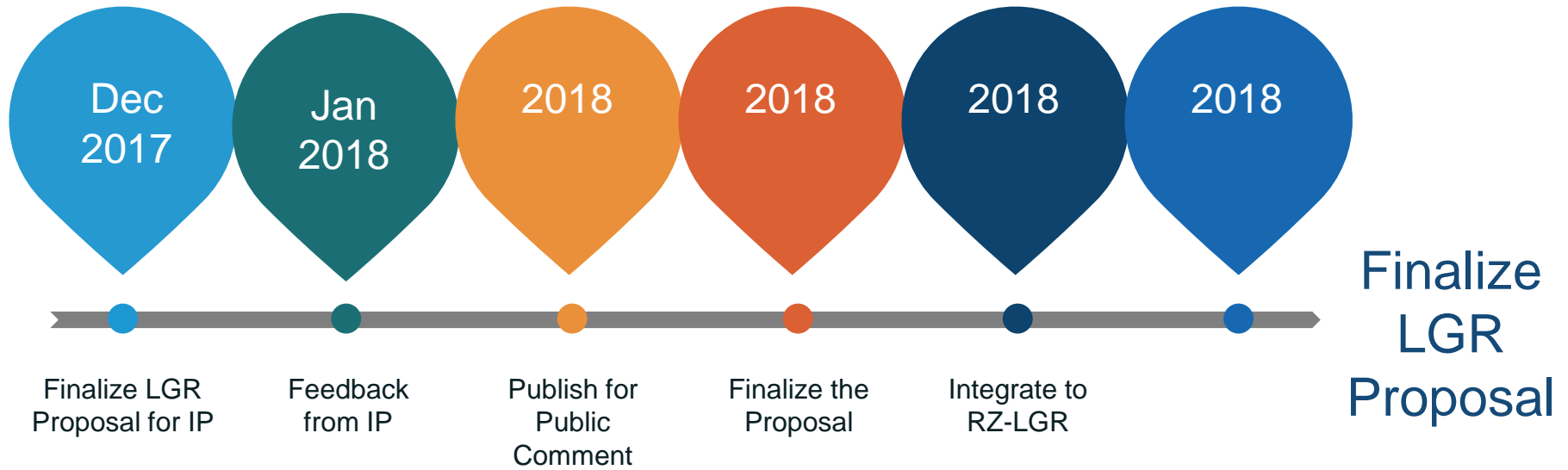
- ⦿ H: must be preceded by C
- ⦿ M: must be preceded by C
- ⦿ X: must be preceded by either of V, C, or M



# Current Status

Section in Google Doc	Status (To be initiated / In Progress / Completed / Reviewed)	Additional Notes
Description of Script and use		
Languages using Script		
Classification of characters		
Repertoire Analysis		
Within-script Variants		
Cross-script Variants		Devanagari
WLE Rules		
References		

# Plan Till Completion



## To Summarize

# Thank You and Question

---

## Contributors:

Name	Address	Email
Dr. Shanmugam R.		shanmugamr@cdac.in