Proposal for a Tamil Script Root Zone Label Generation Rule-Set (LGR)

LGR Version: 3.0

Date: 2019-03-04

Document version: 2.11

Authors: Neo-Brahmi Generation Panel [NBGP]

1 General Information/ Overview/ Abstract

This document lays down the Label Generation Rule Set for the Tamil script. The three main components of the Tamil Script LGR, Code point repertoire, Variants, and Whole Label Evaluation Rules have been described in detail here. These components have been incorporated in a machine-readable format in the accompanying XML file named "Proposal-LGR-Taml-20190304.xml".

In addition, a document named “Taml_Test_Labels_20190304.txt” has been provided. It provides a list of valid and invalid labels as per the Whole Label Evaluation laid down in Section 7 of this document. In addition, a set of labels which can produce variant labels is laid down in Section 6 of this document. The labels have been tagged as valid and invalid under the specific rules[1].

2 Script for which the LGR is proposed

ISO 15924 Code: Taml

ISO 15924 Key N°: 346

[1] The categorization of invalid labels under specific rules is given as per the general understanding of the LGR Tool used by the NBGP. During testing with a specific LGR tool, whether a particular label gets flagged under the same rule or the different one may depend on the order of evaluation and therefore on the internal implementation of the LGR Tool. In case of discrepancy, only the fact that it is an invalid label should be considered.

ISO 15924 English Name: Tamil

Latin transliteration of native script name: tamil̲

Native name of the script: தமிழ்

Maximal Starting Repertoire [MSR] version: 3

3 Background on Script and Principal Languages Using It

Tamil is one of the oldest Dravidian languages which has a continuous history since the age of tolkāppiyam. The earliest known inscriptions in Tamil date back to 2,200 BC. Tamil literature emerged in around 300 BC, and the language used from then until the 700 AD is known as Old Tamil. From 700-1600 AD the language is known as Middle Tamil, and since 1600 the language has been known as Modern Tamil. Tamil is mainly spoken in the southern part of India, known as Tamilnadu. It is also spoken in Pondycherry, Andaman and Nicobar islands and other states of India. It is one the official languages of Sri Lanka and Singapore. A Tamil-speaking community is found in countries such as Malaysia, Mauritius, South Africa, Myanmar, the UK, Canada, the USA, France and Réunion.

3.1 The Evolution of the Script

Tamil was originally written with a version of the Brahmi script known as Tamil Brahmi, and from 3rd century to 10th century AD this script had become more rounded and developed into the *vaṭṭel̲uttu* [1004] script. Over time the script has changed somewhat, and it was simplified in the 19th and 20th centuries. The image below shows how Brahmi transformed as *vaṭṭel̲uttu* and Tamil letters[2].

[2] https://ta.wikipedia.org/s/jt1

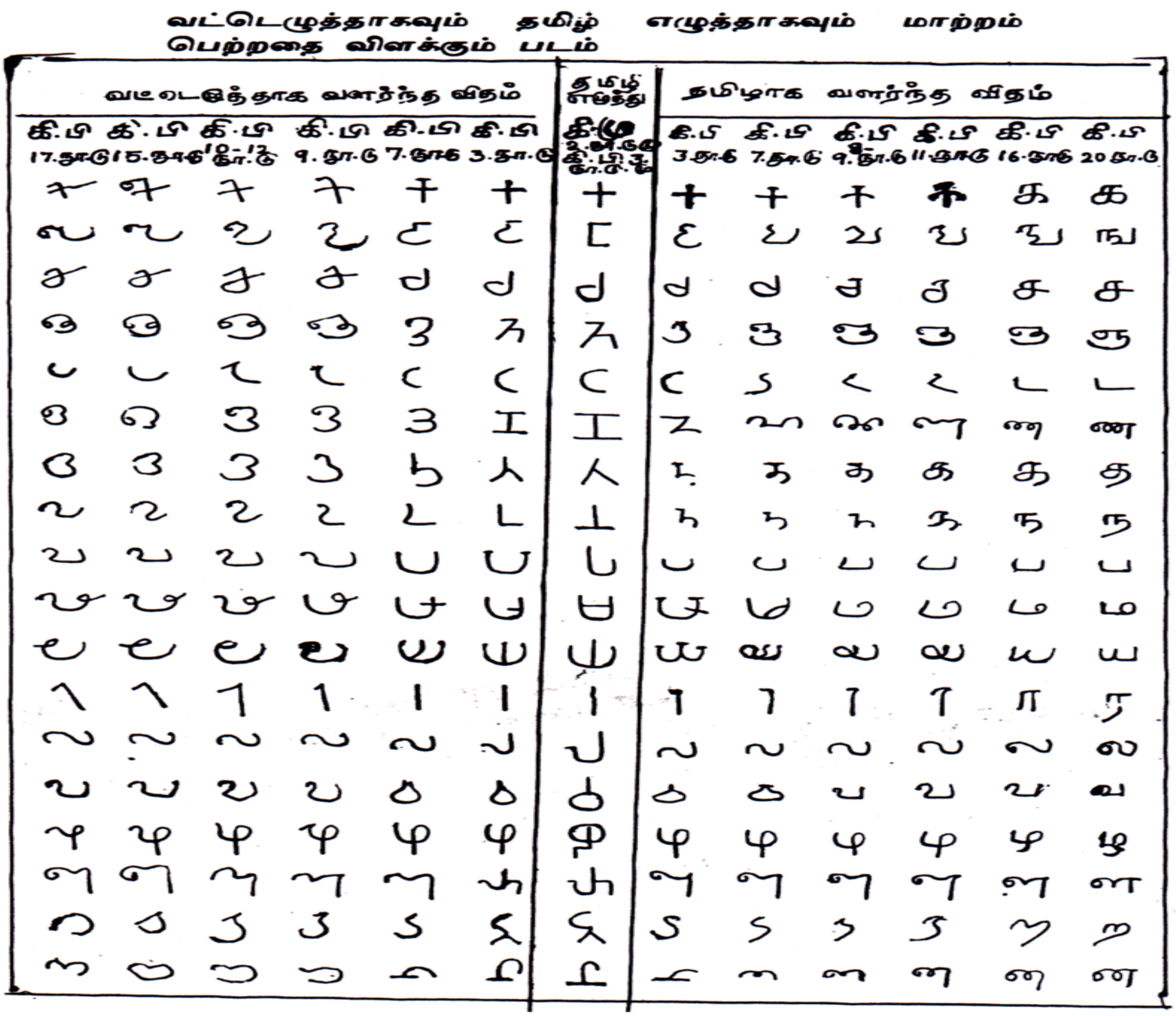

வட்டெழுத்தாகவும் தமிழ் எழுத்தாகவும் மாற்றம் பெற்றதை விளக்கும் படம்
வட்டெழுத்தாக வளர்ந்த விதம்
தமிழ் எழுத்து
தமிழாக வளர்ந்த விதம்

Figure 1: *vaṭṭeḻuttu and Tamil letters transformation of Brahmi*

The central column of the above image indicates (oldest) Tamil Brahmi characters, diverging to *vaṭṭeḻuttu* towards left, and to Tamil towards the right. Tamil is also written with a version of the Arabic script known as Arwi by Tamil-speaking Muslims.

3.2 Languages considered

The Tamil script is mainly used to write the Tamil Language. However, there are some tribal languages such as Badaga, Irula, Kurumba Betta, Kurumba Kannada, Paniya, and Saurashtra, which also use the Tamil script; but since the EGIDS [EGIDS] value of those languages is above four they have not been considered in the present analysis.

EGIDS Scale 1	EGIDS Scale 2	EGIDS Scale 3	EGIDS Scale 4
Tamil (Sri Lanka, Singapore)	Tamil (India)		Tamil (Malaysia)

Table 1~~1~~: Languages considered under Tamil LGR

3.3 The structure of written Tamil

The Tamil script is an alphasyllabary and the heart of the writing system is the *Akshar*. It is this unit, which is instinctively recognized by users of the script. To understand the notion of Akshar, a brief overview of the writing system is provided in this Section and the Akshar itself will be treated in depth in Section 5.4.

The writing system of Tamil could be summed up as composed of the following:

3.3.1 The Consonants

As per traditional grammar classification, Tamil consonants have been categorized in three groups according to their phonetic properties (especially in terms of place and manner of articulation with voiced and voiceless nature). They are Stops (vallin̲am), Medial (iṭaiyin̲am) and Nasal (mellin̲am). Tamil also has five Grantha consonants. It should also be noted that as per Tamil traditional grammar, "Tamil Consonant" is ideally a combination of consonants (as defined in Unicode) + Virama combination. E.g. க் (TAMIL LETTER KA + TAMIL SIGN VIRAMA) is actually a consonant in Tamil grammar. On the other hand, what Unicode designates as consonant is termed as Vowel-Consonant in Tamil Traditional grammar. However, for the sake of uniformity across all the LGRs under NBGP the Unicode naming convention has been followed.

The Unicode Consonant set of Tamil comprises the following characters:

STOP	க TAMIL LETTER KA (U+0B95)	ச TAMIL LETTER CA (U+0B9A)	ட TAMIL LETTER TTA (U+0B9F)	த TAMIL LETTER TA (U+0BA4)	ப TAMIL LETTER PA (U+0BAA)	ற TAMIL LETTER RRA (U+0BB1)
NASAL	ங TAMIL LETTER NGA (U+0B99)	ஞ TAMIL LETTER NYA (U+0B9E)	ண TAMIL LETTER NNA (U+0BA3)	ந TAMIL LETTER NA (U+0BA8)	ம TAMIL LETTER MA (U+0BAE)	ன TAMIL LETTER NNNA (U+0BA9)
MEDIAL	ய TAMIL LETTER YA (U+0BAF)	ர TAMIL LETTER RA (U+0BB0)	ல TAMIL LETTER LA (U+0BB2)	வ TAMIL LETTER VA (U+0BB5)	ழ TAMIL LETTER LLLA (U+0BB4)	ள TAMIL LETTER LLA (U+0BB3)
GRANTHA	ஸ TAMIL LETTER SA (U+0BB8)	ஷ TAMIL LETTER SSA (U+0BB7)	ஜ TAMIL LETTER JA (U+0B9C)	ஹ TAMIL LETTER HA (U+0BB9)	ஶ TAMIL LETTER SHA (U+0BB6)	

Table 22: Group classification of consonants

The IPA of Tamil Consonants is as follows:

	Bilabial	Lab-Dental	Dental	Alv	Post-Alv	Retroflex	Palatal	Velar	Uvu	Glottal
Plosive	p (ப) b (ப)		t̪ (த) d̪ (த)			ʈ (ட) ɖ (ட)		k (க) g (க)		
Nasal	m (ம)		n̪ (ந)	n(ன)		ɳ (ண)	ɲ (ஞ)	ŋ (ங)		
Tap/Flap				ɾ̪ (ர)						
Trill				r (ற)						
Fricative				s (ச)						ɦ (க)
Approx		ʋ (வ)				ɻ (ழ)	j(ய)			

Lat Approx				l̲(ல)		ɭ(ள)				
Affricate							tʃ (ச) dʒ (ஜ)			

Table 3~~3~~: IPA classification of Tamil consonants

3.3.2 Virama[3]/Pulli

All consonants contain an implicit vowel (a) within them. A special sign is needed to denote that this implicit vowel is stripped off. This is known as the virama "்" (U+0BCD). The virama thus joins two adjacent consonants. In Tamil, unlike other scripts under Neo-Brahmi GP, there are only three instances where this results in the formation of conjunct. Example 1 shows the conjuncts and Example 2 shows the non-formation of conjunct.

Example 1

க் + ஷ	TAMIL LETTER KA TAMIL SIGN VIRAMA+ TAMIL LETTER SSA	க்ஷ
ஸ் + ரீ	TAMIL LETTER SA TAMIL SIGN VIRAMA+ TAMIL LETTER RA TAMIL VOWEL SIGN II	ஸ்ரீ
ஶ் + ரீ	TAMIL LETTER SHA TAMIL SIGN VIRAMA+ TAMIL LETTER RA TAMIL VOWEL SIGN II	ஶ்ரீ

Example 2

க் + க	TAMIL LETTER KA TAMIL SIGN VIRAMA+ TAMIL LETTER KA	க்க

3.3.3 Vowels

Separate symbols exist for all vowels that are pronounced independently either at the beginning or after a vowel sound. To indicate a vowel sound other than the implicit one, a

[3] Unicode (cf. Unicode 3.0 and above) prefers the term Virama. In this report both the terms have been used to denote the character that suppresses the inherent vowel.

vowel sign (Matra) is attached to the consonant. Since the consonant has a built-in ‘a’ sound, there are equivalent Matras for all vowels except the அ (VOWEL LETTER A).

The correlation is shown in the table below:

Vowel	Corresponding vowel sign (Matra)
அ U+0B85	
ஆ U+0B86	◌ா U+0BBE
இ U+0B87	◌ி U+0BBF
ஈ U+0B88	◌ீ U+0BC0
உ U+0B89	◌ு U+0BC1
ஊ U+0B8A	◌ூ U+0BC2
எ U+0B8E	ெ◌ U+0BC6
ஏ U+0B8F	ே◌ U+0BC7
ஐ U+0B90	ை◌ U+0BC8
ஒ U+0B92	ெ◌ா U+0BCA
ஓ U+0B93	ே◌ா U+0BCB
ஔ U+0B94	ெ◌ள U+0BCC

Table 44: Vowels with corresponding Matras

3.3.4 Visarga / Aytham (ஃ - U+ 0B83)

The Visarga is also used in Tamil and represents a sound very close to /x/.

As per Tamil grammar, a Visarga must always be preceded by a short vowel and followed by a stop consonant e.g. அஃறிணை (Non-human) /akṟiṇai/ (U+0B85 U+0B83 U+0BB1 U+0BBF U+0BA3 U+0BC8).[4] Hence, in Tamil grammar Visarga + Visarga combination is not allowed.

In modern Tamil, Visarga is also used to represent some foreign sounds by combining it with certain consonants e.g., Fa is generated using Pa, as shown in word ஃபாரின் *(Foreign)* /fawr-in/(U+0B83 U+0BAA U+0BBE U+0BB0 U+0BBF U+0BA9 U+0BCD) Za is generated using Ja, as shown in word ஃஜிராக்ஸ் *(Xerox).* /zeer-oks/ (U+0B83 U+0B9C U+0BBF U+0BB0 U+0BBE U+0B95 U+0BCD U+0BB8 U+0BCD)

These combinations are originally borrowed from “arwi” which is an Arabic Tamil language coined by Tamil speaking Muslims. To facilitate this modern usage apart from barring Visarga – Visarga combination, the above-mentioned rules have not been strictly enforced in the WLE section.

4 Overall Development Process and Methodology

Under the Neo-Brahmi Generation Panel, there are many different scripts belonging to separate Unicode blocks. Each of these scripts will be assigned a separate LGR; however, the Neo-Brahmi GP ensures that the fundamental philosophy behind building those LGRs are all in sync with all other Brahmi derived scripts.

The Tamil script LGR proposal was published for public comment to allow those who had not participated in the NBGP to make their views known. The NBGP analyzed all comments received to finalize the proposal. The analysis of public comments can be accessed online given at [1005].

[4] Appendix C: An image of Visarga rule with its translation

4.1 Guiding Principles

The NBGP adopts the following broad principles for selection of code-points in the code-point repertoire across the board for all the scripts within its ambit.

4.1.1 Inclusion principles:

4.1.1.1 Modern usage:

Every character proposed should be in the everyday usage of a particular linguistic community. Characters which have been encoded in Unicode for transcription or archival purposes only will not be considered for inclusion in the code point repertoire.

4.1.1.2 Unambiguous use:

Every character proposed should have an unambiguous understanding among the linguistic community about its usage in the language. However MSR has already restricted these characters.

4.1.2 Exclusion principles:

The main exclusion principle is that of External Limits on Scope. These comprise of protocols or standards which are pre-requisites to the Label Generation Rulesets. All further principles are in fact subsumed under these limitations but have been spelt out separately for the sake of clarity.

4.1.2.1 External Limits on Scope:

The code point repertoire for root zone being a very special case, up the ladder in the protocol hierarchies, the canvas of available characters for selection as a part of the Root Zone code point repertoire is already constrained by various protocol layers beneath it. Following three main protocols/standards act as successive filters:

i. The Unicode Chart:

Out of all the characters that are needed by the given script, if the character in question is not encoded in Unicode, it cannot be incorporated in the code point repertoire. Such cases are quite rare, given the elaborate and exhaustive character inclusion efforts made by Unicode consortium.

ii. IDNA Protocol:

Unicode, being the character encoding standard for providing the maximum possible representation of a given script/language, has encoded as far as possible all the possible characters needed by the script. However, domain names, being a specialized case, are governed by an additional protocol known as IDNA (Internationalized Domain Names in Applications). The IDNA protocol excludes some characters of the Unicode repertoire from being part of domain names.

Example: TAMIL NUMBER TEN "௰" (U+0BF0) is not allowed to be a part of domain name.

iii. Maximal Starting Repertoire:

The Root-zone LGR being a repertoire of the characters which are going to be used for creation of the root zone TLDs, which in turn are an even more specialized case of domain names, the ROOT LGR procedure introduces additional exclusions on IDNA allowed set of characters.

Examples: TAMIL OM "ௐ" (U+0BD0) and TAMIL SIGN ANUSVARA (U+0B82), even if allowed by IDNA protocol, are not permitted in the Root Zone Repertoire as per the [MSR].

To sum up, the restrictions start off with admitting only such characters as are part of the code-block of the given script/language. This is further narrowed down by the IDNA Protocol and finally an additional filter in the form of Maximal Starting Repertoire restricts the character set associated with the given script even more.

4.1.2.2 No Punctuation Marks:

The TLDs being identifiers, punctuation markers present in Brahmi based languages such as Danda "।" (U+0964) and double Danda "॥" (U+0965) will not be included.

4.1.2.3 No Symbols and Abbreviations:

Abbreviations, weights and measures and other such characters like Tamil Debit Sign "௶ " (U+0BF6) etc. will not be included.

4.1.2.4 No Rare and Obsolete Characters:

AU LENGTH MARK "◌ௗ" (U+0BD7) is a character in Tamil which has been added to Unicode for technical reasons and is not used in Tamil. As it is not used by the language community the same character will not be included in the proposed repertoire. This is in compliance with the Conservatism principle as laid down in the Root Zone LGR procedure.

4.1.2.5 No Stress Markers of Classical Sanskrit and Vedic:

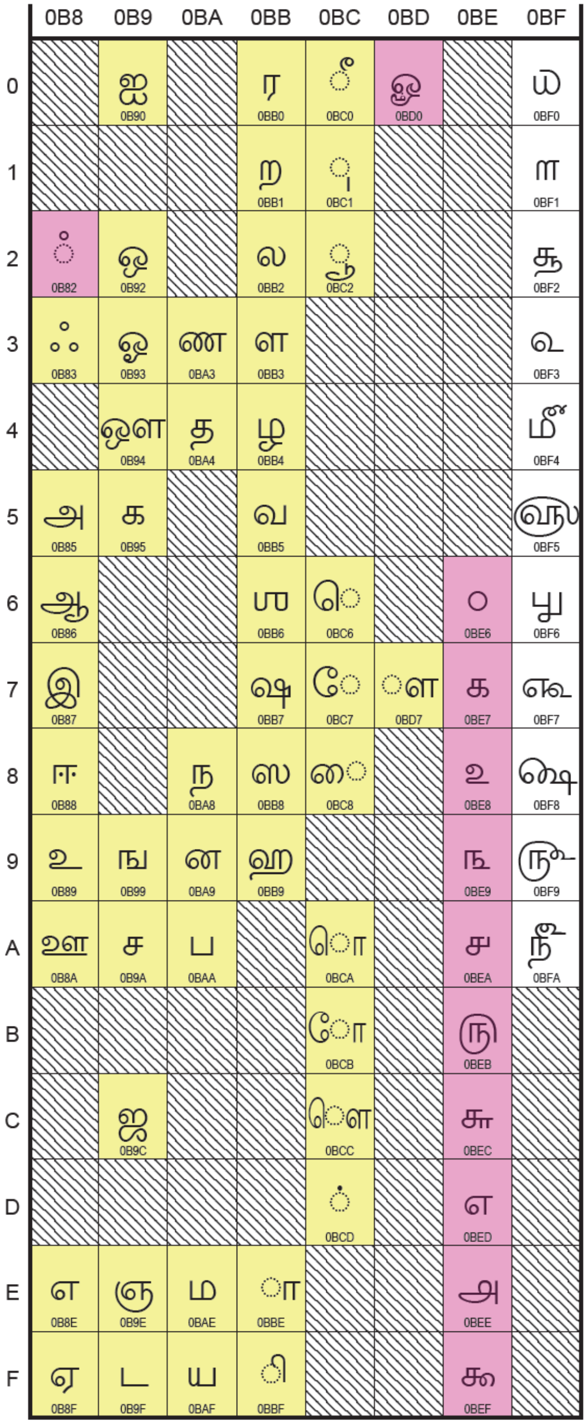
Stress markers for classical Sanskrit e.g. DEVANAGARI STRESS SIGN UDATTA "◌॑" (U+0951) and DEVANAGARI STRESS SIGN ANUDATTA "◌॒" (U+0952) will not be included. Since Tamil has no stress, there are no such cases found in Tamil. This is also in compliance with the Letter principle as laid down in the Root Zone LGR Procedure.

5 Repertoire

Section 5.1 shows the section of the [MSR] applicable to the Tamil script on which the Tamil code-point repertoire is based.

Section 5.2 details the code-point repertoire that the Neo-Brahmi Generation Panel [NBGP] proposes to be included in the Tamil LGR.

5.1 Tamil section of Maximal Starting Repertoire [MSR] Version 4



	0B8	0B9	0BA	0BB	0BC	0BD	0BE	0BF
0		ஐ 0B90		ர 0BB0	◌ீ 0BC0	ௐ 0BD0		௰ 0BF0
1				ற 0BB1	◌ு 0BC1			௱ 0BF1
2	◌ஂ 0B82	ஒ 0B92		ல 0BB2	◌ூ 0BC2			௲ 0BF2
3	ஃ 0B83	ஓ 0B93	ண 0BA3	ள 0BB3				௳ 0BF3
4		ஔ 0B94	த 0BA4	ழ 0BB4				௴ 0BF4
5	அ 0B85	க 0B95		வ 0BB5				௵ 0BF5
6	ஆ 0B86			ஶ 0BB6	◌ெ 0BC6		௦ 0BE6	௶ 0BF6
7	இ 0B87			ஷ 0BB7	◌ே 0BC7	◌ௗ 0BD7	௧ 0BE7	௷ 0BF7
8	ஈ 0B88		ந 0BA8	ஸ 0BB8	◌ை 0BC8		௨ 0BE8	௸ 0BF8
9	உ 0B89	ங 0B99	ன 0BA9	ஹ 0BB9			௩ 0BE9	௹ 0BF9
A	ஊ 0B8A	ச 0B9A	ப 0BAA		◌ொ 0BCA		௪ 0BEA	௺ 0BFA
B					◌ோ 0BCB		௫ 0BEB	
C		ஜ 0B9C			◌ௌ 0BCC		௬ 0BEC	
D					◌் 0BCD		௭ 0BED	
E	எ 0B8E	ஞ 0B9E	ம 0BAE	◌ா 0BBE			௮ 0BEE	
F	ஏ 0B8F	ட 0B9F	ய 0BAF	◌ி 0BBF			௯ 0BEF	

Color convention[5]**:**

All characters that are included in the [MSR] - Yellow background

PVALID in IDNA2008 but excluded from the [MSR] - Pinkish background

Not PVALID in IDNA2008 - White background

Figure 2: Tamil Code from [MSR 4]

[5] This document needs to be printed or viewed in colour for this to be read correctly.

5.2 Code Point Repertoire:

For each of the code points, language references have been given in the last column titled "Reference". The examples chosen for referencing, together cover all the code-points required for Tamil Language that NBGP has considered as given in 3.2.

Sr. No.	Unicode Code Point	Glyph	Character Name	category	Example language(s) using the code-point (Not exhaustive list)	Language with lowest EGIDS scale using the code point	Reference
1	0B83	ஃ	TAMIL SIGN VISARGA	Visarga	Tamil	Tamil	[1003]
2	0B85	அ	TAMIL LETTER A	Vowel	Tamil	Tamil	[1001]
3	0B86	ஆ	TAMIL LETTER AA	Vowel	Tamil	Tamil	[1001]
4	0B87	இ	TAMIL LETTER I	Vowel	Tamil	Tamil	[1001]
5	0B88	ஈ	TAMIL LETTER II	Vowel	Tamil	Tamil	[1001]
6	0B89	உ	TAMIL LETTER U	Vowel	Tamil	Tamil	[1001]
7	0B8A	ஊ	TAMIL LETTER UU	Vowel	Tamil	Tamil	[1001]
8	0B8E	எ	TAMIL LETTER E	Vowel	Tamil	Tamil	[1001]
9	0B8F	ஏ	TAMIL LETTER EE	Vowel	Tamil	Tamil	[1001]
10	0B90	ஐ	TAMIL LETTER AI	Vowel	Tamil	Tamil	[1001]
11	0B92	ஒ	TAMIL LETTER O	Vowel	Tamil	Tamil	[1001]
12	0B93	ஓ	TAMIL LETTER OO	Vowel	Tamil	Tamil	[1001]

13	0B94	ஔ	TAMIL LETTER AU	Vowel	Tamil	Tamil	[1001]
14	0B95	க	TAMIL LETTER KA	Consonant	Tamil	Tamil	[1002]
15	0B99	ங	TAMIL LETTER NGA	Consonant	Tamil	Tamil	[1002]
16	0B9A	ச	TAMIL LETTER CA	Consonant	Tamil	Tamil	[1002]
17	0B9C	ஜ	TAMIL LETTER JA	Consonant	Tamil	Tamil	[1002]
18	0B9E	ஞ	TAMIL LETTER NYA	Consonant	Tamil	Tamil	[1002]
19	0B9F	ட	TAMIL LETTER TTA	Consonant	Tamil	Tamil	[1002]
20	0BA3	ண	TAMIL LETTER NNA	Consonant	Tamil	Tamil	[1002]
21	0BA4	த	TAMIL LETTER TA	Consonant	Tamil	Tamil	[1002]
22	0BA8	ந	TAMIL LETTER NA	Consonant	Tamil	Tamil	[1002]
23	0BA9	ன	TAMIL LETTER NNNA	Consonant	Tamil	Tamil	[1002]
24	0BAA	ப	TAMIL LETTER PA	Consonant	Tamil	Tamil	[1002]
25	0BAE	ம	TAMIL LETTER MA	Consonant	Tamil	Tamil	[1002]
26	0BAF	ய	TAMIL LETTER YA	Consonant	Tamil	Tamil	[1002]
27	0BB0	ர	TAMIL LETTER RA	Consonant	Tamil	Tamil	[1002]
28	0BB1	ற	TAMIL LETTER RRA	Consonant	Tamil	Tamil	[1002]
29	0BB2	ல	TAMIL LETTER LA	Consonant	Tamil	Tamil	[1002]
30	0BB3	ள	TAMIL LETTER LLA	Consonant	Tamil	Tamil	[1002]

31	0BB4	ழ	TAMIL LETTER LLLA	Consonant	Tamil	Tamil	[1002]
32	0BB5	வ	TAMIL LETTER VA	Consonant	Tamil	Tamil	[1002]
33	0BB6	ஶ	TAMIL LETTER SHA	Consonant	Tamil	Tamil	[1002]
34	0BB7	ஷ	TAMIL LETTER SSA	Consonant	Tamil	Tamil	[1002]
35	0BB8	ஸ	TAMIL LETTER SA	Consonant	Tamil	Tamil	[1002]
36	0BB9	ஹ	TAMIL LETTER HA	Consonant	Tamil	Tamil	[1002]
37	0BBE	◌ா	TAMIL VOWEL SIGN AA	Matra	Tamil	Tamil	[1002]
38	0BBF	◌ி	TAMIL VOWEL SIGN I	Matra	Tamil	Tamil	[1002]
39	0BC0	◌ீ	TAMIL VOWEL SIGN II	Matra	Tamil	Tamil	[1002]
40	0BC1	◌ு	TAMIL VOWEL SIGN U	Matra	Tamil	Tamil	[1002]
41	0BC2	◌ூ	TAMIL VOWEL SIGN UU	Matra	Tamil	Tamil	[1002]
42	0BC6	ெ◌	TAMIL VOWEL SIGN E	Matra	Tamil	Tamil	[1002]
43	0BC7	ே◌	TAMIL VOWEL SIGN EE	Matra	Tamil	Tamil	[1002]
44	0BC8	ை◌	TAMIL VOWEL SIGN AI	Matra	Tamil	Tamil	[1002]
45	0BCA	ெ◌ா	TAMIL VOWEL SIGN O	Matra	Tamil	Tamil	[1002]
46	0BCB	ே◌ா	TAMIL VOWEL SIGN OO	Matra	Tamil	Tamil	[1002]

47	0BCC	◌ௌ	TAMIL VOWEL SIGN AU	Matra	Tamil	Tamil	[1002]
48	0BCD	◌்	TAMIL SIGN VIRAMA	Matra	Tamil	Tamil	[1002]

Table 5~~5~~: Code point repertoire

5.2.1 Code Point Sequence:

The following sequences have been defined for the purpose of variant. (see Section 6.1.3)

1.	U+0BB6 U+0BCD U+0BB0 = U+0BC0	ஶ ◌் ர ◌ீ [ஶ்ரீ]	TAMIL LETTER SHA TAMIL SIGN VIRAMA TAMIL LETTER RA TAMIL VOWEL SIGN II
2.	U+0BB8 U+0BCD U+0BB0 = U+0BC0	ஸ ◌் ர ◌ீ [ஸ்ரீ]	TAMIL LETTER SA TAMIL SIGN VIRAMA TAMIL LETTER RA TAMIL VOWEL SIGN II

Table 6~~6~~a: Code point sequence

5.2.2 Code Point variants pair 1

The following variants pair have been defined for the purpose of variant. (see Section 6.1.1 and 6.1.2)

1.	U+0B94	ஔ	TAMIL LETTER AU
2.	U+0B92 U+0BB3	ஒள	TAMIL LETTER O + TAMIL LETTER LLA

5.2.3 Code Point variants pair 2

1.	U+0BCC	◌ௌ	TAMIL VOWEL SIGN AU
2.	U+0BC6 U+0BB3	◌ெள	TAMIL VOWEL SIGN E +TAMIL LETTER LLA

5.3 Code points not included:

The following code points have not been included in the repertoire.

Sr. No.	Unicode Code Point	Glyph	Character Name	Reason for exclusion
1.	U+0BD7	◌ௗ	TAMIL AU LENGTH MARK	Not in modern usage. Excluded as per conservatism principle.

Table 7~~7~~: Code points not included

5.4 Structural Formation of Tamil:

All the languages written in any Brahmi-derived scripts follow a particular way of formation of their words, known as Akshar. In the next section, there are detailed Akshar formation rules applicable to the representation of the Tamil language when written in the Tamil script.

5.5 Akshar formation rules for Tamil:

This section details the Akshar formation rules as applicable to Tamil. The first section lists the categories of the characters in the form of variables. In the rules, instead of their full descriptive names, abbreviated variable names are used. The second section lists four operators along with their functions which are assumed while specifying the rules. The final two sections describe the formation of the two major categories of Akshar; the first of which begins with the vowels and the second one with the consonants.

5.5.1 Variables involved

Dash → Hyphen -
Digit → Indo-Arabic digits [0-9]
C → Consonant
M → Matra
V → Vowel
X → Visarga / Aytham
H → Virama / Pulli

5.5.2 Operators used:

Symbol	Function
|	Alternative
[ ]	Optional
*	Variable Repetition
( )	Sequence Group

Table 88: Symbol functions

In what follows, the vowel sequence and the consonant sequence pertinent to Tamil, when used to write Tamil, are given.

5.5.3 Vowel Sequence

A vowel sequence begins with a vowel. It may be optionally followed by a Visarga (X). The number of X which can follow a V in Tamil is restricted to one.

The vowel sequence in Tamil is therefore V [X]

Examples:

Sequence Description	Sequence	Example	Constituting characters
Vowel	V	அ /a/ U+0B85	
Vowel + Visarga	V[X]	அஃ /ak̠/ U+0B85 U+0B83	அ ஃ U+0B85 U+0B83

Table 9~~9~~: Vowel sequence

5.5.4 Consonant Sequence

A consonant sequence begins with a consonant. It may be optionally followed by a Matra (M), Visarga (X) or a Virama/Pulli (H). The number of instances of these characters occurring after a consonant is restricted to one. There is a possibility of further extension of the consonant sequence after the M and H. Each of these has been discussed in the following sections:

1. A single consonant (C)

Examples:

Sequence Description	Sequence	Example	Constituting characters
Consonant	C	க /ka/ U+0B95	<single character>

Table 10~~10~~: Single consonant sequence

2. A consonant optionally followed by dependent vowel sign/Matra [M], Visarga [X] or Virama/Pulli [H]

C [M| H|X]

Examples:

Sequence Description	Sequence	Example	Constituting characters
Consonant + Matra	C[M]	கி /ki/	க ◌ி 0B95 0BBF

Consonant + Virama/Pulli	C[H]	க் /k/ (Pure Consonant)	க ◌் U+0B95 U+0BCD
Consonant + Visarga	C[X]	கஃ / kk̠ /	க ஃ U+0B95 U+0B83

Table 11~~11~~: Consonant sequences with Matra, Visarga or Virama

2.A. A CM sequence can be optionally followed by X

(CM)[X]

Example:

Sequence Description	**Sequence**	**Example**	**Constituting characters**
Consonant + Matra + Visarga	CM[X]	முஃ /muk/	ம ◌ு ஃ U+0BAE U+0BC1 U+0B83

Table 12~~12~~: Consonant sequence with Matra and Visarga

3. A sequence of consonants (up to 3) joined by Virama/Pulli *2(CH)C

Example:

Sequence Description	**Sequence**	**Example**	**Constituting characters**
Consonant + Virama/Pulli + Consonant + Virama/Pulli + Consonant	CHCHC	ழ்த்த /l̠t̠t̠a/	ழ ◌் த ◌் த U+0BB4 U+0BCD U+0BA4 U+0BCD U+0BA4

Table 13~~13~~: Sequence with multiple consonants

Subsets:

3. A. The combination may be followed by M or X

Example:

Sequence Description	**Sequence**	**Example**	**Constituting characters**
Consonant + Virama/Pulli + Consonant + Matra	CHC[M]	க்கு /kku/	க ◌் க ◌ு U+0B95 U+0BCD U+0B95 U+0BC1

Consonant + Virama/Pulli + Consonant + Visarga	CHC[X]	க்கஃ /kka*k*/	க ் க ஃ U+0B95 U+0BCD U+0B95 U+0B83

Table 14~~14~~: Sequence with multiple consonants with Matra or Visarga

3. B. *3(CH)CM may be followed by an X

Example:

Sequence Description	**Sequence**	**Example**	**Constituting characters**
Consonant + Virama/Pulli + Consonant + Matra + Visarga	CHCM[X]	ம்முஃ /kkīḥ/	ம ் ம ு ஃ U+0BAE U+0BCD U+0BAE U+0BC1 U+0B83

Table 15~~15~~: Sequence with multiple consonants with Matra and Visarga

These are the basic Akshar rules on which the overall Tamil LGR is based. There are some additional finer aspects to these rules as one takes into account the digits, punctuations and special standalone characters like Avagraha. Those aspects are not discussed here as the [MSR] on which the LGRs are supposed to be based, excludes those characters. The usage of Visarga can be found in section 3.3.4

6 Variants

There are some characters/character sequences in Tamil that can be created by using the characters permitted as per the [MSR] and that look alike. The NBGP categorizes these confusingly similar characters in three groups:

- Group 1: Confusing due to being exact homoglyphs
- Group 2: Confusing due to partial similarity
- Group 3: Confusing due to similar appearance but actually not valid as per Akshar formation rules

6.1 Group 1: Confusing due to being exact homoglyphs

Cases which belong to Group 1 are proposed to be considered as variants. There are three such cases.

6.1.1 TAMIL LETTER AU with TAMIL LETTER O followed by TAMIL LETTER LLA:

This variant pair involves the pure vowel TAMIL LETTER AU (ஔ U+0B94) which looks exactly similar to the vowel + Consonant TAMIL LETTER O + TAMIL LETTER LLA (ஒள U+0B92 U+0BB3) combination. These two cases can cause confusion even to a careful observer and hence are being proposed as variants.

Variant 1	Variant 2
ஔ U+0B94	ஒள U+0B92 U+0BB3

Table 16~~16~~: Proposed Variants - Set 1

6.1.2 TAMIL VOWEL SIGN AU with TAMIL VOWEL SIGN E followed by TAMIL LETTER LLA:

This variant pair involves the split Matra TAMIL VOWEL SIGN AU (◌ௌ U+0BCC) having left and right side catenators which sit on the preceding consonant. It looks exactly alike to a combination of Matra TAMIL VOWEL SIGN E (ெ◌ U + 0BC6) followed by consonant TAMIL LETTER LLA (ள U+0BB3).

Variant 1	Variant 2
◌ௌ U+0BCC	ெ◌ள U+0BC6 U+0BB3

Table 17~~17~~: Proposed Variants - Set 2

However, it must be noted that the above variant pair needs a preceding consonant to make it a valid Akshar formation.

6.1.3 Alternate representation for Shri

This variant pair involves forming "Shri" ligature by inputting two different consonants. Prior to Unicode 4.1, the best mapping to represent the ligature Shri was to the sequence ஸ் + ரீ TAMIL LETTER SA TAMIL SIGN VIRAMA + TAMIL LETTER RA TAMIL VOWEL SIGN II <U+0BB8 U+0BCD+ U+0BB0 U+0BC0>. Unicode 4.1 in 2005 added the character U+0BB6 TAMIL LETTER SHA and as a consequence, the best mapping became TAMIL LETTER SHA TAMIL SIGN VIRAMA +TAMIL LETTER RA TAMIL VOWEL SIGN II <U+0BB6 U+0BCD U+0BB0 U+0BC0>.

Due to slow updates to implementations, both representations are widespread in existing text. Therefore, in the present situation, Unicode recommends treating both representations as equivalent sequences. All the Tamil fonts which support both the combinations, represent both the sequences in exactly similar form (glyph).

Thus these representations should be treated as allocatable variants of each other as they don't cause any semantic change of the labels and also the display of the labels would remain the same in both cases, they are being proposed as allocatable variants. A brief description of these variants is in Table 16 , Table 17 and Table 18.

This is also being a case of ~~c0-~~allocatable variant, as required by the Conservatism principle, it is being restricted by a Whole Label Evaluation rule which will try to minimize the cases of unwarranted labels. One such case is of mixing of both the instances of this variant in a single label. A rule to this effect has been introduced in the section 7. Whole Label Evaluation Rules (WLE).

Code Point Sequence 1	Code Point Sequence 2
ஶ ் ர ீ = ஶ்ரீ U+0BB6 U+0BCD U+0BB0 U+0BC0	ஸ ் ர ீ = ஸ்ரீ U+0BB8 U+0BCD U+0BB0 U+0BC0

Table 18~~18~~: Proposed Variants - Set 3

6.2 Group 2: Confusing due to partial similarity

This happens with the partial similarity of the characters appearance of TAMIL LETTER JA "ஜ" (U+0B9C) with TAMIL LETTER AI "ஐ" (U+0B90). However, no cases belonging to Group 2 are proposed, as there is another panel (String similarity assessment panel) entrusted to deal with such cases.

Code Point 1	Code Point 2
ஜ U+0B9C	ஐ U+0B90

Table 19~~19~~: Not Proposed as Variants - Set 1

6.3 Group 3: Confusing due to similar looking but actually not valid as per Akshar formation rules.

This happens with wrong formation of consonant followed by two continuous Matras. The TAMIL VOWEL SIGN O "ொ" (U+0BCA) looks exactly same as TAMIL VOWEL SIGN E "ெ" (U+0BC6) followed by TAMIL VOWEL SIGN AA "ா" (U+0BBE). However, as the formation is not valid as per Akshar formation rules, this case is not proposed as variant.

Code Point	Code Point Sequence
ொ (U+0BCA)	ெ ா (U+0BC6) (U+0BBE).

Table 20~~20~~: Not Proposed as Variants - Set 2

6.4 Cross script variants:

A cross-script variant label, also sometimes referred to as "Whole Label confusable", is the variant case where one label in one script can be composed in such a way that it can resemble an entire label in a different script. Tamil script has a set of possible cross-script variants only with the Malayalam script. Table 22: Proposed Cross-script variants

lists the variants that are proposed as cross-script variants between Tamil and Malayalam. It is to be noted that none of the combinations listed in Table 22: Proposed Cross-script variants

are termed to be equivalents of each other semantically or otherwise. They are only grouped based on possible visual confusability. Here are some of examples of variant labels.

Tamil label	Malayalam label
வமி U+0BB5 U+0BAE U+0BBF	ഖഥി U+0D16 U+0D25 U+0D3F
ஜெமி U+0B9C U+0BC6 U+0BAE U+0BBF	ജെഥി U+0D1C U+0D46 U+0D25 U+0D3F

Table 21~~21~~: Cross-script variant label examples

A label can be considered to have a cross-script variant label only if "all" the constituent characters/Aksharas have an equivalent confusable in the other script. If there is even one single character/Akshara which does not have an equivalent visual confusable in another script, it essentially provides a visual distinction and hence a non-confusable string.

The following table gives the set of proposed cross-script variants between Tamil and Malayalam.

Tamil	Malayalam
ஜ U+0B9C	ജ U+0D1C
வ U+0BB5	ഖ U+0D16
ம U+0BAE	ഥ U+0D25
◌ி U+0BBF	ി U+0D3F
ெ◌ U+0BC6	െ U+0D46
ே◌ U+0BC7	േ U+0D47

Table 22~~22~~: Proposed Cross-script variants

In addition to the above cases, Tamil and Malayalam scripts have a possible set of code points which look similar but not similar enough to be recommended as cross-script variants. They are listed in Table 22: Tamil and Malayalam Confusable Code Points based on pure visual similarity, in Appendix A.

6.5 Variant Disposition:

6.5.1 Blocked variant

Variants mentioned in Table 16 and Table 17 are cases of homoglyphs and hence it is proposed that these be "blocked" variants.

There is no preference among these variants. Whichever label containing either of these variants is chosen earlier, the other one equivalent variant label should be “blocked”.

6.5.2 Allocatable variants

The variant “Shri” described in section 6.1.3 is a case of variant where exactly same visual form is rendered with two distinct sequences. Also, in the minds of the user, regardless of which sequence they choose to input, both are intended to be the same Akshar i.e. “Shri”. Hence, it is imperative that both the sequences be treated as the same in terms of variant analysis and any label formed with either form should be made available to the same entity. This variant pair is thus being proposed as an “allocatable” variant.

7 Whole Label Evaluation Rules (WLE)

This section provides the WLE rules that are required by Tamil language mentioned in section 3.2 when written in Tamil script. The rules have been drafted in such a way that they can be easily translated into the LGR specification.

Below are the symbols used in the WLE rules, for each of the "Indic Syllabic Category" as mentioned in the Table 5: Code point repertoire.

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

Below are the specific WLE rules:

1. H: must be preceded by C
2. M: must be preceded by C
3. X: cannot be preceded by X

7.1 No mixing of instances of ~~co~~-allocatable variants within a single label:

As elaborated in section 6.1.3 Alternate representation for Shri says that the "Shri" can be written in the following two ways.

U+0BB6 U+0BCD U+0BB0 U+0BC0	ஶ ◌் ர ◌ீ = ஶ்ரீ
U+0BB8 U+0BCD U+0BB0 U+0BC0	ஸ ◌் ர ◌ீ = ஸ்ரீ

Table 23 two representations of Shri

As is evident from the above table, despite clear differences in the constituting code-points, the final ligatures assume the same shape, thereby making it a case of variant. Out of the two ways, there is no clear favorite among the user community and both the sequences are used by different set of user communities. This makes it necessary make it a case of allocatable variant as given in Alternate representation for Shri , However, one particular user does not use both the form in general, more so within the same label. Hence, it is being proposed that, within a single label, if it contains more than one instances of either of the instances of writing "Shri", they need to be the same. In case there is a label which contains more than one instances of "Shri" which are different from one another, that label will be termed as invalid. This is in consonance with the Conservatism Principle as laid down in the LGR Procedure. The below table shows the things in detail.

S.No	Sequences which cannot co-occur within a label	Character representation	Example
1.	U+0BB6 U+0BCD U+0BB0 U+0BC0	ஶ ◌் ர ◌ீ = ஶ்ரீ	ஶ்ரீலஷ்மிஸ்ரீ
	U+0BB8 U+0BCD U+0BB0 U+0BC0	ஸ ◌் ர ◌ீ = ஸ்ரீ	

Table 24 Sequences which cannot co-occur within a label

8 Contributors

NBGP Co-chairs: Dr. Uday Narayan Singh, Mr. Mahesh D Kulkarni and Dr. Ajay Data

Following is the full list of NBGP members with their Language expertise.

Name	**Language Expertise**
Udaya Narayana Singh	Bengali, Maithili, Hindi, English
Ajay Data	Hindi
Mahesh D. Kulkarni	Marathi, Hindi
Anupam Agrawal	Hindi, Bengali
Akshat S. Joshi	Hindi, Marathi
Abhijit Dutta	Bengali, Hindi
Neha Gupta	Hindi
Nishit Jain	Hindi
Prabhakar Pandey	Hindi
Raiomond Doctor	English, Hindi, Marathi, Gujarati

N. DeivaSundaram	Tamil
Shantaram S. Warde Walawalikar	Konkani
Bal Krishna Bal	Nepali
Ganesh Murmu	Santali
Balaram Prasain	Nepali
Rajib Chakraborty	Bangla (Bengali)
Gurpreet Singh Lehal	Panjabi
Saroja Bhate	Sanskrit
Shambhu Kumar Singh	Maithili
Swarna Prabha Chainary	Bodo
Ghanashyam Nepal	Nepali
Kalyan Vasudeo Kale	Marathi
Shashi Pathania	Dogri
Santhosh Thottingal	Malayalam, Sourashtra, Tamil
Uma Maheshwar G	Telugu
Girish Chandra Mishra	Odia
K. C. Tikayat ray	Odia
Debajit Sharma	Assamese
Basanta Kumar Panda	Odia
Arvind Bhandari	Gujarati
Harish Chowdhary	Hindi
Chitrita Chatterjee	Multiple languages represented by members of IAMAI
U.B. Pavanaja	Kannada
Hempal Shrestha	Nepali, Newari
Suraj Adhikari	Nepali

Gangadhar Panday	Telugu
Vinay Murarka	Hindi
Mukesh Saini	Hindi
Jay Paudyal	Hindi
Pawan Chitrakar	Nepali
Nirajan Parajuli	Nepali
Uttam Shrestha Rana	Nepali
Dev Dass Manandhar	Nepali, Newari
Bhim Dhoj Shrestha	Nepali, Newari
Rajiv Kumar	Hindi
Shubham Saran	Hindi
Anivar A. Aravind	Malayalam
Shanmugam R	Tamil
Prasad PK	Malayalam
Cinnathambi Shanmugaraja	Tamil
K. Sarweswaran	Tamil
S.Maniyam	Tamil

In addition, following members externally gave inputs to NBGP for the respective languages/scripts.

Name	**Language/Script Expertise**
Ajit Kumar	Awadhi, Braj Language
Basil Baa	Sadri Language
Basil Kiro	Kharia Language
Biswa Limbu	Limbu Language

Devendra Kumar Devesh	Bhojpuri Language
Dinbandhu Mahto	Panchpargania Language
Dr. Birendra Kumar Soy	Mundari Language
Dr. Dinesh Kumar Shrivastav	Magahi Language
Dr. Harvinder Kaur	Gurmukhi Script
Dr. Laxmi Prasad Khatiwada	Nepali Language
Jagannath Singh	Panchpargania Language
Narendra Kumar Negi	Kinnauri Language
Prateek Harshwal	Wagdi and Dhundhari Language
Rayem Olem Dungdung	Sadri Language
Tej Man Angdembe	Limbu Language

Full updated list of NBGP members is available at:

https://community.icann.org/display/croscomlgrprocedure/Neo-Brahmi+GP

9 References

[MSR] Integration Panel, "Maximal Starting Repertoire — MSR-4 Overview and Rationale", 25 Jan 2019 https://www.icann.org/sites/default/files/packages/lgr/msr/msr-4-wle-rules-25jan19-en.html

[EGIDS] Expanded Graded Intergenerational Disruption Scale,

https://www.ethnologue.com/about/language-status (Accessed on 13th Nov. 2017)

[NBGP] Neo-Brahmi Generation Panel
[gTLD] generic Top Level Domain

[1001] Omniglot, Tamil, http://www.omniglot.com/writing/tamil.htm (Accessed on 05th.July 2018)

[1002] Unicode 11.0.0, South and Central Asia-I, Page 488-493, R5 and R5a, https://www.unicode.org/versions/Unicode11.0.0/ch12.pdf (Accessed on 05th July. 2018)

[1003] Tamil, https://www.charbase.com/0b83-unicode-tamil-sign-visarga (Accessed on 27th Nov. 2017)

[1004] Title: *vaṭṭeḻuttu,* (Description and history of Tamil writing system *vaṭṭeḻuttu*) ,Tamil, https://ta.wikipedia.org/s/jt1 (Accessed on 28th Nov. 2018, Contents of this page are in Tamil)

[1005] Public comment feedback for Malayalam, Tamil Script LGR Proposals https://docs.google.com/document/d/1Am1qJXSYPpuUifcfUWT01uwCV-LCAe3XgBsnJvM5tHs/edit

10 Books, articles and webographies consulted

Following is a thematically sorted set of documents, books, articles and webographies consulted in the drafting of this report

1. Karunakaran K [2000], Simplified grammar of Tamil. Suvitha Publishers.
2. Kothandaraman Pon [1997]., A Grammar of contemporary Literary Tamil. International Institute of Tamil Studies.
3. Kothandaraman Pon [2001]., Tamil studies. Ambuli publications
4. Kothandaraman Pon [2002]., Ikkālat Tamiḻ ilakkaṇam. Pūmpoḻil publications
5. Meenakshi Sundaranar Te.Po [1965]., A History of Tamil Literature. Annamalai University
6. Vaiypuripillai [1988], Vaiyapuripillai's History of Tamil language and literature. New Century Book House
7. Varadharajan Mu. [1988], History of Tamil Literature. Sahitya Akademi.
8. Tamil Script Evolution http://www.virtualvinodh.com/wp/tamil-script-evolution/ (Accessed on 28th Nov. 2018)

11 Appendix A: Cross-script Confusable Code Point

As discussed earlier, Tamil script has a set of possible cross-script confusables with the Malayalam script and considered as variant code points. Table 21 lists them. In addition, the following code points could be considered similar but not variants of each other.

Tamil	Malayalam
ஸ U+0BB8	സ U+0D38

Table 22: Tamil and Malayalam Confusable Code Point

The following code points were discussed and the NBGP concluded that they are distinguishable

Tamil	Malayalam	Resolution
ய U+0BAF	ധ U+0D27	distinguishable
க U+0B95	ക U+0D15	distinguishable

Table 23: NBGP resolutions for Tamil and Malayalam

12 Appendix B: A NOTE ON ZERO WIDTH NON-JOINER

This note is pertinent to the use of Zero Width Non Joiner (ZWNJ) as used in Tamil. ZWJ (U+0200D) and ZWNJ (U+0200C) are code points that have been provided by the Unicode standard to instruct the rendering of a string where the script has the option between joining and non-joining characters. Without the use of these control codes, the string may be rendered in an alternate form from what is intended.

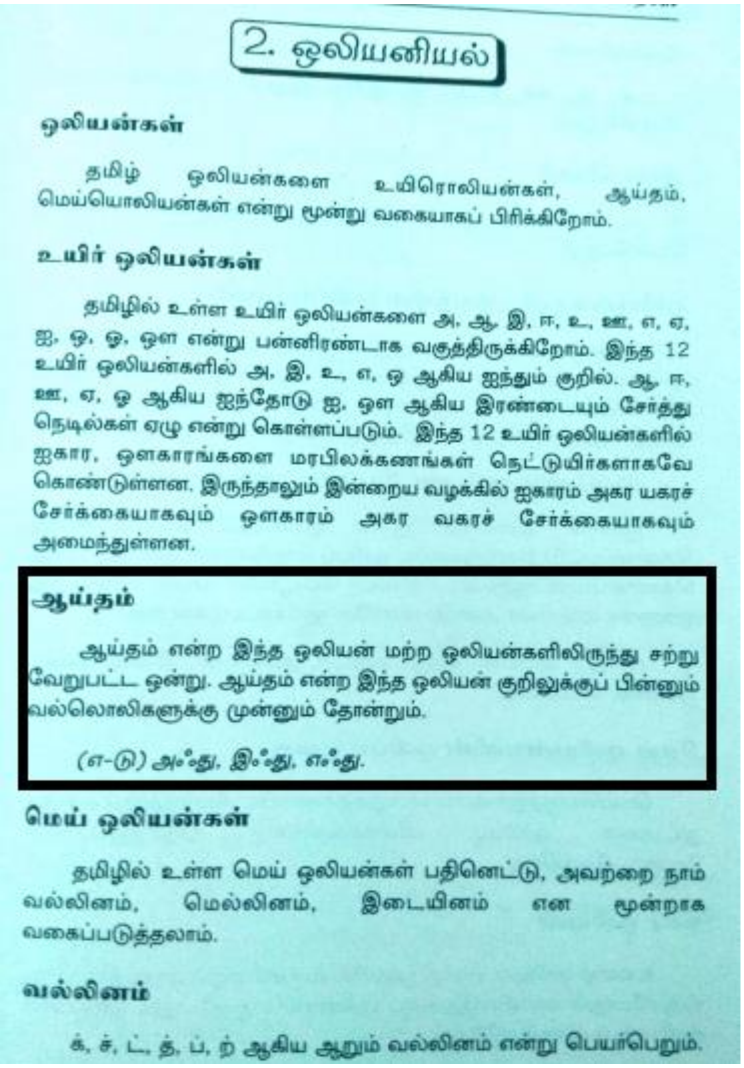
In the case of Tamil, ZWJ does not play an important role insofar as functionality is concerned. But ZWNJ plays a role in the following combinations for example: ஸ்ரீ/srɪ/(U+092C U+0922 U+093C), க்ஷ /kshə/(U+0B95 U+0BCD U+0BB7 U+0BAF).

The word "அக்‌ஷய்" /əkshəy/(U+0B85 U+0B95 U+0BCD U+200C U+0BB7 U+0BAF U+0BCD) can be written with the Unicode values:
U+0B85 U+0B95 U+0BCD U+200C U+0BB7 U+0BAF U+0BCD (அக்‌ஷய் with ZWNJ)
as well as U+0B85 U+0B95 U+0BCD U+0BB7 U+0BAF U+0BCD (அக்ஷய் without ZWNJ).

Insofar as Tamil is concerned ZWNJ is used to render alternate rendering of ligatures. The use of ZWNJ in Tamil is restricted to representing a dead consonant within a string. Thus to show the combination of க்+ஷ /k+shə/(U+0B95 U+0BCD U+0BB7) as a single word and retain the shape of the consonant followed by the Virama; ZWNJ is used. This practice is followed to represent Sanskrit loan words or proper names demanding a "dead" consonant. As ZWNJ is not part of the MSR, representing the above words in the specific forms would not be possible.

13 Appendix C: An image of Visarga rule with its translation

An attached image is a first page of Chapter 2 from Dr. Ponkothandaraman’s book titled “Ikkālat Tamiḻ ilakkaṇam” (Contemporary Tamil grammar).



2. ஒலியனியல்

ஒலியன்கள்

தமிழ் ஒலியன்களை உயிரொலியன்கள், ஆய்தம், மெய்யொலியன்கள் என்று மூன்று வகையாகப் பிரிக்கிறோம்.

உயிர் ஒலியன்கள்

தமிழில் உள்ள உயிர் ஒலியன்களை அ, ஆ, இ, ஈ, உ, ஊ, எ, ஏ, ஐ, ஒ, ஓ, ஔ என்று பன்னிரண்டாக வகுத்திருக்கிறோம். இந்த 12 உயிர் ஒலியன்களில் அ, இ, உ, எ, ஒ ஆகிய ஐந்தும் குறில். ஆ, ஈ, ஊ, ஏ, ஓ ஆகிய ஐந்தோடு ஐ, ஔ ஆகிய இரண்டையும் சேர்த்து நெடில்கள் ஏழு என்று கொள்ளப்படும். இந்த 12 உயிர் ஒலியன்களில் ஐகார, ஔகாரங்களை மரபிலக்கணங்கள் நெட்டுயிர்களாகவே கொண்டுள்ளன. இருந்தாலும் இன்றைய வழக்கில் ஐகாரம் அகர யகரச் சேர்க்கையாகவும் ஔகாரம் அகர வகரச் சேர்க்கையாகவும் அமைந்துள்ளன.

ஆய்தம்

ஆய்தம் என்ற இந்த ஒலியன் மற்ற ஒலியன்களிலிருந்து சற்று வேறுபட்ட ஒன்று. ஆய்தம் என்ற இந்த ஒலியன் குறிலுக்குப் பின்னும் வல்லொலிகளுக்கு முன்னும் தோன்றும்.

(எ-டு) அஃது, இஃது, எஃது.

மெய் ஒலியன்கள்

தமிழில் உள்ள மெய் ஒலியன்கள் பதினெட்டு, அவற்றை நாம் வல்லினம், மெல்லினம், இடையினம் என மூன்றாக வகைப்படுத்தலாம்.

வல்லினம்

க், ச், ட், த், ப், ற் ஆகிய ஆறும் வல்லினம் என்று பெயர்பெறும்.

Translation of the highlighted part:

Aytham

The Aytham in Tamil is slightly different from other sounds. It can come after the short vowels and always be followed by stop consonants

(e.g.) அஃது, இஃது, எஃது

/akthu/,/ikthu/, /ekthu/