

SCRIPT GRAMMAR FOR TAMIL LANGUAGE

Prepared by

Technology Development for Indian Languages (TDIL) Programme of DIT, GoI in co-ordination with

C-DAC, GIST Pune.

Following guidelines have been followed while preparing this script grammar.

1. This template will contain information especially as to shapes which will need to be filled out by hand.
2. Please print out the template and fill it in completely.
3. Once it is complete, please have it validated by an expert.
4. Subsequent to the validation, please get the document checked and validated by the State Government or the statutory certifying body in your State.

CLARIFICATION:

The final statutory body will be the State Government which will validate all Script Grammar documents. The State Government may delegate the evaluation to a committee or a Normative body for certification

Where no such body exists the State Govt. shall name a committee or members for the evaluation process.

The final approval of the state is a must for certifying the document

5. Insofar as Section 8 is concerned, please note the following:

1. LIGATURES:

Dead ligatures i.e. ligatures which are dysfunctional in the language will not be used CHC cases will be tested and checked.

2. **VARIANTS:** Variants shall be handled in the script grammar and where two variants exist concurrently for the same shape and are deemed as viable, one of them shall be entered in the template and the other shall be provided separately as a variant. Uniformity shall be maintained i.e. all stacked variants will be bunched together whereas the non-stacked variant will be grouped together.

6. Items such as the History of the Language and the evolution of the script shall be supplied in shape of an Appendix.

7. Other points: ZWJ/ZWNJ Consortium members were requested to determine exactly where these 2 characters are to be used and send a list for onward transmission to DIT

Name(s) of Expert(s) –

Dr. S. Velayuthan, Asst. Professor in Tamil,
Asst. Professor in Tamil, Periyar University,
Salem – 636011, Tamil Nadu

Name(s) of Evaluator(s)

1. Name of the language and its representation in the 3 letter mnemonic

Name of the Language:TAMIL

Alpha-3 code: tam

2. Name of the statutory board governing the language

The name and address /tel number/email of the statutory body

A scanned/hard copy of the statutes laid down. (Please append)

3. Identification of the writing system(s) used to inscribe the given language

The name(s) of the script system(s) used.

TAMIL SCRIPT

4. Short Historical Picture of the Language and the Script used.

PLEASE PROVIDE THE DATA IN APPENDIX.

5. Modifications brought to the writing system by a given language in terms of addition of characters and deprecation of other characters.

You need to enumerate here the character set of the language preferably as per the sorting order. Please write the exact shape of the character:

CONSONANTS

க ச ட த ப ற , ந ன ஞ ண ந ம ன

ய ர ல ள வ ழ

Grantha Consonants

ஜ
ஷ
ஸ
ஹ
சஷீ

VOWELS

அ ஆ
இ ஈ
உ ஊ
எ ஏ
ஐ
ஒ ஓ
ஒள

MATRAS

ா ீ ூ ெ ே ை ொ ோ ௌ

DIACRITICS (Visarga)

ஃ (Ayutham)

NUMERALS

Digits

Tamil Numerals

0BF0 Ten - ஓ

0BF1 One Hundred - ழ

0BF2 One Thousand - சூ

Ten thousand - ஐத

OTHERS

Tamil Symbols



0BF3 Tamil Day Sign -
0BF4 Tamil Month Sign
0BF5 Tamil Year Sign
0BF6 Tamil Debit Sign
0BF7 Tamil Credit Sign
0BF8 Tamil As Above Sign
0BFA Tamil Number Sign

Currency Symbol is not listed since it has been replaced by the standardized Currency Sign.

A copy of the TAMIL Unicode Page is appended for your kind information.

6. The structure of the writing system of the language

Tick whichever is appropriate: Abjad √Abugida.

7. Rule ordering of the characters within the syllable (only for abugidas)

NO description needed. Unless the script does not obey ISCII syllable rules

8. Script Pertinent Description of the syllabic clusters

8.1. BASIC SET OF CHARACTERS

The basic set of characters has been provided in this inventory.

These are arranged as per their class: CONSONANT / VOWEL / MATRAS / DIACRITICS

The allographs are presented at the end.

INSTRUCTIONS

- In case you do not see any issues just tick the VALID box. In case you see issues tick invalid and provide the necessary correction for the combination in question.

- In case a particular character is not used in your script, please cross it out
- In case you feel a particular character from your script has been left out, please specify the same.

8.1.1. CONSONANT SET: VALID / INVALID

Basic Consonants arranged as per their vargas

க			ங	
ச		ஜ	ஞ	
ட			ண	
த			ந	ன
ப			ம	
ய	ர	ற	ல	ள
வ	ழ	ஷ	ஸ	ஹ

8.1.2. VOWEL SET:

அ ஆ இ ஈ உ ஊ எ ஏ ஐ ஒ ஓ ஔ

8.1.3. MATRA SET

ா ி ீ ு ூ ெ ே ை ோ ௌ ்ள

Active Catenator(s) i.e. Displaced Matra(s):

CATENATOR	POSITION	EXAMPLE
ெ	To left of the consonant	கெ
ே	To left of the consonant	கே
ை	To left of the consonant	கை
ொ	To left and right of the consonant	கொ
ோ	To left and right of the consonant	கோ
ௌ	To left and right of the consonant	கௌ

8.1.4. DIACRITICS

◌̣ Visarga (Ayatham)

8.1.5.1. ALLOGRAPHS

Please mention below with substantiating evidence.

8.1.6. PUNCTUATION MARKERS

Please specify the punctuation markers specific to the character set omitting the markers taken from the Latin set such as . , ; : “ u () [] etc.

8.1.7. NUMERALS/DIGITS

Please specify the numbers for your script. Is the following VALID/ INVALID DIGITS:

0 1 2 3 4 5 6 7 8 9

NUMERALS

Digits

0 க ௨ ௩ ௪ ௫ ௬ ௭ ௮ ௯

Tamil Numerics

0BF0 Ten - ஓ

0BF1 One Hundred - ஸ

0BF2 One Thousand - சூ

Ten thousand - ஓத

If not valid please give the correct form/forms

Please specify if the English (Latino-Arabic set: 0,1,2,3,4,5,6,7,8,9) is used in official communications

Tamil Fraction

க 1 ஒன்று

ஏ ¾ முக்கால்

வ ¼ கால்

2 ½ அரை

8.1.8. OTHER SYMBOLS (religious, currency markers etc. included in Unicode) OTHERS

Tamil Symbols:

0BF3 Tamil Day Sign
0BF4 Tamil Month Sign
0BF5 Tamil Year Sign
0BF6 Tamil Debit Sign
0BF7 Tamil Credit Sign
0BF8 Tamil As Above Sign
0BFA Tamil Number Sign

Currency Symbol is not listed since it has been replaced by the standardized Currency Sign



8.2. CONSONANT+MATRA COMBINATIONS

This set is divided into three parts:

CM:The combination of Consonant and Matra

CMD(Anunasika) i.e. Consonant+Matra+Anunasika

- In case you do not see any issues just tick the VALID box. In case you see issues tick invalid and provide the necessary correction for the combination in question.
- Please do not forget that some combinations are dead clusters but are still needed by the font designer to generate out the grammar.
- In case you feel a particular Consonant+Matra combination has been left out, please specify the same.
- In case a particular character combination is not used in your script, please cross it out

8.2.1. CM: VALID / INVALID

	க	ங	ச	ஞ	ட	ண	த	ந	ன
ா	கா	ஙா	சா	ஞா	டா	ணா	தா	நா	னா
ி	கி	ஙி	சி	ஞி	டி	ணி	தி	நி	னி
ீ	கீ	ஙீ	சீ	ஞீ	டீ	ணீ	தீ	நீ	னீ
ு	கு	ஙு	சு	ஞு	டு	ணு	து	நு	னு
ு	கூ	ஙூ	சூ	ஞூ	டூ	ணூ	தூ	நூ	னூ
ெ	கெ	ஙெ	செ	ஞெ	டெ	ணெ	தெ	நெ	னெ
ே	கே	ஙே	சே	ஞே	டே	ணே	தே	நே	னே
ை	கை	ஙை	சை	ஞை	டை	ணை	தை	நை	னை
ொ	கொ	ஙொ	சொ	ஞொ	டொ	ணொ	தொ	நொ	னொ
ோ	கோ	ஙோ	சோ	ஞோ	டோ	ணோ	தோ	நோ	னோ
ௌ	கௌ	ஙௌ	சௌ	ஞௌ	டௌ	ணௌ	தௌ	நௌ	னௌ

	ப	ம	ய	ர	ற	ல	ள	வ	ழ	ஷ	ஸ	ஹ
ா	பா	மா	யா	ரா	றா	லா	ளா	வா	ழா	ஷா	ஸா	ஹா
ி	பி	மி	யி	ரி	றி	லி	ளி	வி	ழி	ஷி	ஸி	ஹி
ீ	பீ	மீ	யீ	ரீ	றீ	லீ	ளீ	வீ	ழீ	ஷீ	ஸீ	ஹீ
ு	பு	மு	யு	ரு	று	லு	ளு	வு	ழு	ஷு	ஸு	ஹு
ு	பூ	மூ	யூ	ரூ	றூ	லூ	ளூ	வூ	ழூ	ஷூ	ஸூ	ஹூ
ெ	பெ	மெ	யெ	ரெ	றெ	லெ	ளெ	வெ	ழெ	ஷெ	ஸெ	ஹெ
ே	பே	மே	யே	ரே	றே	லே	ளே	வே	ழே	ஷே	ஸே	ஹே
ை	பை	மை	யை	ரை	றை	லை	ளை	வை	ழை	ஷை	ஸை	ஹை
ொ	பொ	மொ	யொ	ரொ	றொ	லொ	ளொ	வொ	ழொ	ஷொ	ஸொ	ஹொ
ோ	போ	மோ	யோ	ரோ	றோ	லோ	ளோ	வோ	ழோ	ஷோ	ஸோ	ஹோ
ெள	பெள	மெள	யெள	ரெள	றெள	லெள	ளெள	வெள	ழெள	ஷெள	ஸெள	ஹெள

Combinations with Visarga are not provided since the Visarga just sits to the right of the syllable.

8.3. CONSONANT+CONSONANT CLUSTERS

8.3.1. CHC

Tamil is basically linear and hence has very clusters. Instead of a Tabular distribution, the clusters are provided below:

ப்ப் – படுப்பின்	ர்த் – உணர்தல்
த்த் – எடுத்து	ர்க் – அவர்கட்கு
ற்ற் – இயற்றல்	ர்ம் – கூர்மை
ட்ட் – ஊட்டல்	ர்வ் – ஆர்வம்
ச்ச் – அச்சம்	ழ்ப் – உறழ்பு
க்க் – ஆக்கல்	ழ்த் – சூழ்தல்
ம்ம் – தும்மல்	ழ்வ் – கதழ்வு
ந்ந் – முந்நீர்	ல்ப் – இயல்பு
ண்ண் – பண்ணை	ல்க் – அல்குல்
ஞ்ஞ் – அய்ஞ்நூறு	ள்ப் – கொள்ப
ய்ய் – எய்யாமை	ல்வ் – கல்வி
ல்ல் – முல்லை	
ள்ள் – தள்ளாடு	
ற்ப் – ஏற்பின்	ண்ட் – கண்டு
ற்ச் – அகற்சி	ண்ம் – எண்மை
ற்க் – ஏற்கும்	ங்ச் – அஞ்ச
ட்க் – கேட்கம்	வ்க் – அவ்கி
ட்ச் – மாட்சி	ய்ப் – செய்பு

ம்ப் – அம்பல்

ம்க் – எம்கய்யர்

ந்த் – ஏந்து

ன்ப் –என்பு

ன்ற் – என்றல்

ன்ம் – என்மனார்

ய்த் – செய்த

ய்க் – செய்கை

ய்ம் – சேய்மை

ய்வ் – காய்வு

ர்ப் – உணர்ப

9.0 COLLATION ORDER OF THE CHARACTERS: LEXICAL / DICTIONARY SORTING ORDER

List all the basic characters of the language in the expected sort-order. A sample sort order is provided below. Please provide an exhaustive collation order for your language. If there is any change in the sort order, please specify:

ஃ: அ ஆ இ ஈ உ ஊ எ ஏ ஐ ஒ ஓ ஔ கங ச ஜ ஞ ட ண த ந ப ம ய ர ற ல ள வ ழ ஷ ஸ ஹ ா ி ீ ு ூ ௃ ெ ே ை ொ ோ ௌ

HOMOGRAPHIC IDENTITIES WITHIN THE CHARACTER SET.

Please provide a list of look alikes. Each set of homographs will be proposed as a pair. In extreme cases even three homographs are permissible Add more columns if so required.

Unique characters

HOMOGRAPH1	HOMOGRAPH2	HOMOGRAPH3
ஆ	கவி	நாகம்
ஈ	கரி	யூகம்
மா	படி	ஊக்கி
கோ	ஓடு	அக்கி
பை	அகல்	மாலை
தை	ஓசி	கம்பம்
கை		

Conjunct characters.

COMPOSING CONSONANTS	RESULTING HOMOGRAPH	COMPOSING CONSONANTS	RESULTING HOMOGRAPH

Compliance with Unicode.

- Is the character set compliant with Unicode: YES / NO
- If not identify the characters which should be proposed to the Unicode consortium with substantiating evidence.

ZWJ/ZWNJ

Please provide all such cases where you feel that ZWJ/ZWNJ is a must e.g.

NONE