

Analyzing Tamil

A Project in Linguistics

K S Saisankalp Davey

UG -1

CLD

IIIT - Hyderabad



Introduction to Linguistics - 1

Date: 11th October 2024

Contents

1 Introduction

2 Phonetic Transcription:

- 2.1 Text - 1: கல்வி [kəlvɪ]
- 2.2 Text - 2: கல்லாமை [kəllamaɪ]
- 2.3 Transcription of Text - 1
- 2.4 Transcription of Text - 2

3 Syllabification:

4 Phonology:

- 4.1 IPA Charts:
- 4.2 Phonological Analysis:
- 4.3 Vowel Elision:
- 4.4 Consonant Doubling:

5 Morphological Analysis:

6 Syntactic Analysis:

- 6.1 Simple Constituency Grammar Rules
- 6.2 Sentences:

7 Historical Context:

8 Writing Systems and Orthography:

9 References:

1 Introduction

Tamil is the official language in the State of Tamil Nadu, and the Union Territory of Puducherry in India. There are also a significant number of speakers in the neighbouring states of Kerala, Karnataka, and Andhra Pradesh. It is also an official language in Sri Lanka and Singapore and is recognized as a minority language in the countries of South Africa and Malaysia. Belonging to the Dravidian family, Tamil is one of the two oldest classical languages in the Indian Subcontinent, with works appearing as early as the 2nd century BCE. Tamil is generally classified into two forms, Senntamil (செந்தமிழ்) [seɪnt̪əmɪɻ], the literary variant, and Kotuntamil (கொடுந்தமிழ்) [koɻuɪnt̪əmɪɻ], the spoken variant. While both forms are mutually intelligible, the Tamil script is particularly ill-suited for transcribing Kotuntamil. Over the years, Kotuntamil has become the favoured variant, and Centamil is now used only in certain places like formal debates and textbooks. Tamil was the lingua franca of early maritime traders around the Indian Subcontinent due to the strong influence of the Tamil kingdoms of South India. There are 77,026,881 speakers in India, of which 69,026,881 are first-language Tamil speakers. Most of the earliest works in Tamil come from the Sangam times and are classified broadly into two groups, the பதினெண்மேல்கணக்கு [paɻɪmɛɳməɻkəɳaku] (The eighteen greater works) and the பதினெண்கீழ்கணக்கு [paɻɪmɛɳkiɻkəɳaku] (The eighteen lesser works). I have chosen an excerpt from the திருக்குறள் [t̪iruk̪uraɻ], which belongs to the latter.

2 Phonetic Transcription:

Since Tamil does not have a native latin-based script, a romanized transcription is provided after the couplets, followed by the IPA transcription of the same.

2.1 Text - 1: கல்வி [kəɻvi]

1. கற்க கசடறக் கற்பவை கற்றபின்
நிற்க அதற்குத் தக

katrka	kasadara	karpavai	katrapin
learn.IMP	without-errors.INF	learn-things.INF	after-learning.PART
nirka	athatrkū	thaka	
stand.IMP	for-it.DAT	suitable.ADJ	

"Let a man learn thoroughly whatever he may learn, and let his conduct be worthy of his learning."

2. எண்ணென்ப ஏனை எழுத்தென்ப இவ்விரண்டும்
கண்ணென்ப வாழும் உயிர்க்கு

ennaenpa aenai ezhuthaenpa ivvirandum kannanba
numbers.NOM and.ADJ letters.NOM these-two.DEM.DUAL eyes.NOM
vaazhum uyirkku
living.ADJ life.NOM

"Letters and numbers are like the two eyes of man."

3. கண்ணுடையர் என்பவர் கற்றோர் முகத்திரண்டு
புண்ணுடையர் கல்லா தவர்

kannutaiyer enbavar katror mukathirandu
with-eyes.NOM who-is.COP learned.NOM in-face-two.LOC.DUAL
punnutaiyer kallaa thavar
sores.NOM unlearned.NEG people.NOM

"The learned are said to have eyes, but the unlearned have two sores in their face."

4. உவப்பத் தலைக்கூடி உள்ளப் பிரிதல்
அனைத்தே புலவர் தொழில்

uvappa thalaikoodi ulla pirithal anaithae
joyfully.ADV meet.INF pleasantly.ADV part.INF such-is.COP
pulavar thozhil
scholar.NOM profession.NOM

"You meet with joy, with pleasant thought you part. Such is the scholar's wondrous profession!"

5. உடையார்முன் இல்லார்போல் ஏக்கற்றுங் கற்றார்
கடையரே கல்லா தவர்

utaiyaarmun illaarpol ekatrun katraar
in-front-wealthy.LOC destitute.SIM begging.ADV learned.NOM
kataiyarae kallaathavar
indeed-inferior.EMPH unlearned.NOM

"The unlearned are inferior to the learned, before whom they stand begging, as the destitute before the wealthy."

6. தொட்டனைத் தூறும் மணற்கேணி மாந்தர்க்குக்
கற்றனைத் தூறும் அறிவு

thottanai thoorum manarkaeni maantharkku katrranai
depth.ACC yields.V sandy-well.NOM men.DAT learning.ACC
thoorum arivu
flows.V knowledge.NOM

”As a sandy well yields water in proportion to the depth it is dug, so does knowledge flow to men in proportion to their learning.”

7. யாதானும் நாடாமால் ஊராமால் என்னொருவன்
சாந்துணையுங் கல்லாத வாறு

yaathanum naataamaal ooraamaal
any-place.INDEF without-seeking.NEG without-belonging.NEG
ennoruvan santhunaiyun kallaatha vaaru
any-person.NOM to-his-end.ADV without-learning.ADJ state.NOM

”How is it that any one can remain without learning, even to his death, when (to the learned man) every country is his own, and every town his own?”

8. ஒருமைக்கண் தான்கற்ற கல்வி ஒருவற்கு
எழுமையும் ஏமாப் புடைத்து

orumaikkan thaankattra kalvi oruvarku
one-birth.LOC self-learned.ADJ knowledge.NOM person.DAT
ezhumaiyum emaa putaithu
seven-lives.LOC pleasure.NOM yields.V

”The learning which a man has acquired in one birth will yield him pleasure during seven births.”

9. தாமின் புறுவது உலகின் புறக்கண்டு
காமுறுவர் கற்றறிந் தார்

thaamin puruvathu ulakin purakkandu kaamuruvar
self.NOM joy.NOM world-of.GEN outwardly-seeing.V love-develops.V
katrrarinthaar
learned-ones.NOM

”The learned will long (for knowledge), when they see that while it gives pleasure to themselves, the world also derives pleasure from it.”

10. கேடில் விழுச்செல்வம் கல்வி யொருவற்கு
மாடல்ல மற்றை யவை

kaetil vizhuchelvam kalvi yoruvarku
without-destruction.ADV excellent-wealth.NOM learning.NOM person.DAT
maadalla matrrai yavai
not-material.NEG other.ADJ things.NOM

"Learning is excellence of wealth that none destroy; to man, nought else affords reality of joy."

2.2 Text - 2: கல்லாமை [kollamai]

1. அரங்கின்றி வட்டாடி யற்றே நிரம்பிய
நூலின்றிக் கோட்டி கொளல்

arangindri vattadi yatre nirambiya noolindri
without-stage.LOC circle-play.V only.ADV full.ADJ without-thread.LOC
kotti kolal
drawing.V taking.V

"Without foundational knowledge, to argue is like drawing conclusions without a guiding thread."

2. கல்லாதான் சொற்கா முறுதல் முலையிரண்டும்
இல்லாதாள் பெண்காமுற் றற்று

kalladhan sorka muruthal
unlearned-person.NOM speech-affection.NOM breasts-both.NOM
mulairandum illathal penkamu tratra
not-having.ADJ woman-beauty-lacking.NOM

"The speech of the unlearned is as unpleasing as the beauty of a woman without breasts."

3. கல்லா தவரும் நனிநல்லர் கற்றார்முன்
சொல்லா திருக்கப் பெறின்

kalla thavarum naninalla katrarmun
unlearned.NOM persons-also.NOM respectable.ADJ learned-before.LOC
sollathu irruka perin
speak-not.ADV remain-INF if.COND

"Even the unlearned may seem respectable if they remain silent in the presence of the learned."

4. கல்லாதான் ஓட்பம் கழியநன் றாயினும்
கொள்ளார் அறிவுடையார்

kallathan otpam kazhiyanan
unlearned-person.NOM natural-knowledge.NOM very-good-even-if.COND
trayinum kollar arivu utaiyar
accept-not.V wisdom-having-persons.NOM

"Although the natural knowledge of an unlearned man may be very good,
the wise will not accept it as true knowledge."

5. கல்லா ஒருவன் தகைமை தலைப்பெய்து
சொல்லாடச் சோர்வு படும்

kalla oruvan thakaimai thalaipeythu
unlearned.NOM person.NOM self-conceit.NOM head-receive.V
sollada sorvu padum
speak-attempt.INF fade.V experience.V

"The self-conceit of an unlearned man will fade away, as soon as he speaks
in an assembly (of the learned)."

6. உளரென்னும் மாத்திரையர் அல்லால் பயவாக்
களரனையர் கல்லா தவர்

ularennum mathiraiyar allal payava
exist-said.V only-in-appearance.ADV except.CONJ worthless.ADJ
kalaranaiyar kalla thavar
barren-land-resemble.V unlearned-persons.NOM

"The unlearned are like worthless barren land: all that can be said of them
is, that they exist."

7. நுண்மாண் நுழைபுலம் இல்லான் எழில்நலம்
மண்மாண் புனைபாவை யற்று

nunmaan nuzhaipulam illan ezhilnalam
subtle.NOM penetrating-sense.NOM not-have.V large-beauty.NOM
manmaan punaipaavai yatru
clay.NOM decorated-puppet.V false.V

"The beauty and goodness of one who is destitute of knowledge by the
study of great and exquisiteworks, is like that of a painted earthen doll."

8. நல்லார்கண் பட்ட வறுமையின் இன்னாதே
கல்லார்கண் பட்ட திரு

nallarkan patta varumaiyin innathe
goodly-wisdom-to-men.NOM struck-by.V poverty-GEN evil.NOM
kallarkan patta thiru
unlearned-men.NOM struck-by.V fortune.NOM

"Wealth, gained by the unlearned, will give more sorrow than the poverty
which may come upon the learned."

9. மேற்பிறந்தார ஆயினும் கல்லாதார் கீழ்ப்பிறந்தும்
கற்றார் அனைத்திலர் பாடு

merpiranthar ayinum kallathar kizhpirandhum
noble-born.NOM even-if.CONJ unlearned-men.NOM low-born.NOM
katrar anaithilar paadu
learned-men.NOM in-all.NOM lower.V

"The unlearned, though born in a high caste, are not equal in dignity to
the learned, though they may have been born in a low caste."

10. விலங்கொடு மக்கள் அனையர் இலங்குநூல்
கற்றாரோடு ஏனை யவர்

vilangodu makkal anaiyar ilangunool
beast-with.NOM men.NOM excel.V learning's-grace.NOM
katraarotu enai yavar
learned-with.NOM others.NOM men.NOM

"As beasts by the side of men, so are other men by the side of those who
are learned in celebrated works."

2.3 Transcription of Text - 1

1. karkə kəsədərə kərpəvai kətrəpm
nirkə əḍərkui θəgə
2. eṇeṇpə enai eṇu θenbə
irvīrəṇḍum kaṇeṇbə vaṇum ɕjīrīkui
3. kəṇṣṭṭaijər enbəvər kətror mukəθīrəṇḍu
pṇṇṣṭṭaijər kəlla θəvər

4. svæppæ þælaikudr ulla þræðæl
ænaiþe þulævær þoþl
5. vðaijarmun ilarpol ekætrum kætrar
kætaijære kælla þævær
6. þotþænai þurum mænærkeni manþærkkur
kætrænai þurum ærivur
7. jaðanum naðamal uramal enoruvæn
sænþunaijun kælla þævar
8. orumaikæn þankætræ kælví oruværkur
eþumaijum ema þutaiður
9. þamín þuruvæþu vlækin þurakkaþur
kamuruvær kætrærin þær
10. ketl vutþælvæn kælví joruvarkkur
maðællæ mætrai jævai

2.4 Transcription of Text - 2

1. ænmdri vætþaðr jætre nrambijæ
nulmdri kottí kolæl
2. kællaðan sorka mæruþæl mælajirændum
illaðal þenkamur trætrur
3. kælla þæværum nænneðlæ kætrarman
solla þirukkæ þerm
4. kællaðan oþram kaþjænæn trajmum
kollar arivutaijar
5. kælla oruvæn þækaimai þælaiþeþur
sollaður sorvur þæðum
6. vlærennum maðiraijær ællal þæjæva
kælærænaijær kælla þævær
7. nænman nuparþælæn illan eþilnælæn
mænman þunarpavar jætrur

8. η əllarkəŋ pət̪tə vərəmaɪjɪm mnaθɛ
kəllarkəŋ pət̪tə θɪru
9. mɐrpɪrəŋθar aɪmum kəllaθar kiɭprəŋðum
kətrar ənaɪθɪlər paɖu
10. vɪlaɪodʊ məkkəl ənaɪjar ɪlaɪʊŋɪl
kətrarotʊ enaɪ jəvər

3 Syllabification:

A few sentences have been chosen at random from the dataset to demonstrate the syllabic structure of Tamil. Each entry in the list below consists of two lines of IPA; the first is the excerpt from the dataset, while the second is the same sentence, but with each syllable separated by a space in the word. In the case of doubled/geminate consonants, the pair is split into two, with each syllable receiving one instance of the now non-geminate consonant.

1. kərkə kəsədərə kərpəvai kətrəpɪn nɪrkə əɖərkɪu θəgə
kər kə kə sə də rə kər pə vai kə trə pɪn nɪr kə əɖ ər kɪu θə gə
2. kəŋʊɖaijər enbəvər kətror mʊkəθɪrəŋɖu pʊŋʊtaijər kəlla θəvər
kəŋ ʊ ɖai jər en bə vər kətr or mʊ kə θɪ rəŋ ɖu pʊŋ ʊ tai jər kəl la θə vər
3. ʊɖaijarmʊn ɪllarpol ekətrum kətrar kətaijərə kəlla θəvər
ʊ ɖai jar mʊn ɪl lar pol ekət trum kət rar kə tai jə re kəl la θə vər
4. jaθanum ɲaɖamal uramal enorʊvən səŋθʊŋaijʊn kəlla θəvar
ja θa num ɲa ɖa mal ur am al en or ʊv ən səŋ θʊ ŋaij ʊn kəl la θə var
5. θamin pʊrʊvəθʊ ʊləkm pʊrakkaŋɖu kamʊrʊvər kətrərm θar
θam ɪn pʊ ru və θʊ ʊl ək ɪn pʊ rak kaŋ ɖu kam ʊr ʊv ər kət rər ɪn θar
6. kəllaðan sorka mʊrʊθəl mʊləjɪrəŋɖum ɪllaθal pɛŋkamɪ trətru
kəl la ðan sor ka mʊ ru θəl mʊ ləj ɪr əŋ ɖum ɪl la θal pɛŋ kam: trə tr:
7. kəllaθan oɖpam kaɪjəŋən trajmum kollar arɪvʊtaijar kəl la θan oɖ pam ka
ɪ jə ŋən tra jɪm um kol lar ar ɪv ʊ taij ar
8. ʊləɾennum maθɪraɪjər əllal pəjəva kələrənaijər kəlla θəvər
ʊ ləɾ en num ma θɪ rai jər əl lal pə jə va kə lə rə nar jər kəl la θə vər
9. η əllarkəŋ pət̪tə vərəmaɪjɪm mnaθɛ kəllarkəŋ pət̪tə θɪru
 η əl lar kəŋ pət̪ tə və rə maɪ jɪm ɪn na θɛ kəl lar kəŋ pət̪ tə θɪ ru

10. vīḷaḷoḍu mēkkēḷ ēnāṭṭar ilaṭṭuṇṇuḷ kēṭṭarotṭu enāṭṭ jēvēr
vī ḷaṭṭ o ḍu mēk kēḷ ēn aṭṭ jar il aṭṭ ṭṭuḷ kē ṭṭa ro ṭṭu en aṭṭ jē vēr

From this, we can see that while the syllabic structure in Tamil is very similar to that of English, it is much more simpler in Tamil than in English. Null codas and null onsets are common sights in Tamil, and most vowels take one of three forms, CV, VC or CVC, which is more constrained compared to English's wide variety, ranging from V to CCCVCCC.

4 Phonology:

Throughout this report, the [ɯ] symbol has been used to represent the unrounded close back vowel sound. This sound is not represented by a single symbol in Tamil, and usually occurs when the terminal consonant in a word is followed by either the 'உ' or the 'ஊ' sounds. [Ex: கு in அதற்கு]

4.1 IPA Charts:

Consonants	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Glottal
Nasal	ம்		ந்	ன	ஞ்	ண		ங்	
Plosive	ப்		த்			ட்		க்	
Fricative				ச ஸ*	ஸ்*	ஷ*			ஹ*
Affricate					ச் ஜ்*				
Tap/Flap				ர்					
Trill				ற்					
Approximant	வ்	வ்		ய்	ய்	ழ்			
Lateral Approximant				ல்		ள			

Table 1: Pulmonic Consonants with Tamil script

NOTE: All the letters marked with an asterisk (*) were not originally part of the tamil script, but these were borrowed from the Grantha to represent phonemes that were not initially present during the time of the தொல்காப்பியம் [tolkappiyam] [θolkappijəm].

Consonants	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Glottal
Nasal	m		n	n	ɲ	ɳ		ŋ	
Plosive	p b		θ ð			ʈ		k g	
Fricative				s z	ʃ ʒ	ʂ			h
Affricate					tʃ tʒ				
Tap/Flap				r					
Trill				r					
Approximant	w	v			j	ɻ			
Lateral Approximant				l		ɭ			

Table 2: Pulmonic Consonants with IPA symbols

Vowels	Front	Near-Front	Central	Near-Back	Back
Close	இ				உள
Near-Close		ஈ		உ	
Close-Mid		ஏ			ஔ ஓ
Mid		எ	அ		
Open-Mid					
Near-Open					
Open		ஆ			

Table 3: Vowels with Tamil script

Vowels	Front	Near-Front	Central	Near-Back	Back
Close	ɪ				u
Near-Close		i		ʊ	
Close-Mid		e			o ɔː
Mid		ɛ	ə		
Open-Mid					
Near-Open					
Open		a			

Table 4: Vowels with IPA script

The Tamil language has only two diphthongs, namely ஐ [ai] and ஔ [au].

4.2 Phonological Analysis:

In Standard Spoken Tamil, the voiced and the voiceless variants of the plosives form pairs of allophones of the same phonemes. For example, [b] and [p] are used interchangeably, depending on the accent of the speaker. Similarly, in most places, the plosive [k] softens to the fricative [ɣ] without any change in the original meaning of the word. Since the Tamil script does not have unique characters for voiced or aspirated variants of sounds, the three are often used interchangeably. With the passage of time, the difference between [r] and [ɾ] in spoken Tamil has reduced so much so that both can be used interchangeably without loss of meaning. Nasality is an established feature in Tamil, and just like in English where the vowel sound preceding a nasal sound becomes nasalised. Finally, in many instances, the lateral approximant is converted into trill or tap depending on the phonetic environment it occurs in.

Ex: mel + pirəṇṇəɾ = merpirəṇṇəɾ

4.3 Vowel Elision:

In Tamil, when two or more words/syllables/morphemes combine to form a larger word, the vowel sound of the initial word/syllable/morpheme is dropped, and the vowel sound of the one following it gets higher precedence. If the following morpheme

is devoid of an initial vowel, the trailing vowel of the first morpheme is not dropped. No vowel sound is inserted if neither morpheme had a vowel in their first or last position respectively.

Examples:

1. mʊkəθu + ɪɾɐɳɖu
mʊkəθɪɾɐɳɖu
2. θan + kətrə
θankətrə

4.4 Consonant Doubling:

Consonant Doubling (also known as Gemination) is a very common phenomenon in Tamil, and is even represented in the script, with the use of the (.) diacritic on top of the consonant in the doubled pair to convey that there is no vowel sound between the two consonant sounds.

Example: In "மக்கள்", the consonant cluster "க்க்" represents the sound [k:] instead of [kək] which is represented by "கக". Gemination usually occurs when a morpheme beginning with a consonant is joined to a morpheme that is terminated with a vowel. In most cases, the vowel that appears right before the geminal consonant is short, while monosyllabic words usually have a long vowel as the nucleus of the monosyllabic word.

Example: The word "வீட்டுக்காரன்" is made of two words "வீடு" and the suffix "காரன்".

5 Morphological Analysis:

The following table provides the full forms of the gloss abbreviations used in the dataset.

Gloss	Full Form
IMP	Imperative
INF	Infinitive
PART	Particle
DAT	Dative
ADJ	Adjective
NOM	Nominative
DEM	Demonstrative
DUAL	Dual
COP	Copula
LOC	Locative
NUM	Number
NEG	Negative
ADV	Adverb
SIM	Similative
EMPH	Emphatic
V	Verb
ACC	Accusative
INDEF	Indefinite
GEN	Genitive
COND	Conditional
CONJ	Conjunction

Table 5: Gloss Abbreviations and Full Forms

Perspective	First Person	First Person	Second Person	Second Person
Case	Singular	Plural or Formal	Singular	Plural or Formal
Nominative	ᲒᲗᲚ	ᲒᲗᲚᲚ	ᲒᲗ	ᲒᲗᲚᲚ
Ergative	ᲒᲗᲚᲚ	ᲒᲗᲚᲚᲚ	ᲒᲗᲚ	ᲒᲗᲚᲚᲚ
Accusative-Dative	ᲚᲗᲚᲚ	ᲚᲗᲚᲚᲚ	ᲚᲗᲚᲚ	ᲚᲗᲚᲚᲚ
Possessive-Genitive	ᲚᲗ	ᲚᲗᲚᲚ	ᲚᲗ	ᲚᲗᲚᲚ
Propriative-Possessive	ᲚᲗᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚ

Table 6: Personal Pronouns

Case	Male Singular		Female Singular		Neutral Singular		Plural or Formal	
Proximity	Distal	Proximal	Distal	Proximal	Distal	Proximal	Distal	Proximal
Nominative	ᲚᲗᲚ	ᲚᲗᲚᲚ	ᲚᲗᲚ	ᲚᲗᲚᲚ	ᲚᲗᲚ	ᲚᲗᲚᲚ	ᲚᲗᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚ
Ergative	ᲚᲗᲚᲚ	ᲚᲗᲚᲚᲚ	ᲚᲗᲚᲚ	ᲚᲗᲚᲚᲚ	ᲚᲗᲚᲚ	ᲚᲗᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚ
Accusative-Dative	ᲚᲗᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚᲚᲚ
Possessive-Genitive	ᲚᲗᲚ	ᲚᲗᲚᲚ	ᲚᲗᲚᲚ	ᲚᲗᲚᲚᲚ	ᲚᲗᲚᲚ	ᲚᲗᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚ
Propriative-Possessive	ᲚᲗᲚᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚᲚᲚᲚᲚ
Locative-Inessive	ᲚᲗᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚᲚ	ᲚᲗᲚᲚᲚᲚᲚᲚᲚ

Table 7: Demonstrative Personal Pronouns

Declinable Pronouns: These pronouns can be declined as normal nouns.

1. Interrogative Pronoun (Which)
εθuu
2. Interrogative Personal Pronoun (Who)
jar
3. Reflexive Personal Pronoun (Oneself)
θane
4. Indefinite Quantifiers
sɪlə = Few
pələ = Many
ɛllam / ənaɪθuu = Everything

Type	Interrogative	Demonstrative				
Location			Proximal		Distal	
Time	εppoθuu	When	ɪppoθuu	Now	əppoθuu	Then
Direction	εn̩θuu vəɽɪ	Which way	ɪn̩θuu vəɽɪ	This way	ən̩θuu vəɽɪ	That way
Place	εɲuu	Where	ɲuu	Here	əɲuu	There
Manner	εppəɽɪ	How	ɪppəɽɪ	Like this	əppəɽɪ	Like that
Quantity	ɛvələvuu	How much	ɪvələvuu	This much	əvələvuu	That much

Table 8: Indeclinable Pronouns

Code	Case	Suffix	Example	Meaning
NOM	Nominative	Ø	ram	Ram (subject)
OBL	Oblique	Ø	ram	Ram (base)
ERG	Ergative	-al	ramal	by Ram (subject)
ACC/DAT	Accusative-Dative	-kkʊ	ramʊkkʊ	to Ram (object)
LOC	Locative-Inessive	-ɪl	ramɪl	in Ram
SUP	Superessive	-mel	ram mel	on Ram
TER	Terminative	-vəraikʊm	ram vəraikʊm	up to Ram
GEN/POS	Genitive-Possessive	-ʊɽajə	ram ʊɽajə	Ram's
PRO/POS	Proprietary-Possessive	-ɪn	ramɪn	belonging to Ram
ADE	Adessive-Sociative	-oɽuu	ramoɽuu	with Ram
VOC	Vocative	-e	rame	O Ram!
INS	Instrumental	-al	ramal	by means of Ram

Table 9: Primary Adpositions in Tamil

Case/Noun	sīṭa	punai	ṇai	punaigəḷ	naigəḷ	peṇ	aṇ
Nominative	sīṭa	punai	ṇai	punaigəḷ	naigəḷ	peṇ	aṇ
Oblique	sīṭa	punai	ṇai	punaigəḷ	naigəḷ	peṇ	aṇ
Ergative	sīṭaval	punaijal	ṇaijal	punaigəḷal	naigəḷal	peṇal	aṇal
Accusative-Dative	sīṭakkū	punaikkū	ṇaikkū	punaigəḷikkū	naigəḷikkū	peṇikkū	aṇikkū
Locative-Inessive	sīṭavil	punaiḷ	ṇaiḷ	punaigəḷil	naigəḷil	peṇḷ	aṇḷ
Superessive	sīṭamel	punamel	ṇamel	punaigəḷmel	naigəḷmel	peṇmel	aṇmel
Terminative	sīṭa vōraikom	punai vōraikom	ṇai vōraikom	punaigəḷ vōraikom	naigəḷ vōraikom	peṇ vōraikom	aṇ vōraikom
Genitive-Possessive	sīṭa oḍaijə	punai oḍaijə	ṇai oḍaijə	punaigəḷ oḍaijə	naigəḷ oḍaijə	peṇ oḍaijə	aṇ oḍaijə
Propriative	sīṭa kaṛən	punai kaṛən	ṇai kaṛən	punaigəḷ kaṛən	naigəḷ kaṛən	peṇ kaṛən	aṇ kaṛən
Adessive-Sociative	sīṭa oḍu	punai oḍu	ṇai oḍu	punaigəḷ oḍu	naigəḷ oḍu	peṇ oḍu	aṇ oḍu
Vocative	sīṭave	punaije	ṇaije	punaigəḷe	naigəḷe	peṇe	aṇe
Instrumental	sīṭaval	punaijal	ṇaijal	punaigəḷal	naigəḷal	peṇal	aṇal

Table 10: Declensions of Tamil Nouns

Case	IPA	Translation
Nominative	əvən pəḷḷikkū selkiran	"He goes to school."
Oblique	əvənukkū ʈeɾɟaɖu	"He doesn't know."
Ergative	əvənəl paṭṭu paɖə pəṭṭəɖu	"The song was sung by him."
Accusative-Dative	ṇan əvənai sən̪ɖiṭən	"I met him."
Locative-Inessive	əvəl viṭṭil iɾukkiral	"She is in the house."
Superessive	kaka mərəθil əməɾɳɖəɖu	"The crow sat on the tree."
Terminative	ṇan salai vərai naɖaṇḍən	"I walked up to the road."
Genitive-Possessive	əvən oḍaijə puṭṭəgəm	"His book."
Propriative	əvən oḍən əvəḷum vən̪ɖəl	"She came along with him."
Adessive-Sociative	ṇan əvən oḍəḷ vən̪θən	"I came with him."
Vocative	o ɳəṇbəre!	"Oh friend!"
Instrumental	penaval eɻuṭəkiren	"I write with a pen."

Table 11: Examples of Primary Adpositions in Tamil

In daily speech, the markers are often dropped since the meaning is easily inferred from the context it appears in. This is easily seen in the last example, where the Nominative marker *ṇan* is dropped from the beginning of the sentence, without any loss in the meaning of the sentence.

Abbr.	Case	Adposition (IPA)	Meaning
SPR	Superessive	mele	above
APU/POE	Apudessive-Possessive	munnaḷ	in front of
ADE	Adessive	aruḡiḷ	near
CON	Concerning Something	pətri	about
SMB	Semblative	pol	like
BEN	Benefactive	kagə	for
SBE	Subessive	kiḷe	below
ELA	Elative	veḷḷe	outside of
PST	Posterior	pinnal	behind
CSL	Causal	kaṇṇamaḡu	because
INS	Instrumental-Comitative	uṭṇaṇ	using (along with)
ABL	Ablative	iṟuṇṇu	from
PTS	Postessive	pṛəḡ:	after
ANT	Antessive	mūṇ	before
INE	Inessive	uḷḷe	inside of
INT	Intrative	nəḍuvil	in the middle of

Table 12: Secondary Adpositions in Tamil (IPA and English)

Examples:

- əvəḷ viṭṭm veḷḷiḷ ṇṛkīraḷ
"She is standing outside the house."
- əvəḷ pəṇəṭṭai əvəṇṇukkag: koḍuṭaḷ
"She gave the money for him."

Numeral	English	Cardinals	Ordinals	Multipliers	Fractional
0	Zero	budḡijəṃ	budḡijəṭṭu	budḡijə muraḷ	
1	One	ōṇḍru	muṭṭaḷ	oru muraḷ	
2	Two	iṟṇḍḍu	iṟṇḍḍavəṭṭu	iṟṇḍḍu muraḷ	paṭi / ərai
3	Three	mūṇḍru	mūṇḍravəṭṭu	mūṇḍru muraḷ	
4	Four	ṇāḷḷu	ṇāḷḷavəṭṭu	ṇāḷḷu muraḷ	kal
5	Five	aṇḍu	aṇḍavəṭṭu	aṇḍu muraḷ	
6	Six	aru	aravəṭṭu	aru muraḷ	
7	Seven	eḷḷu	eḷḷavəṭṭu	eḷḷu muraḷ	
8	Eight	eṭṭu	eṭṭavəṭṭu	eṭṭu muraḷ	
9	Nine	ōṇbəṭṭu	ōṇbəṭṭavəṭṭu	ōṇbəṭṭu muraḷ	
10	Ten	pəṭṭu	pəṭṭavəṭṭu	pəṭṭu muraḷ	
100	Hundred	ṇuru	ṇuravəṭṭu	ṇuru muraḷ	
1000	Thousand	ajrəṃ	ajrəṃavəṭṭu	ajrəṃ muraḷ	
100000	Hundred Thousand	ləṭḷəṃ	ləṭḷəṭṭu	ləṭḷəṃ muraḷ	
1000000	Ten Million	koḍi	koḍijəṭṭu	koḍi muraḷ	

Table 13: Counting System

6 Syntactic Analysis:

Just like most other Indian languages, Tamil too has a very flexible word ordering in sentences. The most commonly used ordering is Subject-Object-Verb [SOV]. While

other orderings are possible, they are comparatively very rare. Tamil uses case markers extensively to preserve the meaning of the sentence, and hence the strictness of the ordering is relatively loose.

For the syntactic analysis, constituent trees and dependency trees have been used. Each sentence has been followed by its constituency tree (in green and blue), and its dependency tree.

The abbreviations used have been listed below.

Abbreviation	Description
S	Sentence
NP	Noun Phrase
VP	Verb Phrase
PP	Postpositional Phrase
AdjP	Adjective Phrase
AdvP	Adverbial Phrase
N	Noun
V	Verb
Adj	Adjective
Adv	Adverb
Post	Postposition

Table 14: Constituency Grammar Components for Tamil

6.1 Simple Constituency Grammar Rules

1. $S \rightarrow NP VP$
2. $VP \rightarrow (NP) (PP) (AdvP) V$
3. $NP \rightarrow (AdjP) N (PP)$
4. $AdjP \rightarrow (Adv) (Adj) (AdjP)$
5. $AdvP \rightarrow Adv$
6. $PP \rightarrow (NP) P$

6.2 Sentences:

- பெரிய மரம் பச்சை நிறத்தில் இருக்கிறது
 pɛɾijə məɾəm pətʃai ɳɪɾəθɪl ɪɾukɪɾəθɪɯ
 (The big tree is green.)

(a) Phrase Structure Tree:

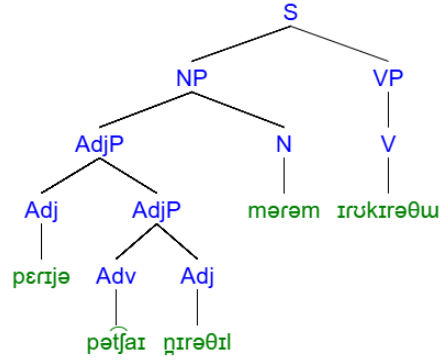


Figure 2: Phrase Structure Tree

(b) Dependency Tree:

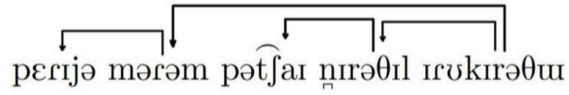


Figure 3: Phrase Structure Tree

2. குழந்தை பாலைக் குடிக்கிறது
 குடிந்஠ை palai kuṭṭikīṙṙṙṙ
 (The child is drinking milk.)

(a) Phrase Structure Tree:

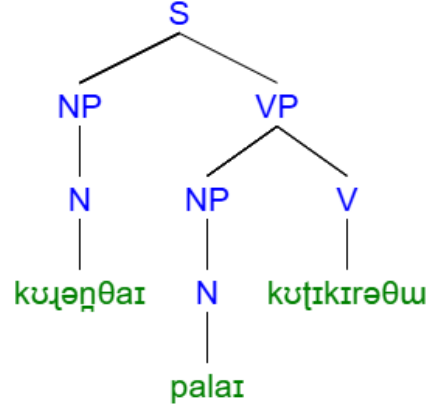


Figure 4: Phrase Structure Tree

(b) Dependency Tree:

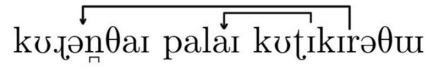


Figure 5: Phrase Structure Tree

3. வீட்டில் நாய் உறங்குகிறது
 viṭṭil nai ũṛṅṅukirṇṇu
 (The dog is sleeping in the house.)
 (a) Phrase Structure Tree:

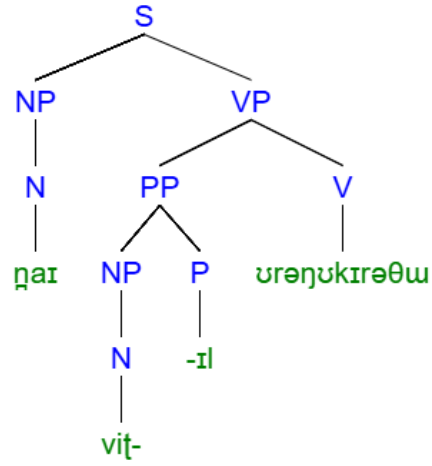


Figure 6: Phrase Structure Tree

- (b) Dependency Tree:

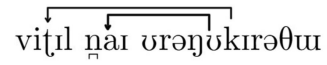


Figure 7: Phrase Structure Tree

4. மாணவர்கள் கல்வியில் தீவிரமாக ஈடுபட்டனர்
 maṇavarkəl kəlvijil θivirəmakə iṭupəṭəṇər
 (The students were deeply engaged in academics.)

(a) Phrase Structure Tree:

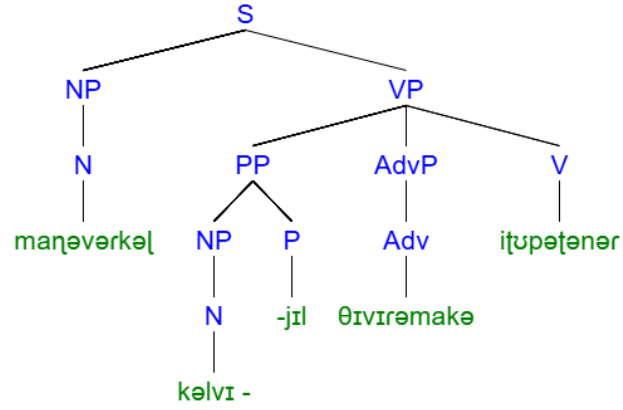


Figure 8: Phrase Structure Tree

(b) Dependency Tree:

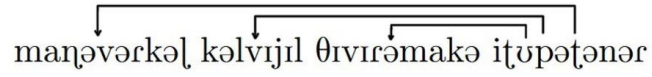


Figure 9: Phrase Structure Tree

5. சமூகத்தில் கல்வியின் முக்கியத்துவத்தை ஆசிரியர் வலியுறுத்தினார்
 s̥mukəθil kəlvij̃m mukijəθ̥vəθai asirij̃ər vəl̥ij̃s̥r̥θ̥mar
 (The teacher emphasized the importance of education in society.)

(a) Phrase Structure Tree:

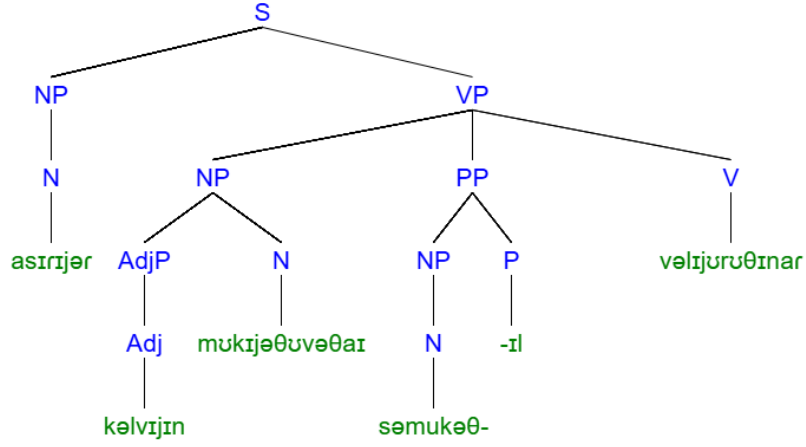


Figure 10: Phrase Structure Tree

(b) Dependency Tree:

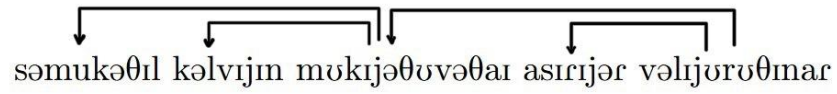


Figure 11: Phrase Structure Tree

6. பாரம்பரிய கலைகளை அடுத்த தலைமுறைக்கு பாதுகாக்க வேண்டும்
 parəmbərijə kəlaigəlaɪ əɖʊθə θəlaɪmʊraikʊ paʊgaka vɐɖʊm
 (We must preserve traditional arts for the next generation.)

(a) Phrase Structure Tree:

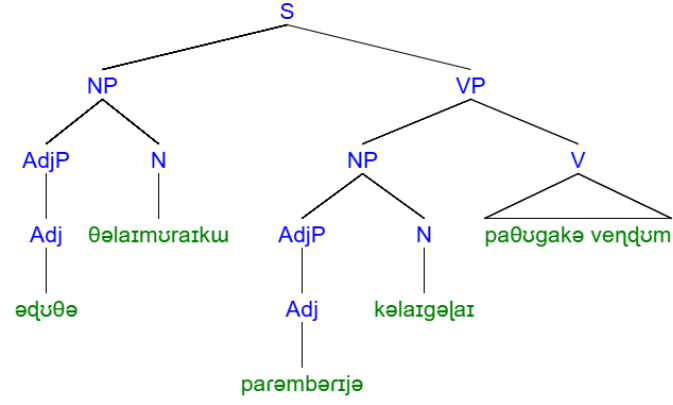


Figure 12: Phrase Structure Tree

(b) Dependency Tree:

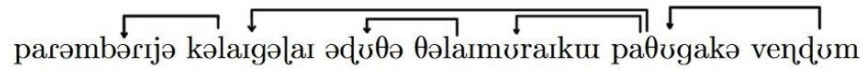


Figure 13: Phrase Structure Tree

7. இயற்கையின் அருமை மனிதர்களுக்கு பாடம் சொல்கிறது
 ijerkaɪjɪn əɾumaɪ məniθərkəɭɪkɐ paɖəɪ solkiɾəθɐ
 (The beauty of nature teaches humans a lesson.)

(a) Phrase Structure Tree:

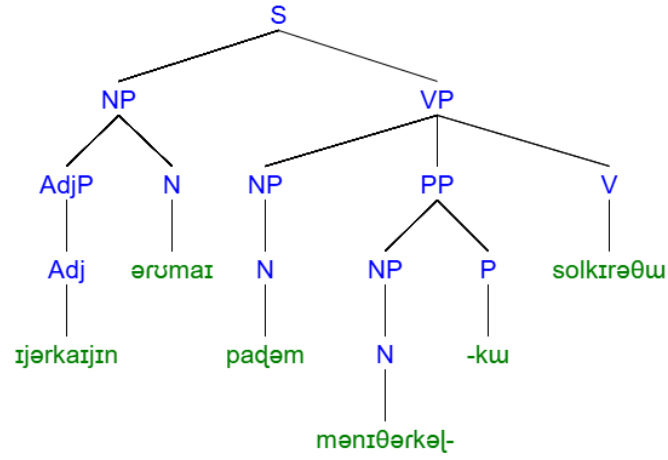


Figure 14: Phrase Structure Tree

(b) Dependency Tree:

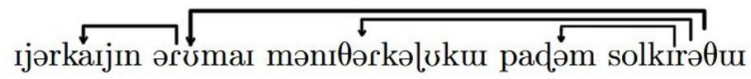


Figure 15: Phrase Structure Tree

8. கலாச்சார மரபுகள் சமுதாயத்தின் அடிப்படை தூண்கள் ஆகும்
 kəlatʃarə məɾəpukəl səmʊθajəθɪn əɖɪpəɖar ʈuŋgəl akum
 (Cultural traditions are the fundamental pillars of society.)

(a) Phrase Structure Tree:

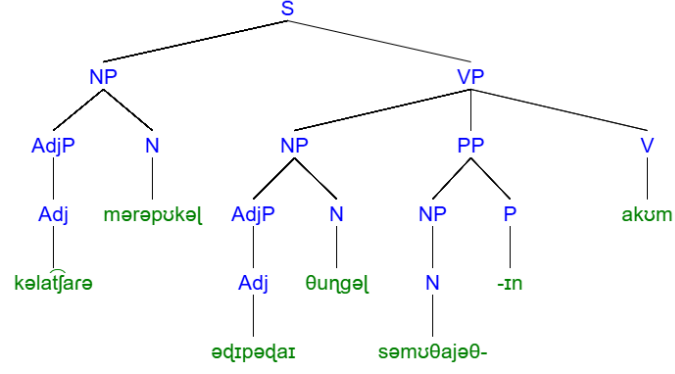


Figure 16: Phrase Structure Tree

(b) Dependency Tree:

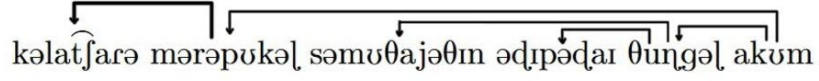


Figure 17: Phrase Structure Tree

9. தொழிலாளி மேடையில் வேலை செய்கிறார்

தொழிலாளி meḍaijil velai sejkirar

(The worker is working on the stage.)

(a) Phrase Structure Tree:

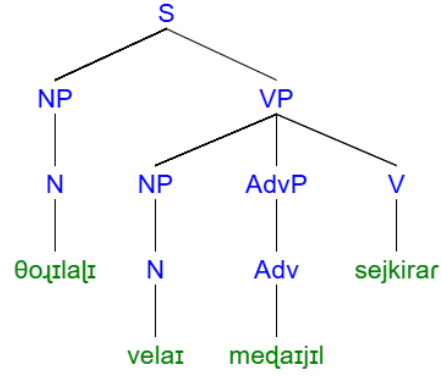


Figure 18: Phrase Structure Tree

(b) Dependency Tree:

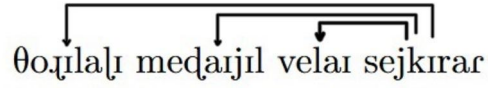


Figure 19: Phrase Structure Tree

10. தமிழ் எழுத்தாளர் சமீப சமூக மாற்றங்களைப் பற்றி எழுதுகிறார்
 தமிழ் எழுத்தாளர் சமீப சமூக மாற்றங்களைப் பற்றி எழுதுகிறார்
 (The Tamil writer is writing about recent social changes.)

(a) Phrase Structure Tree:

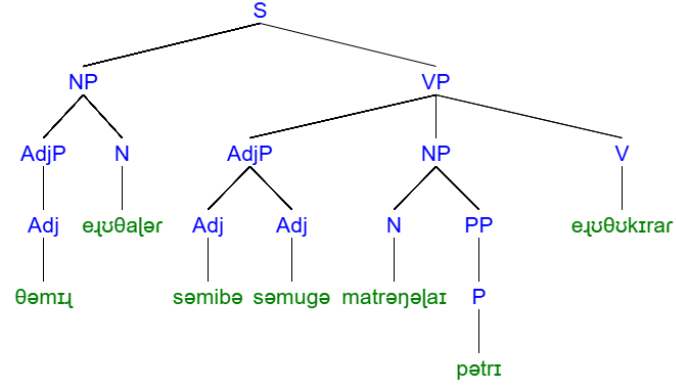


Figure 20: Phrase Structure Tree

(b) Dependency Tree:

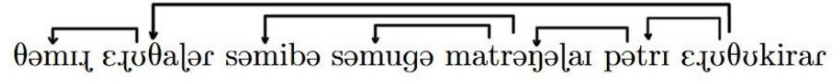


Figure 21: Phrase Structure Tree

7 Historical Context:

Tamil does not belong to the Indo-European family, but rather to the Dravidian family of languages. For the greater part of its history, the influence of foreign languages was quite limited. Tamil extensively borrowed only from Sanskrit and Arabic due to their growing presence in the Indian Subcontinent. Influences of the neighbouring languages of Malayalam, Kannada and Telugu was limited only to the borderlands. However, in the last few centuries, this has undergone a drastic change, especially due to the arrival of the European powers. Due to this, the spoken Tamil language saw the introduction of a wide variety of new words and expressions. This is clearly visible in the case of the influence of English, where many words of the language were directly integrated into spoken Tamil. In the last century, there was a very strong wave of Linguistic Purism and Nationalism, which aimed to promote the usage of "pure" Tamil, and tried to replace the borrowed phrases with native equivalents. But interestingly, most of these changes affected only the spoken variant, and the written variant remained quite untouched. Indeed, Tamil actually exhibits the linguistic phenomenon known as diglossia, which is when a single language has two forms, a "high" variant which is used only in certain situations like in literature, or in formal education. This is in contrast with the case of Hindi and Urdu, which can be considered simply as two scripts for the same language, which is known as Digraphia. The 'High' variant of Tamil is known as செந்தமிழ் [seṇṭamiɻ] while the 'Low' variant is known as கொடுத்தமிழ் [koṭuṭṭamiɻ]. Around 92.36% of the total Tamils in India live in Tamil Nadu, followed by 3.05% in Karnataka, 1.59% in Puducherry, and 0.72% in Kerala.

Bilingualism in Tamil Nadu is around 25.45% while trilingualism is around 3.17%, which do not show a large deviation from India's averages at 26% and 7%. The slight negative deviations could be due to the fact that Tamils tend not to learn other languages since Tamil is the single dominant language in Tamil Nadu.

Outside India, there are a sizeable number of Tamil speakers in countries like Sri Lanka, Singapore, Malaysia and South Africa. There is also a marked presence of people of Tamil origin in countries like Réunion, Guyana, Fiji, Suriname, and Trinidad and Tobago.

8 Writing Systems and Orthography:

The Tamil script is an abugida script that is used to represent the languages of Tamil, Saurashtra, Badaga, Irula. Unlike English and many other languages, Tamil is a syllabic script, and is written from the left to right. The modern Tamil script is derived from the Tamil-Brahmi script, from where another script, known as the vattezhuthu [vəɽɛ.ɻuθu] also descended. This vattezhuthu script was initially used to represent the Tamil language, until the 4th century when the Tamil script was imposed by the Pallavas. The vattezhuthu script was still used in certain parts of the state until the 11th century after which the Chola empire conquered the rest of the state and imposed the Tamil script on the citizens. The use of palm leaves as a medium to store written information ensured that there were no hard strokes in the letters. This also had the side effect of rendering certain symbols like the ூ obsolete. Simultaneously, the Grantha alphasyllabary was formed, which encompassed all the symbols from Tamil, alongside the symbols from the vattezhuthu for sounds that could not be transcribed using the Tamil script. This Grantha influenced the Tamil script to the extent where certain symbols were borrowed from the Grantha to represent sounds like /dʒ/. The

Tamil script has 12 vowels, which are also known as the "உயிரெழுத்து" [ujirɐɻuθu], and 18 consonants known as the "மெய்யெழுத்து" [mɐjɐɻuθu]. There is also another letter that is not often used these days, known as the "ஆய்த எழுத்து" [ajɐθə ɐɻuθu] (ஃ). This symbol was initially used to mark the short /ɐ/ sound, but it fell into disuse. In modern Tamil, it is used to represent phonemes borrowed from other languages which do not have a native representation in the Tamil script [For example: ஃப is used to represent the phoneme /f/ which is not present in modern Tamil].

Script:

க	ங	ச	ஞ	ட	ண	த	ந	ப	ம	ய	ர	ல்	வ	ழ	ள	ற்	ன்
k	ŋ	s	ɲ	t̪ d̪	ɳ	t̪ θ ð	n̪	p b	m	j	r	l	w v	ɻ	ɭ	r	n

Table 15: Consonants

ஜ்	ஸ்	ஷ்	ஸ்	ஹ்	க்ஷ்
dʒ	ʃ	ʂ	s	h	kʂ

Table 16: Grantha Consonants

அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ	ஔ	ஃ
ə	a	i	iː	u	uː	ɛ	e	aɪ	o	oː	aʊ	ək

Table 17: Vowels

9 References:

- [1] Tamil | ethnologue free, . URL <https://www.ethnologue.com/language/tam/>.
- [2] Tamil language and alphabet, . URL <https://www.omniglot.com/writing/tamil.htm>.
- [3] Census of india - 2011, . URL <https://censusindia.gov.in/nada/index.php/catalog/10191/download/13303/DDW-C16-STMT-MDDS-0000.XLSX>.
- [4] Thirukkural text, . URL <https://thirukkural.io/>.
- [5] Thirukkural audio - 1, . URL https://www.youtube.com/watch?v=vZ8zPetRtKw&ab_channel=TamilFire.
- [6] Thirukkural audio - 2, . URL https://www.youtube.com/watch?v=DsBYTs0-lyQ&ab_channel=TamilFire.
- [7] Classical and extended diglossia, . URL <https://ccat.sas.upenn.edu/~haroldfs/messeas/diglossia/node6.html>.