

## Tamil Morphological Features and Annotation Guidelines

Morphological features:

Lexical features*	Inflectional features*	
	<i>Nominal*</i>	<i>Verbal*</i>
<a href="#">PronType</a>	<a href="#">Gender</a>	<a href="#">VerbForm</a>
<a href="#">NumType</a>	<a href="#">Animacy</a>	<a href="#">Mood</a>
<a href="#">Poss</a>	<a href="#">Number</a>	<a href="#">Tense</a>
<a href="#">Reflex</a>	<a href="#">Case</a>	<a href="#">Aspect</a>
<a href="#">Foreign</a>	<a href="#">Definite</a> (DET)	<a href="#">Voice</a>
<a href="#">Abbr</a>		<a href="#">Polarity</a>
<a href="#">Typo</a>		<a href="#">Person</a>
		<a href="#">Polite</a>
		<a href="#">Clusivity</a> (PRON)

	<b>Features</b>	<b>Values</b>
Pronominal type	PronType	Personal (Prs), Reciprocal (Rcp), Article (Art), Interrogative (Int), Relative (Rel), Exclamative determiner (Exc), Demonstrative (Dem), Total (Tot), Negation (Neg), Indefinite (Ind)
Numeral Type	NumType	Cardinal (Card), Ordinal (Ord), Fraction (Frac)
possessive	Poss	Yes
reflexive	Reflex	Yes
foreign word	Foreign	Yes
abbreviation	Abbr	Yes
Wrong spelling	Typo	Yes
gender	Gender	Masculine (Masc), Feminine (Fem), Neuter (Neut)
animacy	Animacy	Animate (Anim), Human (Hum), Inanimate (Inan)
Number	Number	Singular (Sing), Plural (Plur), Dual number (Dual), Trial number (Tri), Collective (Coll)
Case	Case	Nominative (Nom), Accusative (Acc), Instrumental (Ins), Dative (Dat), Ablative (Abl), Allative (All), Benefactive (Ben), Comitive (Com), Locative (Loc), Genitive (Gen), Vocative (Voc)
definiteness	Definite	Definite (Def), Indefinite (Ind)
Verbal form	VerbForm	Finite (Fin), Infinite (Inf), Participle (Part), Gerund (Ger), Verbalnoun (Vnoun)
mood	Mood	Indicative (Ind), Imperative (Imp), Conditional

		(Cnd), Potential (Pot), Desiderative (Des), Necessity (Nec)
tense	Tense	Present (Pres), Past (Past), Future (Fut)
aspect	Aspect	Progressive (Prog), Perfective (Perf), Prospective (Prosp)
voice	Voice	Active (Act), Passive (Pass), Causative (Cau)
polarity	Polarity	Positive (Pos), Negative (Neg)
person	Person	1, 2, 3
polite	Polite	Formal (Form)
clusivity	Clusivity	Inclusive (In), Exclusive (Ex)

**Positional tagging order for morph features:**

1. POS
2. Sub POS
3. Case
4. Tense
5. Person
6. Number
7. Gender
8. Voice
9. Negation

If spelling errors are found, then Typo=Yes

If a word is wrongly split, then (i) Typo=Yes

(ii) the second word is tagged 'goeswith' with the first word

If a word is abbreviated, then Abbr=Yes

**Morph features for each POS in Tamil:**

**1. ADJ**

- **NumType (Numeral type)**

From the syntactic point of view, some numtypes behave like adjectives.

- (i) **Ord**

Ordinal numbers are classified under ADJ in Tamil. Ordinal numbers represent the position or rank of an object or a person.

Example:

இரண்டாவது அறை

## 2. ADV

- **AdvType (adverb type)**

Semantic subclasses of adverbs.

- (i) **Man**

Adverb of manner

Example:

எப்படி

- (ii) **Loc**

Adverb of location

Example:

இங்கே, அங்கே

- (iii) **Tim**

Adverb of time

Example:

இப்பொழுது, அப்பொழுது

- (iv) **Deg**

Adverb of quantity of degree

Example:

மிகவும்

- (v) **Cau**

Adverb of cause

Example:

அதனால்

- **Deixis**

Deixis is typically a feature of demonstrative pronouns, determiners, and adverbs. Its value classifies the location of the referred entity with respect to the location of the speaker or of the hearer.

- (i)**Prox**

The entity is close to the reference point (e.g., to the speaker).

Example:

இங்கே

- (ii)**Remt**

The entity is far away from the reference point (e.g., from the speaker).

Example:

அங்கே

- **PronType**

Pronominal adverbs have this feature of PronType.

- (i)**Tot**

Total (collective) pronoun, determiner or adverb

Example:

எப்பொழுதும்

### 3. INTJ

- **Echo**

Echo or reduplicative words are found in Hindi and other Indian languages. It is found in Tamil too.

- (i)**Rdp**

The word is a copy of a previous word.

Example:

**தனி தனி** 'separately'

**(ii)Echo**

The word rhymes with a previous word but it is not identical to it and typically it does not have any meaning of its own

Example:

**பல்லி கில்லி**

**4. NOUN**

- **Animacy**

Animacy is usually a lexical feature of nouns that mark agreement with the nouns.

**(i)Anim**

Human beings, animals, fictional characters, names of professions etc. are normally animate. Even nouns that are normally inanimate can be inflected as animate if they are personified.

Example:

**பையன் , விலங்குகள்**

**(ii)Inan**

Nouns that are not animate are inanimate.

Example:

**வீடுகள்**

**(iii)Hum**

A subset of animates where the prototypical member is a human being but not an animal.

Example:

**பையன்**

- **Gender**

Gender is usually a lexical feature of nouns

**(i) Masc**

Nouns denoting male persons are masculine.

Example:

**பையன்**

**(ii) Fem**

Nouns denoting female persons feminine.

Example:

**பெண்**

**(iii) Neut**

Nouns denoting gender that are neither masculine nor feminine

Example:

**குழந்தை**

- **Case**

Case is an inflectional feature of nouns in Tamil.

**(i) Nom (Nominative case)**

Nom marks the syntactic subject/ irrational objects of the sentence. No case marker is added.

Example:

**விலங்குகள்** கூண்டினுள் இருந்தன  
**குமார்** பழம் சாப்பிட்டான்

**(ii) Acc (Accusative case)**

Acc marks the object of the sentence in Tamil. -ஐ is the case marker.

Example:

நான் **புத்தகத்தைக்** கொடுத்தேன்.  
குமார் **பழம்** சாப்பிட்டான் - obj position

**(iii) Dat (Dative case)**

Dat marks the indirect object of the sentence. -க்கு is the case marker.

Example:

நான் **அப்பாவுக்கு** கடிதம் கொடுத்தேன்.

**(iv) Gen (Genitive case)**

Gen marks one's possession. -உடைய is the case marker.

Example:

ராமு அப்பாவுடைய பேனாவை எடுத்தான்.

**(v) Voc (Vocative)**

The vocative case is a special form of noun used to address someone.

Example:

அம்மா! இங்கே வாருங்கள்.

**(vi) Ins (Instrumental case)**

Noun is used as an instrument to do something. -ஆல் is the case marker.

Example:

சாவியால் கதவைத் திறந்தேன்.

**(vii) Com (associative)**

Nouns referring to comitative or associative case, marked with -ஓடு/-உடன் in Tamil.

Example:

அவனுடன்/அவனோடு சென்றேன்

**(ix) Ben**

The benefactive case corresponds to the English preposition *for*. In Tamil, it is marked with -க்காக / க்கான

Example:

நான் மகளுக்காக வாங்கினேன்  
நான் குமாருக்கான சட்டையைக் கொடுத்தேன்

**(x) Loc**

The locative case often expresses location in space or time. -இல் is the locative case marker.

Example:

நான் வீட்டில் உள்ளேன்.

**(xi) Abl**

Ablative case denotes direction from some point. -இலிருந்து



Example:

நான் வங்கியிலிருந்து பணம் எடுத்தேன்

**(xii)All**

The allative case expresses direction to something

Example:

நான் கடைக்குப் போனேன் .

- **Verbform=Gdv**

Gerunds in Tamil are considered as noun

Example:

உடற்பயிற்சி செய்தல் மனதுக்கு புத்துணர்ச்சி தரும்  
உடற்பயிற்சி செய்வது மனதுக்கு புத்துணர்ச்சி தரும்

- **Number**

The number is an inflectional feature, indicating the number of nouns present in the context of the sentence.

**(i)Sing**

A singular noun denotes one person, animal or thing.

Example:

பையன்

**(ii)Plur**

A plural noun denotes several persons, animals or things.

Example:

பையன்கள்

**(iii)Dual**

A dual noun denotes two persons, animals or things.

Example:

இருவர்

#### (iv)Tri

A trial noun denotes three persons, animals or things.

Example:

மூவர்

#### (v) Coll

Collective or mass or singulare tantum is a special case of singular. It applies to words that use grammatical singular to describe sets of objects, i.e. semantic plural.

Example:

ஆட்டு மந்தை வந்தது.

- If a word is possessive, then Poss=Yes
- Person[subj] (Person agreement with subject)

#### (i) Third person subject

Subject agrees with the third person in case of nouns.

Example:

மழை பெய்தது.

### 5. PROPN

- **Case**, as seen in NOUNs, all the same cases apply to PROPn as well.
- **Animacy**, as seen in NOUNs, all the same cases apply to PROPn as well.
- **Gender**, as seen in NOUNs, all the same cases apply to PROPn as well.
- NameType= (i)Geo (ii)Prs (iii)Sur (iv)Nat (v)Com (vi)Pro (vii)Oth
- If a word is possessive, then Poss=Yes

#### Person[subj] (Person agreement with subject)

#### (i) Third person subject

Subject agrees with the third person in case of proper nouns.

Example:

குமார் வாங்கினான்

### 6. VERB

- Animacy

Animacy is usually a lexical feature of nouns that mark agreement with the nouns.

**(i)Anim**

Human beings, animals, fictional characters, names of professions etc. are normally animate. Even nouns that are normally inanimate can be inflected as animate if they are personified.

Example:

வந்தான், வந்தது

**(ii)Inan**

Nouns that are not animate are inanimate.

Example:

வந்தது

**(iii)Hum**

A subset of animates where the prototypical member is a human being but not an animal.

Example:

வந்தான்

Aspect= (i)Imp (ii)Prog (iii)Perf

Manual? Clusivity /(psor)/subj/obj= (i)In (நாம்-I+you) (ii)Ex (நாம்-I+they)

Verbform= (i)Fin (ii)Inf (iii)Part

Gender= (i)Mas (ii)Fem (iii)Com (iv)Neut

Manual? VerbType= (i)Cop (ii)Light

- **Voice**

Voice expresses relation between the participants in a narrated event

**(i)Act**

Subject is the doer of the action and the verb can take a direct object in active voice constructions.

Example:

சீதா வேலையைக் செய்தாள்

**(ii)Pass**

In passive voice constructions, the subject expresses the theme/ becomes the patient of the verb. -படு is the passive marker.

Example:

வேலை சீதாவால் செய்யப்பட்டது.

**(iii)Cau**

Action is made to be done by another person in causative constructions.

Example:

சீதா ராமுவை வேலை செய்வித்தாள் .

Mood= (i)Ind (ii)Cond (iii)Jus (iv)Des (v)Int

Number= (i)Sing (ii)Plur

Polarity= (i)Pos (ii)Neg

**Person**

Person is morphologically seen in Tamil verbs

**(i)1**

First person

Example:

வந்தேன்

**(ii)2**

Second person

Example:

வந்தாய்

**(iii)3**

Third person

Example:

வந்தான்

## 7. ADP

AdpType (adposition type)

The adposition comes after a noun in Tamil.  
= Post | Case= Inst

## 8. AUX

Animacy= (i)Anim (ii)Inan (iii)Hum (iv)Nhum

VerbType= (i)Aux (ii)Mod

Aspect= (i)Hab (ii)Imp (iii)Perf (iv)Prosp (v)Prog (vi)Perf

Voice= (i)Pass

Mood= (i)Imp (ii)Cond (iii)Pot (iv)Int

### Person

Person is seen in auxiliaries as well

#### (i)1

First person

Example:

செய்ய வருகிறேன்

#### (ii)2

Second person

Example:

செய்ய வருகிறாய்

#### (iii)3

Third person

Example:

செய்ய வருகிறான்

Polarity= (i)Pos (ii)Neg

## 9. CCONJ

### ConjType

Under CCONJ, mathematical operators are seen

#### (i)Oper

Operators that are expressed using words

Example:

ஒன்று கூட்டல் இரண்டு

## 10. DET

?Animacy= (i)Anim (ii)Inan (iii)Hum (iv)Nhum

Manual? Deixis= (i)Prox (ii)Remt

Number= (i)Pauc (சில) (ii)Grpl (பல)

PronType= (i)Dem (ii)Tot (iii)Rcp (எல்லாரும் எல்லாரையுமே)

If reflexive, then Reflex=Yes

## 11. NUM

### NumForm

Different forms of numbers are found in the texts. The following types are noted.

#### (i)Combi

Combined form of numerals and Tamil alphabets

Example:

15ஆம் நூற்றாண்டு

#### (ii)Word

Numerals are written in word forms.

Example:

பத்து பேனாக்கள்

#### (iii)Digit

Numerals/ Numbers like 1234...

Example:

3 மணி

#### (iv)Roman

Roman numbers/ scripts

Example:

கிங் சார்லஸ் IV

NumType= (i)Card → Frac

### **Num [subj]: Number agreement with the subject**

Finite verbs in Tamil agree in number with the subject of the sentence.

#### **(i)Sing (singular subject)**

Only one noun/pronoun/proper noun is found in the subject position and it agrees with the verb's plurality.

Example:

ராமு அப்பாவுடைய பேனாவை எடுத்தான்.

#### **(ii)Plur (Plural subject)**

Two or more nouns/pronouns/proper nouns are found in the subject position and it agrees with the verb's plurality.

Example:

பறவைகள் பறந்தன.

## **12. PART**

### **PartType**

Types of particles are found in various tagsets and are highly language-specific. The following are found in Tamil:

#### **(i) Emp (particle of emphasis)**

Emphatic marker is marked with -ஏ in Tamil.

Example:

குமார் தானாகவே வந்தான்.

#### **(ii) Inf (Infinitive marker)**

Infinitive marker is marked with -அ in Tamil.

Example:

நான் வர வேண்டும்

#### **(iii) int (question particle)**

Question marker is marked with -ஆ at the end of the word in Tamil for yes/no types.

Example:

குமார் வந்தனா?

### NEW ADDITION:

(iv) incl (inclusive particle)

Inclusive pronouns / nouns are denoted by -உம் in Tamil

Example:

நானும் நீயும் போனோம்.

## 13. PRON

Clusivity/(psor)/subj/obj

(i)In

Inclusive of the second person

Example:

(நாம்-I+you)

(ii)Ex

Exclusive of second person

Example:

(நாம்-I+they)

Animacy= (i)Anim (ii)Inan (iii)Hum (iv)Nhum

Case= (i)Nom (ii)Acc (iii)Dat (iv)Gen (v)Voc (vi)Ins (vii)Com (associative) (viii)Cau

(செய்வதற்கு) (ix)Ben (x)Loc (xi)Abl (xiii)All

Deixis= (i)Prox (ii)Remt

DeixisRef= (i)1 (ii)2

**Number**

(i)Sing

One pronoun in context

Example:

நான்

(ii)Plur



Two or more pronouns in context

Example:

**நாம்**

**Person**

**(i)1**

First person

Example:

**நான்**

**(ii)2**

Second person

Example:

**நீ**

**(iii)3**

Third person

Example:

**அவன்**

PronType= (i)Prs (ii)Rcp (iii)Int

(iv)rel

(v) Dem

(vi)

**Politeness**

If reflexive, then Reflex=Yes

**Person[subj] (Person agreement with subject)**

**(i) First person subject**

First person agreement with subject

Example:

**நான் வந்தேன்**

### (ii) Second person subject

Second person agreement with subject

Example:

நீ வந்தாய்

### (iii) Third person subject

Third person agreement with subject

Example:

அவன் வந்தான்

- If a word is possessive, then Poss=Yes

## 14. SCONJ

### ConjType (Conjunction types)

Under SCONJ, the following types are seen:

#### (i)Pred

Subordinating conjunction introducing a secondary predicate

Example:

அவன் வரவில்லை என்று சொன்னான்.

#### (ii)Comp

Comparing conjunction

Example:

நூலைப் போல சேலை

## 15. PUNCT

### PunctType

Punctuation type is classified in details seen below:

#### (i) Peri

Period at the end of the sentence or clause

Example:

மழை பெய்தது.

## (ii) Elip

Ellipsis

Example:

நான் வருவேன்... என்று சொன்னான்.

## (iii) Qest

Question mark

Example:

அவன் எங்கே?

## (iv) Excl

Exclamation mark

Example:

ஆஹா!

## (v) Quot

Quotation marks including single and double quotes.

Examples:

"வா இங்கே" என்று சொன்னாள். - double quotes (marks conversations)

இந்த புத்தகத்தின் பெயர் 'அக்னிச் சிறகுகள்'. - single quotes (marks specific book/poem/text names and headlines)

## (vi) Brck

Bracket- abbreviations, marks something similar to previously stated/ something unimportant.

Example:

ஐக்கிய நாடு (ஐநா)

## (vii) Comm

Comma- separates list of words

Example:

நான் ஆப்பிள், ஆரஞ்சு, திராட்சை வாங்கினேன்.

**(viii) Colo**

Colon- enlists

Example:

முக்கனிகள்: மா, பலா, வாழை

**(ix) Semi**

Semicolon- separates clauses

Example:

அவன் போனான்; நான் வந்தேன்.

**16. SYM**

PunctType

Punctuation type under symbols are only these two as seen below:

**(i) Dash**

Dash/hyphen

Example:

உப்பு-தேவையான அளவு

**(ii) Slsh**

Slash/backslash

Example:

இட்லி/தோசை

---

[Abbr=Yes](#)

?? AdvType

Should look

Style= (i)Arch (ii)Coll (iii)Expr (iv) Form (v)Rare (vi)Sing (vii)Vrnc (viii)Vulg

This is applicable to any wordform if the following characters are found.

### **NOUN:**

(i) Gender : Fem(f), Masc (m), Neut (n) , Com (mf)

(ii) Animacy: Anim, Hum, Inan, ??Nhum

(In Tamil, animacy feature for the syntactic analysis (esp. Agreement purpose) not required as the Gender info itself predicts the Animacy (but not Anim Vs Inan, Nhum). However, in semantic analysis, we can re-look.

(iii) Number: Sing, Plur

(iv) Case:

Core: Nom, ACC

Non-Core: Ben, Com, Dat, Gen, Ins, Voc

Local: Abl, Loc, ??All

Derived Nouns:

Verb-Noun:

Tense:

---

---

### **VERB:**

(i) Verb Form: Fin, Inf, Part, Conv, Ger, ??Vnoun

(ii) Mood: Imp, Cnd,Pot, opt, Nec, ??Int,

(iii) Tense: Fut, Past, Pres

(iv) Aspect: Prog, Perf (auxiliaries), ??Hab (future), (AUX) (vanwu+koVNtiru) \*(vanwu+koVIY+iru), marked on AUX,

(v) Voice: ?Act, Pass (AUX) patu/?peVrYu, Caus (on verbs (ceVyvi, patippi etc.. “vi and pi” endings), Reciprocity

(vi) polarity: Neg, Pos (-a- AUX, VERB)

(vii) Person: 1,2,3 (AUX, VERB)

(viii) Politeness ??

(ix) Clusivity: Ex, In (1,pl) (

---

---

**ADP:** AdpType=Post|Case=Acc

---

