# Word order with psych verbs in Tamil language

Abstract:

This project analyzes the word order pattern for psych verbs in Tamil language. Psych verbs in general especially Experiencer Object (EO) verbs are considered to be peculiar among linguists as they do not follow the rules that has been applied for verbs in the grammar theory. One of their main peculiarities is that they deviate from the preferred word order. This has been already proved for German language. Given the German EO verbs, the object which act as an experiencer is preferred to occur at the first position in the sentence. Following these facts, this project addresses the nature of psych verbs in Tamil and determine if there is any deviation with the word order. Due to the dominance of light verb construction in modern Tamil, the availability of EO verbs in the form of main verbs is very less. Owing to this limitation, this work has taken both experiencer subject (ES) and experiencer object (EO) into account and examine them in the name of psych verbs instead of focusing only on EO. Two annotation datasets were created for the purpose of finding word order, the first is for psych verbs and the second is for non-psych verbs. The second one is just to compare the word order and their effects between psych and non-psych properties. The analysis was carried out using generalized models through R. The results show that there is no deviation or preference for the with psych verb properties in Tamil. In fact, subject predominantly occurs at the first position in Tamil sentences, though it is free word order language i.e., subject before object and vice versa is acceptable.

Introduction:

According to linguistics, the verbs can be classified broadly into psychological (psych) and non-psychological (non-psych) verbs. The former is the one where there is no physical movement involved but only the emotional state and the later deals with physical movement. (1) and (2) are the examples for psych and non-psych verbs respectively. In (1), the verb ‘fears’ denotes a certain mental state of Clara, whereas in (2) the verb ‘moved’ denotes the physical motion. This work mainly focused on the psych verbs and their properties.

(1) Clara *fears* the noise.

(2) He *moved* the chair.

In psych verb, one of the arguments i.e., subject or object is in certain mental state. The argument being in the mental state is called experiencer and the other which stimulates the mental state is called stimulus. It further leads to two sub types in psych verbs namely Experiencer Subject (ES) and Experiencer Object (EO). This work focused on both types. (3) and (4) are the examples for ES and EO respectively. In (3), the argument which experiences the emotion is a subject ‘I’. In (4), the argument which experiences the emotion is an object ‘everyone’.

(3) *I* love nature.

(4) His jokes amused *everyone*.

Since decades, the psych verbs particularly EO verbs are quite among linguistic researchers due to their special nature. According to (Belletti & Rizzi, 1988), (Landau, 2010) and many others, the psych verbs behave syntactically and semantically unusual, in contrast to non-psych verbs. The work of (Temme, 2019) proved that the experiencer generally occurs first in German language and mainly dative EO psych verbs in German language prefers first position in the sentence. In fact, the analysis of syntactic pattern of German EO verbs by (Masloch et al., 2021) is being the main motivation for the existence of this work. Since there is no relevant research work on psych verbs in Tamil so far, that again encouraged to initiate this work. The main difference between the previous related works and this work is the later includes both ES and EO for the word order analysis, due to the limitation of EO as main verbs in modern Tamil.

Background:

Tamil and its grammar:

Tamil (Thamizh) is one of the most oldest languages in the world, which is surviving till date. Tamil is dated to approximately 200 BCE (Steever, 2019). Yet, the origin of this language is still unknown. It belongs to the family of Dravidian languages and is natively spoken by the people of Tamil Nadu and Puducherry in India and in some northern parts of Sri Lanka. Being the oldest language, it can be classified into three stages such as Old Tamil (c .300 BCE to 700 CE) , Middle Tamil (700 to 1600 CE) and Modern Tamil (1600 CE to the present) (Steever, 2019). This work deals only with the written form of formal modern Tamil.

Tholkaapiyam is the most ancient literature of Tamil dated to 200 BCE and it is the source for Tamil grammar. Nannool dated to 1000 CE derived from tholkappiyam is considered the second biggest source of grammar. In Tamil, there are 12 vowels and 18 consonants. They form the basic letters.

Vowels

அ ஆ இ ஈ உ ஊ எ ஏ ஐ ஒ ஓ ஔ

*a aa e ee u oo ye yae i o oa ow*

Consonants

க் ங் ச் ஞ் ட் ண் த் ந் ப் ம் ய்

*ik ing ich in it in ith inth ip im iy*

ர் ல் வ் ழ் ள் ற் ன்

*ir il iv izh il ir in*

The vowels and consonants are combined and forms the compound letters. There are 216 compound letters i.e., 12 vowels \* 18 consonants = 216 compound letters. Also, there is 1 special letter ‘ஃ (ak)’. Therefore, there are totally 247 letters in Tamil. 12 vowels + 18 consonants + 216 compound letters + 1 special letters = 247 letters.

Rules of Tamil grammar:

Tamil is agglutinative in nature and explicitly suffix the particles at its end. Though there are plenty of rules in Tamil grammar, only certain important rules (Arden, 1891), (இராச திருமாவளவன், n.d.) that is necessary for this work are explained here.

Rule 1: The verb always stands at the end and concludes the sentence. In (5), the verb படித்தான் (*padithaan -* learnt) lies at the end.

(5) பையன் பாடத்தை படித்தான்.

*paiyan paadathai padithaan.*

boy lesson learnt.

Boy learnt the lesson.

Rule 2: The verb should agree with the subject. Tamil verbs indicate the tense, the gender of subject including its honor (formal or informal) and the number (singular or plural) of subject. When the subject is a first person pronoun நான் (*naan* - I) or second person நீ (*nee* - you), it is not necessary to include the subject in sentence as the verb defines them. In (6), the verb வியந்தேன் (*viyandhen – I wondered*) itself indicates that the noun is a first person subject நான் (*naan* - I) and hence it is not necessary to mention the subject.

(6) வியந்தேன் = விய + ந்து + ஏன்

*= viya* + *indhu* + *yaen*

(verb lemma) (past tense) (first person singular)

= viyandhen - I wondered

Rule 3: The nouns are classified into 2 major classes such as rational and irrational. Rational nouns denote human beings and Irrational nouns denote animals and inanimate things.

Rule 4: There are no prepositions in Tamil, but only postpositions. These postpositions affix with nouns and pronouns at the end and inflect them in order to denote the new meaning. In (7), the postposition இல் (*il*) suffixed with the noun and denotes a new phrase.

(7) பசுவில் = பசு + இல்

= *pasu* + *il*

(noun) (postposition)

= cow + in

pasuvil = in a cow

Rule 5: There are no articles in Tamil.

Rule 6: The subordinate clause which expresses purpose, reason and condition always precedes the main clause. In (8), the reason (because) occurs at the initial position and the main clause comes after it.

(8) இதனால் 1\_லட்சத்து\_80\_ஆயிரம் பயணிகள் தங்கள் idhanaal 1\_lachathu\_80\_aayiram payanigal thangal

because.of.this 1.lakh.80.thousand passengers their

இடங்களுக்குச் செல்ல முடியாமல் தவித்தனர்.

idangalukku sella mudiyaamal thavithanar.

places to.go not.able.to were.suffered.

Because of this, 1 lakh 80 thousand passengers were suffered by not able to go to their places.

Rule 7: There should be only one finite verb in a sentence and the other verbs are in the form of non-finite in terms of either infinitive or verbal participles. These non-finite verbs usually end the subordinate clause. In (9) there is only one finite verb வியந்தனர் (*viyandhanar -* wondered) and the another verbs பார்த்து (*paarthu –* having seen) is in the non-finite verbal participle form. This non-finite verb ends the subordinate class as per the rule.

(9) ஒரு சிறுவன் அதில் கலந்து \_கொண்டதைப் பார்த்து

*oru siruvan adhil kalandhu\_kondadai paarthu*

one little.boy in.it having.participated having.seen

எல்லோரும் வியந்தனர்.

*ellorum viyandhanar.*

everyone wondered.

Everyone wondered on seeing one little boy having participated in it.

Rule 10: There are 8 cases in Tamil and these cases are affixed with nouns or pronouns.

Case1: Nominative. In this case, it is the noun itself. For e.g., நரி (*nari* - a Jackal)

Case2: Vocative - ஏ (*yae*)

(10) நரி + ஏ = நரியே

*nari* + *yae* = *nariyae*

jackal + O = O Jackal

Case3: Genitive - இன் (*in*)

(11) புலி + இன் = புலியின்

*puli* + *in = puliyin*

tiger + of = of a tiger

Case4: Dative – கு (*ku*)

(12) புலி + கு = புலிக்கு

*puli* + *ku* = *pulikku*

tiger + to = to a tiger

Case5: Accusative - ஐ (i*)*

(13) புலி + ஐ = புலியை

*puli* + *i* = *puliyai* - a tiger

Case6: Locative – இல் (*il*)

(14) புலி + இல் = புலியில்

puli + il = puliyil

tiger + in = in a tiger

Case7: Instrumental - ஆல் (*aal*)

(15) புலி + ஆல் = புலியால்

*puli* + *aal* = *puliyaal*

tiger + by = by a tiger

Case 8: கண் (*kan*). It is specific only to Tamil and there is no English equivalent for this case.

Study of word order through annotation:

Tamil is a free word order language where both SOV and OSV is allowed with one strict rule that verb should be at the end in both cases. Poetic literatures of Tamil is an exception to the final verb position, where verb can come in any of the three positions so as to enhance the aesthetic. All the other forms of Tamil except poetics should have final verb position. There are several factors which leads to the scrambling of SO and OS including case of object, length of subject and focus etc. In order to determine the word order and its factors, it is necessary to annotate the sentences. As there was no annotated dataset available for Tamil psych verbs, creating manual annotation following the rules of Tamil grammar became one of the major parts of this work. As a first step, 10 psych verbs in Tamil were found out. They are விய (*viya* - to wonder), மகிழ் (*magizh* - to be happy), விரும்பு (*virumbu* - to like), களி (*kali* - to ecstasize), நெகிழ் (*negizh* - to delight), அஞ்சு (*anju* -to fear), திகை (*thigai* - to shock), தவி (*thavi* - to suffer), வருந்து (*varundhu* - to be sad), வெறு (*veru* - to hate). For comparison purpose, two non-psych verbs such as படி (padi - to read / to study), தேடு (thaedu - to search / to find) were also taken.

Leipzig corpora (*Download Corpora Tamil*, n.d.) is the data source for this work, from where the sentences containing these verbs were taken. This corpus is a set of modern written Tamil collected from community, news, newscrawl, web and Wikipedia for the years of 2011, 2016, 2017, 2019, 2020 and 2021. In modern Tamil, the way words are uttered in spoken is different from the way it is written. For e.g., the word அவர்கள் (*avaragal*) is used in written to express ‘they’, but the same has been uttered as ‘அவங்க’ in spoken. Apart from the utterance variation, all the other grammatic, syntax and semantic concepts are same between them. This work focused only on the formal written Tamil. Tamil has various geographical dialects, yet the major distinction can be found between the dialects of Indian Tamil and Sri Lankan Tamil. However, the grammatical concepts are exactly similar between them other than dialect. This corpus includes both Indian and Sri Lankan Tamil. Sentences were extracted from this corpus with the help of Pandas tool using Python language. The initial plan was to annotate 200 sentences for each verb. Due to data constraint, the no. of sentences varies for every verb. Especially, the count for களி (*kali* - to ecstasize) and நெகிழ் (*negizh* - to delight) is very less.

Certain limitations were followed to perform the annotation. Sentences where there is no subject or object were excluded. The usage of light verbs to express the psych properties is quite high in modern Tamil. In light verb construction, the actual verb denoting the psych meaning comes in a noun form and is followed by the light verb. Yet, both combines and deliver the same psych meaning.

These light verbs cannot standalone and give meaning. Among these 10 verbs, only விரும்பு (*virumbu* - to like) has no light verb construction in any of its sentences.

The common light verbs used to express the psych properties are அடை (*adai –* to attain), அளி (*ali –* to give), திளை (*thilai* – to indulge), ஆழ் (*aazh*, to immerse) and செய் (*sei* – to make).

In (16), the actual psych verb occurs in noun form வியப்பை (*viyappai,* wonder) and the light verb அளித்தன (*alithana –* gave) comes after it.

(16) அந்தப் பெண்ணின் பிடிவாதமும், வீம்பும் எனக்கு வியப்பை

*andha pennin pidivaadhamum, veembum enakku viyappai*

that women’s adamant, stubbornness me wonder

அளித்தன.

*alithana.*

gave.

That women’s adamant and stubbornness gave me wonder.

The light verb usage is very vast in EO type than ES. (16) is also an EO with light verb where the experiencer is an object எனக்கு (*enaku*, me). Certain verbs like மகிழ் (*magizh* - to be happy), நெகிழ் (*negizh* - to delight), அஞ்சு (*anju* -to fear), வருந்து (*varundhu* - to be sad) are having their own EO form. The particle called இ (*e*) will be suffixed to the verb lemma in order to denote the EO form. With these EO form of verbs, light verb construction is absent. But still, light verb construction is possible among these verbs without their EO forms. Yet, the count of light verb construction is high for these verbs though they have EO forms.

Similar to light verbs, the usage of dative subjects is also very common in modern Tamil. According to Tamil grammar, the subject should be nominative without any particle inflection. But today, subject in dative form is widely used and this violates the subject verb agreement. When the subject is dative, the verb comes in the form of noun and still delivers the meaning of verb. In (17), the subject எனக்கு (*enakku* - I) is in dative and the verb is in noun form வருத்தம் (*varutham -* sad). Though, they deliver the meaning of ‘I feel sad for her.’

(17) எனக்கு அவள்மீது வருத்தம்.

*enakku avalmeedhu varutham.*

I her.on sad (noun).

Through this work, it was identified that in modern Tamil the light verb construction and dative subjects are used only for psych verbs. They are not used in any non-psych verbs. Sentences containing light verbs and dative subjects were excluded from annotation work, as they modify the original form of psych verbs which actually playing a central role in this work. Sentences with passive voice were also excluded.

Sentences containing both subject, object as well as the psych verb as main finite verb are selected for the annotation. Annotation work determined the type of psych verb, word order and factors affecting word order such as case of subject, case of object, animacy of subject, animacy of object, pronoun of subject, pronoun of object, subject syllable, object syllable and focus. There are totally 1412 annotated sentences. In this annotation, subjects with and without phrase are considered as subject. In Tamil, there is only one finite verb and all the other verbs will be non-finite either in the form of verbal participle or infinite. The finite verbs are considered as verb. The object cases are broadly classified into clausal and non-clausal. The clause that ends with non-finite verbs are considered as object. When there is no non-finite verbs, then the elements denoting the object is considered as non-clausal object.

(18) பல்லாயிரக்கணக்கானவர்கள் இந்தப் புடவையைப் பார்த்து வியக்கின்றனர்.

*pallaayirakanakaanavargal indha pudavaiyai paarthu viyakkinranar.*

thousands.of.people this saree having.seen are.wondering.

Thousands of people are wondering by having seen this saree.

Psych verbs are generally defined in terms of experiencer and stimulus, where the experiencer can be a subject and the stimulus can be an object or vice versa. In (18), the subject is பல்லாயிரக்கணக்கானவர்கள் (*pallaayirakanakaanavargal* – thousands of people), verb is வியக்கின்றனர் (*viyakkinranar* – are wondering). Here, the subject is experiencing the verb and hence subject is an experiencer. The clause இந்தப் புடவையைப் பார்த்து (*indha pudavaiyai paarthu* - having seen this saree) is acting as the stimulus. In the sense, this clause stimulates the experiencer subject in order to wonder i.e., thousands of people are wondering by having seen this saree. Therefore, the entire clause has been considered as an object. Most of the sentences have clausal object and very less sentences have non-clausal object with psych verbs. Among these clausal objects, verbal participle cases is more prevalent and is followed by special verbal participle. In (18), the word பார்த்து (*paarthu*, having seen) is the verbal participle form of verb பார் (*paar* - to see). Verbal participles namely கண்டு (*kandu,* having seen), கேட்டு (*kaetu* – having heard) and special verbal participle especially என்று (*endru* ), என (*ena*) are very common and more frequently used with Tamil psych verbs. Apart from subject and object, the other additional factors are animacies and pronouns of subject and object. The frequent animacy of subject and object is rational and none respectively. The occurrence of pronouns are also very less in both subject and object. These animacies and pronouns are having less impact towards word order in Tamil.

One of the main factors is focus and syllables. Generally, in Tamil the most common way to emphasize subject, object and verb is affixing particles like தான் (*thaan*), ஏ (*yae*) or ஓ (*o*) to them. The other way is to place the elements before finite verb. When the elements occurs before verb, then it is considered to be more emphasized. It applies only for subject and not for object, because the standard word order of Tamil is SOV and by default object lies before verb meaning it is very common and doesn’t mean to emphasize.

The syllable of elements which is lengthy i.e., the high no. of words is preferred to occur at the first position in a sentence. But this is not compulsory. As per the annotation, it shows that when object is having lengthy syllable compared to subject, object always occurs first. The same applied for subject also. In this annotation, the SOV word order is more than the OSV.

Analysis of word order using Generalized models:

In order to make the analysis more effective and efficient, the word order was analyzed using generalized models through R programming. Before proceeding with the models, some basic pre-processing of dataset was carried out. Likely, removing the unwanted columns, converting categorical variables into factors and numerical variables into logs for getting the good results.

References:

Arden, A. H. (1891). *A Progressive Grammar of Common Tamil*. Society for Promoting Christian Knowledge.

Belletti, A., & Rizzi, L. (1988). Psych-verbs and θ-theory. *Natural Language and Linguistic Theory*, *6*(3), 291–352. https://doi.org/10.1007/BF00133902

*Download Corpora Tamil*. (n.d.). Retrieved August 25, 2023, from https://wortschatz.uni-leipzig.de/en/download/Tamil#tam\_newscrawl\_2011

Landau, I. (2010). *The locative syntax of experiencers*. MIT Press.

Masloch, S., Poppek, J. M., Robrecht, A., & Kiss, T. (n.d.). *Syntactic Pattern Distribution Analysis of Experiencer-Object Psych Verbs*.

Steever, S. B. (2019). *The Dravidian Languages*. Routledge.

Temme, A. (2019). *The peculiar nature of psych verbs and experiencer object structures*. https://doi.org/10.18452/19889

*வேற்றுமைத் தொகை | தமிழ் இணையக் கல்விக்கழகம் TAMIL VIRTUAL ACADEMY*. (n.d.). Retrieved August 22, 2023, from https://www.tamilvu.org/ta/courses-degree-a021-a0214-html-a0214211-6615

Appendix:

A.1. Annotation guidelines:

## **Dataset summary:**

The name of the dataset is ‘annotation\_tamil\_psych\_verbs’. It is in excel format (.xlsx). This dataset is a collection of annotations of Tamil sentences containing the psych verbs. The main purpose of creating the dataset is to analyze the word order pattern of Tamil psych verbs and to determine which pattern is most prevalent, whether it is SOV or OSV.

The dataset contains 1412 rows and 19 columns. Rows represents the data and columns represents the annotation patterns. These annotation patterns are the main factors affecting the word order.

## **Annotation:**

Below are the column names and their annotation guidelines.

1. **ID:**

This column contains the unique ID, which identifies the sentence. The values of the column are finite integers that starts from 1 to 1412.

1. **Verb:**

This column contains the lemma of psych verbs. There are 10 psych verbs in this column and they are ‘விய (viya - to wonder), மகிழ் (magizh - to be happy), விரும்பு (virumbu - to like), களி (kali - to ecstasize), நெகிழ் (negizh - to delight), அஞ்சு (anju - to fear), திகை (thigai - to shock), தவி (thavi - to suffer), வருந்து (varundhu - to be sad) and வெறு (veru - to hate)’.

1. **Sentence:**

This column contains the sentences in Tamil. (1) is the sample sentence from the dataset.

1. அங்கே அவர்கள் லோகாம்பிகையைக் கண்டு மகிழ்ந்தனர்.
2. **EO / ES:**

This column defines the type of psych verbs. Psych verbs are classified into 2 major groups such as EO and ES. The values of the column are ‘EO’ and ‘ES’. EO means Experiencer Object and ES means Experiencer Subject. When the object of a sentence is an experiencer, then the value is EO. When the subject of a sentence is an experiencer, then the value is ‘ES’.

In (2), object (Chidhambaram) is an experiencer and hence the value of this column is ‘EO’.

1. சிவந்தியின் கோபம் **சிதம்பரத்தை** வெகுவாக

*Sivanthiyin kobam* ***Chidhambarathai*** *veguvaaga*

of.Sivanthi anger (S) **Chidhambaram**(O) fastly

அச்சுறுத்தியது.

*achurithiyadhu.*

frightened (V).

The anger of Sivanthi frightened Chidhambaram fastly.

1. **Word\_order:**

This column contains the word order pattern of psych verb sentences. The values are ‘SOV’ and ‘OSV’. SOV means Subject Object Verb and OSV means Object Subject Verb. When the sentence is having SOV order, then the value is ‘SOV’. Else, it will be ‘OSV’.

In (3), the order of sentence is OSV and hence the value of this column is ‘OSV’.

1. அணில் வந்த\_காரணத்தை அறிய அனைவரும் விரும்பினர்.

*anil vandha\_kaaranathai ariya* *anaivarum virumbinar.*

squirrel reason.of. arrival to.know(O) everyone (S) wanted (V).

Everyone wanted to know the reason of squirrel’s arrival.

1. **Phrasal\_subject:**

This column indicates whether the subject contains more than one word or not. The values of the column are ‘yes’ and ‘no’. In a sentence, the subject can be a single word: அரசர் (*arasar*-King) or pronoun: அவன் (*avan* - He) or more than one word: (4) is the subject with more than one word.

When the subject contains more than one word, then the value is ‘yes’. There is an exception to it. Sometimes, a single name has more than one word. For e.g. பால் தாக்கரே (Pal Thakare). In these cases, the value is ‘no’. In addition, when the subject has only one word, then also the value is ‘no’.

In Tamil when the temporal and locative occurs before the subject, then they belongs to the subject. Subject with temporal: சங்க காலத்தில்(temporal) மக்கள்(subject) (*sanga\_kaalathil makkal -* people in sangam period) and subject with locative: மெரினாவில்(locative) பல்லாயிரக் கணக்கில் மக்கள்(subject) (*merinavil pallaayira kanakkil makkal* – thousands of people in Merina beach). These two examples are also considered as subjects.

1. காட்டு\_விலங்குகளை வேட்டையாடச் சென்ற ஒரு சோழ\_அரசன்

*kaatu\_vilangugalai vaetaiyaada sendra oru chozha\_arasan*

wild.animals to.hunt went one chola.king

One chola king who went to hunt the wild animals.

1. Clausal\_object:

This column indicates whether the object is clausal or not. The values of this column are ‘yes’ and ‘no’. In general, an object can be a single-word noun: பெண்ணை (*pennai -* women), pronoun: என்னை (*ennai -* me), phrasal noun: அரசியல் தலைவர்களை (*arasiyal thalaivargalai -* government officers) or clausal noun. Clausal noun is the one which has a verb in it.

In (5), bold words are the clausal noun that is an object. It has its own verbal form ‘**கண்டு** (*kandu* – by seeing)’ and hence it is a clausal noun. So, the column’s value is ‘yes’.

If the object of a sentence is clausal noun, then the value is ‘yes’. The prevalence of clausal object is high in Tamil psych verb sentences. If the object is other than clausal i.e., single-word noun or pronoun or phrasal noun, then the value is ‘no’.

1. **உங்கள் திறமையை கண்டு** அடுத்தவர்கள் வியப்பார்கள்.

***ungal thiramaiyai kandu*** *aduthavargal viyapargal.*

**your talent by.seeing (O)** others (S) will.wonder (V).

Others will wonder **by seeing your talent**.

1. **Case\_of\_subject:**

This column contains the cases of subject. In general, subject is always a nominative case. The values of the column are ‘nominative’ and ‘nominative (phrasal)’. When the subject is not phrasal, this column value will be ‘nominative’. Else, the value will be ‘nominative (phrasal)’.

1. **Case\_of\_subject\_inTamil:**

This column contains the Tamil grammatical terms for subject’s case. In English, the case of subject is called nominative. In Tamil, it is called முதல் வேற்றுமை (*1st vaetrumai*) or எழுவாய் வேற்றுமை (*yezhuvai vaetrumai*). The values of the column are ‘1st vaetrumai (yezhuvai)’, ‘ennummai’ and ‘mutrummai’.

For ‘1st vaetrumai (yezhuvai)’, no particles are affixed to the noun. When the subject is

affixed with particle ‘உம் (*um*)’ like ‘அனைவரும் (*anaivarum - everyone*)’, ‘எல்லோரும் (*ellorum - all*)’, ‘பலரும் (*palarum* - many)’ etc., then the value is ‘mutrummai’. If the subject is more than one like ‘அவனும் அவளும் (*avanum avalum* – He and she)’ or in other words if the subject has more than one ‘உம் (*um*)’, then the value is ‘ennummai’.

1. **Case\_of\_object:**

This column contains the cases of object. There are totally 18 values in this column and the most frequent ones are ‘verbal participle clause’, ‘special verbal participle ‘endru’ and ‘accusative’.

When the object is not clausal, then the value can be ‘accusative’, ‘accusative (phrasal)’, ‘dative’, ‘dative (phrasal)’ or ‘adverbial phrase’. When the object is clausal, the value can be ‘verbal participle clause’, ‘special verbal participle’, ‘negative verbal participle clause’, ‘special negative verbal participle’, ‘infinite verbal clause’, ‘accusative participle’, ‘dative participle ’or ‘adverbial clause'. The special verbal participle is divided into 5 groups such as ‘endru’, ‘ena’, ‘aai’, ‘aaga’ and ‘patri’. The special negative verbal participle is divided into 2 groups such as ‘illamal’ and ‘indri’.

Owing to the fact that the adverbials are acting as the stimuli of subject for psych verbs,

it was considered as one of the cases of object in this annotation though they are actually not objects according to linguistics.

In Tamil, the accusative case has particle ‘ஐ (*i*)’ at the end. In (6), bold words are the object and it is the ‘accusative (phrasal)’ case. The dative case has particle ‘கு (*ku*)’ at the end. (7) is the dative case. The adverbial phrase is formed by affixing ‘ஆல் (*aal*)’, or ‘இல் (*il*)’ to the accusative.

This verbal participle is formed by affixing ‘உ (*oo*)’ to the non-finite verb. In (8), bold words are the verbal participle clause as it contains ‘இழந்து (*izhandhu* - having lost)’ verbal participle. The negative verbal participle is having ‘அல் (*al*)’ or ‘ஆது (*aadhu*)’ at the end of non-finite verb. In (9), bold words are the negative verbal participle clause.

When ‘ஆல் (*aal*)’, or ‘இல் (*il*)’ is affixed to the verbal participle, it becomes adverbial clause - reason. When ‘போது (*podhu*)’, ‘பொழுது (*pozhudhu*)’, ‘உம் (*um*)’ is affixed, it becomes adverbial clause - timing. In some cases, ‘இல் (*il*)’ also defines the timing. In (10), the bold words are the adverbial clause (reason).

The infinite verbal clause is formed by affixing ‘அ (*a*)’ to the non-finite verb. In (3), ‘அணில் வந்த\_காரணத்தை அறிய (*anil vandha\_kaaranathai ariya -* to know the reason of squirrel’s arrival)’ is the infinite verbal clause. Sometimes, the infinite form act as adverbial clause.

All the remaining cases such as ‘special verbal participle’ and ‘negative special verbal participle’ comes in between the clause and the verb. (11) is a sample special verbal participle ‘endru’ and it is similar to the English subordinate clause ‘that / about’.

1. இந்த நாட்டில் பெரும்பான்மையோர் **அடிமைத்தன ஒழிப்பை**

*indha naatil perumbaanmaiyor* ***adaimaithana ozhippai***

this in.country majority.people(S) **slavery abolishment (O)**

விரும்புகிறார்.

*Virumbigiraargal*.

wants (V).

Majority people in this country wants the abolishment of slavery.

1. புலிக்கு (*pulikku*, to a tiger / for a tiger)
2. நிலங்களை வாங்கியவர்கள் **பணத்தை இழந்து, நிலத்தையும்**

*nilangalai vaangiyavargal* ***panathai izhandhu nilathaiyum***

lands those.who.bought (S) **money having.lost land**

**இழந்து** தவிக்கிறார்கள்.

***izhandhu*** *thavikiraargal.*

**having.lost (O)** are.suffering (V).

**Having lost the money and land**, those who bought the lands are suffering.

1. அப்போது சம்பந்தர், **தந்தையைக் காணாமல்**

*appodhu sambandhar* ***thandhaiyai kaanaamal***

that.time Sambandhar (S) **father not.having.found (O)**

திகைத்தார்.

*thigaithaar.*

was.shocked (V).

**Not having found the father**, Sambandhar was shocked that time.

1. **பணம் காணாமல் போனதால்** அனைவரும்

***panam kaanaamal ponadhaal*** *anaivarum*

**money missing because.went (O)** everyone (S)

செய்வதறியாது திகைக்கிறார்கள்.

*seivadhariyaadhu thigaikiraargal*.

not.having.known.to.do is.shocked (V).

Everyone is shocked **because the money went missing** and don’t know what to do.

1. சிலர் **இவ்வளவு செழிப்பான இலக்கியம் தமிழில்**

*silar* ***ivvalavu sezhipaana ilakiyam thamizhil***

someone(S) **how rich literature in.tamil**

**இருக்கிறதா என்று** வியந்துபோவார்கள்.

***Irukiradhaa endru*** *viyandhupovaargal*.

**is.there that (O)** will.wonder (V).

Someone will wonder **that how rich literature is there in Tamil**.

1. **Case\_of\_object\_inTamil:**

This column provides the Tamil grammatical terms for object cases. There are totally 5 values in this column. They are ‘2nd vaetrumai 'ஐ', ‘4th vaetrumai 'கு', irandha kaala vinaiyecham, ethir kaala vinaiyecham and idaichol thodar’.

In Tamil, the accusative case is called ‘2nd vaetrumai’. When the ‘Case\_of\_object’ column value contains ‘accusative’, then this column’s value is ‘2nd vaetrumai 'ஐ'’. The dative case is called ‘4th vaetrumai’. When the ‘Case\_of\_object’ column value contains ‘dative, then this column’s value is ‘4th vaetrumai 'கு'’.

The verbal participle clause and negative verbal participle clause are called ‘irandha kaala vinaiyecham’. The infinite verbal clause is called ‘ethir kaala vinaiyecham’.

All the remaining cases such as adverbial phrase, adverbial clause, special verbal participle and negative special verbal participle are called ‘idaichol thodar’ in Tamil.

1. **Animacy\_of\_subject:**

This column defines the animacy of the subject. The values are ‘rational’, ‘animate’ and ‘inanimate’. When the subject is human-being, then the value is ‘rational’. If it is animal, then the value is ‘animate’. If it is a non-living thing, then the value is ‘inanimate’.

In (9), the subject (Sambandhar) is a human-being and hence the value of this column will be ‘rational’.

1. **Animacy\_of\_object:**

This column defines the animacy of the object. The values are ‘rational’, ‘animate’, ‘inanimate’ and ‘none’. When the object is human-being, then the value is ‘rational’. If it is animal, then the value is ‘animate’. If it is a non-living thing, then the value is ‘inanimate’.In case if the object is clausal, then the value should be ‘none’.

In (5), the bold words are clausal object and hence the value of this column will be ‘none’.

1. **Pronoun\_subject:**

This column defines whether the subject is pronoun or not. The values are ‘yes’ and ‘no’. If the subject is a pronoun, then the value is ‘yes’. Else, it will be ‘no’.

In (12), the subject (நான் - I) is a pronoun and hence the value of this column will be ‘yes‘.

1. **நான்** கரையில் நின்று **அவரை** வியக்கின்றேன்.

***naan*** *karaiyil ninru* ***avarai*** *viyakkinren.*

**I** (S) in.shore by.standing **him** (O) wonder (V).

I wonder him by standing on the shore.

1. **Pronoun\_object:**

This column defines whether the object is pronoun or not. The values are ‘yes’ and ‘no’. If the object is a pronoun, then the value is ‘yes’. Else, the value will be ‘no’.

In (12), the object (அவரை - him) is a pronoun and hence the value will be ‘yes’.

1. **Syllable\_of\_subject:**

It contains the no. of words in the subject. The values are finite integers that starts from 1. In (11), the subject is ‘சிலர்’ and that is one word, hence the syllable is 1.

1. **Syllable\_of\_object:**

It contains the no. of words in the object. The values are finite integers and starts from 1. In (11), the object is ‘இவ்வளவு செழிப்பான இலக்கியம் தமிழில் இருக்கிறதா என்று’ and they are 6 words. Hence, the syllable is 6.

1. **Focus:**

This column indicates which part is focused more on a sentence, whether it is subject or object or verb. The values of this column are ‘none’, ‘subject’, ‘subject (emphasized)’, ‘object (emphasized)’, ‘subject (emphasized), object(emphasized)’ and ‘verb (emphasized)’.

In Tamil, particles like ‘தான் (*thaan*)’, ‘கூட(*kooda*)’ and ‘உம் (*um*)’ are affixed to the subject or object inorder to emphasize them.

In Tamil, the part which occurs before verb is considered to be more focused. But, it doesn’t apply for the SOV order, as this is the standard word order of Tamil. Suppose if the order is OSV, it means subject is more focused as it occurs before verb. In (5), the word order is OSV and hence the value of this column is ‘subject’.

If there are no emphasized particles and the order is SOV, then the value is ‘none’. Despite the word order, if the subject or object is emphasized, then the value is either ‘subject (emphasized)’ or ‘object (emphasized)’ accordingly. In case if both the subject and object are emphasized, then the value is ‘subject (emphasized), object(emphasized)’. In rare scenario, verbs are emphasized with particles like 'போ (*po*)', 'விடு (*vidu*)’. In those cases, the value is ‘verb (emphasized)’.

1. **Psych:**

This column indicates whether the verb in a sentence denotes the emotional (psych) meaning or not. This column has only one value ‘psych’. Because all verbs of this dataset denote only the emotional meaning in their sentences and there is no non-emotional or ambiguity occurrences.