

# Analysis of Uvama Urubugal in Tamil Sangam Literatures

Subalalitha C.N

Department of Computer Science and Engineering  
SRM Institute of Science and Technology  
Chengalpatu District, Tamil Nadu, India  
subalaln@srmist.edu.in

## Abstract

Uvama urubugal in Tamil are used to explain a particular context by citing another equivalent context. This is referred to as “Uvamaiyani” in Tamil Grammar rules as stated in Tholkappiam. The uvama urubu is called as simile in English. Similes bring out many beautiful poetic contexts. Automatic extraction of such similes can help to build better Natural Language Generation applications such as, story generation systems and lyric suggestion systems. This paper attempts to automatically extract the uvama urubugal from Tamil Sangam Literatures. Natrinai and Mullai Pāṭu have been used for the analysis. There are 12 uvama urubugal in Tamil as per Nanool and this paper has attempted to analyze the usage of these 12 uvama urubugal in Sangam Literatures and compares their usage distribution in the Tamil film songs data set comprising of 4215 songs. It was found that only two uvama urubugal were used in the current-day Tamil film songs. This comparison was done to reveal the diminishing usage of these beautiful uvama urubugal by the current generation and the urge to use them again.

## 1 Introduction

Similes bring out different perception to a context and they add beauty to a text. In Tamil Language, similes are termed as uvamai and follows the grammar rules called, “Uvamayani”. The “uvama urubugal” are those words that are used to frame the simile. There are 12 uvama urubugal in Tamil. The Tamil uvama urubugal and their respective English Transliteration is given in Table 1. The rest of this paper makes use of the English transliteration of uvama urubugal.

S.No	Uvama Urubu (Tamil)	Uvama Urubu (English Transliteration)
1	பூவல்	Pōla
2	புரைய	puraiya
3	ஓப்பு	oppa
4	உறழ்	urala
5	மான	māna
6	கடுப்பு	katuppa
7	இயைய	iyaiya
8	ஏய்ப்பு	ēyppa
9	நேர்	nēr
10	நிகர்	nigar
11	அன்ன	anna
12	இன்ன	inna

Table 1: English Transliteration of Tamil Uvama Urubugal

The uvama urubu (singular form of uvama urubugal) Pōla is widely being used by the current generation . This paper attempts to dig and explore the usage and the distribution of these categories in Sangam literatures and compares with the Tamil film songs to show the decreasing usage of uvama urubugal.

Tamil language is one of the ancient languages in the world and the epigraphic attestations of Tamil have been written from the 6th century BC (Southworth, 2004) (Jesudasan, 2019) . Since then, Tamil language has transformed a lot in terms of its written form and speech form. These transformations have also reflected the beauty of the language followed during every time period through the word usages. This paper focuses on the language usage followed in Sangam literature era that spanned from 600

BCE to 300 CE (Singh, 2008).

This paper tries to understand the simile usage in the two Sangam literatures namely, Mullai Pāṭu and Natrinai and has compared with 4215 Tamil film songs dataset (Siva Subramanian, 2020). The aim of this experiment is to increase the usage of these similes by the current Tamil generation. Also, this paper lays a foundation to explore more on applications such as, Automatic Lyric generation system and Automatic story generation system that can be built on top of an automatic simile detection system. The word, "Sangam Literature" used in the rest of the paper denotes only Natrinai and MullaiPāṭu.

The contribution of this paper is two-fold.

1. Analysing the usage distribution of twelve types of similes in two Sangam Literatures namely, Natrinai and Mullai Pāṭu.

2. Comparison of the simile usage in 4215 Tamil film songs dataset with the Sangam Literature.

Three types of data sets have been used for the analysis of uvama urubugal namely, 400 Natrinai songs, 103 lines of Mullai Pāṭu and 4215 Tamil film songs(Siva Subramanian, 2020). <https://www.kaggle.com/sivaskvs/tamil-songs-lyrics-dataset>.

The rest of the paper is organized as follows. Section 2 describes the Literature Survey, Section 3 describes the experimental set up that extracts and analyses the uvama urubugal and section 4 describes the conclusion and future work.

## 2 Literature Survey

This section focuses on three dimensions of the state-of -the art works namely, existing works on simile, works on Sangam Literature, and works on lyric generation. Since the paper is about type of similes, the existing works that have focused on simile analysis in other languages are discussed. Furthermore, the work that are done pertinent to Sangam Literature are also discussed. The automatic simile detection can lead to building lyric generation system and hence the works done in Tamil Lyric generation system are discussed.

(Manjusha and Raseek, 2018) have attempted to classify the similes using Convolution Neural Networks (CNN) and Machine Learning algorithms for English language.

(Chakrabarty et al., 2020) have attempted to generate similes using pretrained model BART which is a sequence-to-sequence models for English language.

(Zeng et al., 2020) have attempted to recognize similes for English language using BERT models.

The paper currently identifies the types of simile (Uvama urubugal) in Sangam literature. Deep learning models must be used as the extension of this work when the similes and their equivalent explanations are automatically extracted using the types of uvama urubugal. This kind of a deep learning-based classification can be used as the foundation for building automatic lyric generation system.

(Sridhar et al., 2018) have attempted to generate lyric in Tamil based on given tune and situation. These kinds of works could be improved if similes in Tamil are given much importance.

(Subalalitha, 2019) has attempted to extract information from Kurunthogai which is also a Sangam literature poem. In this work, the information such as flora, fauna, food, the vessels, and the water bodies mentioned in Kurunthogai are automatically extracted. The sangam literature depicts the life of the Tamils lived during the Sangam era portraying their culture and the five types of landscapes namely, kurunji, mullai, marutham, neithal and palai where they lived. These landscapes represent the mountains, sea, forest, agricultural land and desert encompassing the information about people lived in such different landscapes. This is the reason why in this paper, the analysis of types of simile used by Tamils is done on Sangam Literatures.

## 3 Experimental Set-up for Analysing Uvama Urubugal

### 3.1 Data set

Three types of data sets have been used for the analysis of uvama urubugal

namely, 400 Natrinai songs, 103 lines of Mullai Pāṭu and 4215 Tamil film songs (Siva Subramanian, 2020). <https://www.kaggle.com/sivaskvs/tamil-songs-lyrics-dataset>.

### 3.2 The Experiment

The Natrinai and Mullai Pāṭu songs are scraped from the web page <https://sangamtranslationsbyvaidehi.com/>.

This web page consists of all Sangam Literature poems in Tamil, their Translations, their explanation in Tamil and English. The experiment consists of Pre-processing, Uvama Urubu Classification and analysis. Pre-processing involves removing the HTML tags and separating out only the Natrinai and Mullai Pāṭu songs in Tamil. In Uvama Urubu Classification, the Sangam literature songs and Tamil film songs are checked for all 12 types of uvama urubugal and grouped under their appropriate classes. If the uvama urubu is a free morpheme then an exact match of it is looked for. If the uvama urubu is a bound morpheme, a partial match is found. Human judgement is used as the evaluation metric to check the correctness which is done by six domain experts.

Each uvama urubu has its own peculiarity in usage. These peculiarities are analysed and recorded here.

Fig 1 shows the experimental set up for detecting the uvama urubugal.

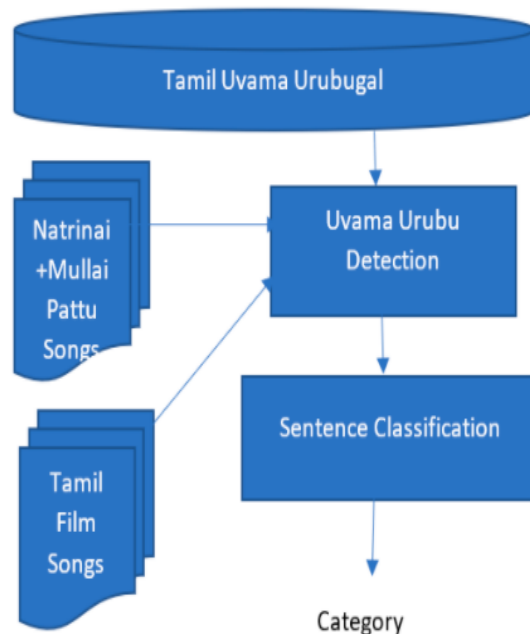


Fig 1: Uvama Urubu Classification

It was observed that the uvama urubu Pōla has been used so frequently compared to the other uvama urubugal both in Sangam literature and Tamil film songs. Tamil film songs had only Pōla and māṇa. The interesting fact is that māṇa was not found in Sangam literature as per the experiment. While observing the uvama urubu puraiya, it was found both as a free morpheme and a bound morpheme attached along with a suffix யும் (yum). The uvama urubu oppa was found along with the bound morpheme இன் (ஓப்பின்). The uvama urubu urāla was not identified by the experiment. The uvama urubu kaṭuppa has found to have been used relatively frequent and it has been used as a free morpheme. The uvama urubu iyaiya has occurred only once in Sangam literature as per the experiment. The uvama urubu ēyppa was found as a free morpheme and was found in least counts. nēr was found as a free morpheme. The uvama urubu, nikar was found as nikarpa attaching itself with a bound morpheme and was only found only twice. The uvama urubu anna is also a frequently used like pōla), and is used as a free morpheme. The uvama urubu inna was not found by the experiment.

Table 2 shows the distribution of the uvama

urubu in Sangam literature and Tamil film songs.

S.No	uvama urubu	Count in Sangam Literature	Count in Tamil Film songs
1	போல(Pōla)	102	32
2	புரைய(puraiya)	10	0
3	ஓப்ப (oppa)	6	0
4	உறழ் (urāḷa),	0	0
5	மான((māṇa)	0	2
6	கடுப்ப (kaṭuppa)	16	0
7	இயைய(iyaiya)	1	0
8	ஏய்ப்ப (ēyppa)	3	0
9	நேர்(nēr)	13	0
10	நிகர்(nikar)	2	0
11	அன்ன( anna)	56	0
12	இன்ன( inna)	2	0

Table 2:Uvama Urubugal Count Comparison  
Fig 2 shows the graph visualization of the comparison of uvama urubugal count in sangam literature and Tamil film songs.

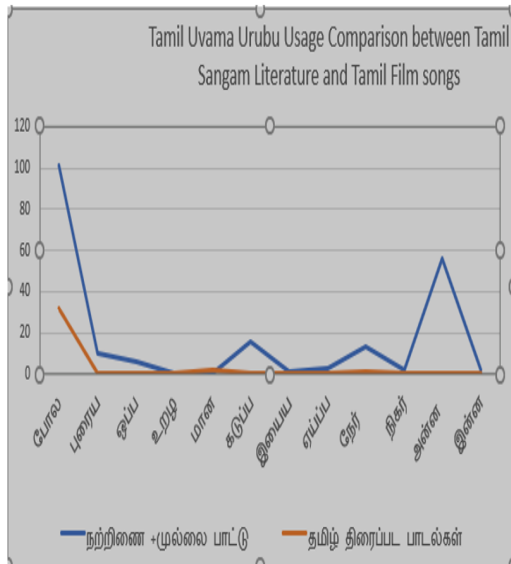


Fig 2:Comparison between Sangam Literature and Tamil film songs

The uvama urubu count obtained by the experiment from Natrinai songs and Mullai Pāṭu is not the exact count of uvama urubugal and needs to be validated by the Tamil experts. However, the uvama urubugal identified by

the experiment are correct as per the human expert judgements. It was observed that mullai Pāṭu did not contribute towards the count and had only one uvama urubu as it is the shortest Sangam literature song containing 103 lines. The calculation of recall can only reveal the exact distribution of all uvama urubugal in the sangam literature songs considered. Apart from the bound morphemes identified by the experiment discussed in this paper many other variants of the uvama urubugal need to be explored by conducting the experiment with all Sangam literature songs. The experiment also revealed that the usage of similes has drastically reduced and should be used again to renovate the beauty of the language.

#### 4 Conclusion and Future work

Similes capture the different yet beautiful dimensions of the language. Uvama urubu in Tamil is used to express the similes. Tamil language has seen many transformations in terms of it's word usage both in text and speech forms since 2nd century BC. There are 12 uvama urupugal in Tamil language and currently Pōla and māṇa are only used. As this is an initial experiment it is tested only with Natrinai and Mullai Pāṭu. It has to be tested with rest of the Sangam songs to get a wider view.

This work has many directions in which it can be extended. The experiment discussed can be viewed as a multi-classification problem and can be implemented using Machine learning and deep learning algorithms to automatically classify the similes and their equivalent contexts. The contextual features must be identified for each uvama urubu and the model needs to be trained with many Tamil literature texts. The extraction of these similes and their contexts can throw a lot of light in creative usage of Tamil language mainly in Natural language generation-based applications such as lyric generation and story generation.

#### References

Tuhin Chakrabarty, Smaranda Muresan, and Nanyun Peng. 2020. Generating similes< effortlessly> like a pro: A style transfer

approach for simile generation. *arXiv preprint arXiv:2009.08942*.

Dennis S. Jesudasan. 2019. Keezhadi excavations: Sangam era older than previously thought, finds study. *The Hindu*.

PD Manjusha and C Raseek. 2018. Convolutional neural network based simile classification system. In *2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR)*, pages 1–5. IEEE.

Upinder Singh. 2008. *A history of ancient and early medieval India: from the Stone Age to the 12th century*. Pearson Education India.

Franklin Southworth. 2004. *Linguistic archaeology of south Asia*. Routledge.

Rajeswari Sridhar, Subasree Venkatsubhramaniyen, and S Subha Rashmi. 2018. Automatic singable tamil lyric generation for a situation and tune based on causal effect. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1152–1159. IEEE.

CN Subalalitha. 2019. Information extraction framework for kurunthogai. *Sādhanā*, 44(7):1–6.

Jiali Zeng, Linfeng Song, Jinsong Su, Jun Xie, Wei Song, and Jiebo Luo. 2020. Neural simile recognition with cyclic multitask learning and local attention. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9515–9522.