# Simile Generation

**Elanchezhiyan K, Tamil Selvi E, Revathi N., Shanthi G.P,
Shireen S. & Madhan Karky.**

*Karky Research Foundation*
*E-mail: inbox@karky.in*

## Abstract
A simile is a figure of speech that expresses a thought through an explicit comparison between two entities. It reveals unexpected likeliness between two seemingly disparate things. Poems and lyrics are enriched with features such as rhyme, simile, metaphor, etc. In this paper, we present our work in progress towards automatic generation of simile. Simile relies on specific patterns that make them recognizable in any phrase. In order to extract patterns, sample of 2500 Tamil lyrics were analyzed and with specific words used in similes we identified the set of patterns used in aforesaid set of lyrics. The identified pattern is then used to generate new similes by randomly comparing one entity with another entity of different kind that matches with one or more attributes. The attributes associated with each entity such as colour, shape, actions and other descriptors are extracted from web using a focussed crawler.

## 1. Introduction
Tamil is one of the Dravidian Languages, spoken mainly in South India. It is a classical language, containing a large body of literature. Sangam Literature is one such ancient literature; Mostly Literature gives information about ancient cultures, kings, dynasty, nature etc. While describing those concepts, poets or authors uses similes to compare an object with another object having same attributes. A simile is a type of metaphor in which the comparison is made with the use of the words like போல, போன்று, போல் or its equivalent. Simile compares two unlike things having same attributes of those things, like actions, sizes, colors, features etc. Basically Tamil Similes are classified into three categories, which are பண்பு உவமை (paNbu uvamai), தொழில் உவமை (thozhil uvamai), பயன் உவமை (payan uvamai). PaNbu uvamai is used to compare the attributes likes shapes, color and etc between two objects. Thozhil uvamai is used to compare the action between two objects. Payan uvamai is used to compare usage between two attributes [1].

In this paper we present our work, Simile generator that can be used to generate large number of unique similes, which can be used as in Automatic Tamil poetry or lyric Generation. This paper aim to generate unique similes based on similar attributes and its functionality between two entities.

## 2. Background Work
In "Learning to understand figurative language: From similes to metaphors to irony" [2] the authors have described a computational approach to simile and metaphor that takes the career-of-metaphor hypothesis as its starting point. They have described how the category-defining knowledge required by metaphor can be acquired from exposure to

explicit similes, and have demonstrated that this knowledge offers a richer and more diagnostic picture of category structure than that acquired from alternate sources.

## 3. How Attributes are classified and generated?
We classify entity attributes into two kinds: lyrical and non lyrical. Lyrical attributes exaggerates the entity. Non lyrical attributes provide the actual information about the entity. In our work, Similes are generated based on patterns and attributes. Patterns are ordering of words or sentence structures.

### 3.1 Exploiting Attributes for Simile Generation
Here Attributes refers to a characteristic of an object. Here we consider similes based on six features: Actions, Colour, Size, Shape, Adverbs and Adjectives related to a thing or entities.

For Example
" குயில் போல பாடும் குழந்தை / *kuyil poala paadum kuzhandhai*"
குயில் and குழந்தை are two entities. குயில் (*kuyil*) can chirp (கூவும்), but it cannot sing. But குழந்தை can sing but it can't chirp. The above simile tells that Child can sing like a cuckoo. Here the action of child is exaggerated. We call this exaggeration as lyrical attribute for a simile.

" நிலா போன்ற முகம் / *nilaa poandRa mugam* "
Here நிலா and முகம், both are round in shape and white in colour. The Shape and colour of these objects are same. Here nothing has been exaggerated in the comparison. We call these types of attributes as non lyrical ones.

### 3.2 Exploiting pattern for Simile Generation
Tamil Language's sentence patterns are more complex in nature. Here patterns refer to sentence structure. Initially we were able to find more than 90 unique Simile patterns from 2500 lyrics. We can create new similes based on different patterns.
For example
" ஒட்டகம் போன்ற மனிதன் / *ottagam poandRa manidhan*"
In above example     and     are entities and     is Preposition. Sentence pattern of above example is NPN.

" ஆமை போல தவழும் குழந்தை / *aamai poala thavazhum kuzhandhai*"
Here ஆமை and குழந்தை are entities, போல is Preposition and தவழும் is action of a குழந்தை. Sentence pattern of above example is NPVN

### 3.3 Conceptual Similes
We can create unique conceptual similes by using UNL techniques [3]. We have linked similar concepts together through which we can create a large number of unique similes.

For example
" நிலா போன்ற முகம் / *nilaa poandRa mugam*"

" சந்திரன் போன்ற முகம் / *chandhiran poandRa mugam*"
" மதி போன்ற வதனம்  / *madhi poandRa vadhanam*"

Here  நிலா,  சந்திரன் and மதி are equivalent concepts. This helps us to generate more similes.

## 4. Difficulties in Generating Similes

In Tamil, Sentence Structures are complex. Patterns are based on sentence structure. As Tamil is a partial free word order language, we can write a sentence in different ways, without affecting the meaning of the sentence. But English follows a predefined set of patterns - subject comes first, verb second and the object third (SVO). For Example In tamil   முயலை போல வேகமாக ஓடினாள் மருதம் (*muyalai poala vaegamaaga oadinaaL marudham*) and மருதம் முயலை போல வேகமாக ஓடினாள் (*marudham muyalai poala vaegamaaga oadinaaL*) gives the same meaning but have two different sentence patterns. In English the above example can be said as 'marutham runs like a rabbit' and 'rabbit runs like marutham' has totally another meaning.

### 4.1 Extracting Object/Entity Features

While extracting the features of entities from structured documents such as Wikipedia, we get more irrelevant information to an object and find difficulties in choosing correct URL to extract correct information. Moreover since this generator generates similes for lyrics, we need some imaginary information like "நிலா பூ கொண்டு வா / *nilaa poo konduvaa*". In this example  நிலா (*Nila*) is not capable of doing such actions. Now we extract imaginary information about entities from lyrics and we manually enter attribute for entity.

### 4.2 Normalizing Sentence Patterns

We have identified the patterns from Lyrics and thirukkural vurai. These sentence patterns are based on post position and preorder position from the Preposition like (      - pola, pondra). As a result we obtained around 900 unique patterns from 2500 lyrics and 1330 thirukkural interpretations. Some of the patterns are N-P-V-N-P-V, N-P-V-N-N, N-N-V-P-V-V-N etc. Where P denotes Preposition, N denotes Noun, V denotes Verb, Adj Denotes Adjective and Adv denote Adverb.

## 5. Results

Sample results of our simile generator are shown below.

" திருவிழா போல சந்தோஷ படுத்தும் உறவு ."

(thiruvizhaa poala chandhoasha paduththum uRavu.)

" நீர்த்துளி போல விழும் கண்ணீர் ."

(neerththuLi poala vizhum kaNNeer)

" நதி போன்ற தடையற்ற வாழ்க்கை ."

(nadhi poandRa thadaiyatRa vaazhkkai)

" மயில் போல கர்வமுள்ள பெண் ."

(mayil poala karvamuLLa peN)

" ஓடை போல் ஓடும் பெண்."

<div align="center">

(oadai poal oadum peN)

" செம்பருத்தி போல் சிவப்பான சதை."

(chembaruththi poal chivappaana chadhai)

" கீதம் போல அற்புதமான வாழ்வு."

(keedham poala aRpudhamaana vaazhvu)

"கவிதை போல இனிமையான காதல் ."

(kavidhai poala inimaiyaana kaadhal)

"பூ போல மென்மையான மேனி ."

(poo poala menmaiyaana maeni)

"குழந்தை போல பிடிவாதமான காதல்."

(kuzhandhai poala pidivaadhamaana kaadhal)

</div>

## 6. Conclusion and Future work

This paper presents an automatic simile generator by using attributes and patterns which can create a large number of unique similes. These similes aid in character description in auto poetry generation or lyric Generation etc. Enhancing the Automatic Attribute extraction feature will help to generate more and better similes. Visualizing the attributes on a time line will enable us to determine various patterns of words, combinations and thoughts used over time. Normalizing the patterns for the generation of new similes will enhance the system.

## 7. Reference

[1] http://www.tamilvu.org/courses/degree/d031/d0314/html/d03141l2.htm Last Access Date 09th August 2014

[2] T. Veale and Y. Hao, "Learning to understand figurative language: From similes to metaphors to irony," in *Proceedings of CogSci, Nashville, USA, 2007.*

[3] Foundation, U., and the Universal Networking Language (UNL) Specifications Version 3 3ed. December 2004: UNL Computer Society, 2004. 8(5). Centre UNDL Foundation