

# ISOLATING WORD LEVEL RULES IN TAMIL LANGUAGE FOR EFFICIENT DEVELOPMENT OF LANGUAGE TOOLS

Suriyah M, Aarthi Anandan, Anitha Narasimhan and Madhan Karky

Karky Research Foundation, India

## ABSTRACT

*With the advent of social media, the amount of text available for processing across different natural languages has become enormous. In the past few decades, there has been tremendous increase in the number of language processing applications. The tools for natural language computing of various languages are very different because each language has its own set of grammatical rules. This paper focuses on identifying the basic inflectional principles of Tamil language at word level. Three levels of word inflection concepts are considered – Patterns, Rules and Exceptions. How grammatical principles for word inflections in Tamil can be grouped in these three levels and applied for obtaining different word forms is the focus of this paper. These can be made use of in a wide variety of natural language applications like morphological analysis, morphological generation, word level translation, spelling and grammar check, information extraction etc. The tools using these rules will account for faster operation and better implementation of Tamil grammatical rules referred from [தொல்கொப்பியம் | tholgaappiyam] and [நன்னூல் | nannool] in NLP applications.*

## KEYWORDS

*Natural language processing, Rule based approach, word level rules, Tamil tool, language tools*

## 1. INTRODUCTION

The number of social media platforms for various purposes like blogging, micro-blogging, photo sharing, social networking etc., has risen steadily. The number of people who are members of the same has also increased. For instance, the micro-blogging website Twitter has about 330 million users (on 1st January 2018) and about 500 million tweets (on 24th January 2017) are sent on an average per day [1]. This in turn, has paved way for collection and processing of various forms of data – text, image, video etc., generated with high velocity and veracity. Focusing on the processing of text, the posts of users in different natural languages have to be analysed; this avenue of NLP has many challenges. One of them is to analyse text and interpret the meaning conveyed. This is not possible without a tool which has knowledge of all the grammatical rules followed in the natural language considered. Different tools need to be used for processing different natural languages because of the differences that exist in terms of the syntax and grammatical rules. In this work, a three level grouping of grammatical principles for word inflections in Tamil language is proposed for better grammatical analysis.

Inflection in grammar refers to the modification of a root word to convey various characteristics like tense, gender, number, person, case etc. Inflection expresses one or more grammatical categories with a prefix, suffix or infix, or another internal modification such as a vowel change [2].

For example, consider the Tamil word [கிளிஞ்சல்கள் | kiLinjalgaL | sea shell <plural>]. Here, [கிளிஞ்சல் | kiLinjal | sea shell] is the root word which is a noun in singular form. The suffix [கள் | kaL ] has been added to obtain the inflected version of the word in the plural form.

[கிளிஞ்சல்களொல் | kiLinjalgaLaal | sea shell <plural><instrumental case>] is another inflected form of the same noun where the root word கிளிஞ்சல் is added with two suffixes – [கள் | kaL ] and [ஆல் | aal ] to denote plural and instrumental case respectively.

[கிளிஞ்சல்களிடமிருந்து | kiLinjalgaLidamirundhu | sea shell <plural><locative case><ablative case>] is yet another form of the same root word [கிளிஞ்சல் | kiLinjal | sea shell] with three suffixes. The suffixes are [கள் | kaL ] denoting plural, [இடம் | idam ] denoting locative case and [இருந்து | irundhu ] denoting ablative case.

#### EXAMPLE 1: SAMPLE NOUN INFLECTIONS

- 1.1 கிளிஞ்சல்கள் ← கிளிஞ்சல் <root word> <noun> + கள் <plural>  
kiLinjalgaL ← kiLinjal + kaL  
sea shell<plural> ← sea shell + plural suffix
- 1.2 கிளிஞ்சல்களொல் ← கிளிஞ்சல்<root word><noun> + கள்<plural> + ஆல்<instrumental case>  
kiLinjalgaLaal ← kiLinjal + kaL + aal  
sea shell <plural><instrumental case> ← sea shell + plural suffix + instrumental case suffix
- 1.3 கிளிஞ்சல்களிடமிருந்து ← கிளிஞ்சல் <root word> <noun> + கள் <plural> + இடம் <locative case> + இருந்து <ablative case>  
kiLinjalgaLidamirundhu ← kiLinjal + kaL + idam + irundhu  
sea shell<plural><locative case> <ablative case> ← sea shell + plural suffix + locative case suffix +ablative case suffix

#### Example 2: Sample verb inflections

- 2.1 பாடுவாய் ← பாடு<verb> + வ் <future tense> + ஆய் <second person> <singular>  
paaduvaay ← paadu + v + aay  
sing <imperative> ← to sing + future tense + second person + singular
- 2.2 பாடிக்கொண்டிருந்தாள் ← பாடு<verb> + இ<past tense marker> + க்<sandhi> + தகொண்டிரு<auxiliary verb> + ந்த்<past tense marker> + ஆள் <third person feminine singular suffix >  
paadikkoNdirundhaaL ← paadu + i + k + koNdiru + nth + aaL she was singing ← to sing + past continuous + third person + singular + feminine

- 2.3 பாடுவார்கள் ← பாடு<verb> + ஓ < future tense marker > + ஆர் < pronominal third person singular > + கள் < plural suffix >  
paaduvaargaL ← paadu + v + aar + kaL they will sing ← to sing + future tense + third person + plural

In this work, three levels of word inflection rules are considered – Rules, Patterns and Exceptions. How grammatical rules for word inflections in Tamil can be grouped in these three levels and applied for obtaining different word forms is the focus of this paper. For the sake of simplicity, in this paper we call a morphological ending, a pattern. A Pattern can be a number or case ending associated with a noun or a person-number-gender marker associated with a verb. We have a set of Rules and Exceptions associated with each pattern. A Rule will define how the pattern can add to a root word or how it can separate from a morphological variation to obtain the root. An Exception to a rule will handle special cases where the rule cannot be applied directly. These three levels of grammatical forms discussed in this paper facilitate simple and efficient implementation of grammatical rules in basic NLP tools like morphological analyser, morphological generator, word level translator etc.

## 2. LITERATURE REVIEW

The three main types of morphological relations are Inflection, Derivation, and Compounding. Inflectional morphology modifies the properties of lexemes, while maintaining the basic meaning of the lexeme [3].

Singh and Sarma aim at reducing the number of affixes and word generation and re-adjustment rules describing nominal inflection by using inflectional class information to analyse various forms of Hindi nouns. They propose that the inflectional behaviour of nominal forms should be used entirely in the formation of nominal classes in Hindi [4].

To Ramscar, it appears that inflection is carried out through analogical reminding based on semantic and phonological similarity and that a rule-based route is not necessary to account for past tense inflection [5].

Agglutination is a process of morphology in which complex words are formed by joining morphemes together without changing the spelling or phonetics [6]. Prominent south Indian languages like Tamil, Telugu, Malayalam, Kannada and Tulu are agglutinative [7].

Reddiah suggests that in addition to having similar sounds, all Dravidian languages are related closely in grammatical aspects too. There are similarities in prefix, suffix, numbers, gender and syntax and that all the Dravidian Languages are formed from one main Dravidian language, which may be Tamil by most scholars' opinions [8].

Anand kumar et al propose a machine learning based approach for morphological analysis using SVM. This is a supervised model which needs voluminous training data to get good results [9]. But rule based approaches do not have this overload of having enormous tagged training data. TamilComputingLab morphological analyser is a conventional rule based one [10]; Piripori is the Tamil morphological analyser built on the principles discussed in this paper [11]. It was observed from the working of the above two morphological analysers that when morphological rules are

grouped and streamlined following the approach mentioned in this work, the speed of processing and the accuracy with respect to precision and recall are increased compared to the conventional rule based approach. The grouping of inflectional rules in Tamil accounts for faster and more accurate operation of language tools; hence becoming necessary for better tools and applications.

### 3. DESCRIPTION OF INFLECTIONAL RULES IN TAMIL

Tamil script has 12 vowels [உயிரெழுத்துக்கள் | uyirezhuthukkaL], 18 consonants [மெய்யெழுத்துக்கள் | meyyezhuthukkaL] which combine to form 216 combinational letters [உயிர்மெய்யெழுத்துக்கள் | uyirmeyyehuthukkaL] summing up to 247 letters. Tamil has more than 4 lakh words. These words are classified as [பெயர்ச்சொல் | peyarchol | noun], [வினைச்சொல் | vinaichol | verb], [இடைச்சொல் | idaichol | preposition and post position], [உரிச்சொல் | urichol | adjective], [இலக்கியச்சொல் | ilakkiyachol | literary words]. Apart from the classes mentioned, Tamil has the ability to adapt words from other languages; these words are called [வடசொல் | vadasol | words adapted from Sanskrit], [திரைச்சொல் | thiraichol | word adapted from other languages like Urdu, Portuguese, Telugu etc] [12]. These words combine with various prefixes and suffixes to yield their different inflectional forms. When these combinations happen, certain alphabet transformations may (or may not) happen. This work focuses on the collection of different prefixes and suffixes added to Tamil root words (which are nouns or verbs) and the different grammatical transformations which may happen during inflection. [தொல்கொப்பியம் | tholgaappiyam] and [நன்னூல் | nannool] are the sources of the inflectional rules discussed in this paper.

## 4. PROPOSED WORK

### 4.1 PATTERNS

These are suffixes added to root words to get the different morphological variations of the words considered. These patterns contain morphological information of number and case for nouns; tense, number, case, person and gender for verbs.

Example 3 (Inflectional forms of a noun): The Tamil root word, [மலர் | malar | flower], has been taken as a sample to demonstrate some of the inflectional forms of nouns in Table 1. [மலர் | malar | flower] is a noun whose inflected forms are obtained by mere suffixing of appropriate patterns. Some other nouns may require some transformation, addition or removal of certain letters. This is discussed in the Rules section.

Table 1. Different inflectional forms of the noun [மலர் | malar | flower].

Pattern	Implication of the pattern	Inflectional split of the pattern	Inflected form of the noun	Description
[கள்   kaL ]	<plural>	[கள்   kaL   <plural> ]	[மலர்கள்   malargaL   flower<plural> ]	[மலர் + கள்   malar + kaL ]
[களின்   kaLin ]	<plural> <ablative case>	[கள்   kaL   <plural> ] + [இன்   in   <ablative case>]	[மலர்களின்   malargaLin   flower <plural> <ablative case> ]	[மலர் + கள் + இன்   malar + kaL + in ]
[களால்   kaLaal ]	<plural> <instrumental case>	[கள்   kaL   <plural> ] + [ஆல்   aal   <instrumental case> ]	[மலர்களால்   malargaLaal ]	[மலர் + கள் + ஆல்   malar + kaL + aal ]
[களை   kaLai ]	<plural> <accusative case>	[ கள்   kaL   <plural> ] + [ ஐ   ai   <accusative case> ]	[மலர்களை   malargaLai ]	[மலர் + கள் + ஐ   malar + kaL + ai ]
[களிலிருந்து   kaLilirundhu ]	<plural> <ablative case>	[ கள்   kaL   <plural> ] + [இல்   il   <Locative case> ] + [ இருந்து   irundhu   <ablative case> ]	[மலர்களிலிருந்து   malargaLilirundhu ]	[மலர் + கள் + இல் + இருந்து   malar + kaL + il + irundhu ]
[களிடம்   kaLidam ]	<plural> <locative case>	[ கள்   kaL   <plural> ] + [இடம்   idam   <locative case> ]	[மலர்களிடம்   malargaLidam ]	[மலர் + கள் + இடம்   malar + kaL + idam ]
[களிடமிருந்து   kaLidamirundhu ]	<plural> <locative case> <ablative case>	[ கள்   kaL   <plural> ] + [இடம்   idam   <locative case> ] + [ இருந்து   irundhu   <ablative case> ]	[மலர்களிடமிருந்து   malargaLidamirundhu ]	[மலர் + கள் + இடம் + இருந்து   malar + kaL + idam + irundhu ]
[களுக்கு   kaLukku ]	<plural> <dative case>	[ கள்   kaL   <plural> ] + [ உக்கு   ukku   <dative case> ]	[மலர்களுக்கு   malargaLukku ]	[மலர் + கள் + உக்கு   malar + kaL + ukku ]
[ களுடன்   kaLudan ]	<plural> <associative case>	[ கள்   kaL   <plural> ] + [ உடன்   udan   <associative case> ]	[மலர்களுடன்   malargaLudan ]	[மலர் + கள் + உடன்   malar + kaL + udan ]
[களில்   kaLil ]	<plural> <locative case>	[ கள்   kaL   <plural> ] + [இல்   il   <locative case> ]	[மலர்களில்   malargaLil ]	[மலர் + கள் + இல்   malar + kaL + il ]

Example 4 (Inflectional forms of a verb): The Tamil root word, [பார் | paar | see], has been taken as a sample to demonstrate different inflectional forms of verbs when affixed with different patterns denoting number and cases.

The noun and verb considered in Table 1 and Table 2 are such that their inflectional forms can be obtained by mere suffixing of the patterns given. But there are other words which would require

International Journal on Natural Language Computing (IJNLC) Vol.8, No.1, February 2019  
transformation of one or more alphabet to obtain the inflected form when combining with a  
pattern. Such examples are discussed in the “Rules” part.

Table 2. Different inflectional forms of the verb [பார் | paar | see]

Pattern	Implication of the pattern	Inflectional split of the pattern	Inflected form of the verb	Description
[ப்பாய்   ppaay ]	<second person> <singular> <future tense>	ப்(சந்தி) + ப்(எதிர்கொல இடைநிலை) + ஆய் (முன்னிலை ஒருனை)	[பார்ப்பாய்   paarppaay   you will see ]	[பார் + ப் + ப் + ஆய்   paar + p + p + aay]
[ப்பார்கள்  ppaargaL]	<third person> <plural> <future tense>	ப்(சந்தி) + ப் (எதிர்கொல இடைநிலை) + ஆர்கள் (படர்க்கை பலர்பால்விகுதி)	[பார்ப்பார்கள்   paarppaargaL   they will see ]	[பார் + ப் + ப் + ஆர்கள்   paar + p + p + aargaL]
[ப்பாள்  ppaaL]	<third person> <singular> <future tense> <female>	ப்(சந்தி) + ப்(எதிர்கொல இடைநிலை) + ஆள் (பெண் பெல் விகுதி)	[பார்ப்பாள்   paarppaaL   she will see]	[பார் + ப் + ப் + ஆள்   paar + p + p + aaL ]
[ப்பீர்  ppeer]	<second person> <plural> <future tense>	ப்(சந்தி) + ப்(எதிர்கொல இடைநிலை) + ஈர் (முன்னிலை பன்மை விகுதி)	[பார்ப்பீர்   paarppeer   you will see ]	[பார் + ப் + ப் + ஈர்   paar + p + p + eer ]
[ப்பார்  ppaar]	<third person> <plural> <future tense>	ப்(சந்தி) + ப்(எதிர்கொல இடைநிலை) + ஆர் (படர்க்கை பலர்பால் விகுதி)	[பார்ப்பார்   paarppaar   they will see]	[பார் + ப் + ப் + ஆர்   paar + p + p + aar]
[ப்பேன்  ppaen]	<first person> <singular> <future tense>	ப்(சந்தி) + ப்(எதிர்கொல இடைநிலை) + ஏன்(தன்மை ஒருமை)	[பார்ப்பேன்  paarppaen  i will see]	[பார் + ப் + ப் + ஏன்   paar + p + p + aen]
[ப்போம்  ppoam]	<first person> <plural> <future tense>	ப்(சந்தி) + ப்(எதிர்கொல இடைநிலை) + ஒம் (தன்மை பன்மை விகுதி)	[பார்ப்போம்  paarppoam  let us see]	[பார் + ப் + ப் + ஒம்   paar + p + p + oam ]
[ப்பான்  ppaan]	<third person> <singular> <future tense> <masculine>	ப்(சந்தி) + ப்(எதிர்கொல இடைநிலை) + ஆன் (ஆண்பால் விகுதி)	[பார்ப்பான்   paarppaan   he will see ]	[பார் + ப் + ப் + ஆன்   paar + p + p + aan]
[ப்பீர்கள்  ppeergaL]	<second person> <plural> <future tense>	ப்(சந்தி) + ப் (எதிர்கொல இடைநிலை) + ஈர்கள் (முன்னிலை பன்மை விகுதி)	[பார்ப்பீர்கள்   paarppeergaL   you will see]	[பார் + ப் + ப் + ஈர்கள்   paar + p + p + eergaL]

#### 4.2. RULES

Each pattern has a set of Rules associated with it which records the list of all possible transformations which may occur when that pattern combines with different words to form the inflected forms of the same.

When a root word combines with a specific pattern (suffix with some morphological information), the former may have a change in one or more letters, or may have one or more letters added to it to form the final inflected form. These transformations are governed by grammatical principles described in [13], [14] and [15]. These are classified as “Rules” under each pattern i.e., each pattern has an associated set of Rules governing the transformations which may occur when the pattern combines with a word to form the inflected form of the latter.

The relationship between patterns and rules is illustrated in the Figure 1.

Let

$n$  - number of patterns

Rules( $x$ ) - the set of rules for the pattern ‘ $x$ ’

Rule  $x,y$  - Rule number ‘ $y$ ’ of the pattern ‘ $x$ ’

$np_x$  - number of rules for pattern ‘ $x$ ’

Each of the ‘ $n$ ’ patterns have a set of rules associated with it. A pattern is a suffix which can be affixed with a word to form its morphological variants. Each of the rules under a pattern are numbered with a unique number also denoting the pattern it is associated with.

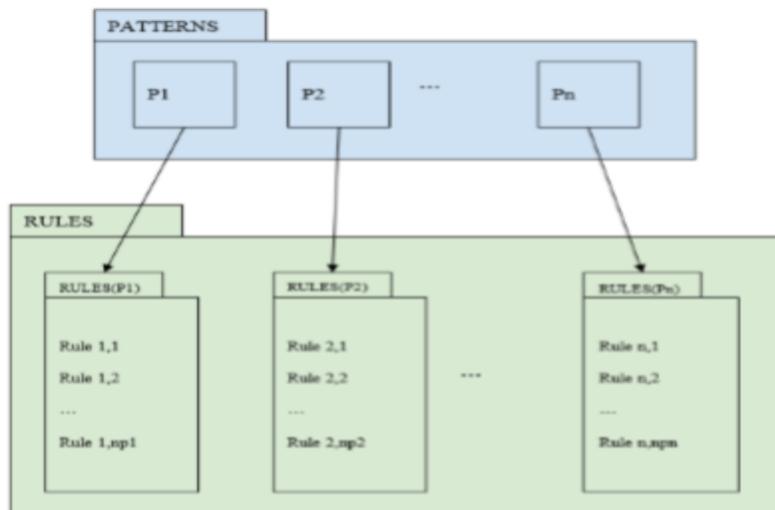


Fig 1. Relationship between patterns and rules

Example 5 (Rules associated with a noun pattern): The following table illustrates some of the rules which are associated with the pattern [கள்|kaL] which denotes plural.

Table 3. Rules associated with the pattern [கள் | kaL]

Pattern	Root noun	Rule	Inflected form	Inflectional split
[கள் kaL]	[மரம்   maram   tree]	[ங் > ம்   ng > m ]	[மரங்கள்  marangaL   tree<plural>]	[ மரம் + கள்   maram + kaL ]
	[வாழ்த்து  vaazhthu  greeting]	[க் > -   k > -]	[வாழ்த்துக்கள்   vaazhthukkaL  greeting <plural>]	[வாழ்த்து + கள்  vaazhthu + kaL]
	[முள்  muL  thorn]	[ட் > ள்   t > L]	[முட்கள்   mutkaL  thorn<plural>]	[முள் + கள்   muL + kaL]
	[சொல்  chol  word]	[ற் > ல்   R > l ]	[சொற்கள்   choRkaL  word<plural>]	[சொல் + கள்   chol + kaL]

In Table 3, the noun [மரம் | maram | tree], has [ம் | m] as the last letter. When [மரம்| maram | tree] is pluralized with the suffixing of the pattern [கள் | kaL], the “ம்” of “மரம்” has to be transformed to [ங் | ng]. [வாழ்த்து | vaazhthu | greeting] has a [க் | k] attached to it when pluralized with the suffixing of the pattern [கள் | kaL]. [முள் | muL | thorn] ends with the alphabet [ள் | L]. When suffixed with the pattern [கள் | kaL], the “ள்” of “முள்” changes to [ட் | t].

Example 6 (Rules associated with a verb pattern): Coming to Rules for verb inflections, some of the rules of the pattern [பட்டிருக்க | pattirukka] when combining with verbs are studied.

Table 4. Rules associated with the pattern [பட்டிருக்க | pattirukka]

Pattern	Root verb	Rule	Inflected form	Inflectional split
	[செய்   chey   do]	[ய > -   ya > - ]	[செய்யப்பட்டிருக்க   cheyyappattirukka   ]	[செய் + பட்டிருக்க   chey + pattirukka ]
	[வழங்கு  vazhangu  give]	[க > கு   ka > ku ]	[வழங்கப்பட்டிருக்க   vazhangappattirukka   ]	[வழங்கு + பட்டிருக்க   vazhangu + pattirukka ]
[பட்டிருக்க   pattirukka ]	[மூடு  moodu  close]	[ ட > டு   ta > tu ]	[மூடப்பட்டிருக்க   moodappattirukka  ]	[மூடு + பட்டிருக்க   moodu + pattirukka ]

[நிறுத்து   niRuthu  stop]	[ >து   tha >thu ]	[நிறுத்தப்பட்டிருக்க   niRuthappattirukka ]	[நிறுத்து + பட்டிருக்க   niRuthu + pattirukka ]
[பேசு  paesu  speak]	[ ச >சு   cha >chu ]	[பேசப்பட்டிருக்க   paesappattirukka ]	[பேசு + பட்டிருக்க   paesu + pattirukka   ]
[பரப்பு  parappu  spread]	[ப>பு   pa>pu ]	[பரப்பட்டிருக்க   parappattirukka   ]	[பரப்பு + பட்டிருக்க   parappu + pattirukka ]
[உதறு  udhaRu  shake ]	[ ற >று   Ra >Ru]	[உதறப்பட்டிருக்க  udhaRappattirukka ]	[உதறு + பட்டிருக்க   udhaRu + pattirukka ]

In Table 4, the root words end with the letters [கு | ku], [சு | chu], [டு | tu], [து | thu], [பு | pu], [று | Ru] which are formed by the combination [வல்லிை எழுத்துக்கள் | vallina ezhuthukkaL | hard consonants] with the vowel [உ | u]. The vowel [உ | u] changes to [அ | a] in these transformations.

### 4.3. EXCEPTIONS

These are transformations which do not follow any regular structure with respect to the patterns discussed earlier.

Example 7 (Some sample exceptions for verbs): The following table illustrates some exceptions which occur with verbs.

Table 5. Sample Exceptions

Root verb	Pattern	Implication of the pattern	Inflected form of the word	Intermediate inflectional split	Inflectional split
[வா   vaa   come]	[ந்தான்   ndhaan ]	<third person> <singular> <past tense> <masculine>	[வந்தான்  vandhaan  he came]	[வ + ந்தான்  va + ndhaan]	[வா +ந்தான்   vaa +ndhaan]
[வா   vaa   come]	[கிறான்  kiRaan]	<third person> <singular> <present tense> <masculine>	[வருகிறான்  varugiRaan  he comes]	[வரு + கி றான்  varu + kiRaan]	[வா +கிறான்   vaa +kiRaan]
[நில்  nil  stand]	[ஆன்   aan]	<third person> <singular> <past tense> <masculine>	[நின்றான்  ninRaan   he stood]	[நின்ற + ஆன்   ninRa + aan]	[நில் + ஆன்   nil + aan]

[கொள்   koL   have ]	[ஆன்   aan]	<third person> <singular> <past tense> <masculine>	[கொண்டான்   koNdaan   he had]	[கொண்ட + ஆன்   koNda + aan]	[கொள் + ஆன்   koL + aan]
[சாப்பிடு   chaappidu   eat ]	[ஆன்   aan]	<third person> <singular> <past tense> <masculine>	[சாப்பிட்டான்   chaappittaaan he ate]	[சாப்பிட்ட + ஆன்   chaappitta + aan]	[சாப்பிடு + ஆன்   chaappidu + aan]

The verb [வா | vaa | come] combines with the pattern [ந்தான் | ndhaan] to form [வந்தான் | vandhaan | he came] i.e., the alphabet [வா | vaa] changes to [வ | va]. The same verb combines with the pattern [கிறான் | kiRaan] to form [வருகிறான் | varugiRaan | he comes]. In this inflection, the alphabet [வா | vaa] changes to [வ | va] and the alphabet [ரு | ru] gets added as well.

## 5. RESULTS AND DISCUSSION

280 noun patterns and 880 verb patterns were collected. The number of rules for nouns, verbs and compound words are 50, 124 and 54 respectively. Each pattern is given a specific ID. All the rules applicable for a specific pattern are grouped with the ID allotted for the pattern. These inflection rules form a repository which can be used to form a variety of NLP tools like morphological analyser, morphological generator, word level translator, spelling and grammar checker etc.

The grouping and classification of inflectional rules in Tamil discussed in this paper have been applied to build Piripori [11], a Tamil morphological analyser. A list of 232292 Tamil words were taken from different sources and were input to Piripori and Tamil Computing Lab analyser. In Table 6, the analysis times for 137144 Tamil words for Tamil Computing Lab analyser [10] and Piripori [11] are tabulated. The time taken by Piripori to analyse the 232292 words given as input is lesser than that taken by Tamil Computing Lab analyser. The graphical representation of the data is given in Figure 2.

Table 6. Analysis times of TaCoLa and Piripori analysers

	Analysis time in nanoseconds	Analysis time in minutes
<b>TaCoLa analyser</b>	4.65535E+11	7.75891543
<b>Piripori</b>	44597119480	0.74328532

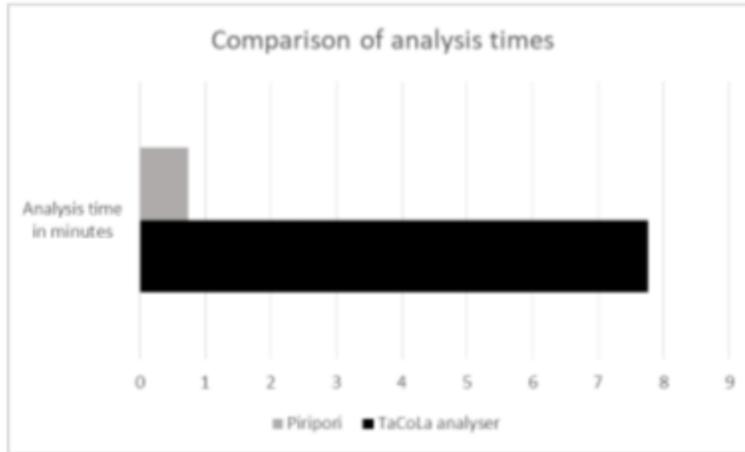


Fig 2. Comparison of analysis times of Tamil Computing Lab analyser and Piripori

To compare precision and recall of the two analysers, a standard POS tagged document [16] was taken from LDCIL website as test data. The file consisted 164 words out of which 121 were nouns and 43 were verbs. Precision and recall of analysis of noun and verbs in the abovementioned file were calculated separately. The results of the same are tabulated in Table 7.

Table 7. Data for precision and recall calculation

	Tacola	Piripori
No. of nouns correctly classified	85	110
No. of verbs correctly classified	20	25
No. of nouns classified as verbs	9	0
No. of verbs classified as nouns	12	5

Precision and Recall are calculated using the following formulae.

$$\text{Precision}_{\text{noun}} = (\text{No. of nouns correctly analysed}) / (\text{No. of words analysed as nouns})$$

$$\text{Precision}_{\text{verb}} = (\text{No. of verbs correctly analysed}) / (\text{No. of words analysed as verbs})$$

$$\text{Recall}_{\text{noun}} = (\text{No. of nouns correctly analysed}) / (\text{No. of nouns})$$

$$\text{Recall}_{\text{verb}} = (\text{No. of verbs correctly analysed}) / (\text{No. of verbs})$$

The results of the calculation are tabulated in Table 8 and represented graphically in Fig 3.

Table 8. Precision and Recall of nouns and verbs

	TaCoLa	Piripori
Precision of Nouns	0.87628866	0.9565217
Recall of Nouns	0.702479339	0.9090909
Precision of Verbs	0.689655172	1
Recall of Verbs	0.465116279	0.5813953

The precision and recall values of nouns and verbs of Piripori, which is based on the grouping of grammatical rules discussed in this paper are higher than those of Tamil Computing Lab analyser.

Thus with the application of the grouping of grammatical rules into Patterns, Rules and Exceptions as discussed in this paper has led to a better performing Tamil morphological analyser in terms of analysis time, precision and recall compared to another analyser following the traditional rule based approach.

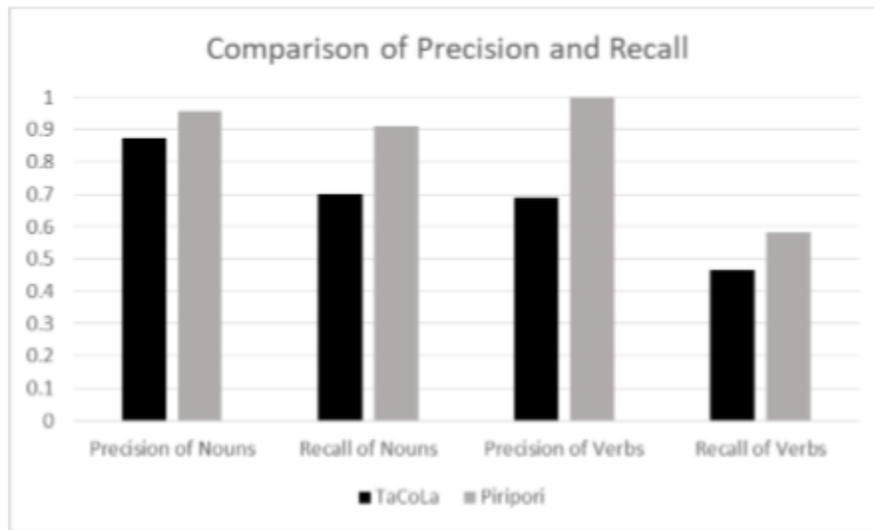


Fig 3. Comparison of Precision and Recall

## 6. CONCLUSION

Three levels of Tamil word inflection rules are considered in this work – Rules, Patterns and Exceptions. How grammatical rules for word inflections in Tamil can be grouped in these three levels and applied for obtaining different word forms is the focus of this paper.

Efficient application of syntax and grammatical rules of a natural language in a natural language application or tool is very important for minimal error. This work focuses on the collection of inflectional rules in Tamil grammar referring [13], [14] and [15] thus forming a repository which can be used in various NLP tools and applications. There has been great improvement in retrieval speed and accuracy with respect to precision and recall when these rules were used to build a morphological analyser compared to the traditional rule based approach.

## ACKNOWLEDGEMENTS

The authors would like to thank Subalalitha C.N, Kanniammal.R, Manju Bashini P, Elanchezhiyan K, Rajapandian C, Swathi P.G and Sindhuja G for their guidance and contribution to this work

## REFERENCES

- [1] Omnicore.[Online]. Available: <https://www.omicoreagency.com/twitter-statistics/>
- [2] L.J.Brinton, *The Structure Of Modern English: A Linguistic Introduction*. Amsterdam, Philadelphia, PA: John Benjamins, 2000.
- [3] UC Sandiego Linguistics Department.[Online]. Available: [http://grammar.ucsd.edu/courses/lign120/08-intro\\_rev.pdf](http://grammar.ucsd.edu/courses/lign120/08-intro_rev.pdf)
- [4] S. Singh And V. M Sarma, “Hindi Noun Inflection And Distributed Morphology” In *Proceedings Of The 17th International Conference On Head-Driven Phrase Structure Grammar*, 2010, Pp. 307321
- [5] M. Ramscar , “The Role Of Meaning In Inflection: Why The Past-Tense Does Not Require A Rule,” *Cognitive Psychology*, Vol. 45, No. 1, Pp. 45–94, 2002.
- [6] Wikipedia.[Online]. Available: <https://en.wikipedia.org/wiki/Agglutination>
- [7] Wikipedia.[Online]. Available: [https://en.wikipedia.org/wiki/Agglutinative\\_language](https://en.wikipedia.org/wiki/Agglutinative_language)
- [8] S. C. Reddaiah. “Dravidian Languages And Its Fundamental Grammar,” *Indian Journal Of Research*, Vol. 3, No. 2, Pp. 164-166, 2014.
- [9] Anand Kumar M, Dhanalakshmi V, Soman K.P And Rajendran S, “A Sequence Labeling Approach To Morphological Analyzer For Tamil Language”, *International Journal On Computer Science And Engineering*, Vol. 2, No. 6, Pp. 1944 – 1951, 2010
- [10] P. Anandan, K. Saravanan, R.Parthasarathi And T. V. Geetha, “Morphological Analyzer For Tamil” In *Proceedings Of International Conference On Natural Language Processing*, 2002
- [11] Suriyah M, Aarthi Anandan, Anitha Narasimhan And Madhan Karky, “Piripori - Morphological Analyser For Tamil” In *International Conference On Artificial Intelligence, Smart Grid And Smart City Applications*, 2019.
- [12] [கலஞ்சியம் | Kalanjiam].[Online]. Available: <http://store.tamillexicon.com>
- [13] Maanikkavaasakan, *Tholkaappiyam*, Chennai, TN : Uma Padhippagam, 2010
- [14] A. Manikkam, *Nannool Kaandigaiyurai*, Chennai, TN : Poompuhar Padhippagam, 1988
- [15] Seeni Naina Muhammad, *Nalla Tamizh Ilakkanam*, CITY, TN : Adayalam Padhippagam, 2013
- [16] Linguistic Data Consortium For Indian Languages. [Online]. Available: <http://www.ldcil.org/standardstextpos.aspx>

**AUTHORS**

Suriyah M is a post graduate in Computer Science and Engineering with 4 years' experience in teaching graduate engineering students. Now she is a research writer with Karky Research Foundation.

Aarthy Anandan is a post graduate in Computer Applications. Now she is working as Software Developer with 4 years' in Karky Research Foundation.

Anitha Narasimhan is a post graduate in Computer Applications. Now she is working as Software Developer with 4 years' in Karky Research Foundation.

Madhan Karky is a Ph.D in Computer Science. He is the Founder and Research Head of Karky Research Foundation, a nonprofit educational research organization which focuses on language computing and language literacy.