International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

Noun Identification for Tamil Language using Morphophonemic Rules

M. Mercy Evangeline, K. Shyamala

Abstract*: In any language, words are considered as the basic or the smallest element with a distinctive meaning. Words can be categorized into several types, depending upon their use and functions. Basically how a word changes its form to express itself in grammatical notation defines its type. The process of categorizing a word to a particular type depending upon its grammatical notation is termed as Part of Speech tagging. In this paper, an attempt has been made to identify part of speech tagging for words in Tamil language, particular to noun inflections. An algorithm has been proposed for Noun Identification for Tamil Language using Morphophonemic Rules (NIMR). A Rule based suffix stripping approach has been adopted for this implementation. The approach proposed here identifies the root word by applying various morphophonemic rules particular to suffixes. It removes the various inflections based on the set of grammatical rules available for Tamil Language and tags the word identified as a Noun. It is proposed to explore the traditional way of categorizing words in Tamil language, avoiding the influence of English grammars.

Keywords*: Tamil, Part of Speech tagging, Tamil Grammar, Text Mining, NLP, word identification, Noun, Morphophonemic rules

I. INTRODUCTION

Part of speech tagging (பேச்சின் பகுதி) is generally described as the process of reading a text in a language and assigning part of speech to each word or tokens. The part of speech generally assigns every token in the given text as noun, verb, adjective, etc. generally based on its context and definition. This process is not a straightforward method, as the token may have different part of speech on the context of its usage.

Tamil is a Dravidian Language spoken by more than 75 million people across the world [14]. It is one of the oldest languages. Tamil is an agglutinative language, so the words are formed by combining several morphemes. Morpheme is the smallest meaningful unit of representation in any language. But these entire morphemes remain unchanged even after union. Generally, a word is a combination of root word and other grammatical accretions. The root word can be traced up to monosyllabic level by removal of suffix accretions, irrespective of the complexity, length and type of the word. English language follows a SVO (Subject – Verb – Object) structure, where as Tamil follows a SOV (Subject – Object - Verb) structure.

The vowels are called as உயிர் எழுத்துக்கள். The consonants are called as மெய் எழுத்துக்கள். Tamil consonants are categorized into three categories, வல்லினம், மெல்லினம் and இடையினம். Tamil script consists of 12 vowels, 18 consonants and one special character, ஃ (aytam). Tamil vowels and consonants along with aytam for 216 compound characters, bringing the total number of characters to 247 (12 + 18 + 1 + (12 × 18)). The vowels are categorized as 5 short vowels (அ, இ, உ, எ, ஒ) and 7 long vowels (ஆ,ஈ,ஊ,ஏ,ஐ,ஓ,ஔ). The 18 consonants are divided into a category of 3, namely, hard sounding letters - வல்லினம் (க, ச, ட, த, ப, ற), soft sounding letters மெல்லினம் (ங, ஞ, ண, ந. ம, ன) and middle sounding letters இடையினம் (ய, ர, ல, வ, ழ, ள). Tamil Part of Speech generally included four types, Nouns, Verbs, Particles and Adjectives. Pronouns are included with nouns, Preposition, Conjunction and Interjection within Particles and Adverb with Adjectives [5].

Traditionally, a Tamil word is divided into a maximum of six parts, namely பகுதி(prime-stem), சந்தி (junction), விகாரம் (variation), இடைநிலை (middlepart), சாரியை (enunciater) and விகுதி (terminator) in that order.

Different types of words which can be available in any given sentence for Tamil Language will include Nouns (பெயர்ச்சொல்), Verb (வினைச்சொல்), Preposition and postposition (இடைச்சொல்), Adjective(உரிச்சொல்). Apart from this classification, there is another set of classification available under the literary lore. They are இயற்சொல், திரிசொல், திசைச்சொல் and வடசொல். The words combine with prefix and suffix forming different inflections. This work mainly focuses on the inflections formed by adding suffixes for a word. Here the word suffixes are identified for whether it is in number inflection or case inflection.

II. LITERATURE SURVEY

Suriyah & Anandan, et al [4] have identified basic inflectional principles for Tamil Language at word level. They have defined three levels of groups – patterns, rules and exceptions. The different grammatical rules were categorized into these three groups. Using this method, a repository of inflectional rules was created which can be used for various NLP applications.

Manuscript published on November 30, 2019.
* Correspondence Author
M. Mercy Evangeline*, PG & Research Department of Computer Science, Dr. Ambedkar Govt. Arts College(Autonomous), Affiliated to University of Madras, Chennai, Tamil Nadu, India.
Dr. K. Shyamala, Associate Professor, PG & Research Department of Computer Science, Dr.Ambedkar Govt. Arts College (Autonomous), Affiliated to University of Madras, Chennai, Tamil Nadu, India.

This work mainly focuses on how words and their different forms can be obtained based on the different grammatical rules forms the main focus.Another method for extracting individual morphemes using Finite State Automata (FSA) was proposed by Sobha Lalitha Devi et al. [3]. A morpheme analysis by including inflectional patterns as an FSA was performed. These inflectional patterns define the root words and their endings in a particular model. This also included a lexicon of rood words. Anshu Sharma et al. have proposed a stemmer algorithm for Hindi Language, which is based on a Hybrid approach [11]. This approach is a hybrid of Brute force approach, Suffix stripping and substitution method. In this approach, query entered will be searched with words in dictionary using brute force method, then suffix stripping and substitution methods, one at a time. Hindi nouns show inflection only for number and case. It can either be singular or plural in number. In case it can be direct or oblique. For this approach rules were used from previous literature and some were added for improving the accuracy.

Prasad et al. have discussed about methods available for transforming a singular word to its Plural form and methods of obtaining plural forms from the inflections [8]. This work is particular to Telugu Language. Plural forms of a word are obtained by employing different set of rules along with few exceptions. According to the methods adopted in this paper, a considerable amount of plurals were identified for the Telugu language.

A Tamil Morphological generator solution using supervised manner for learning word formation was proposed by Rajan,K., et al [2]. The proposed model learns the sound changing rules that take place when words combine, from the given training dataset. This model is also trained to learn the rules using various learning algorithms. It concludes that the word formation can be learned successfully using supervised manner instead of explicitly describing the rules.

Jancy Joseph and Dr.Babu Anto have proposed a Rule based analyser for Malayalam nouns which uses the technique of suffix stripping with a reverse application of sandhi rules in Malayalam [9]. This stripping method uses a Rule-cum-Dictionary based approach. Here a dictionary of predefined root words was defined for the same. When a word is inputted, it is compared with the dictionary; if available it is tagged as noun. Otherwise suffix stripping is done with rules and the stem word generated is compared. The uniqueness of this approach is that the input and output are handled with Malayalam scripts; they are not transliterated before and after process.

Ganapathiraju et al. have developed a Morphological generator of Telugu [13] which generates all the Noun forms of a given word including case forms, plural forms. It also generates the Verb forms of the word for all numbers and genders. This generation is done automatically except few irregular noun forms. The verb generation is fully automatic

Rupali Deshmukh and Varunakshi Bhojane, have done a survey of Sandhi splitting techniques available for different Indian Languages like Sanskrit, Hindi, Marathi and Malayalam [10]. All the Sandhi splitter taken for comparison work on Rule-Based algorithm for finding the root stem from the given input. Some splitter uses a dictionary of stem words along with the rule based approach. All these splitters have some limitations.

Reji Rahmath et al. have proposed Morphological generator for Nouns using Memory Based Language Processing for Malayalam Nouns [12]. This method uses a training system which generates Nouns which includes number, case and last syllable of the word. It doesn't use rule based or dictionary based implementation working for its identification. A training corpus was created for training the system.

III. IDENTIFICATION OF NOUN INFLECTIONS USING MORPHOPHONEMIC RULES

From morphological point of view, the two major word classes for Tamil are Nouns and Verbs. Noun is a word representing a person, place, animal, thing or an event. A pronoun is a word which takes the place of a Noun in a sentence. When a word talks about the action performed, then it is defined as a Verb.

Nouns are words formed by combining root stem and suffix. Noun stem is a word without any inflectional suffixes. Generally Noun stem in Tamil is generally inflected for case and number. A noun can be generally shown as Noun + (Plural Marker) + Case Marker. Plural Marker indicates the variations in number, whether it is single or many. The Case Markers represents the changes the word undergoes when it is available in a sentence or when it is combined with the adjacent word.

The different forms of inflections for a noun stem have been listed in Table I [1]. A noun stem can be a simple stem or a complex form. A noun stem without any inflections is represented by Nominative form. A complex noun will have a root stem and a derivational suffix. On phonological level, Noun can have four types of suffix - plural suffix, oblique suffix, euphonic suffix and case suffix [6].

i. When a noun is inflected for 'number' variation, a plural suffix is added. For example,

வீடுகள் = வீடு + கள்.

ii. When a noun is inflected for case only, it can be added either directly to the noun stem or oblique stem. An Oblique stem is a noun formed by suffixation of an oblique suffix to the noun or by doubling of a consonant.

மரத்தால் = மரம் + த்த் + ஆல்

(Root + Oblique Increment + Case suffix (Instrumental))

iii. When a noun is inflected for both number and case, the plural suffix will be added first, followed by euphonic increment which is optional and ending with case suffix.

வீடுகளை = வீடு + கள் + ஐ

(Root + Plural suffix + Case Suffix (Accusative))

சாவியினால் = சாவி + இ (இன்) + ஆல்

(Root + Euphonic Increment + Case Suffix (Instrumental))

A. Identification of Nouns with Case Suffix

General case suffixes which are added with a noun word are specified as rules below.

In Tamil Language, there are eight case markers available for a word in suffixed forms. These represent the general possible inflections a word can have when they combine with other words in a sentence.

1. If the suffix begins in a vowel sound and the word ends in இ, ஈ, ஏ, or ஐ, insert a ய் in between.
2. If the suffix begins in a vowel sound and the word ends in அ, ஆ, ஊ, ஒ, ஓ, or ஔ, insert a வ் in between.
3. If the suffix begins in a vowel sound and the word ends in உ sound, and...
   a. If the word is made of 2 short letters, insert a வ் in between.
   b. If the word is not made of 2 short letters, drop the -உ and add the suffix.
4. If the suffix begins in a vowel sound and the word is made of 2 short letters, with the 2nd letter being a consonant, then double the 2nd letter and add the ending.

Table- I: Possible Noun Inflections

		1	2	3	4
Category	**Suffix**	**ய் inserted**	**வ் inserted**	**Doubling**	**த்த் inserted (ம்)**
Nominative	-	தடி	பசு	கண்	மரம்
Accusative	-ஐ	தடியை	பசுவை	கண்ணை	மரத்தை
Instrumental	-ஆல்	தடியால்	பசுவால்	கண்ணால்	மரத்தால்
Dative	-உக்கு	தடிக்கு	பசுவுக்கு	கண்ணுக்கு	மரத்துக்கு
Locative	-இல்	தடியில்	பசுவில்	கண்ணில்	மரத்தில்
Ablative	-இலிருந்து	தடியிலிருந்து	பசுவிலிருந்து	கண்ணிலிருந்து	மரத்திலிருந்து
Sociative	-உடன்/ ஓடு	தடியுடன்	பசுவுடன்	கண்ணுடன்	மரத்துடன்
Genitive	-இன்/இனுணடய	தடியின்	பசுவினுடைய	கண்ணினுடைய	மரத்தினுடைய
Adjective	-ஆன	தடியான	பசுவான	கண்ணான	மரமான
Adverb	-ஆக	தடியாக	பசுவாக	கண்ணாக	மரமாக

The Fig. 1(a) represents the different inflectional forms for Accusative class form for Noun. State 1 represents the noun stem with a suffix of any one form - ய் inserted or வ் inserted or consonant doubling or த்த் inserted. Depending upon the suffix combination, it is removed for generating the noun root stem.

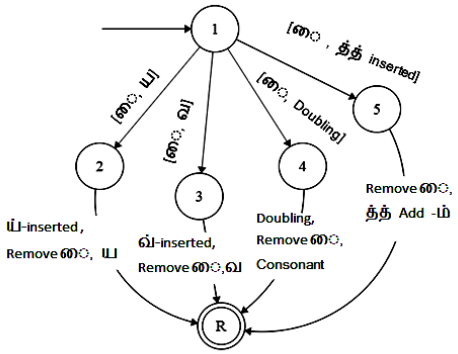

Fig. 1.(a) FSA for Noun Identification with Accusative Rules

The Fig. 1(b) represents the different inflectional forms for Accusative class form for a Noun. The words which are ending with suffix -ஐ are identified. Then the previous characters are checked for a pattern. Depending upon the suffix combination, the corresponding suffixes are removed for generating the noun root stem.

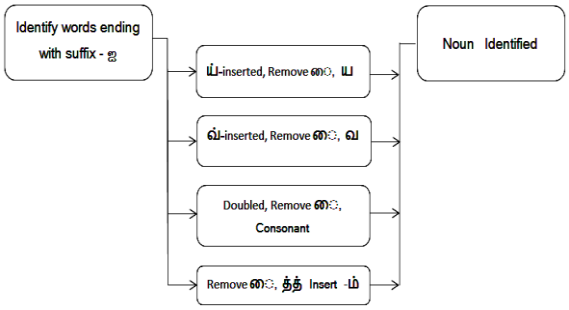

Fig. 1.(b) **Identifying Nouns with Accusative Rules**

The Fig. 2(a) represents the different inflectional forms for Ablative class form for a Noun. State 1 represents a word with a suffix of any one form - ய் inserted or வ் inserted or consonant doubling or த்த் inserted. Depending upon the suffix combination, it is removed for generating the noun root stem. For each state a different character is added or removed.

For example, if the word is தடியை, then it represents that ய் is inserted. So யை is removed and the root word is tagged as Noun. The word is splitted as தடியை → தடி + ஐ. Here ஐ represents the Case suffix for the root word.

Let us consider another word மரத்தை. Here doubling of character த்த் inserted has taken place. For such suffixes, த்த் is removed and ம் is inserted. Then it is tagged as Noun.

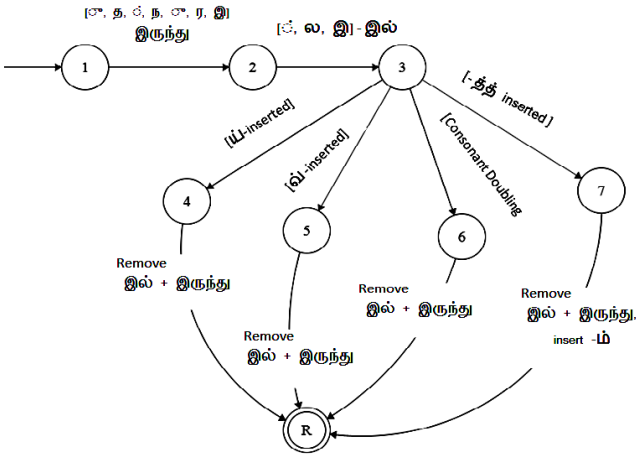

Fig. 2.(a) FSA for Noun Identification with Ablative Rules

The Fig. 2(b) represents the different inflectional forms for Ablative class form for a Noun. The words which are ending with -இலிருந்து are identified. Then the previous characters are checked for the different patters. Depending upon the suffix combination, the corresponding suffixes are removed for generating the noun root stem.

B. Identification of Nouns in Plural Forms

Plural forms are mainly used for representing more than one of the designated things. It is a grammatical form of changing a singular to plural form. Generally, there are two Plural forms – regular and irregular plural form.

For regular forms, there are some standard set of rules which tells how to get the plural form for a given noun.

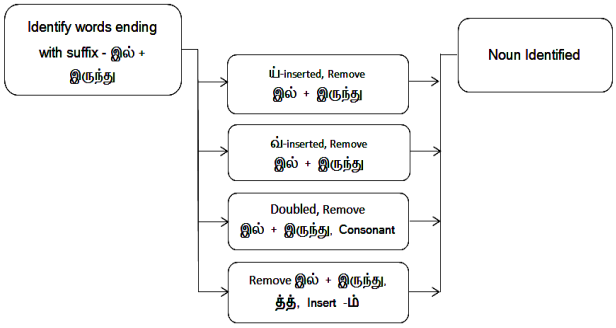

Fig. 2.(b) FSA for Noun Identification with Ablative Rules

For example, in English language, plural form is obtained by adding –s to the noun. (Boy → Boys, Book → Books). Nouns that end with -ch, x, s, z or s-like sounds requires –es for plural form. (Box → Boxes, Witch → Witches). Nouns that end in a consonant + y, drops y and takes ies (Baby → Babies) and those that end with –o takes –es as the plural (Potato → Potatoes).

Tamil nouns which are either human or non-human have number distinction. They are singular and plural. Always the singular property is left unmarked. Only the plural property is marked with suffixes.

In Tamil Language the suffix added for changing a singular form to plural is by adding -கள். Apart from this, some words ending with a specific letters change while becoming plural. Table III gives a set of morphophonemic rules for Plural formation of a given noun. Fig. 3. (a) represents the different morphophonemic rule change which can be identified if a word is in its plural form. The Fig. 3.(b) gives a simple Finite State Automata for suffixes being to a noun for its plural form.

A word which is considered for tagging is checked for its suffixes. If it ends with –கள், it is checked for the next combination. Check for the suffix, whether it has only – கள் or – க்கள், if so drop it and tag the word as Plural noun. If not check for next suffix character , whether it is (ங்,ற், ட்). If so drop them and change with the corresponding characters (ம், ல் , ள்).

Table - II discusses about the various forms taken by a noun when it is represented in plural form. In Tamil language, a noun can be as singular or plural form. The plural form is marked by a suffix of – கள் and – க்கள். A plural form is always added as a suffix to a noun stem and it is never added to an oblique stem. The plural suffix – க்கள் is always added with nouns ending with long vowels and nouns with two syllables and ending with –உ. The plural suffix – கள் occurs with all other nouns.

Generally, the plurals are identified by suffix – கள். It does not start with a vowel, so any of the special rules for having suffixes doesn't apply on this. Some words that end in certain letters change while becoming plural.

Table - II Various Plural forms for a noun stem

கல் + கள்→கற்கள் பல் + கள்→பற்கள்	Consonant change ல்-ற்
முள் + கள்→முட்கள் நாள் + கள்→நாட்கள்	Consonant change ள்-ட்
படி + கள்→படிகள்	Remove – கள்
படம் + கள்→படங்கள்	
பூ + கள்→பூக்கள்	Remove – க்கள்

Some rules which are adopted for Plural suffix are given below.

- If a word ends in – ம், then it is dropped as – ங்கள் is added
- If the word is single syllable word:
  - ✓ And ends in a long vowel, add – க்கள்
  - ✓ Has a short vowel sound and ends in :
    - - ல், ற் is added by dropping ல்
    - - ள், ட்கள் is added by dropping ள்
- For all other nouns add - கள்

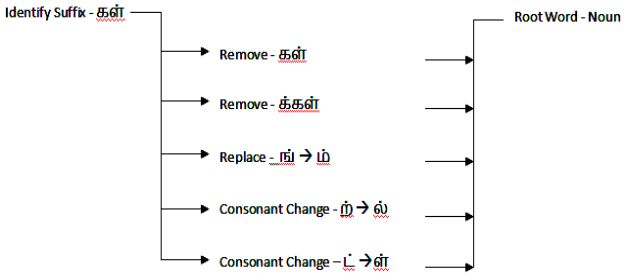

Fig. 3.(a) Identification of Noun Root by Suffix Removal from its Plural form

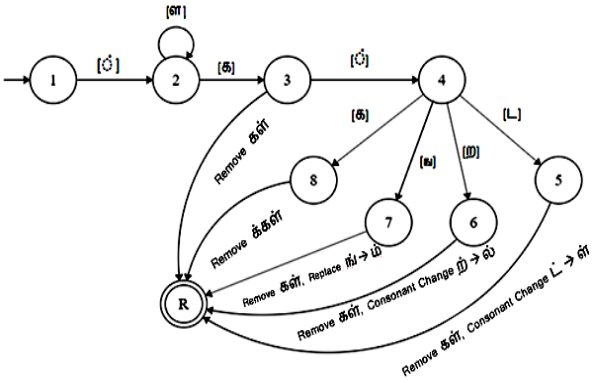

Fig. 3.(b) FSA for Noun Identification in Plural forms

Almost, many words fall into the last category. The one syllable words with long vowel sound are words with one letter. Examples include தீ, பூ, and ஈ. They are very few. There are not so many words with short vowel sound and ending in ல் or ள்.

Table- III Plural Inflections for a Noun

Suffix	Word	Root Noun	Morphophonemic Rule	Inflected Form	Word Split
[கள்]	படிகள்	படி	Remove கள்	படிகள்	படி + கள்
	பூக்கள்	பூ	Remove க்கள்	பூக்கள்	பூ + கள்
	படங்கள்	படம்	Remove கள், Replace ங்→ம்	படங்கள்	படம் + கள்
	சொற்கள்	சொல்	Remove கள், Consonant Change ற்→ல்	சொற்கள்	சொல் + கள்
	நாட்கள்	நாள்	Remove கள், Consonant Change ட்→ள்	நாட்கள்	நாள் + கள்

IV. NOUN IDENTIFICATION IN TAMIL USING MORPHOPHONEMIC RULES (NIMR)

The proposed algorithm, NIMR gets an intermediate file which has been pre-processed. Pre-processing involves removal of punctuation marks, tokenizing words and removal of stop words from the input text file. The Input file is a text file saved in UTF-8 format. This input text file is pre-processed to generate a text file with a collection of words, which is an output available after tokenization and stop word removal. The stop word is removed by using Dictionary Based Stop Word Removal algorithm (DBSWRA) proposed earlier [7]. This algorithm is dictionary based method, where words considered as stop words are stored in a dictionary and is used for pre-processing [7]. The output text file generated at the second level is given as input to the proposed NIMR algorithm. After the identification it generates a text file with list of words identified as nouns.

In this Rule-based approach, the Sandhi Rules for Tamil are taken into consideration. Generally, in Tamil language, Sandhi addition is understood using Vetrumai Urubugal. There are eight variations of urubugal available for Tamil Language. They are முதல் வேற்றுமை (எழுவாய் வேற்றுமை) – Nominative Case, இரண்டாம் வேற்றுமை – Accusative Case, மூன்றாம் வேற்றுமை – Instrumental/Social Case, நான்காம் வேற்றுமை – Dative Case, ஐந்தாம் வேற்றுமை – Ablative Case (from inanimate object/ animate object, ஆறாம் வேற்றுமை – Genitive Case, ஏழாம் வேற்றுமை – Locative Case, எட்டாம் வேற்றுமை (விளி வேற்றுமை) – Vocative Case.

In this Rule based approach, the first and the last case have not been taken into consideration. Other 6 case rules have been included for identifying nouns. These cases represent a set of grammatical relation a stem word can undergo when it is combined with another word.

Fig. 4. Represents the different steps involved in identification of noun for a given input file. Each block performs a specified task and gives an output text file saved in UTF-8 format.

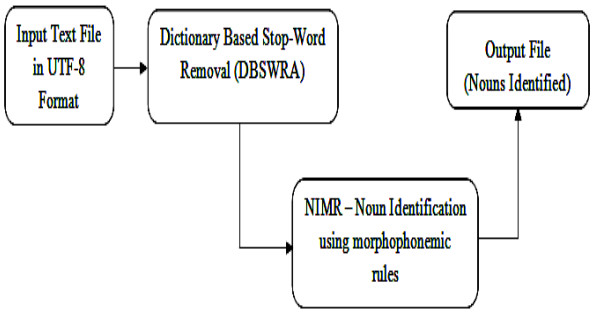

Fig. 4.Noun Identification in Tamil Language using Morphophonemic Rules (NIMR)

Algorithm: Noun Identification using Morphophonemic Rules: (NIMR)

Input: A text file with list of words which has been pre-processed.

Output: An array of words identified as nouns.

Step 1: The input text file is compared with a list of pronouns and are removed

- *This list include words like* நீ நான்

நாங்கள் அவர் அவர்கள் நீங்கள் அவன்
அவள் நாம் அவை

Step 2: The output generated from the previous step goes through a module for checking whether it is a noun or not. This module identifies nouns in plural form and nouns which has undergone transition with Suffix addition rules. These rules include the following steps:

For each word in the list of words
Rules are checked for the word suffix
If satisfied
Tag the word as noun

Step 3: The output is generated as a list of words identified as nouns.

The proposed NIMR algorithm takes a list of words from a text file which is pre-processed. Each word is analysed with the set of rules defined based on the grammatical representation of nouns in Tamil. Noun suffixes are identified first. The suffixes are identified using their Unicode character values. The novelty of this method is, the input and output words does not undergo transliterations and re-transliteration. Each character is checked for its Unicode value and checked for the combination.

The different rules are framed depending upon the reverse rule of suffix addition. The output generated is compared with nouns which were identified manually.

Manually, nouns are checked in each input file and are created as a separate text file. This text file is array of words which are nouns. It does not include nouns which are proper nouns, nouns which haven't undergone transition and verbs. The text file contains only words which are nouns undergoing transitions due to Suffix addition. It also includes nouns which are in their plural forms.

The output of NIMR includes only the nouns which have gone transition according to the reverse rule of splitting the Sandhi. Words are generally combination of Morphemes. Some morphemes appear independently and some of them are bounded to be part of words known as bound morpheme. They appear in conjunction with root and with other bound morphemes. The ways in which these morphemes appear are defined as morphological rules. One important characteristic of morphemes is that they change when they are grouped together. These changes are known as morphophonemic changes.

Reverse rule of splitting Sandhi

Each word is checked for its suffix, starting from the last character. Patterns are matched from the reverse. If it abides by the rule then corresponding characters are removed or replaced with the characters according to the rule discussed earlier. Here the process of stripping the suffix character is known as the reverse process of Sandhi.

V. RESULTS AND DISCUSSION

For the implementation of NIMR algorithm, text files in UTF-8 format were considered. First the text was pre-processed using DBSWRA. The output generated was a list of words. This output was passed through NIMR for identification of nouns. Table. IV gives a review of files considered for implementation, the percentage of words identified correctly by NIMR compared to the manual method.

Table- IV: NIMR implementation for Noun identification

File Name	Total No. of words	Nouns identified		Non-Nouns identified by NIMR	% of Words correctly identified
		Manually	NIMR		
File 1	586	48	30	3	62.50
File 2	532	36	18	3	50.00
File 3	678	57	21	15	36.84
File 4	641	59	22	8	37.29
File 5	368	25	13	5	52.00
File 6	509	44	25	6	56.82
File 7	492	68	26	11	38.24

Fig. 5 shows the graphical representation of the words identified as nouns by NIMR. The algorithm has done the identification using the rules based on vetrumai urubugal. Out of the 8 vetrumai urubugal rules, it has excluded the first and last vetrumai urubu as the definition of nouns by these rules are very vast. NIMR algorithm has considered all the other rules for identification. It also identifies the words which are in the plural forms. In the manual identification of nouns from the text files, pronouns and proper nouns have been removed. It has included only the words which have gone transition by the case rules. During this process of identification, some words which are not a noun have been identified as nouns by NIMR. This is because of the transition they have in their suffix. But the percentage of these words is very minimal.

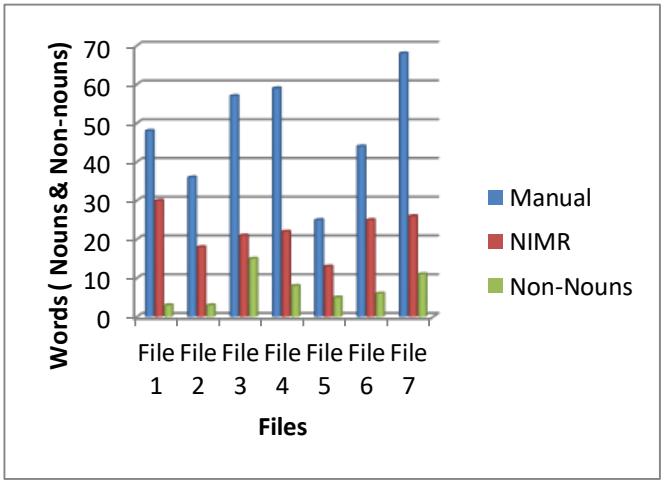

Fig. 5.Graphical Representation of Noun Identification in Tamil Language using Morphophonemic Rules (NIMR)

VI. CONCLUSION

NIMR algorithm has been designed with modules pertaining to different grammatical rules available for Tamil Language. The rule specifications have been taken from the book of Nannool. For obtaining the rood word, the suffixes attached with them are removed. The grammatical rules considered for removing these suffixes include the rule defined according to morphophonemic change taken by a word when it combines with adjacent morphemes. During the implementation, every word is analyzed for its suffixes.

If it falls under a category of morphophonemic rule, the suffix is stripped and the word is tagged as Noun. The input considered for this implementation includes story files in Tamil Language. They have been downloaded and saved in UTF-8 format. The output consists of list of words tagged as Noun. This output is compared with manually computed list of words for the corresponding files. The output shows that half the percent of noun words are identified correctly. Some words were prejudicially identified as nouns because of the transition undergone by them and some noun words were left unclassified.

REFERENCES

1. Rajan, K., V. Ramalingam, and M. Ganesan. "Machine Learning of Sandhi Rules for Tamil." Proceedings of the 11th International Conference INFITT, 2012.
2. Rajan, K. & Ramalingam, Dr & Ganesan, Dr., Machine Learning of Phonologically Conditioned Noun Declensions For Tamil Morphological Generators, 2014.
3. Devi, SobhaLalitha et al. “Morpheme Extraction in Tamil using Finite State Machines (FIRE-2013 - Morpheme Extraction Task).”, 2013.
4. M, Suriyah & Anandan, Aarthy & Narasimhan, Anitha & Karky, Madhan. (2019). “Isolating Word Level Rules in Tamil Language for Efficient Development of Language Tools”, International Journal on Natural Language Computing. 8. 29-42. 10.5121/ijnlc.2019.8103.
5. Charles Theophilus Ewald Rhenius , “A grammar of the Tamil language: with an appendix”, Church Mission Press, 1836, Oxford University
6. Thomas Lehman, “A Grammar of Modern Tamil, Pondicherry, India”, Pondicherry Institute of Linguistics and Culture, ©1989.
7. Mercy Evangeline M & Dr K Shyamala, (2018), “Stop word removal algorithm for Tamil Language”, International Conference on Recent Innovation in Electrical, Electronics & Communication Engineering, “CFP18P8 – PRT: 978-1-5386-5994-6.
8. Prasad, T. & Tiyyabindi, Kameswara Rao & Kumar, T. (2015). “Handling Plural Forms of Telugu Words in Machine Translation”
9. Jancy Joseph & Dr. Babu Anto, “Rule based Morphological Analyser for Malayalam nouns”, Computational Analysis of Malayalam Linguistics, IJIRCCE vol3, special issue 7, OCT 2015.
10. Rupali Deshmukh, Varunakshi Bhojane, *“*Sandhi Splitting Techniques for Different Indian Languages*”,* International Journal of Engineering Technology, Management and Applied Sciences (IJETMAS), December 2014, Volume 2 Issue 7, ISSN 2349-4476
11. Anshu Sharma, Rakesh Kumar, Vibhakar Mansotra, *“*Proposed Stemming Algorithm for Hindi Information Retrieval*”,* International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 6, ISSN(Online): 2320-9801, ISSN (Print) : 2320-9798, June 2016.
12. Reji Rahmath, K and P. C. Reghu Raj. *“*A Memory Based approach to Malayalam noun generation.” *2015 International Conference on Control Communication & Computing India (ICCC)* (2015): 634-637.
13. Ganapathiraju M, Levin L. “TelMore: Morphological Generator for Telugu Nouns and Verbs*”,* Alexandria, Egypt: Proceedings of Second International Conference on Universal Digital Library. 2006; p. 17-19
14. https://en.wikipedia.org/wiki/Tamil_language

AUTHORS PROFILE

Ms. M. Mercy Evangeline is currently pursuing her Ph.D in Computer Science, PG and Research Department of Computer Science, Dr. Ambedkar Govt. Arts College, Vyasarpadi, Chennai, Tamil Nadu, India. She has a Master’s degree in Computer Applications and M.Phil in Computer Science. Her areas of interest include Data Mining, Text Mining, Natural Language Processing and Machine Learning.

Dr. K. Shyamala is working as an Associate Professor in the PG and Research Department of Computer Science, Dr. Ambedkar Govt. Arts College, Vyasarpadi, Chennai, Tamil Nadu, India. She has her Masters degree, M.Phil and Ph.D. in Computer Science. She has 29 years of teaching and research experience. Six candidates have completed Ph.D. under her guidance. She has authored numerous books, published 62 research articles and conducted several conferences. She has also chaired sessions in International conferences. She has served as program committee member and chairman for Board of Studies in various colleges and universities. Her area of specialization includes Data Mining, WBAN, Agent Based Computing and Advanced Computer Networks.