International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Articles available online http://www.ijoer.in

Vol.3., Issue.3, 2015

RESEARCH ARTICLE

ISSN: 2321-7758

MODERN TAMIL COMPUTATIONAL GRAMMATICAL TERMS FOR SENTENCE MAKING IN NLP

Dr.K.NIRMALA[1], M.K.KALPANA[2*]
[1]Associate Professor, [2] Research Scholar
Department of Computer Science
*Quaid-E-Millath Government College for Women(TamilNadu),
Chennai, India.

Article Received:01/06/2015 | Article Revised on:17/06/2015 | Article Accepted on:20/06/2015

ABSTRACT
NLP stands for Natural Language Processing. Recent trends of technologies implementing with natural languages. NLG is stands for Natural Language Generation. It is a subfield of NLP and it is also referred to text generation. Language technology includes sounds, word formations, sentence structures, meanings and understandings. Real world NLP applications are Information retrieval, Information visualization, Speech synthesis and recognition, Optical handwritten/character recognition and Grammar checking systems.

Keywords— Tamil computational linguistics, NLG, Computational Sentence Making, Text-Based Applications and Grammar checking systems.

I. INTRODUCTION

Modern Tamil is a colloquial communication in South Indian region of Tamil Nadu. It is communicate by the style of formal language. Modern Tamil is created for the organizations, government and some institutions document preparation. Tamil help us to improve NLP Research easily for the beginner of the research in this domain. System allowed us to apply Modern Tamil grammatical rules. Each and every rule needs different kinds of methodology.

II. TAMIL COMPUTATIONAL LINGUISTICS

Tamil computational linguistics/Tamil NLP research is doing under some Educational universities, National and State level Government. In this domain, currently Speech synthesis and Recognition, OCR/OHCR and Tamil Grammatical/Linguistics based research is going on. Tamil Computational Linguistics consists of Tamil Unicode, Inter-lingual Transliteration and Methodology of Grammar Rules.

A. Tamil Unicode

Tamil Unicode is having 247 Tamil letters with and without dependent consonants/symbols and 9 North Indian letters named as Sanskrit for a colloquial communication. Totally, Tamil Unicode processing with the 256 letters with Independent vowels, Independent consonants and vowels. In Fig.1. Tamil old type writing keyboard layout is shows the Tamil Unicode independent vowels and consonants and dependent consonant symbols. In this figure, white rounded letters resembles dependent consonants. These marked dependent consonants make a consonant letter like this,

க + ி = கி

B. Inter-lingual Transliteration

The inter-lingual approach was inspired by chomsky's claim that regardless of varying 'surface' syntactic structures, languages share a common 'deep structure' (chomsky's 1965). In interlingua-based machine translation approach, the source language text is converted into a language independent meaning representation called 'interlingua'. An Interlingua represents all sentences

that mean the same thing in the same way regardless of the source language they happen to be in (Jurafsky and Martin 2000). From inter-lingual representation, texts are generated into other languages.

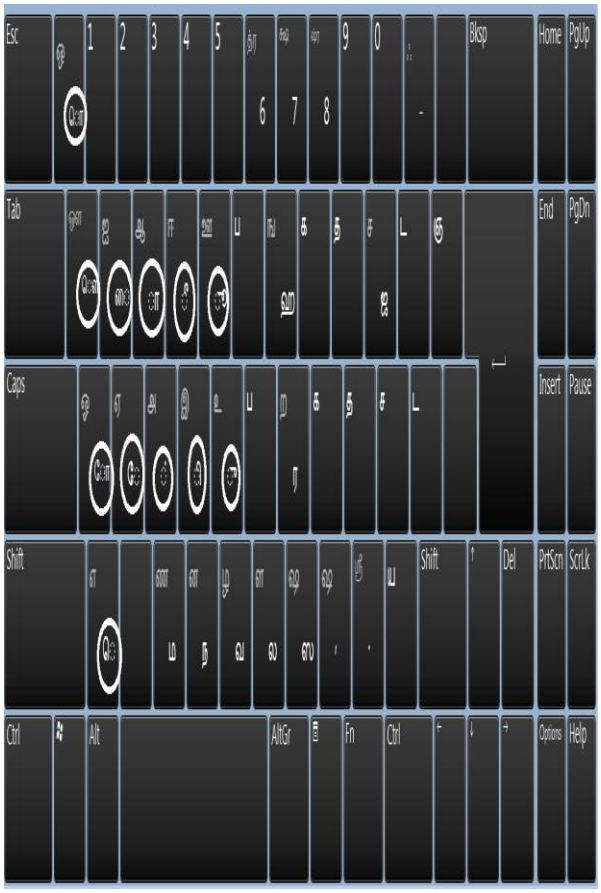

Fig.1.Tamil Keyboard layout

C. Methodology of Tamil Grammar Rules

Figure 2 represents the working flow of the Tamil Grammatical tool using inter-lingua machine translation approach.

The generation process can be viewed as a choice-making process (Reiter and Dale 2000). A pass through the grammar makes a choice regarding lexical and grammatical features. This choice is based upon input specification and information from the knowledge base. The successive recursive passes make a series of increasingly fine-grained choices. Taken together, these choices describe the syntactic characteristics of the sentence.

I. Applications of NLG

The NLG systems have been used to provide natural language interface to many databases, such as airline schedule database, accounting databases or spreadsheets, expert system knowledge bases, etc. The internal representations used by these systems cannot be understood easily by a non-expert user.

Fig. 2. Methodology for Tamil Grammar Rules

III. NATURAL LANGUAGE GENERATION

NLG technology has been fruitfully utilized to present such data in a human-readable form. Its use is not just restricted to databases and spreadsheets; it can be used to summarize graphical and speech data. NLG is also needed for generating the abstractive summary of a textual document. Given here is a list of some of the previous work involving application of the technique.

IV. MODERN TAMIL COMPUTATIONAL SENTENCE MAKING

Each and every comparison of the words collection is forming the sentences. In case, any wrong is occurred in the sentences, Tamil computational sentence making techniques are giving guidelines to solve the problems of sentence making error in Tamil. Sentence making is a vast are to work in both grammar and linguistics levels.

Basically, the research defines the error of the sentence making in a tool for the purpose of organizational purposes.

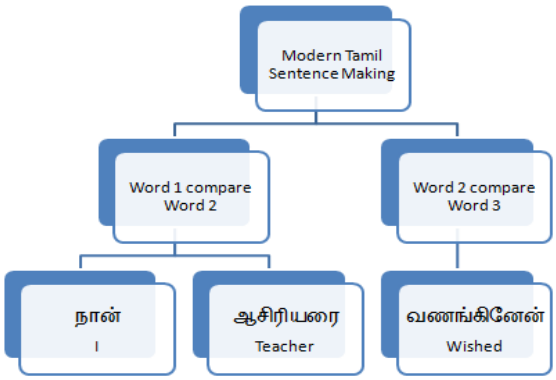

Fig. 3. Modern Tamil Sentence Making using Word-based paradigm.

In this above mentioned hierarchical representation. Word 1 is compared to the word 2 and Word 2 is compared to the Word 3. It is one of the chaining method to making a sentence.

V. TEXT-BASED APPLICATIONS

Tamil application development of the computer is in the level of creating the Grammar/linguistics-based applications like WordNet, Spellchecker, and Search Engines etc., In Tamil language also able to do patterned applications or customized applications. These types of applications are created by the organization. For an example, in the medical field doctors can able to create their own medical text based application using patient case study etc.,

VI. GRAMMAR CHECKING SYSTEM

Generally, Grammar checking system is processing under the Prose, poetry and novel. In this paper, Grammar checking system is following the general document correction terms using simple Modern Tamil grammatical terms. Every grammar checker is having many grammatical features for the use of the users, Author, officers, researchers, poet, government, medical etc.,

Modern Tamil Grammatical Computational terms are sandhi, noun, verb, preposition, etc., for making a sentence the researcher or programmer should be form the words using grammar rules like a fig.3. for the creation of custom grammatical terms, the researcher or programmer should have deep sense of knowledge about the grammar of the implemented language and in the mean time the researcher should identify relevant methodology for making a grammatical application.

Grammar checker created under the methods of untoken and token. This paper is processing with untokenized basis. Untokenized is the general process to making an application. Tokenization is named an entity to create a clear pattern. In this tokenization patterns are viewable for the querying the dictionary. In Untokenised things, the programmer/researcher should be noted about the grammatical terms manually and as per their idea they querying to the given information.

Tokenization may be used in both of the supervised and semi-supervised things. Un Tokenization only for the creation of customized and semi-supervised data. This paper processing in the basis of un tokenized manner. For the individual researcher implementation this methodology handled for the better understandings.

VII. CONCLUSIONS

Modern Tamil sentence making is same like an English language to check the errors of the sentence writing by the users. This sentence making techniques in the tools is used only for the formal behaviors, not for poetry or other type errors detection.

REFERENCES

[1]. G.Palanirajan. “POS Tagging for Classical Tamil Texts” infitt, 2012, 33.pdf.

[2]. Lakshmana Pandian S, and T V Geetha, “Morpheme Components based Parts of Speech Tagging”, International Conference on Natural Language Processing, 2008.

[3]. Rajendran S, (2007) Complexity of Tamil in POS tagging, Language in India, Jan 2007.

[4]. Rajan, K. Corpus analysis and tagging for Tamil. In: Proceeding of symposium on Translation support system STRANS-2002.

[5]. Andronov, M. 1969. The Standard Grammar of Modern and Classical Tamil. Madras: New Century Book House, Pvt. Ltd.

[6]. Arulmozhi. P, Sobha. L, Kumara Shanmugam. B. Parts of Speech Tagger for Tamil. In the Proceedings of the Symposium on Indian Morphology, Phonology & Language Engineering, Indian Institute of Technology, Kharagpur. pp 55- 57, 2004.

[7]. *Anandan, P. K. Saravanan, Ranjani Parthasarathi and T. V. Geetha. Morphological Analyzer for Tamil. In: International Conference on Natural language Processing, 2002.*

[8]. R.Akilan and Prof. E.R.Naganathan, "Morphological Analyzer for Classical Tamil texts: A rule-based approach", infitt, 2013,V7-11.pdf..

[9]. R.Akilan,E.R.Naganathan,"POS TAGGING FOR CLASSICAL TAMIL TEXTS" vol-1, issue-1, Jan-Jun-2012.

[10]. *http://en.wikipedia.org/wiki/Tamil_grammar*.

[11]. ”Natural Language Processing and Information Retrieval”, Tanveer siddique and U.S.Tiwary. Oxford Edition.