Solving the Noun Phrase and Verb Phrase Agreement in Kannada Sentences

B.M. Sagar, Shobha G and Ramakanth Kumar P.

Abstract— **This paper proposes a way of producing context free grammar for solving Noun and Verb agreement in Kannada Sentences. In most of the Indian languages including Kannada a verb ends with a token which indicates the gender of the person (Noun/ Pronoun). This paper shows the implementation of this agreement using Context Free Grammar. It uses Recursive Descent Parser to parse the CFG. Around 200 sample sentences have taken to test the agreement.**

Index Terms— **Context Free Grammar, Noun Phrase, Verb Phrase, Recursive Descent Parser.**

I. INTRODUCTION

Kannada has been estimated to be over 2, 500 years old, ranking as the 3rd oldest language after Sanskrit and Tamil. Kannada is one of the 5 major Dravidian languages. When Computational Linguistic is concern Kannada is lagging far behind compared to Telugu and Tamil. Kannada is a highly inflected language with three gender forms, masculine, feminine, neutral or common, and two number forms, singular and plural. The number forms interestingly shows inflection based on the gender, number and tense, of the commodity of reference, among other factors.

The application of being solving the Noun and Verb agreement lies in Natural Language Processing. Natural Language Processing (NLP) is an area of research and application that explores how computers can be used to understand and manipulate natural language text to do useful things. The goal of Natural Language Processing is to build computational models of natural language for its analysis and generation. These computational models provide a better insight into how humans communicate using natural language and also help in building of intelligent computer systems such as machine translation systems, text analysis etc. The Paper is organized as given below: In the II section background study is discussed. Section III about Noun phrase in Kannada. Section IV about the Verb Phrase and Verb Suffix extraction. Section V about the Context Free Grammar

and the Programming language Python is discussed. Section VI discusses reading Testing grammar using NLTK tool and finally VII concludes the paper.

II. PREVIOUS WORK

Very few Researchers are worked in Kannada Natural Language Processing. The following section shows the survey on various Language (Tamil, Bangla, Hindi Etc) Parsers with the type of sentences, methods used.

Bala sundara Raman L, Ishwar S, Sanjeeth Kumar Ravindranath, **(2003)** implemented Natural Language constructs for Venpa class of Tamil Poetry using Context Free Grammar. Push Down Automata is the parser used to parse the CFG. **[5]**

Ayesha Binte Mosaddeque & Nafid Haque, **(2004)** written Context Free Grammar for Bangla for 12 sentennces that have taken from a newspaper. Recursive Descent Parser used to parse the CFG. **[6]**

G.V. Singh and D.K. Lobiyal, **(1994)** checked the grammar for the sentences with compound, conjunct or complex verb phrases. **[7]**

The paper titled 'Parsing Natural Language with Context Free Grammars' discusses the implementation of Simple English Sentences parsed by writing grammar in CFG and used Dynamic Programming for the parser.

Rama Sree, R.J., Kusuma Kumari P. **(2005)** Implemented a POS Tagger for Telugu for allmost 12000 words using voting algorithm. **[8]**

Selvam M, Natarajan. A M, and Thangarajan R **(2008)** Implemented a Structural Parsing of Tamil using Phrase Structure Hybrid Language Model for allmost 700 sentences.

III. NOUN PHRASE

A noun (ನಾಮಪದ) is a word used as the name of the person, place, animal or a thing.

Sagar is name of the person, Delhi name of the place and cow is name of the animal. Sagar, Delhi and cow are all noun. Since noun involves person, animal then it may be male, female or neuter.

A noun that denotes a male living being is called **Masculine gender**. Example: Rama, Peter, father etc. A noun that denotes a female living being is called **Feminine gender**. Example: Sita, Diana, mother etc. A noun that denotes a thing that is either male or female without life is said to be **Neuter gender**. Example: Table, box etc. A noun that denotes either a male or female is said to be **Common gender**. Example: Child, pupil, student, pupil etc. Singular Noun: A noun that denotes one person or thing is said to be in the singular number. Example: Book, Pen, Boy etc. Plural Noun: A noun that denotes more than one person or thing is said to be in the plural number. Example: Books, Pens, Boys etc

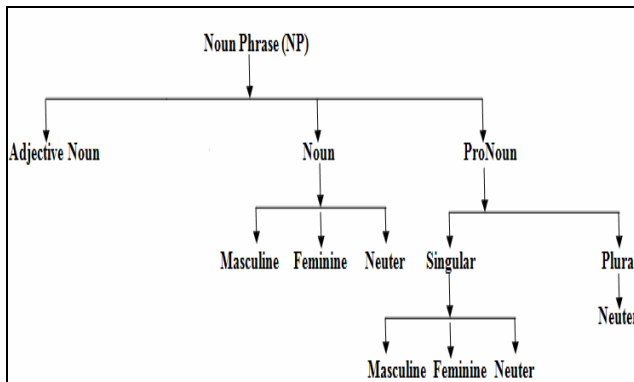For our Implementation different types of Noun is classified as below:

Fig: 3.1: Noun Classification

IV. VERB PHRASE

The verb is the most important part of the sentence. A verb asserts something about the subject of the sentence and express actions, events, or states of being. Words of actions, events and relations are categorized in verbs. Kannada verbs always have to match with their subject. They depend on the person, gender, number, and the form of the verb changes. Kannada verbs are inflected through the use of suffixes. The suffix of the verb will indicate person, number, mood, tense and voice.

A. Suffix Extraction from the Verb

As soon as the verb is identified from the grammar, its suffix is extracted to check for the masculine, feminine and plural verb endings. The Implementation uses the Kannada Unicode format, and then it will be easy to verb suffix from the verb. As we see the Grammatical and Lexical Productions it clearly says the classification.

V. CONTEXT FREE GRAMMAR

Context Free Grammars is a notation that has been used extensively for defining the syntax of languages. In this work we have used CFG to write the grammar. A context-free grammar is a formal system that describes a language by specifying how any legal text can be derived from a distinguished symbol called the sentence symbol. It consists of a set of productions, each of which states that a given symbol can be replaced by a given sequence of symbols. Figure 5.1 and Figure 5.2 shows the Grammatical and Lexical Productions respectively.

A. Python

Python is a simple yet powerful programming language with excellent functionality for processing linguistic data. Python is used in this work because it has a shallow learning curve, its syntax and semantics are transparent, and it has good string-handling functionality. As an interpreted language, Python facilitates interactive exploration. As an object-oriented language, Python permits data and methods to be encapsulated and re-used easily. As a dynamic language, Python permits attributes to be added to objects on the fly, and permits variables to be typed dynamically, facilitating rapid development. Python comes with an extensive standard library, including components for graphical programming, numerical processing, and web connectivity.
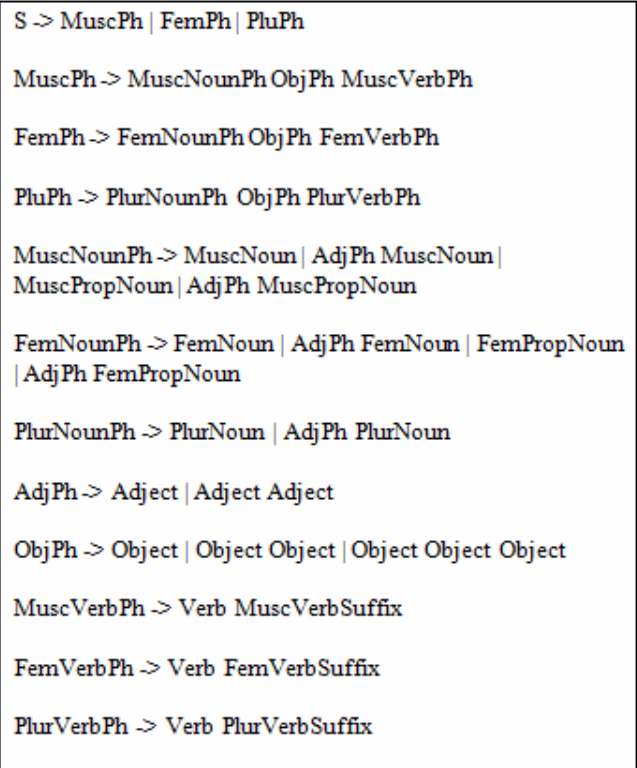

Fig: 5.1: Grammatical Productions

Fig: 5.2: Lexical Productions

FIG: 6.1 TREE FOR THE SENTENCE 1

VI. TESTING THE GRAMMAR

After generating the context-free grammar given in the previous section, the Recursive Descent Parser is used to test the grammar. Recursive Descent Parser of NLTK (Natural Language Tool Kit) was used to test the grammar. Recursive Descent Parser is the simple Top-Down parser. Parser is a program that performs syntax analysis. A syntax analyzer takes tokens as input and output error message if the program syntax is wrong. A top-down parser starts with the root of the parse tree. It is labeled with the start symbol or goal symbol of the grammar. It picks a production and tries to match the input. It may require backtracking. Top-down parsers cannot handle left-recursion in a grammar. Top-Down Parsing is an attempt to find a left-most derivation for an input string. The following tree shows the Parse tree for the sample sentences. (Refer Appendix)
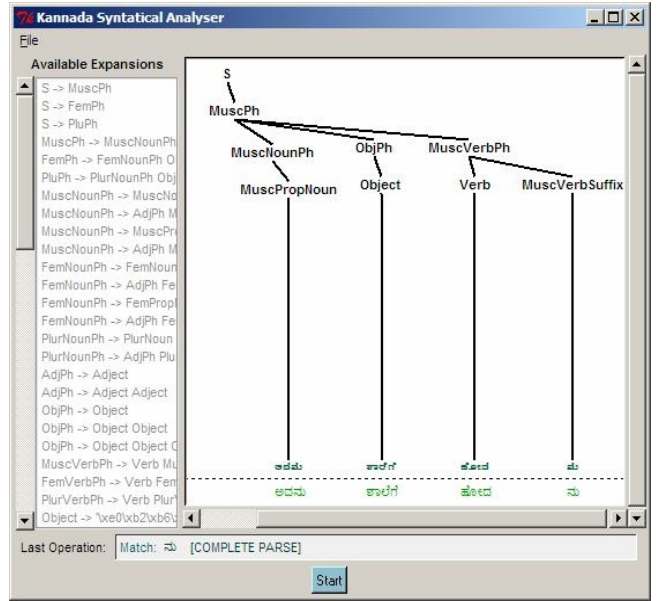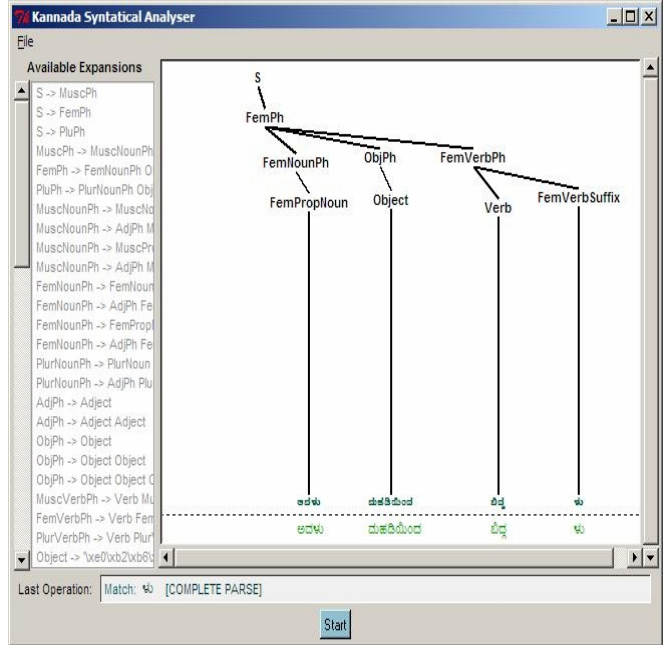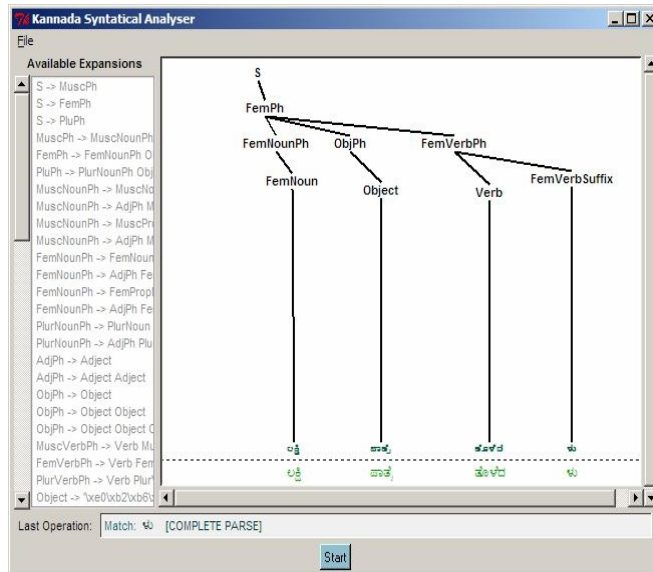

FIG: 6.2 TREE FOR THE SENTENCE 2
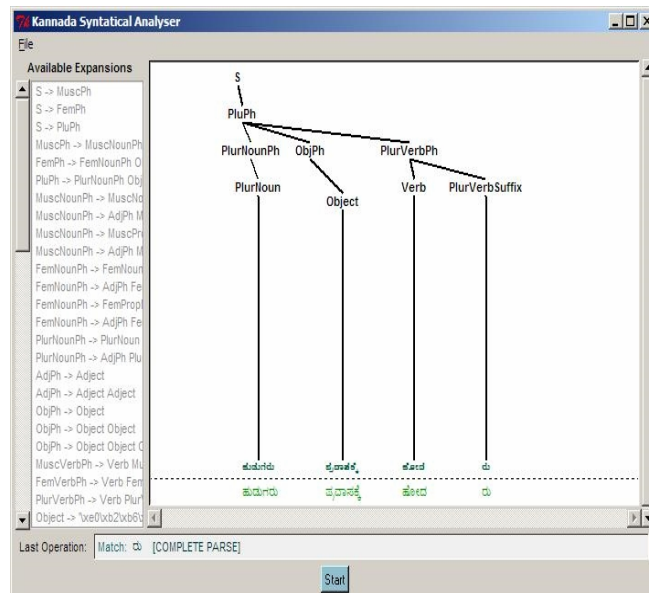
FIG: 6.3 TREE FOR THE SENTENCE 7

Fig: 6.4 Tree for the Sentence 9

For the sentence 10, the sentence is syntactically wrong where it does not show the parse tree.

VII. CONCLUSION

The grammar proposed by this paper is limited to the scope of resolving Noun and Verb agreement. 200 sample sentences have taken to test the agreement. Given a sentence parser says whether the sentence is syntactically correct or wrong depending upon the Noun and Verb agreement. As a future improvement if we can write grammar for all types Kannada sentences to parse and say the sentence is syntactically right or wrong.

APPENDIX

Sample Kannada Sentences

Systems & Networking. She has guided more than 45 undergraduate and 09 post graduate projects.

Dr. Ramakanta Kumar, P was awarded Doctorate from Mangalore University, has teaching experience of around 14 years in academics and Industry. His area of research is on Artificial Intelligence, Pattern recognition. He has to his credits 03 National Journals, 02 International Journals, 12 Conferences and 15 Research Publications. He is guiding 14 M.Tech students and 03 PhD students.

ACKNOWLEDGMENT

REFERENCES

[1] T V Ashwin and P S Sastry, A font and size-independent OCR system for printed Kannada documents using support vector machines, Sadhana Vol. 27, Part 1, February 2002, pp. 35–58.
[2] H.S.K. Vishweswariah, 2008 Edition Standard Kannada – English Grammar
[3] T.N. Sreekantaiya, 2008 Edition Kannada madhyama Vyakarana
[4] Dr. A.N. Narasimhia, Kannada prathama Vyakarana, 2008 Edition, ISBN: 81-89818-91-0
[5] Bala sundara Raman L, Ishwar S, Sanjeeth Kumar Ravindranath, Context Free Grammar for Natural Language Constructs - An Implementation for Venpa class of Tamil Poetry, Tamil Internet 2003, Chennai, India
[6] Ayesha Binte Mosaddeque & Nafid Haque, Context-Free Grammar for Bangla, Dhaka, Bangladesh
[7] G.V. Singh And D.K. Lobiyal, A Computational Grammar For Hindi Verb Phrase, IEEE transactions, 1994
[8] Rama Sree, R.J., Kusuma Kumari P., Combining Pos Taggers For Improved Accuracy To Create Telugu Annotated Texts For Information Retrieval, Dept. of Telugu Studies, Tirupathi, India
[9] www.nltk.org

B.M. Sagar, Assistant Professor of Department of Information Science and Engineering. He obtained his Master's Degree in Computer Science & Engineering from VTU and B.E. in Computer Science & Engineering from Kuvempu University. His research interests are Pattern Recognition. He has guided more than 25 under graduate projects and 3 post graduate projects. He has presented and published papers at national conference / International Conference and 2 Research Publication.

Dr. Shobha G., Director, Master of Computer Applications. She has been awarded Ph.D for her thesis titled "Knowledge Discovery in Transactional Database Systems" from Mangalore University, Mangalore. She obtained her M.S. degree in Software Systems from BITS, Pillani and BE in Computer Science from Gulbarga University. Her research interests are Data Mining, DBMS, and Operating