International Journal of Latest Trends in Engineering and Technology
IJLTET Special Issue- ICRACSC-2016 , pp.201-204
e-ISSN: 2278-621X

THE IMPACT OF NOUN AMBIGUITY IN TELUGU LANGUAGE SENTENCES USING CSG RULES: A MODERN APPROACH

Dr.J.Sreedhar.,M.Tech.,PhD[1], E.S.Madhan.,M.E.,(PhD)[2] & P.Jayaram., M.Tech[3]

Abstract: Context Sensitive Grammar [CSG], which is used to reduce the ambiguity in a language sentences. In the existing literature we aware some of the approaches are used to condense the ambiguity, to our knowledge there is no CSG approach to squeeze the ambiguity in Telugu language. Henceforth, we introduced a new methodology to identify the impact of a Noun ambiguity in Telugu Language Sentence with the help of CSG.
Keywords: CSG, NLP, Word Sense Disambiguation, Left Most Derivation, Right Most Derivation, Parse Trees.

1. INTRODUCTION

Telugu is a major Dravidian language of the Central Dravidian subgroup. It is included in the Eighth Schedule of the Indian Constitution along with the other seventeen languages. About 80 million people (66,017,615 in 1991) speak this language in India. This figure makes it the largest spoken tongue among all the Dravidian languages. Due to influence of local languages the words in Telugu have been undergone ambiguity, various techniques and approaches have been innovated to resolve the ambiguity of words in Telugu language. Because of the influence non-English languages and their locality, religious and cultural aspects and resulted in the ambiguity of sense and meaning of Telugu. Concentrating on Telugu language, adaptability of the language by the people is similar by everyone at higher level quite different when we dig into deeper. The usage of sense of a word drastically changed in oral communication as well in written.

There are mainly four major regional dialects in Telugu, the Coastal Andhra dialect spoken in the six coastal districts of Nellore(partially),Prakasam,Guntur,Krishna,EastGodavari and WestGodavari; the Kalingandhra dialectof visaka patnam visakhapatnam,Vijayanagaram and Srikakulam districts; the Rayalaseema dialect spoken in the four Rayalaseema districts of Cuddpah, Kurnool, Anantapur and Chittoor; and finally the Telangana dialect spoken in the ten districts of Telangana state, namely Adilabad, Nizamabad, Medak, Karimnagar, Warangal, Mahaboobnagar, Nalgonda, Khammam Mettu, Hyderabad and Rangareddi. There are several other dialects in state boundaries that influence Telugu. As mentioned above, there are several dialects in Telugu language; they influence oral communication as well as to writing. The sense of a word varies from one dialect to another dialect. Due to this variation in the senses, ambiguity started in the language through disambiguating the words is necessary. So processing of natural language and disambiguating the word senses are highly required to increase understanding. Natural Language Processing [NLP][1,24] is a field of computer science, artificial intelligence[5,43], machine learning[21,29,34,35,41] and linguistics concerned with the interactions between computers and natural languages. NLP is related to the area of human–computer interaction. Challenges in NLP[11,40,42,49] involve natural language understanding[36,37], that is, enabling computers to derive meaning from human or natural language input and involve in natural language generation. Word Sense Disambiguation[WSD][2,6,9,16,25,26,27,38,39,44,46,47,50,51] in Computational linguistics[7,8,20,22,23,28,45,48] is an open problem of natural language processing and ontology, which governs the process of identifying the sense of meaning a word in a sentence, when the word has multiple meanings. One of the solution to this problem is Context Sensitive Grammar [CSG][3,4], which is used to decrease the ambiguity in a language sentences.

The organization of this paper is as follows: Section II describes the CSG and followed by its methodology, Section III shows the case study followed by derivations and parse trees, Section IV explores Conclusion and Future Research Direction, Section V shows the Acknowledgements followed by references.

2. CONTEXT SENSITIVE GRAMMAR

Here we describes a new methodology which is known as Context Sensitive Grammar [30,31,32,33], that can be used to condense the noun ambiguity in telugu language. We also considered the impact of nouns in this connection.

[1] *Department of ComputerScience and Engineering, SVEC, TIRUPATHI*
[2] *Department of ComputerScience and Engineering, SVEC, TIRUPATHI*
[3] *Department of ComputerScience and Engineering, SVEC, TIRUPATHI*

2.1 Methodology:
Let G be a Context Sensitive Grammar for which the production rules are : S => NP VP
NP => NP NN | NN NP | NST NP | NNP NP | PRP NP | QC NP | CC NP | JJ NP | SYM NP
VP =>VP NP | NP VP|VGF VP|VGNF VP|VGINF VP|VGNN VP |RBP VP|JJP VP|SYM VP |CCP VP
VGF VGNF VGINF VGNN = > VM | VAUX | SYM
Figure 1: Context Sensitive Grammar
Where S is a Sentence, NP is a Noun Phrase, VP is a Verb Phrase, NN is a Noun, NST is a NLoc, NNP is a Proper Noun, PRP is a Pronoun, VM is a Verb Finite, VMUX is a Verb Aux, QC is a Cardinal, CC is a Conjuncts, JJ is a Adjective, RBP is Adverb Phrase, JJP is a Adjective Phrase CCP is a Conjunctive Phrase, SYM is a Sym.

3. CASE STUDY

Consider the telugu language sentences and we observe the impact of the noun kAru
1. rEwulu edAxiki mUdu kArla paMtalu paMdiswAru .
2. AkASaMlo kAru mabbulu kanapadawAyi .

Given Input 1: rEwulu edAxiki mUdu kArla paMtalu paMdiswAru .
This sentence will be converted through our methodology like :

NN NN QC NN NN VM SYM

Where
NN is a Noun, QC is Cardinal, VM is a Verb Finite, SYM is a Sym

Here Derivations generating the sentences of a language. It has left most and right most derivations. If one chooses the leftmost non-terminal in a given sentential form then it is called left most derivation[LMD][17,18,19]. If one chooses the rightmost non-terminal in a given sentential form then it is called right most derivation[RMD]. Any language construct can be defined by the CSG[12,13,14,15].

3.1 Derivations and Parse Trees
Hereafter the above sentences will be derived like:

S => NP VP

=>NN NP VP

= >NN NN NP VP

=> NN NN QC NP VP

=> NN NN QC NN NP VP

=> NN NN QC NN NN NP VP

=> NN NN QC NN NN VP

=> NN NN QC NN NN VP

=> NN NN QC NN NN VGF VP

=> NN NN QC NN NN VGF SYM VP

=> NN NN QC NN NN VGF SYM
Figure 2 : Derivation for the input sentence 1.

A parse tree represents a derivation graphically or pictorially. It is an internal structure created by the parser while parsing some language sentences. Parsing is also known as syntax analysis.
Based on the earlier derivations we constructed the parse tree[10] in figure 3.
Given Input 2 . AkASaMlo kAru mabbulu kanapadawAyi .

This sentence will be converted through our methodology like:

NN NN NN VM SYM
Where
NN is a Noun, VM is Verb Finite, SYM is Sym.
Hereafter the above sentences will be derived like:

S => NP VP
=>NN NP VP
= >NN NN NP VP
=> NN NN NN NP VP
=> NN NN NN VP
=> NN NN NN VP
=> NN NN NN VGF VP
=> NN NN NN VM VP
=> NN NN NN VM SYM VP
=> NN NN NN VM SYM
Figure 4: Derivation for the input sentence 2

Based on the earlier derivations we constructed the parse tree in Figure 4

4. Conclusion and Future Research Direction

In this research paper we analyzed the impact of noun ambiguity in telugu language sentences. We empirically analyzed our methodology and also calculated the statistical measures such as Precision and Recall. Through this we observed our results are significant when compared to existing approaches. In the future there may be a scope for further research on verbs, adjectives and adverbs to measure the impact of verbs, adjectives and adverbs and so on.

5. ACKNOWLEDGEMENTS

We are very thankful to all the esteemed authors in a reference list to make this research article in a better shape and in right direction.

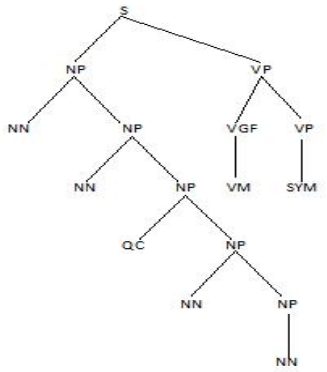

Figure 3: Parse tree for Input sentence 1

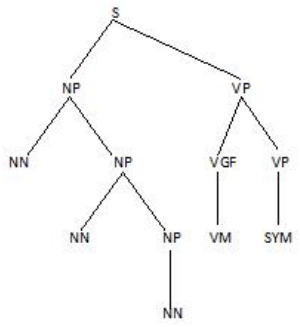

Figure 5: Parse tree for Input sentence 2.

6. REFERENCES

[1] Agirre, E. and Martinez, D. Learning class-to-class selectional preferences. In Proceedings of the 5th Conference on Computational Natural Language Learning (CoNLL, Toulouse, France). 15–22, 2001.

[2] Agirre.E, G. Rigau, Word sense disambiguation using conceptual density, in: Proceedings of the 16th International Conference on Computational Linguistics, Copenhagen, pp. 16–22, 1996.

[3] Aho, A.V., and Johnson, S. C. , and Ullman, J. D. [1975]. "Deterministic parsing of ambiguous grammars," *Comm*. ACM 18:8, 441-452

[4] Bar Hillel, Y., Perles, M., and Shamir, E. [1961]. "On formal properties of simple phrase structure grammers," Z. Phonetik, Sprachwissenschaft und Kommunikationsforschung 14, pp. 143-172.

[5] Banerjee.S,T. Pedersen, Extended gloss overlaps as a measure of semantic relatedness, in: Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, pp. 805–810, 2003.

[6] Brody, S.,Navigli, R., and Lapata,M. Ensemble methods for unsupervised WSD. In Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics joint with the 21st International Conference on Computational Linguistics (COLING-ACL, Sydney, Australia). 97–104, 2006.

[7] Boser, B. E., Guyon, I. M., and Vapnik, V. N. A training algorithm for optimal margin classifiers. In Proceedings of the 5th Annual Workshop on Comput-ational Learning Theory (Pittsburgh, PA). 144– 152, 1992.

[8] Bresnan, Joean, Ronald M. Kaplan, Stanley Peters, and Annie Zaenen. 1982. "Cross-serial dependencies in Dutch. Linguistic Inquiry", 13(4):613–635. Reprinted in (Savitch et al., 1987).

[9] Cottrell, G. W. A Connectionist Approach to Word Sense Disambiguation. Pitman, London, U.K, 1989.

[10] Collins,M. Parameter estimation for statistical parsing models: Theory and practice of distributionfree methods. In New Developments in Parsing Technology, H. Bunt, J. Carroll, andG. Satta, Eds. Kluwer, Dordrecht, The Netherlands, 19–55, 2004.

[11] Chomsky, Noam. 1956. "Three models for the description of language. IRE Transactions on Information Theory" , 2:113–124.

[12] Cohen, D. J., and Gotlieb, C. C. [1970]. " A list structure form of grammars for syntactic analysis", Computing survey 2:1, 65-82.

[13] Demers, A. J. [1975], "Elimination of single productions and merging of non terminal symbols in LR(1) grammars," J. Computer Languages 1:2, 105-119.

[14] Earley, J. [19700. "An efficient context-free parsing algorithms," Comm. ACM 13:2,94-102.

[15] Earley, J.[1975a]. "Ambiguity and precedence in syntax description," Acta Informatica 4:2, 183-192.

[16] Escudero, G., Marquez, L., and Rigau, G.Naive Bayes and exemplar-based approaches to word sense disambiguation revisited. In Proceedings of the 14th European Conference on Artificial Intelligence (ECAI, Berlin, Germany). 421–425, 2000b.

[17] Floyd, R. W. [1964]. "Bounded context syntactic analysis," Comm. ACM 7:2, 62-67.

[18] Ginsburg, S. [1966]. The Mathematical Thory of Context-Free Languages, McGraw-Hill, New York, N. Y.

[19] Gazdar, Gerald, Ewan Klein, Geoffrey Pullman, and Ivan Sag. 1985. "Generalized Phrase Structure Grammar" . Harvard University Press, Cambridge, Massachusetts.

[20] H. Kozima, T. Furugori, Similarity between words computed by spreading activation on an English dictionary, in: Proceedings of the 6th Conference of the European Chapter of the Association for Computational Linguistics, Utrecht, pp. 232–239,1993.

[21] Joachims, T. Text categorization with support vector machines: Learning with many relevant features. In Proceedings of the 10th European Conference on Machine Learning (ECML, Heidelberg, Germany). 137–142, 1998.

[22] J. Cowie, J. Guthrie, L. Guthrie, Lexical disambiguation using simulated annealing, in: Proceedings of the 14th International Conference on Computational Linguistics, Nantes, France, pp. 359– 365,1992.

[23] J. Veronis, N. Ide, Word sense disambiguation with very large neural networks extracted from machine readable dictionaries, in: Proceedings of the 13th International Conference on Computational Linguistics, Helsinki, pp. 389–394,1990.

[24] Jurafsky, D. and Martin, J. Speech and Language Processing. Prentice Hall, Upper Saddle River, NJ, 2000.

[25] Kelly, E. and Stone, P. Computer Recognition of English Word Senses. Vol. 3 of North Holland Linguistics Series. Elsevier, Amsterdam, The Netherlands, 1975.

[26] Keok, L. Y.and NG, H. T. An empirical evaluation of knowledge sources and learning algorithms for word sense disambiguation. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP, Philadelphia, PA). 41–48, 2002.

[27] Klein, D., Toutanova, K., Ilhan, T. H., Kamvar, S. D., and Manning, C. D. Combining heterogeneous classifiers for word-sense disambiguation. In Proceedings of the ACL workshop on Word Sense Disambiguation: Recent Successes and Future Directions (Philadelphia, PA). 74–80, 2002.