ISSN:0975-9646



# Anaphora Resolution for Hindi Language

Priyanka Ahir [1], Bhakti Mehta [1], Riddhi Desai [1], Zinal Patel [1], Elisha Patel [1], Nehal Adhvaryu [2]

[1]*Student of M.Sc.(IT),* [2]*Faculty of M.Sc.(IT)*

[1,2]*Uka Tarsadia University*
*Bardoli, Surat.*

**Abstract -In this paper we explore the issues and types of anaphora resolution by discussing the basic information of anaphora techniques and its algorithms. Here we deeply discuss Gazetteer techniques using algorithm and example.We describe the problem of pronominal anaphora while finding the pronoun and translating statement from English to Hindi.**

**Keywords-Natural Language processing , Anaphora, Gazetteer method, Leppin and leass Algorithm, Pronominal anaphora resolution. Mobile GIS, Natural Language understanding, Machine learning, Algorithm, case markers.**

## 1. INTRODUCTION

Natural Language Processing (NLP) is use of communication between computers and human languages. NLP is a field of computer science and artificial intelligence. NLP is area of human computer interaction. In NLP many types are available such as , Pronominal anaphora, Definite noun phase anaphora , One-anaphora or Quantifier/Ordinal anaphora, Nominal anaphora , Pleonastic 'it' anaphora , Zero anaphora
Anaphora is a sub topic of NLP. Anaphora is used of text conversation to avoid repetition.
Ex. Pronouns *he*, *she*, *it*, and *they*.
Brief summary about different anaphora resolving system using types of anaphora and its described in section II. more useful introduction of many algoritams are described in section III. Resolution system is described in section VI. At last conclusion of review paper is described insection V.

## 2. TYPES OF ANAPHORA

- Pronominal anaphora
- Definite noun phase anaphora
- One-anaphora or Quantifier/Ordinal anaphora.
- Nominal anaphora
- Pleonastic 'it' anaphora
- Zero anaphora

**Pronominal anaphora:**
This is basic type of anaphora and referred a pronoun pronominal anaphora is a problem of determining the noun phrase (NP)that refers to pronoun in sentences.

Ex: "Mehul ne Dhaval ko uski pustak di"
Here uski refers to 'Mehul'. Human can easily work in that sentance .

In complex example
S1: "bacchon ne kele khaye kyunki ve bhukhe the"
S2:"bacchon ne kele khaye kyuki ve pake hue the "
In sentence s1 've' refers to 'bacchon' and sentence s2 've' refers to 'kele'.
This example of pronominal resolution. It is difficult task to handled an anaphora resolution system. proniminal anaphora in that major classifications the first, second and third person pronouns.

**Classification of anaphora and pronoun in Hindi language:**

| Pronominal anaphora in English | Pronominal anaphora in Hindi |
|---|---|
| He | Veh |
| He,She | Ve inhon-ne |
| His,her,its | Us |
| Him,her | Usko |
| That | Veh |
| This | Yeh |
| They<br>These | Ve |
| Them<br>Their | unko/unse<br>unka/unki/unke |

Pronoun in the first ,second and third do not define information about gender. In hindi there is not differentiate between 'he' and , 'she'. 'veh' is used for both gender and dicided for the verb form. In table some forms, like 'usko'(him), 'usne'(he) are singular but some forms can be both singular and plural, like 'unhone'(he)/they, or 'unko'(him)/them.

## ALGORITHMS:

Types of Algorithm:
1. Hobbs Algorithm
Hobbs Algorithm first read file and travers the sentence and generate parse tree and find the anaphora.
2 .Leppin and leass Algorithm
First it read sentence and gives the Weightage of sentence match constarain and evaluate the result.
3. Gazetteer Method

Gazetteer Method is the creates different gazetteer classes for different elements and operations to categorize the features. In our system we produced lists of animistic pronoun, animistic noun, non animistic pronoun, non animistic noun and middle animistic pronoun. This is helps of the system in resolving anaphora by differentiating the co referents on the basis of their classification.

**Advantages of Gazetteer Method**

- The Gazetteer method provides very fast result of Anaphora Resolution System
- Gazetteer method increases the system's accuracy to far extent.

**Disadvantages of Gazetteer Method**

- Spellings must be correct compulsory.
- If all words are not spelled correctly it gives no result.

## LITERATURE REVIEW:

They present anaphora resolution for Hindi language. Anaphora resolution is a key problem in natural language processing. The researching weight of work has been done in English and other European language, and in Hindi language efficient work is pending. And that work completed in different section. That first section of the present paper review of anaphora resolution work in Hindi language. And the second part they solve problem of syntactic and semantic structure of Hindi pronoun. Then after thread part is define source constraint which will form to the task of anaphora resolution. And in last they perform different experiment on different kind of data set. And also they get results is obtain 71%. [1]

They discussed about the Chinese Natural Language Interface for Navigation in Mobile GIS. In that "Mobile GIS" and "Voice technology" both combination has improved the intelligent degree of mobile GIS. Natural language quick convert to GIS commands they can possible because of field of scientific research. This paper study of natural language sentences and GIS command understanding method of mobile voice GIS. Using a machine learning method conversion between natural language and GIS commands. [2]

Author discussed the system's design using two algorithms implemented in the current version of the system for descriptions and pronoun resolution. Develop one tool GUITAR (General Tool for Anaphora Resolution) In this paper they briefly discuss the architecture and implementation of the system, as well as some preliminary evaluation results.[3]

They discuss about the natural language processing or define all the language using different country. There are many NLP applications such as machine translation, question answering system, automatic summarization etc. In this paper there are three techniques are use first is statistical machine translation, second parallel corpus, and third one is phrase-based translation. and also we discuss about Indian language to translate a many language like Gujarati, Hindi, Bengali, Telugu, Tamil, Urdu that all language are translating in Indian language by using the Machine translation technique. [4]

All natural language processing(NLP) like as machine translator, automatic summarization, etc. its required empathy and resolution of anaphora. Almost work is done in English and other European languages in hindi language existent a review of work done in the ground of anaphora resolution in Hindi. This paper is not work committed to pronominal anaphora. [6]

In this paper used anaphoric information in latent semantic analysis LSA) and discuss its application. Anaphoric information is automatically mind a new release of our own anaphora resolution system, GUITAR, is includes proper noun resolution. Anaphoric information is used to check consistency of summary produced our summarizer. [7]

Data are stored in the database and the database is the source of all the information which are used by the human or any all other information. Information is required for do the some work in every human life day-to-day. And this database has an impotent role playing in computer and internet uses. Database management system are used for to accessing, storing and retrieving the data or information which are stored in the system. Whereas all the people are not getting to how to access database because of they have no knowledge of database language. That is why we need to Find new method or technique with the use of NLP (Natural Language Processing)to access the database. The new development of method is called Natural Language Interface to Database (NLIDB).In this method no need to learn any database language those who not accessing the data .So they can give query in their native language such as ,Hindi, Gujrati, English ,etc. And it also give the resonance or result in the same language. [8]

NLP one the challenges is to determine the noun .how they related to with pronoun and how they refer to each other. This is known as Anaphora resolution. Generally there are three main algorithms works for anaphora resolution. That is Hobbs , Centering and Leppin leass Algorithm. This three algorithm are works for hindi language. As hindi language is more difficult to other European language.so many terms and method need to resolving anaphora. [9]

## CONCLUSION

During the review session of various literature reviews and research papers on Natural Language Processing and related areas, we came to conclude that researcher's has done all the work in the field of Anaphora Resolution on different languages like Spanish, European, Nepali, Turkish and many more, but on Indian languages like Hindi, Gujarati, etc. has less work done. During the work we find many techniques like Lappin and Leass, Centering, Hobbs, Gazetteer method. Gazetteer method generate wordlist and traverse the sentence and then compare sentence to the wordlist, after comparing the word will store in index and print the result. For Anaphora Resolution on Hindi language we find Gazetteer method more appropriate.

## FUTURE WORK

We will work on Hindi language using Gazetteer method. In future we will expand the gazetteer method in use in anaphora resolution. It is dedicated text summarization.

## REFERENCES


[1] Priya Lakhmani and Smita Singh, "Anaphora Resolution in hindi language" Department of Computer Science Banasthali University, C-62, Sarojini Marg, C-Scheme, Jaipur, India.

[2] Jiangfan Feng1, Nan Xu2 ,"Using Chinese Natural Language Interfaces for Navigation in Mobile GIS "1 College of Computer Science and Technology, Chongqing University of Post and Telecommunications, Chongqing 400065, China.2 College of Computer Science and Technology, Chongqing University of Post and Telecommunications, Chongqing 400065,China. IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 3, January 2013 ISSN (Print): 1694-0784 | ISSN (Online): 1694-0814 www.IJCSI.org

[3] Massimo poesio and Mijail A. Kabadjov ,"A general purpose off the self-anaphora resolution module :Implementation and preliminary evaluation" University of Essex, Department of Computer Science Wivenhoe Park, Colchester, CO4 3SQ , UK.

[4] Nadeem jadoon khan. ," Statistical machine translation of Indian language: A survey"

[5] Shweta Vikram , "Morphology: Indian Languages and European Languages" Computer Science, Banasthali University, International Journal of Scientific and Research Publications, Volume 3, Issue 6, June 2013 1 ISSN 2250-3153

[6] Triveni Lal Pal and Kamlesh Dutta, "Anaphora Resolution in Hindi: Issues and Challenges" Department of Computer Science and Engineering National Institute of Technology Hamirpur Hamirpur (H.P.)", India jurnol: International Journal of Computer Applications (0975 – 8887) Volume 42– No.18, March 2012

[7] Josef Steinberger a,*, Massimo Poesio b,c, Mijail A. Kabadjov b, Karel Jezˇek a "Two uses of anaphora resolution in summarization" a University of West Bohemia, Univerzitni 8, Pilsen 306 14, Czech Republicb University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdomc Universita´ di Trento, Rovereto, TN 38100, Italjurnol: Received 17 July 2006; received in revised form 8 January 2007; accepted 10 January 2007Available online 6 March 2007

[8] Abhijeet R. Sontakke1, Prof. Amit Pimpalkar2, "A review paper on hindi language graphical user interface to relational database using nlp". Department of Computer Science and Technology, 2G.H. Raisoni Academy of Engineering and Technology, Nagpur, India. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 10, October 2014 3393

[9] Priya Lakhmani1, Smita Singh2, Sudha Morwal3"Performance Analysis of two Anaphora Resolution System for Hindi Language,International Journal of Computer Science and Mobile Computing A Monthly Journal of Computer Science and Information Technology, Department of Computer Science, Banasthali University, India 1 tinalakhmani@gmail.com; 2 smitasingh101@gmail.com; 3 sudha.morwal@yahoo.com , Priya Lakhmani et al, International Journal of Computer Science and Mobile Computing, Vol.3 Issue.3, March- 2014, pg. 576-580 Available Online at www.ijcsmc.com