

# REVIEW ON THE TECHNIQUES OF ANAPHORA RESOLUTION

Shivangi Shekhar<sup>1</sup> and Umesh Kumar<sup>2</sup>

**Abstract :** Natural Language Processing (NLP) is the most interesting field of artificial intelligence. It is the field of study that concentrates on the interaction between humans and computers. Through NLP we can make the computers understand, interpret and derive the meaning of human language in a brilliant way. However there are many challenges that are encountered in NLP and, Anaphora is one such issue. Anaphora arises very frequently in written texts and oral discourse. The process of figuring out the antecedent of corresponding referent is Anaphora Resolution, and is needed in various applications of NLP. Although several works have been reported for anaphora resolution in English and other languages, however very few has been done in Hindi .Through this paper we are detailing the review work accomplished till now in Anaphora Resolution. Also, we will detail the various approaches and techniques that have been developed for Anaphora resolution, like Rule based, Corpus-based Approach, etc.

**Keywords:** Anaphora, Anaphora Resolution, Corpus-Based Approach , Natural Language Processing (NLP), Rule-Based.

## I. INTRODUCTION

Natural Language Processing is the field which includes the interaction between humans and computers. Anaphora Resolution, being the productive area of research in Natural Language Processing has a main motive to produce software, which would be capable of interpreting human languages. Anaphora occurs very often [1],[4] in written texts and spoken discourse. The process of figuring out the antecedent of corresponding referent is termed Anaphora Resolution and is needed in many applications of NLP namely Machine Translation, Automatic Summarization, Co-reference Resolution etc.

Suppose there is a sentence given as: Aditi works for a company where she is working on a project on face recognition. Here 'she' is referring to 'Aditi'. The referring word 'she' is an anaphor and the preceding word 'Aditi' is called an antecedent. So the procedure or process to figure out the antecedent of an anaphora is called Anaphora Resolution.

Anaphora Resolution can be classified into two types [2]: Intrasentential and Intersentential. In Intrasentential the antecedent and anaphor both are found in the same sentence. Whereas in Intersentential both the antecedent and anaphors are found in different sentences.

Basically the problem in Anaphora Resolution is sometimes it becomes difficult to figure out as to which antecedent the corresponding referent is referring to. To understand this let us consider the below given two sentences S1 and S2:

S1: Hall tickets were given to the students just before the commencement of the examination because they were required to enter the examination hall.

S2: Duplicate Hall tickets were given to the students to enter the examination hall because they did could not bring the original ones.

---

<sup>1</sup> *Research Scholar-CSE Government Women Engineering College,Ajmer (Rajasthan),India.*

<sup>2</sup> *Assistant Professor-Department of CSE & IT Government Women Engineering College , Ajmer (Rajasthan ) , India*

Here in the above example the question arises as to whom the word ‘they’ refer in each sentence. In S1 ‘they’ refer to ‘Hall tickets’ whereas in S2 ‘they’ refer to ‘students’. It’s easy for we humans to figure out this but for machines this is not that easy. So resolving Anaphora stands out as a milestone in the field of Natural Language Processing. There should be a proper understanding of syntax, semantics and pragmatics of a particular language. There are many reported works on English Language for Anaphora resolution but as far as Hindi is concerned, very few works have been reported.

This paper begins with the introduction. Section II details the various types of Anaphora. Section III covers the various existing techniques of Anaphora Resolution. Section IV deals with the comparison of the techniques of Anaphora Resolution. Section V discusses the parameters for evaluation of results. Section VI is the conclusion and the references have been outlined in section VII.

## II. TYPES OF ANAPHORS

There are various types of Anaphors categorised on the basis of form of the Anaphor, which are as discussed below [3]:

- a. **Nominal Anaphora:** This emerges when a referring expression be it a pronoun, definite noun phrase or a proper name, has a non-pronominal noun phrase as its antecedent.
- b. **Pronominal Anaphora:** This is a popular anaphora and may consist of sentences which have personal, possessive, reflexive, relative, demonstrative pronouns.
- c. **Lexical Noun phrase Anaphora:** These are also termed as definite description and proper names as they syntactically refer to definite noun phrase.
- d. **Zero Anaphora:** These are invisible and emerge when the anaphoric pronoun is omitted but □ is never understood.

Also Anaphors are categorised on the basis of location of anaphors and antecedent, these are as follows:

- a. **Intrasentential:** In this antecedent and anaphors are present in the same sentence.
- b. **Intersentential:** In this antecedent and anaphors are present in different sentences.

## III. TECHNIQUES /APPROACHES USED IN ANAPHORA RESOLUTION

Research on Anaphora resolution broadly falls into the following categories which are as detailed below:

### 1. Rule Based Approaches

These apply rules (linguistic) and knowledge .They employ linguistic knowledge like syntactic rule or heuristics. This technique commemorates knowledge factors which ignore dissimilar candidates up till we get minimal valid set of candidates. Some of the algorithms that come under Rule based approach are:

- Tree Search Algorithm[6]
- Syntax based approach[8]
- Shallow processing approach[9]
- Multi –strategy approach[10]
- Combination of linguistic and statistical technique[11]

### 2. Corpus Based Approach

This method uses available corpus which is a collection of written texts. In this a set of abstract rule is deduced from a text, for governing a natural language and how that language relates with other languages. Some examples of algorithms under Corpus based Approach are:

- Knowledge-independent Approach[6]
- Machine Learning Approach[12]
- Statistical/Corpus Approach[13]

### 3. Discourse based Approach

In this approach an entity is centralized at a particular point of discourse and it has to be different from all other entities which have been evoked. Also for resolving anaphora, knowledge of word and conclusion are applied. Few examples of algorithms under this approach are:

- LRC Algorithm[14]
- BFP Algorithm[15]
- Centering theory[16]

- S-List Algorithm[16]

#### 4. Knowledge-Poor Approach

This approach is economically fair, fast and very reliable. Their characteristic of not depending on syntax and semantics of a particular language makes them acceptable for vast category of languages. They employ many artificial intelligence techniques like neural networks etc. and can work even without parsing. Algorithms under this are :

- ROSANA[17],[18]
- COGNAIC[19]
- Robust knowledge poor approach[8],[20]

#### 5. Hybrid Approach

Hybrid Approach is basically the combination of all other approach. For the ranking purpose of corresponding antecedents, they combine various knowledge sources like discourse, syntactic, morphological, etc. Lapin and Leass (1994 ) [5] system is one of the popular hybrid based system. In this a model was used to calculate the discourse salience of an entity on the basis of various factors calculated dynamically .This salience measure was then used to rank potential entities.

### IV. COMPARISON OF ANAPHORA RESOLUTION TECHNIQUES

Its usually seen that earlier systems considered only the issue of resolving pronominal anaphora [21]. Whereas Knowledge-poor machine learning techniques takes into consideration the issue of noun phrase coreference resolution. Knowledge-rich techniques make use of manual pre-processed input data. On the contrary knowledge-poor approaches are end to end systems having automated pre-processing stages as that of noun-phrase identification, semantic and morphological class identification. However Mitkov,et.al [6] reports that the systems that have automated pre-processing shows erroneous in their respective pre-processing stages leading to a reduction in the system performance. Thus its quite unfair to compare machine learning techniques using automated pre-processing with knowledge-based techniques or with knowledge-poor techniques.

### V. PARAMETERS FOR EVALUATION OF RESULTS

The two most commonly used parameters for evaluating results obtained in Anaphora Resolution are namely **precision** and **recall**. They are the measure of **accuracy** and **completeness** respectively. The sum of true positive and false positive (FP) signifies the pairs which are incorrectly marked as coreferent and the sum of true positive (TP) and false negative (FN) represent the pairs which should have been marked as coreferent but have not been. F-measure usually is the combination of precision and recall scores. The formula for calculating precision, recall and F-Score are as mentioned below:

- precision =  $\frac{TP}{TP+FP}$

- recall =  $\frac{TP}{TP+FN}$

- F-Score =  $\frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$

Correct Classification		Obtained Classification
TP (True Positive)	FP (False Positive)	
FN (False Negative)	TN (True Negative)	

Fig 1. Classification of a Noun Phrase Pair and Desired Correct Classification

## VI. CONCLUSION

Resolution of Anaphora is a tedious and important millstone in Natural Language Processing. Also, it affects the accuracy of all natural language processing applications. On studying the past works and research in the field of Anaphora Resolution we came to conclude that there have been several works done in different languages like English, Spanish, Turkish etc ,but very few work has been done in Indian Language specially Hindi. The existing techniques of Anaphora Resolution need to be more enhanced so that we could develop better algorithms which can work on all issues like Recency factor, Gender, Number Agreement etc of Anaphora Resolution. at the same time. Also, we need to develop efficient approaches for Indian Languages for resolving Anaphora.

## REFERENCES

- [1] Triveni Lal Pal,Kamlesh Dutta and Pardeep Singh, "Anaphora Resoution in Hindi:Issues and Challenges",International Journal of Computer Applications (0975-8887),Volume 42-No.18,March 2012.
- [2] Ashima and Rajni Mohana, "Anaphora Resolution in Hindi: Issues and Directions", Indian Journal of Science and Technology,Volume-9 (32),August 2016
- [3] Daya Shankar Yadav et.al , "Anaphora Resolution for Indian Languages:The State of the Art", CPUH-Research Journal:2016,I(2),01-07.
- [4] Ishani Sharma and Pradeep Kumar Singh,"A Survey on Anaphora Resolution", International Journal of Computer Applications (0975-8887), Recent Innovations in Computer Science and Information Technology.
- [5] Lapin S.Leass HJ , "An Algorithm for Pronominal Anaphora Resolution ",Computational Linguistics, 1994 Dec 1:20(4):535-61.
- [6] Mitkov,Ruslan,Boguraev,Branimir and Lappin Shalom,2001,"Introduction to the Special Issue on Computational Anaphora Resolution in Computational Linguistics",Volume 27, Number 4 , 2001.
- [7] Hobbs J., "Resolving pronoun references",Lingua 1978 Jan; 44:311-38.
- [8] Mitkov R, "Robust Anaphora with Limited Knowledge", COLING98, Canada 1994.
- [9] Carter DM, "A shallow processing approach to anaphora resolution", PhD thesis, Univ of Cambridge,1987.
- [10] Carbonell JG,Brown RD, "Anaphora Resolution: A Multi-Strategy Approach", Proceedings 12<sup>th</sup> International Conference on Computational Linguistics, Budapest,1988,pg. -96-101.
- [11] Mitkov R, "Anaphora Resolution: A combination of linguistics and statistical approach",Proceedings Discourse Anaphora and Anaphora Resolution, UK,1996.
- [12] Connoly D, Burger JD , Day DS,"A machine learning approach to anaphor reference",Proceedings International Conference, New Methods in Language Processing,UK,255-61.
- [13] Nasukawa T, "Robust method of pronoun resolution using full-text information",COLING '94
- [14] Tetreault JR,"A Corpus-Based Evaluation of Centering and Pronoun Resolution ", Computational Linguistic ,Special Issue on Computational Anaphor Resolution, 2001 Dec;27(4):507-20.
- [15] Brennan S,Friedman M, Pollard C, "A Centering Approach to pronouns", Proceedings 25<sup>th</sup> Annual Meeting of the ACL,USA,1987.
- [16] Kameyama M,"Recognizing referential links: An information extraction perspective", ACL'97/EACL'97 Workshop on intelligent scalable text summarization, Madrid, Spain,1997;3-9.
- [17] Roland S, "Resolving Anaphor references on Deficient Syntactic descriptions ",ACL '97/EACL'97 workshop on operational factors in practical, robust anaphora resolution, Madrid,Spain,1997;30-7.
- [18] Stuckardt R, "Design and Enhanced Evaluation of a Robust Anaphor Resolution Algorithm",Computational Linguistics,2001 Dec;27:445-52.
- [19] Baldwin B.CogNIAC: High precision co-reference with limited knowledge and linguistic resources,ACL'97 workshop on operational factors in practical, robust anaphora resolution, Madrid,Spain,1997;38-45
- [20] Mitkov R, "Towards more consistent and Comprehensive evaluation in anaphor resolution",In Proceedings LREC'2000,Athens,Greece,2000;1309-14.
- [21]Tejaswani Deoskar,"Techniques for Anaphora Resolution: A Survey", CS 674,5/17/2004.