International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/ 10.32628/CSEIT239059

Computational Linguistics: A Research Survey on Ambiguity and Anaphora Resolution

Harjit Singh*[1]

*[1]Department of Linguistics and CSTLs, Indira Gandhi National Tribal University, Amarkantak, Madhya Pradesh, India

ARTICLEINFO

Article History:

Accepted: 05 Sep 2023

Published: 16 Sep 2023

Publication Issue

Volume 9, Issue 5

September-October-2023

Page Number

98-102

ABSTRACT

In this paper we briefly discuss main areas of computational linguistics particularly focus on the problem of ambiguity in computational linguistics. Then we take up the issue of ambiguity with inclusion of term anaphora which is well known in most of the languages and it is seen under pronouns specifically reflexives and reciprocals. In general, we have total five sections in this paper. First section primarily deals with introduction related to the topic and elaborates it in detail. Second section defines ambiguity with its background investigations in relation with computational linguistics. Third section presents detailed study on anaphoric ambiguity with strict sloppy readings at syntax level. Fourth section discusses about anaphora and soon it adds anaphora resolution irrespective of any specific language and most of the work related to anaphora resolution is mentioned from natural language processing point of view. Fifth is the last section where we conclude that ambiguity and anaphora both are difficult tasks for machine translation until unless we don't train machine (computer) to learn as per given instructions and all.

Keywords : Anaphora, Ambiguity, Anaphora Resolution

I. INTRODUCTION

The field of Computational linguistics is that an emerged theoretical as well as practical discipline for Natural language understanding (NLU) which is initially grown in the work of Machine translation late 1950 (when first work was done by computer for interpretation of the natural language and further it is employed in different projects of language translations) and now it becomes as an interdisciplinary field (i.e. which interacts computer science, engineering, linguistics, philosophy of language and logic on the one hand and moreover it is employing in information retrieval and artificial intelligence on the other hand). Assuredly from the beginning, it was a major problem for machine to understand and systematically generate natural language (at all possible levels that we will discuss below. Cited in NLI 23 July 2012) and tried to

diminish this problem by the inter-partnership of computer science, linguistics, engineering and other fields.

Input level: Phonetically the variation in articulation level, the incompatibility of pitch, accent and acoustically reasonable changes presents a problem at the speech recognition level.

Understanding level: The occurrence of ellipsis and anaphora resolution both can be seen at the morphological and syntactic level.

Dialogue level: At this level, the context ambiguity and the encoding of ideas, gestures and gaps. Output level: Despite the fact that the valuable work has been done in speech synthesis but remains the challenge to develop human likes capacity.

Another one of the significant obstacles of ambiguity is a strenuous task for natural language processing which human being solves easily with contextually as well as worldly knowledge. But lack of human like capacity machines cannot do this work easily. Hence, there is a need for a lot of linguistic knowledge and constructing an adequate algorithm for this work and computational linguistics imparted this sort of knowledge. In general, two viable approaches (knowledge-based and statistical-based) employed for this problem. In the further sections, we will scrutinize this problem.

II. THE ISSUE OF AMBIGUITY IN COMPUTATIONAL LINGUISTICS

There is no doubt that ambiguity is a complex feature of natural language that supposed to be seen at every linguistic level and cognitive point of view it is a desirable property of language which motivates both speaker and hearer for better communication. Piantodasi and Gibson (Dec, 2010: 281) discussed two advantages of ambiguity: first where context is informative about meaning unambiguous language is partly redundant and therefore inefficient, and second ambiguity allows re-use of words and sounds that are more easily produced and understand. This view is also considered by Wascow (2005) where he points out that ambiguity might be useful in language contact situation where speakers of both languages should ideally be able to handle words meaning two different things in two different situations (cited in Piantodasi Dec. 2010: 282). In the same way, Abney (1996) described it as, “one often hears in computational linguistics about completely unremarkable sentences with hundreds of parses, and that is in fact no exaggeration”. (Cohen 2006: 1).

However, it is a universal phenomenon of all-natural languages nonetheless by its nature it is an obstacle for natural language parsing. The process of natural language parsing is done at each linguistic level (morphology, phrase, clause and sentence) and also ambiguity occurs at each level but except of it there might be many other decision points (word sense, sub-categorization, complementation pattern, scope of quantification and negation and more) accountable for it. Actually, not only conceptual techniques (Schnak, 1975) but also grammar-based (Pollard and Sag 1987) techniques have seen in natural language analysis. But these techniques were failed when they applied in large vocabulary and grammatical structure data and consequently multiple ambiguities were come up.

Indeed, there are many rudimentary issues in natural language parsing where it is difficult to decide about how much wider grammatical coverage necessary is? How can reduce the frequency of ambiguity at the parallel level? In this context, we can briefly survey of this problem with its background.

In previous work of ambiguity resolution, there was a first problem of ‘unknown words’ that was handled through conceptual parsing (Schnak, 1975) based on syntactic, semantic and pragmatic information. In further, Black (1991) used morphological analysis of new words for using them for speech generation

application and after Weischedel (1993) described POST system for Part-of- speech tagging which included probabilistic model for unknown words. He also reported about 50% error rate of trigram based unknown word tagging. This was corpus-based approach for unknown words. Second type of ambiguity problem (Lexical-syntactic ambiguity) was resolved with different approaches. Klein and Simmons (1963) first of all developed earlier rule-based POS tagger and subsequently Green and Rubin (1971) created TAGGIT tagging program for tagging Brown corpus. Church (1988) notes a similar effect with Donald Hindle's FIDDITH Parser (Hindle 1983). FIDDITH is a deterministic parser with very wide coverage. Thirdly structural ambiguity based on more than one syntactic structure of the sentence. Bever (1970) formulated a number of heuristics for syntactic ambiguity resolution. Primarily, he used syntactic approach. Not only syntactic approach was useful for this problem then Katz and Fodor (1963 'selectional restriction') and others conceptual parsing (Schnak and Abelson 1977) and case frame parsers (Carbonell and Hayes 1987) all researchers employed semantic knowledge domain in their parsing process. On the other hand, Pragmatically Crain and Steedman (1985) proposed a number of pragmatic principles to account for parsers: A prior plausibility, Referential success and Parsimony (Alexander 1996: 4-6, 11-23). In ambiguity resolution framework above approaches were rectified again for getting better outcome but formally it's true that ambiguous nature of natural language as a challenge for natural language parsing. Lack of enough space we cannot discuss more here.

III.ANAPHORIC AMBIGUITY

In this section, our target is to interpret anaphoric ambiguity. In general, anaphoric ambiguity occurs when two or more antecedents will potentially be considered as selected candidates for a single anaphor in a sentence. This sort of information about anaphoric ambiguity dealt in a paper "Analyzing Anaphoric Ambiguity in Natural language requirements" (2011) where primarily two sorts of anaphoric ambiguity (nocuous and innocuous) figure out from the requirement documents (that were written in natural language) by the special concern of different views of different readers (customers, analysts, programmers and on the other hand translators) following the reliable linguistic factor strategy and earlier methodology of automatic identification of nocuous ambiguity of Yang (2005 et al).In general, above mentioned anaphoric ambiguities resolution extended by using classifier technique which is trained on the basis of linguistic knowledge. However, this is a vital part of this paper but we will solely notice these two sorts of anaphoric ambiguity.

We can also scrutinize anaphoric ambiguity by linguistic perspective. In linguistics, specific syntactic-semantic domain is accounted for discerning ambiguous nature of anaphora. For illustration, we can focus on strict-sloppy readings of pronoun that illuminated an ambiguous sense. A pronominal anaphor which when syntactically occurs in VP elliptical construction interpreted between strict identity reading and sloppy identity readings (Gawron & Peters 1990: 4: 6). For example, John left in his car and so did Bill.

In above sentence, pronoun 'his' does not give only deictic expressions but also communicates two different readings: strict-identity reading and sloppy-identity reading. In first reading 'John' and 'Bill' both leave in the same car (John's car) whereas in second reading each (John and Bill) leave in his own car. Hence here above sentence is ambiguous.

A second sort of example, which was found by Partee (1972) first discussed in Roberts work (1967) also evidence of pronominal ambiguity (Gawron & Peters 1990: 74: 105). See below

John expected that he would lose and Bill did too.
Similarly, in this example both types (strict and sloppy identity) readings possible. But here important is that consider Elliptical VP hypothesis which is subject for VP anaphora that was recognized by Partee (1972), Sag (1976), Bach & Partee (1980) and also become source for ambiguity.

IV. ANAPHORA AND ANAPHORA RESOLUTION

This section deals with a term 'anaphor', which etymological come from Greek word(αναφορα) that means 'pointing back' and also applied in the 'repetition sense' of the co-indexation with the previous or successive word (antecedent) in a sentence or discourse. Typologically it interprets as lexical anaphor, pronominal anaphor, verbal anaphor, zero anaphor and others. Mostly it is a contextual property of natural language that presents a challenge for those who works in computational linguistics, Artificial intelligence, philosophy, logic, linguistics etc.

There is a natural fact that the expression of natural language often context-dependent and this dependency is followed by proper noun, nominal and pronominal and verbal. In all of these dependent entities, pronominal most dependent entity might be in particular reflexive and reciprocal cases, where they require antecedent and, in this sense, these dependent expressions are called 'anaphora'. Even dependency expressions have been different types (anaphoric) also and further we can linguistically discuss this typology in a precise way (Posesio et al, 2010: 2).

"The term 'anaphora' is derived from the Greek word (αναφορα) that means 'carrying back'. In contemporary linguistics, it assumed as a relational term between two different linguistic elements whereas one is called (anaphora) and another is its (antecedent) (e.g. Lust 1986b, Wascow 1986, see also Huang 1994: 1 Cited in Huang 2000: 1)". Linguistic elements (those entities who are dependent on others) called anaphor include gaps (or empty categories), pronouns, reflexives, names and descriptions.

In recent years, the scope of anaphora has been wider and attracted not only for philosophers, psychologists but also for cognitive scientists and artificial intelligence workers who have been working in this area. Huang (1994:1) points out some reasons for growing this interest in research as. First it is a very complex phenomenon of natural language that has become a problem for resolution. Secondly anaphora might be as an inquiry of human mind/brain and relevant for understanding about what Chomsky discusses fundamental problem of linguistics (1981, 1982, 1986a, 1995) namely the logical problem of language acquisition. Thirdly anaphora can linguistically be tested by syntactic, semantic and pragmatic factors (Huang 1999: 1-2, cited in Huang: 2000: 1).

By and large Anaphora resolution is a laborious task, which is accomplished with the struggle of linguists, computer scientists, computational linguists etc. To detect the appropriate antecedent (NP, Pronoun, VP, Sentence etc.) for any kind of anaphor is a paramount problem in anaphora resolution. Various sources of information are requisite for automatic anaphora resolution process. For instance, lexical information (head-noun matching, number-gender agreement) can be indicator for co-reference between antecedent and anaphor. Syntactic knowledge employed for defining the boundaries of sentence or clause and in addition to develop rules for anaphora resolution. Semantically, selection restriction is another requisite part for corresponding between antecedent and anaphor. However lexical, syntactic and semantic knowledge is required nonetheless we cannot ignore alternative knowledge like discourse knowledge, which gives a new orientation when neither machine nor humans capable for detecting the antecedent of expletive pronoun 'it'. The application of discourse knowledge is

not only solution of every sort of anaphoric problem. In the context of some another pronominal anaphor, we can smoothly realize the importance of real-world knowledge.

The task of anaphora resolution has a long history of different approaches and different systems that have evolved this process in NLP work. There are numerous phases of this process that might be perceived by contemplating the whole work in a brief way. In general, the inception of this process, when Winograd (1972) developed his SHRLDU system, which partly focused on it in its language module and also juxtaposed with today's systems (Naïve approach by Hobbs (1978). This approach is properly followed syntax and does not give any detail about semantics or pragmatics. Nonetheless Hobbs embraced semantic information also in his previous work. Not even Hobbs gave a positive orientation in this difficult task but other works (of Grosz et al (1983, 1995), Lappin and Leass (1994), Aone and Bennett (1995), Mitkov (1998) etc.) additionally thrived this task in a same or different way.

V. CONCLUSION

The purpose of this work is to provide information about anaphora and anaphora resolution by the brief introductory framework of computational linguistics. First, we focus on the issue of ambiguity and anaphoric ambiguity then consider the term anaphor and its typology. For anaphora resolution, we go through out the historical and current approaches.

IV. REFERENCES

[1]. Cohen, A. (2006). ' Why ambiguity?', (eds.) In Martin H., In Between 40 and 60 Puzzles for Manfred Krifka.

[2]. Cuevas, R. R. and Paraboni, I. (2008). A Machine Learning Approach to Portuguese Pronoun Resolution. Springer, pp. 262-271.

[3]. Dash, N. S. (2000). Language Corpora: Present Indian Need. Proceedings of the SCALLA 2004 Working Conference. Kathmandu.

[4]. Franz, A. (1996). Automatic Ambiguity Resolution in Natural Language Processing- An empirical approach. Tokyo: Springer.

[5]. Freedman, A. (2001). The Computer Desktop Encyclopedia. Osborne: McGraw- Hill.

[6]. Gasperin, C. V. (2009). Statistical Anaphora Resolution in Biomedical Texts. London: University of Cambridge.

[7]. Gawron, J. M. and Peters, S. (1990). Anaphora and Quantification in Situation Semantics, Centre for the study of language and information. Leland: Stanford Junior University.

[8]. Gopal, M. (2011). Dissertation of Computational anaphora and Cataphora in the Sanskrit text Panchtantra, Centre for Linguistics. New delhi: Jawaharlal Nehru University.

[9]. Hicks, G. (2009). The Derivation of Anaphoric Relations. Amsterdem: John Benjamins Publishing Company.

[10]. Huang, Y. (2000). Anaphora: A cross-linguistic study. OUP: Oxford. New York: Oxford University Press.

[11]. Mitkov, R. (2002). Anaphora Resolution. London: Longman.

Cite this article as :

Harjit Singh, "Computational Linguistics: A Research Survey on Ambiguity and Anaphora Resolution", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 5, pp.98-102, September-October-2023.
doi : https://doi.org/ 10.32628/CSEIT239059