INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

A Survey on Anaphora Resolution Toolkits

Seema Mahato[1], Ani Thomas[2], Neelam Sahu[3]

[1]*Research Scholar, Dr. C.V. Raman University, Bilaspur, Chattisgarh, India*
[2]*Dept. of Information Technology, Bhilai Institute of Technology, Durg, Chattisgarh, India*
[3]*Dept. of Information Technology, Dr. C.V. Raman University, Bilaspur, Chattisgarh, India*

Abstract: Anaphora Resolution is one of the fundamental tasks in Natural Language Processing. In foreign language domain, there have been numerous of studies on Anaphora Resolution but still very limited research have been carried out on Indian languages like Hindi, Bengali, Telegu, Malayalam, and Marathi. This study aims to investigate features on which an anaphora resolution system based. However, the systems not meant for Hindi but as per their respective developers the approach and theory they follow, can be modify to re-develop the system for Indian languages. The features of online available anaphora resolution toolkits compared and categorized based on the study of usage of language, corpus, algorithm, preprocessor, etc. in the paper. The paper briefly put forward the investigation of the toolkits developer's in a summarized manner and conclusion drawn for the best suitable model.

Keywords: Anaphora; Anaphora Resolution, Automatic Anaphora resolution tools; Computational strategies; Natural language processing.

I. INTRODUCTION

Large number of researchers involved in developing automatic anaphora resolution (AR) system for different languages by incorporating different approaches and theory. Most of them succeed to do so but accuracy result shows that they are lacking in few aspects. Anaphora as word does not explore much about self except that it may refer to an entity, which could a noun or verb. The need of a successful AR system could be understand by the dependency of few NLP applications such as machine translation system, question answering system, text summarization system, etc. The first edition of fully automatic AR was by Mitkov et al [1] which does not included domain knowledge. The resolution begins with the categorization of words as pronouns and pleonastic pronouns. The pronouns were group into anaphoric and non-anaphoric pronouns. Different techniques have been implements to identify the referent of individual type of pronouns. The preprocessing tools employed were morphological analyzer, POS tagger, lexical noun phrase extractor, proper name recognizer, etc. World or domain knowledge involvement was marked in the knowledge rich anaphora resolution approach. The corpuses annotated using such tools before actual resolution take places.The annotation scheme or model plays important and foundation role in overall resolution process. Wrong annotations lead to low accuracyor failure of a system. After annotation, all noun phrases precedes pronouns are identified and added in a list for possible candidates. The features selection processesfor possible candidates of antecedents vary from algorithm to algorithm. Implementing different filtering rules potential candidates are markfor antecedent on basis of some factors. Factors are broadly group as constraints and preferences whose implementation totally depends on applied approach and theory [2]. Gender and number agreement, person and case agreement, syntactic relation between noun phrase and pronoun comes under constraints. Preference checks the recency of candidates, compare construction of sentences, etc. The algorithms for finding antecedents have been design with a specific search scope. Some algorithm searching scope limits to 1-3 sentences whereas others search beyond 3sentences. This paper provide the basic ground for developing automatic anaphora resolution system by analyzing and determining that blending of which features, approach and theory could boost up the efficiency and performance.

II. ANAPHORA RESOLUTION TOOLS

The few popular anaphora resolution tools listed here.
Out of these, only Mars, Javarap, and Arkref offer option for online demo.

A. Guitar

Guitar comes in three versions. Initially it does not deal with demonstratives or proper nouns but the latest one implemented shallow algorithm for resolving the same[3]. It may take inputs in two formats, XML or text format. A XML file is generated on processing the text input by the LT-XML (Language Technology-XML) tool, which further processed to produce MAS-XML (Minimum Anaphoric Syntax - XML) which contains morphological information. It can resolve pronominal and lexical anaphora. Developers

have evaluated it on two corpuses: Generating Nominal Expressions (GNOME) and Computer-Aided Summarization Tool (CAST). Annotation of this corpuses performed by MMAX (Multi-modal annotation in XML) tool and Charniak parser. In Table I[4] evaluation result of Guitar showed a precision of 69% and recall of 71% for GNOME corpus and precision of 51% and recall of 54% over CAST corpus. In this evaluation, it has found that precision and recall score associated with personal pronouns and possessive pronouns shows the ability of the system to handle both the pronouns.

TABLE I Performance of Guitar 3.2

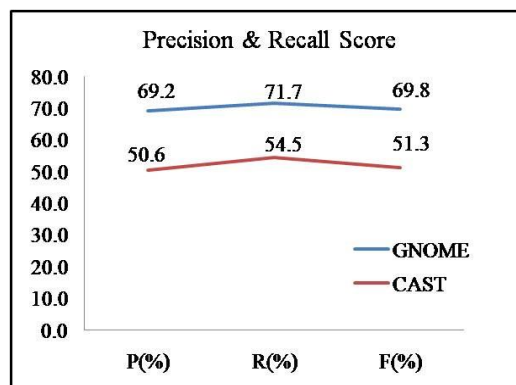| CORPUS	ANAPHOR	P(%)	R(%)	F(%)
GNOME	DD	70.4	63.6	66.8
	PersPro	78.1	77.8	78
	PossPro	79.1	73.3	76.1
	PN	49	72	58.3
		69.2	71.7	69.8
CAST	DD	68.7	54	60.5
	PersPro	44.6	46.9	45.7
	PossPro	54.3	54.1	54.2
	PN	34.7	62.8	44.7
		50.6	54.5	51.3

Figure 1 Precision -Recall Score for Guitar 3.2

In case of proper nouns, the system seems to perform much lower in precision and recall for both the corpus, but at the same time, F1 score was low, which warrants further investigation. Whereas system handled the definite descriptions associated with GNOME preferably good in comparison to CAST corpus. Guitar still needs improvement in order to handle proper nouns.

B. Bart

The best thing about Bart is that it can resolve anaphora as well as co reference. It is a rule-based system built in Java. The flexibility and portability of Bart is due to the language plug-ins to resolve anaphora for more than one language independently like English, German and Italian. It accompanied with modules like preprocessing pipeline, which create markablesand mention factoryto, create mention objects using these markables, feature extraction module for generating classification features, decoder and encoder for training and testing phase. Like Guitar, Bart also takes inputs in two formats, XML or text format and gives output in XML format. It includes Stanford POS tagger, Yam Chachunker, Berkeley parser,MMAX2 annotation tool, Charniak and Johnson's re-ranking parser and supported by two toolkits, WEKA (Waikato Environment for Knowledge Analysis) / ME (Maximum Entropy) machine learning and SVM Light tagger toolkit, where SVM stands for Support Vector Machine. Co reference resolution was evaluated on the training and test corpora from MUC-6[5].Bartwas evaluated on the Sem Eval task 1 corpus by using the SemEval scorer. Bart uses highly informative basic features set which includes distance feature, Pronoun Feature, String Match Feature, Definite Noun Phrase Feature, Semantic Class Agreement Feature, Number Agreement Feature, Appositive Feature, etc. These features are either unary or binary in nature. These features evaluated on basis of F-measure whose value could be zero or

nonzero. It also uses tree kernels representing relation between anaphor and antecedent syntactically to build up extended feature set. Bart shows F-measure of 65.8% and 62.9% on Message Understanding Conference (MUC)-6 and MUC-7 respectively[6].

Figure 2 (data source:[6]) show that Bart results Automatic Content Extraction (ACE)-2 by utilizing a tagger for extracting mentions in ACE corpora, extended feature set with syntactic features and knowledge based features extracted from Wikipedia. ACE contains two sets of data: training and devtest. Each of these sets further divided by source: broadcast news (Bnews), newspaper (Npaper), and newswire (Nwire).
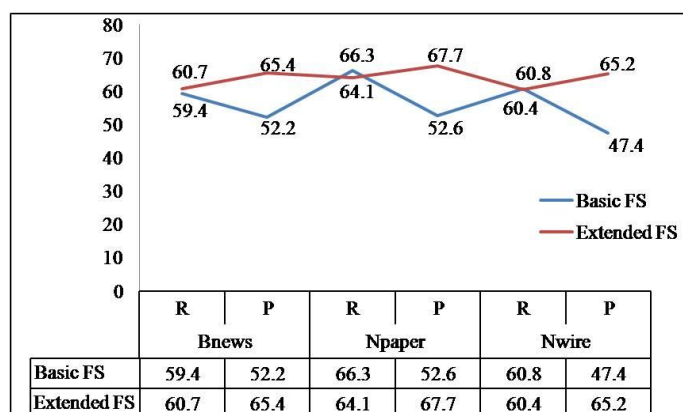

	R	P	R	P	R	P
	Bnews		Npaper		Nwire	
Basic FS	59.4	52.2	66.3	52.6	60.8	47.4
Extended FS	60.7	65.4	64.1	67.7	60.4	65.2

Figure 2 Precision -Recall Score for Bart

Pronoun resolutions using the extended feature set have improved the efficiency.Barthavebeen implemented for German and English languages and showed good performance for both the languages.

C. Mars

Mars is one of earliest fully-fledged automatic anaphora resolution system. It is a knowledge poor multilingual approach includes syntactic and semantic information and able to handle all type of anaphors. This total pronoun resolution integrateddomain and discourse modules apart from heuristically based modulesthat restricted to a sublanguage or genre. It have overcome the burden of manual preprocessing such as pre-editing of the text, removal of pleonastic pronouns, annotating corpora and post editing of outputs. It includes finite numbers of genre independent or genre-specific indicators, the Connexor Functional Dependency (FDG) parser to perform syntactic analysis, modules to recognize instances of nominal anaphors, non-nominal pronominal anaphors, pleonastic pronouns and gender identification automatically in its preprocessing phase. Mitkov [1]evaluated Mars on different technical manuals and achieved asuccess rate of 89.7% for English. The evaluation performed in two manners, one by activating the syntactic, semantic and domain modules and other by adding the discourse modules into these. On combining the syntactic and semantic constraints in its statistical approach shows an improvement. Marsinitially developed and tested for English. With least modification, it shows an accuracy rate of 93.3% for Polish and 95.8% for Arabic languages[1]. The approach tested for Finnish, French and Russian also.

D. Javarap

Javarap is anopen-source, built in Java and platform-portableknowledge based anaphora resolution approach works on algorithm proposed by Lappin and Leass [7]. It can successfully identify third person pronouns, lexical anaphors, and identifies pleonastic pronouns. It take input as plain text, text with XML tags or text with MUC co-reference annotations and give output in the form of anaphor - antecedent pairs. It can identify antecedents of third person pronouns whether it is inter sent entialorintra sentential. It uses Charniak's parser, pleonastic pronoun filter, syntactic filter and anaphor binder module. Apart from these, Sentence Splitter and Anaphora Resolver Evaluator also used as associating tools forthe resolution. Syntactic filter are for identifying third person pronouns whereas anaphor binding algorithm for identifying lexical anaphors. Further internal processing based on group of salience factors like head noun emphasis, subject emphasis, sentence recency, etc. and a weight defined for each factor. Each one in the set of potential candidates assigned with the associated weights of factors to which it belongs. For each candidate sum of the weights calculated and those having highest weight detained as antecedent. In case if multiple candidates attain overall equal weights then the distance between the anaphor and the candidates computed and the nearest one considered as antecedent. Qiu et al. [8] evaluated Javarapon MUC-6coreference task for English language and recorded 57.9% of accuracy.The algorithm has identified

the antecedents in most of the cases. The accuracy of the system gets dropped if the article contains only capitalized letters and it also not able to delimitate sentences if they are all in lower case which may be due to case-sensitivity of sentence splitter.

E. Arkref

Arkrefis a rule-based knowledge-rich system available as an open-source for co reference resolution. The noun phrase co reference resolution approach based on Choonkyu Lee, Smaranda Muresan, and Karin Stromswold works as well described by Haghighi and Klein.Syntactic information exploited from Stanford Parser help out to know the form in which pronoun appear in actual syntactic position. Semantic information obtained from entity recognition component provides the frequency and closeness of noun phrase with anaphor and to group them in matching type, a supersense tagger employed. The shortest path distance mechanism implemented to select correct antecedent from multiple candidates.O'Connor et al. [9] evaluated Arkref on multiple co reference resolution metrics such as Pairwise $F_1$and B^3.The system used the BNC corpora, web corpora and WordNet to identify the NP coreferences among the NPs in the sentences and evaluated on the ACE2004-Roth-Dev and ACE2004-CULOTTA-TEST dataset[9]. The componentssuch as syntactic constraints and semantic compatibility together add on overall evaluation result. The system is deterministic and implemented in Java that can be easily downloaded from the web.
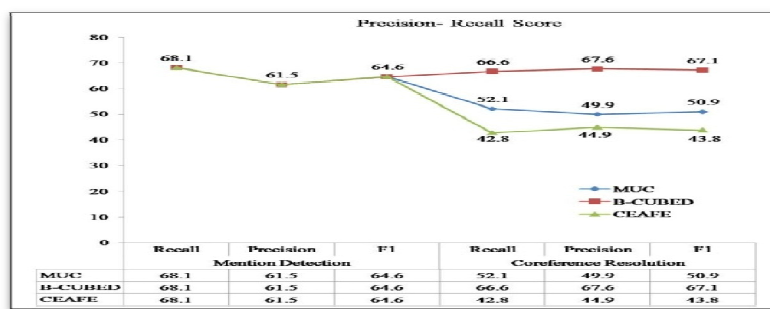
F. Vasisth

Figure 3:Precision -Recall Score ForVasisth

Vasisth is syntax based multilingual AR system without deeply parsing the sentences. It used syntactic knowledge and totally ignored the world knowledge. It may be for all Indo-Aryan, Indo-Dravidian and Indic family of languages after little modification.It deals with all types of pronouns, distributives, gaps and ellipsis.It consists of two separate module such as pronominal resolution module, which works on basis of salience factors, and non-pronominal resolution module to detect non-anaphoric pronouns using machine-learning approach. The data set for training and development contained files from different fields mainly related to news, blogs and magazine articles. Vasisth was evaluated on MUC, B-Cubed and Entity-based Constrained Entity-Alignment F-Measure (a.k.a. CEAFe) metrics[10].Gold standard annotation tool used for identifying actual number of anaphoric and non-anaphoric pronouns. When the non-anaphoric pronoun detection module evaluated, the system shows a high accuracy in identifying anaphoric pronouns in compare to non-anaphoric pronouns. The performance of pronominal resolution module evaluated disjointedly by employing development data. Sobha et al. [10] also evaluated the pronominal resolution module after determining and filtering non-anaphoric pronouns and noted improvement with a high precision score. With or without non-anaphoric pronoun detection, the system was capable to resolve equal numbers of pronouns but without non-anaphoric detection, the system was unable to identify 10% of pronouns. Sobhaet al.[10] tested the system for Malayalam and after minor modification tested it for Hindi that showed an accuracy of 82%. The system not examined the system for longer discourses.

III.COMPARISON AND SUMMARIZATION

Table II highlights the prerequisites like pre-processing tools and name of the anaphora resolved by the tools. Mars, Javarap, Guitar 3.2, Bart, Arkref, and Vasisth are rule-based system. Most of these systems tested for English language except Vasisth. These systems used corpus or datasets from different genre like technical manuals, web blogs, News or magazine articles. Vasisth have treated anaphoric pronominal reference and NP-co reference resolution as separate problems. Building a anaphora resolution system not only require approach and theory, however also necessitate efficient selection of features and factors for mention detection. In addition to variety of pre-processing tool applied for different genre of corpus and testing based on standard evaluation metric, entirely decide a package for a good anaphora resolution.

TABLEII: Comparison of different anaphora resolution systems

Features ➡ Systems ⬇	Processing tools	Purposes
Mars	FDG parser	Third personal pronouns and lexical anaphora
Javarap	McCord's Slot Grammar parser, Charniak parser	Resolves third person pronouns, lexical anaphors, and identifies pleonastic pronouns
Guitar 3.2	Penn Tree Bank tag set used by Charniak's full parser, OpenNLP Tools	Resolves four types of anaphora (Definite Descriptions, Proper Noun, personal and possessive pronoun)
Bart	Charniak's Parser, Carafe/Stanford NER	Focuses more on coreference resolution rather than anaphora resolution properly
Arkref	Stanford Parser, Supersense tagger	Pronominal anaphors, reflexive pronoun
Vasisth	Rule based parser	Resolves all pronominal anaphors, non-pronominal anaphors, gaps and ellipsis

.

Table III summarizes set of constraints and/or preferences used by these systems and compared on basis of capability of Named entity recognition, Salience measurement, Word sense disambiguation, etc. with performance in terms their success rate. As seen in Table III, Mars performed with the highest accuracy of 73.5%, while Javarap produced the lowest accuracy of 57.9%and Vasisth performed with an accuracy of 82%. The low success rate could be because of Javarap only handles pronominal anaphora resolution.

TABLEIIIUsability of Lexical and Semantic information by different ARS

Features ➡ Systems ⬇	Named entity recognition	Semantic analysis	Salience measurement	Word sense disambiguation	Reported success rate
Mars	Yes	Yes	Yes	Yes	89.7
Javarap	No	Yes	Yes	No	57.9
Guitar 3.2	Yes	Yes	Yes	Yes	71.3
Bart	Yes	Yes	Yes	Yes	65.8
Arkref	Yes	Yes	Yes	Yes	$80.5(B^3)$
Vasisth	Yes	No	Yes	No	82

As shown in Table III, the systems that has Named Entity Recognition, Salience Measurement and Word Sense Disambiguation in their system contributes to the high accuracy results, while those has limited features incorporated in order to resolve AR. The authors have also depicted the resulted success rate in graphical manner through Fig. 4.
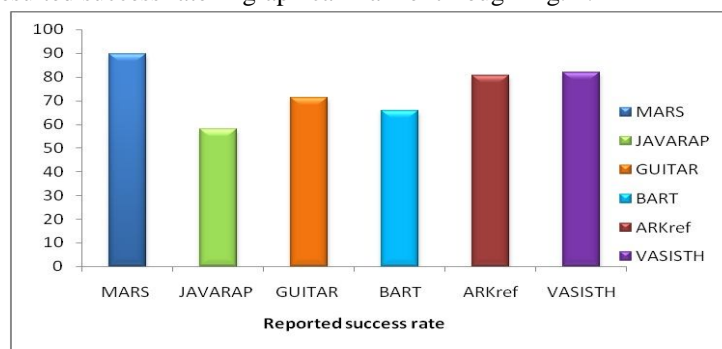

Figure 4 Performance chart of different Anaphora Resolution Tools

IV.CONCLUSIONS

This paper shows that anaphora resolution systems based on common approaches follow uncommon strategies and evaluation metrics, making it hard to compare their performance in absolute and qualitative terms.The approaches discussed in the paper indicate that by utilizing the knowledge source and common set of factors with different computational strategies in an efficient and effective manner could bring out high rate of success. Now-a-days, the availability of pre-processing tools has replaced the manual task of annotation and removal of pleonastic pronoun 'it' that motivated the researchers to rely on approaches based on demand limited knowledge and computational strategies for simplicity and robustness design in comparison to knowledge based systems.

V. ACKNOWLEDGEMENTS

REFERENCES

[1]   Mitkov, R, Evans, R., &Orasanal, C. 2002. A new, fully automatic version of Mitkov's knowledge-poor pronoun resolution method. Lecture Notes In Computer Science, 2276:168-186

[2]   Mitkov, R. 1999. Anaphora Resolution: The State Of The Art.Proceedings of COLING'98/ ACL'98.

[3]   Poesio, M., Kabadjov, M.A. 2004.A General-Purpose, off-the-shelf Anaphora Resolution Module: Implementation and Preliminary Evaluation.Proceedings of International Conference on Language Resources and Evaluation. Portugal

[4]   Steinberger, J.,Poesio, M., Kabadjov, M.A., &Jezek, K. 2006.Two Uses of Anaphora Resolution in Summarization.Information Processing & Management. 43(6):1663-80

[5]   Versley, Y., Ponzetto, S.P., Poesio, M., Eidelman, V., Jern, A., Smith, J., Yang, X., &Moschitti, A. 2008.BART: A Modular Toolkit for Coreference Resolution. ACL:9-12.

[6]   Broscheit, S., Poesio, M., Versley, Y., Ponzetto, S.P., Rodriguez, K.J., Romano, L., Uryupina, O., &Zanoli, R. 2010.BART: A Multilingual Anaphora Resolution System.Proce

[7]   edings of the 5th International Workshop on Semantic Evaluation. Association for Computational Linguistics:104–107.Lappin S., Leass H., J. 1994. An algorithm for pronominal anaphora resolution. Computational linguistics, 20(4):535-561.

[8]   Qiu, L., Kan, M.Y., Chua, T.S. 2004. A public reference Implementation of the RAP Anaphora Resolution Algorithm. Proceedings of International Conference on Language Resources and Evaluation. Portugal.

[9]   O'Connor, B., Heilman, M. 2013. ARKref: a rule-based coreference resolution system.CoRR/abs:1310.1975Pal T., L., Dutta, K., Singh, P. 2012. Anaphora Resolution in Hindi: Issues and Challenges. International Journal of Computer Applications, 42(18):

[10] Sobha, L., Patnaik, B.N. 2000.Vasisth: An Anaphora Resolution System for Malayalam and Hindi. Proceedings of International Conference on Artificial and Computational Intelligence for Decision, Control and Automation In Engineering and Industrial Applications. Monastir, Tunisia.

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ⊙ (24*7 Support on Whatsapp)