STUDY OF VARIOUS ANAPHORA RESOLUTION RELATED ISSUES AND CHALLENGES FOR MARATHI LANGUAGE

Manoj N.Behere [1] *Nita V. Patil*[2] *AjayS.Patil*[3]

[1]Assistant Professor, RCPETS Institute of Management Research and Development, Shirpur
[2] Associate Professor, SOCS, KBCNMU, Jalgaon
[3]Professor, SOCS ,KBCNMU, Jalgaon

1. Introduction

Anaphora resolution is required in various NLP applications such as Machine translation. Information extraction, summarization. Marathi uses many morphological processes to join words together, forming compounds. Marathi language is morphologically rich and relatively free order and syntactic subject-object positions are not always able to explain the varied linguistic phenomena. Pragmatic knowledge is very important for design and development of computational model of Anaphora Resolution System for Marathi Language, because pragmatic level is deals with utilization of contextover the contents of the text for analyzing extra meanings. This requires knowledge, including the understanding of intentions, plans, and goals. Issues and challenges related to anaphora resolution for Marathi language must be considered for design and development of computational model of Anaphora Resolution System for Marathi Language.

2. Literature Survey:

Mehala K. et al[2] presented review paper which focused on event anaphora resolution and its approaches for hindi text. Also elaborated issues and challenges for anaphora resolution. Lakhmani P*etal*.[3] presented review paper which focus on pronominal anaphora resolution for hindi language. Experiment is conducted to determine constraints contribution in which gender and number agreement makes small improvement and animistic knowledge makes significant improvement 71% approx. in accuracy for recency as baseline criteria. Singh P etal[4]. demonstrated analysis of anaphora, cataphora and exaphora for demonstrative Pronoun of hindi using direction of reference as feature and concluded that anaphoric. cataphoric and exaphoric references are 86% , 11%, and 4% for monologue and 87% 4% 8% for dialogue corpus Singh P etal[5] presented paper to automate process of tagging and dealing with semantic information validation .Experiment concluded precession 74% , recall 71%and F-measure 72%. Bharati A. etal[7] elaborated natural language processing with paninian perspective. Umale Y.[8] presented dependency framework for marathi parser. Tidake C. etal[9] presented research paper on Inflection Rules for English to Marathi Translation . MujadiaV.etal[10] elaborated paninian grammar based hindi dialogue anaphora aesolution. B. Uppalapuetal [11] presented research paper on pronoun resolution for hindi.

3. Investigate pragmatic knowledge of Marathi language for anaphora resolution.

NLP applications may consider inference modules. Humans usually utilize all of inference modules conveys different types of meaning. NLP systems use different levels of linguistic analysis. Pragmatic ambiguity considers any instance of two different speech acts performed by a linguistic phrase analyzed by its effect rather than semantic meaning.

Ex. 1रुची, खिडकीउघडीआहे. तिच्यापाशीपालआहे. In Ex.1 possible antecedent of तिच्यापाशी are रुची or खिडकी Depending upon the situation, it can be merely a statement or request to close the window. But as per pragmatic aspect, in Ex. 1तिच्यापाशी must refers to खिडकी . Such levels of ambiguity are not active in a easy approach.As Marathi Language is free order, it is not easy to describe the precise type and level of ambiguity.

Ex. 2 i. पार्वतीसपुजलेसुमनाने. तीदेवताआहे.
ii. पार्वतीसपुजलेसुमनाने. तीसुवासिकहोती
iii. पार्वतीसपुजलेसुमनाने. तीश्रद्धाळूआहे.
iv. पार्वतीसपुजलेसुमनाने. त्याविचारांनीदेवताप्रसन्नझाल्या.
v. पार्वतीसपुजलेसुमनाने. त्याएकाचवर्गातआहेत.

In Ex. 2 i. , ii , iii possible antecedent of ती (female, singular) are पार्वती , सुमन . But as per pragmatic aspect, In Ex. 2 i. ती must refers to पार्वती(Goddess) because of देवता. In Ex. 2 ii. ती must refers to सुमन(Flowers) because of सुवासिक . In Ex. 2 iii. ती must refers to सुमन(Girl/Women) because of श्रद्धाळू. In Ex.2 iv , v possible antecedent of त्या(female, Plural) are पार्वती , सुमन. But as per pragmatic aspect, In Ex. 2 iv. त्या must refers to सुमन(Good Mind) because of विचारांनी . In Ex. 2 v. त्या must refers to both पार्वती and सुमन(Girls/Women) because of एकाचवर्गात.In Ex. 2 all five sentences are vague as well as the sub–classificationof verb form पुजले and for noun phrase सुमनानेis also ambiguous. This is the example of grammatical ambiguity , lexical ambiguity, and ambiguity at word level and at sentence level.Ex. 3 घ्याझालीपूजायाघराची.In Marathi language, the word पूजाhas meaningsi. given name of a female ii. worship iii. plural imperative form of verb पुजणे. Generally ambiguity inth lexical level is resolute at the sentence level. But disambiguation process of the sentence in Ex. 3 is difficult. Syntactic items, namely, घराचीपूजाहोणे (worship of a house) and पूजाघराचीहोणे (Puja becomes member of family) in this linguistic phrase, there are three ambiguous entities, a lexeme as पूजा and a phrase as घराचीहोणे and morpheme as ची . Mainly ambiguous morpheme in Marathi is ची.Varioustacticsrequired to disambiguate various types of linguistic entities as there are various levels of complication.

4. Study of various anaphora resolution related issues and challenges for Marathi language

There are certain issues and challenges which are needed to be considered while performing anaphora resolution.

4.1 Encoding in standard form: Greatquantity of information is existing in Marathi on www (on electronic document form). But this information is encoded in various fonts. There is difficulty in encoding the document in a few standard form. Resolving anaphora in Marathi is a difficultjob. Unicode may be a solution to this problem of standardization.

4.2 Requirement of Unicode based tools for Marathi: The difficulty with Unicode based font is that Unicode based tools may not sustain Marathi. This lack of standardization restricts the utilize of documents in development of corpus. Existing language processing tools are either not complete or restricted to definite domain only.

4.3 Pronoun Inflection: Pronoun forms are inflected for case (विभक्ती) in Marathi language. Cases in Marathi language indicate relation of pronoun or noun with neighbouring word. Table 1 shows Marathi Language Case ending (विभक्ती) for pronoun suffix**.** Also in Marathi language, possessive reflexive pronouns are inflected with some Marathi words like कडे, बाजूला, कडून,तर्फे, जवळ, पासून,

मुळे, पाशी, शी, द्वारे, समोर, मागे, पुढे, संदर्भात, विषयी etc. These inflected words are mostly Subject, sometime object in the sentence

4.4 Study and investigation of grammatical model : Every language consists of anaphoric expressions in its discourse where anaphora belongs to anyone of the categories like NominalAnaphora, Pronominal Anaphora, Lexical Noun phrase Anaphora, Zero Anaphora.

Study and investigation of grammatical model is essential for understanding linguistic phenomenon of Marathi language and design and development computational model of Anaphora Resolution System for Marathi Language.

By considering issues and challenges, it is essential to investigate morphological features agreement using dependency grammar formalism of Marathi Language for Anaphora ResolutionSystem. Computational Paninian Grammar suits well for morphological rich Indian languages, so we choose dependency grammar formalism for computational work of Anaphora Resolution System for Marathi Language. For Anaphora resolution morphological and lexical information about the language can be obtained by specifying formal grammar rules or by maintain a dictionary of lexicons Morphological agreement is means for sinking ambiguity by eliminating noun phrases which are not agree with pronoun from the list of possible antecedents .For Morphological agreement, study of morphological/lexical, syntactic, semantic, pragmatic knowledge of Marathi language is necessary for Anaphora resolution. We studied syntactic and semantic knowledge of Marathi language using dependency grammar approach of Computational Panini grammar The dependency grammar formalism captures the straightword level relation in the phrase. Marathi has six *karaka*, nominative, accusative, instrumental, dative, ablative and location. Ergative subject occurs with *ne* or *ni postposition* in Marathi Ergative subject does not show agreement feature with verb. As per the dependency guidelines, we distinct them as *k1, k2, k3, k4, k5* and *k7*. After studying dependency grammar approach of Computational Panini grammar, we find that case (*karaka*) *k1, k2, k3, k4, k5* and *k7* shows a direct relation between nouns to verb. In Anaphora resolution for Marathi language, mostly Pronoun act as Anaphor and Noun act as Antecedent, So by dependency grammar formalism, we analyze sentences which includes reflexive, locative, relative, personal pronouns as Anaphor. We found that, correct antecedent can be find for these anaphors like reflexive, locative, relative, personal pronouns. We consider some examples sentences. By analyzing these sentences we find newly morphological agreement.

4.5 Marathi **language Corpus Annotation:**

Corpus annotation is technique of accumulationof interpretative linguistic information to a corpus. Most common type of annotation is the accumulation of tags, or labels, representing the word class to which words in a text fit in. If a word in a text is spelt *present*, it might be a noun (= 'gift'), a verb (= 'give someone a present') or an adjective (= 'not absent'). The meaning of these same-looking words is dissimilar, and also there is a distinction of pronunciation, since the verb *present* has stress on the final syllable. These three words may be annotated as follows:

*present*_NN1 (singular Common Noun)
*present*_VVB (base form of a lexical verb)
*present*_JJ (general adjective)

Consider Marathi Example, घ्याझालीपूजायाघराची. In this sentence पूजा(Girl/Woman) is Noun and पूजा(Worship) is General Adjective .

So can be written as

पूजा_NN1(singular Noun)

पूजा_JJ(general adjective)

POS tagging is the process of annotating a word in a text/corpus which corresponds to a particular POS. The annotation is carried out on its definition and its context i.e., its relationship with adjacent and related words in a phrase, sentence, or paragraph. POS-tagged versions of major English language corpora such as the Brown Corpus, the LOB Corpus and the British National Corpus have been circulated widely in the world . Apart from part-of-speech (POS) tagging, there are other types of annotation, corresponding to different levels of linguistic analysis of a corpus or text

1. Phonetic annotation
2. Semantic annotation
3. Pragmatic annotation
4. Discourse annotation
5. Stylistic annotation
6. Lexical annotation

4.6 Marathi language and HPSG

Meaning of syntactic structures is not considered by Chomsky Binding conditions, but for the purpose of anaphora resolution in Marathi language both syntactic and semantic information is equally important, Head Phrase structure grammar (HPSG) allows not only representing syntactic as well as semantic information within the grammar structure.HPSG entails a constraint based grammar which does not use derivational transformation of one grammatical structure into other, but allows any grammatical structure to be well formed provided all the constraints imposed by the grammar are satisfied .HPSG formalism is well suitable for free word order language like Marathi where sentence is based on the notion of phase structure built around the concept of a lexical head. The basic unit of information for a word or lexicon is represented as a head. With each lexicon, information is represented in terms of features, dependency relation and semantic information as well. The formalism adopted by HPSG helps in projecting noun phrase from noun and sentences from verbs and is cable of dependency structure.

The lexical entries are maintained within multiple inheritance hierarchy, which leads to efficient organization of lexicons

There are two assumptions in HPSG

- Languages are basically organization of linguistic objects at a diversity of levels of abstraction not just compilation of sentences
- Grammars are symbolize as process neutral systems of declarative constraints. grammars consist of an inherence hierarchy.

HPSG theory maintains three principals as proposed by Chomsky. These principals are as follows

1. A locally anaphor must be locally o-bound
2. A personal pronoun must be locally o-free
3. A non-pronoun must be o-free

Example 4.

तनयनेजयलाआपलेपुस्तकदिले.

तनयनेजयलात्याचेपुस्तकदिले.

तनयनेजयलाआपलेस्वतःचेपुस्तकदिले.

तनयनेजयलात्याचेस्वतःचेपुस्तकदिले.

After analysis of above sentences, it is essential to use of HPSG formalism as it led to semantic aspects.

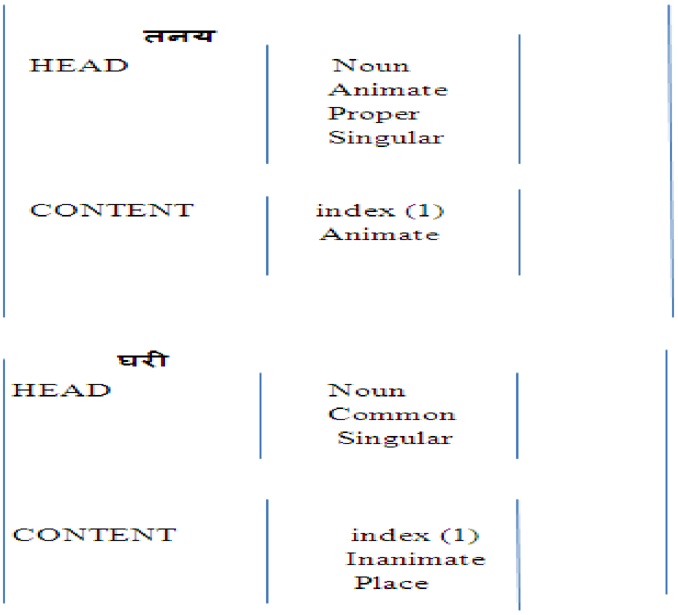
Example 5
तनयनेजयलास्वतःपुस्तकदिले.
जयलातनयनेस्वतःपुस्तकदिले.
In the above example 2 irrespective of the position of the subject “**तनयने**” , anaphor “**स्वतः**“ bounds to subject “**तनयने**” only.
By studying such sentences we must include following formalism in HPSG

- An o-subject is an entity that is first on some ARG-ST list
- o-subject oriented anaphor must be o-bound by an o-subject

Also for development of HPSG specification for Marathi language we must have following information.

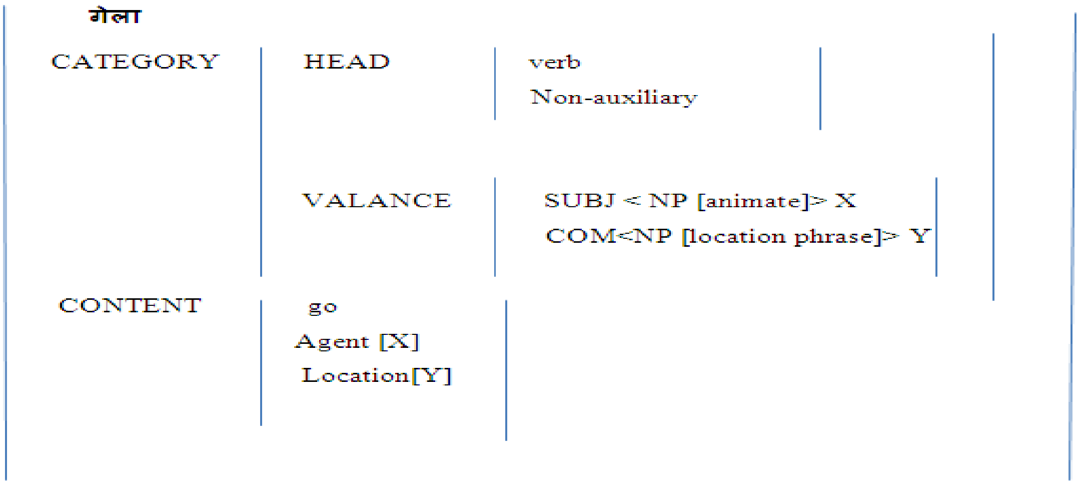
1. Lexicon containing the atomic objects and their properties like different types of pronouns
2. All verbs must be stored in the lexicon with information about the number and type of arguments required by them
3. All semantic actions and events must be listed with the objects.

HPSG representation of a single sentence is demonstrated in the example 6
Example 6 **तनयघरीगेला**



तनय	
HEAD	Noun Animate Proper Singular
CONTENT	index (1) Animate

घरी	
HEAD	Noun Common Singular
CONTENT	index (1) Inanimate Place



5. Conclusion

The aim of this research work is to investigate pragmatic knowledge of Marathi language for anaphora resolution. Issues and challenges related to anaphora resolution for Marathi language are analyzed.Marathi language corpus annotation and HPSG formalism analyzed for anaphora resolution in Marathi language

6 ResearchRelated References

1. Daya Shankar Yadav , Kamlesh Dutta , Pardeep Singh(2016) , and Preetika Chandel Anaphora Resolution for Indian Languages: The State of the Art , Proceedings of the National Conference on Recent Innovations in Science and Engineering (RISE-2016) CPUH-Research Journal: 2016, 1(2), 01-07 ISSN (Online): 2455-6076
2. Komal Mehla, Karambir and Ajay Jangra, 2015. Event Anaphora Resolution in Natural Language Processing for Hindi text, IJISET - *International Journal of Innovative Science, Engineering & Technology*, 2.
3. Lakhmani p. and Smita S., 2013. Anaphora Resolution in Hindi Language, *International Journal of Information and Computation Technology*. 3(7), 609-616.
4. Singh P., and Dutta K., 2015. Analysis of Anaphora, Cataphora and Exaphora for Demonstrative Pronoun of Hindi, International Journal of Advanced Research in Computer Science and Software Engineering, 5(7), 108-112.
5. Singh P., and Dutta K., 2015. Automation and Validation of Annotation for Hindi Anaphora Resolution, International Journal of Advanced Computer Science and Applications (IJACSA), 6, (10).
6. Rajashri Pandharipande, "Marathi", Rutledge Publications ISBN – 0-415-00319-9 London.
7. Akshar Bharati, Vineet Chaitanya, Rajeev Sangal "Natural Languages Processing A Paninian Perspective" Prentice-Hall of India New Delhi ISBN 81-203-0921-9,.
8. Yogesh Vijay Umale "Dependency Framework for Marathi Parser" Language in India www.languageinindia.com ISSN 1930-2940 16:1 January 2016.
9. Charugatra Tidke, Shital Binayakya, ShivaniPatil, RekhaSugandhi "Inflection Rules for English to Marathi Translation" IJCSMC, Vol. 2, Issue. 4, April 2013, pg.7 – 18

10. Vandan Mujadia, Darshan Agarwal, Radhika Mamidi, DiptiMisra Sharma “Paninian Grammar Based Hindi Dialogue Anaphora Resolution” Language Technology Research Centre, IIIT – Hyderabad
11. B. Uppalapu and D. M. Sharma, “Pronoun resolution for Hindi.” in DAARC-7, 2009.