

RESOLVING ANAPHORA IN MARATHI LANGUAGE

KALPANA B. KHANDALE

Research Fellow, Dept. of Computer Science & IT Dr. B. A. M. University, Aurangabad, India.
Email: kalpanakhandale1788@gmail.com

C. NAMRATA MAHENDER

Assistant Professor, Dept. of Computer Science & IT Dr. B. A. M. University Aurangabad, India.
Email: cnamrata.csit@bamu.ac.in

Abstract

NLP is a sub-domain of artificial intelligence (AI) that focuses on how computers interact with human languages. It is a research field that focuses on the interaction between human language and computers. Anaphora resolution (AR) is one of the most exciting areas of research in natural language processing. Anaphora is a common occurrence in both written and spoken texts. It is evident that the text is not just the sentence strings; there could be numerous utterances with a variety of nouns and pronouns, each with its own set of references. Anaphora is resolved in a variety of foreign languages, including English, Spanish, and French, as well as in Indian languages like Hindi, Tamil, Telugu, Kannada, and others, but there has been very little work done in Marathi. Marathi is very difficult language due to its structure. One word has multiple meaning, for finding the correct tag of each word we have developed our own Marathi POS tagger to resolve anaphora because the existing POS tagger failed to resolve anaphora in Marathi text when tested with our corpus. We generated a corpus for the Marathi text because the standard corpus for the Marathi anaphora resolution was unavailable. This paper presents the rule based approach for the resolution of anaphora in Marathi language on the basis of Marathi transformational grammar. We also concentrate on the binding theory for the resolution of anaphora in Marathi. This system aids in the interpretation of Marathi text and attempts to resolve anaphora from the given text.

Keywords: Anaphora Resolution, Antecedents, NLP, Binding Theory, Discourse Anaphora, Gender and Number Agreement, Marathi

INTRODUCTION

Humans have a remarkable ability to communicate with each other. The human language system is a sophisticated and incredibly potent means of communication. It not only allows users to signal, but it also allows them to convey an infinite variety of new meanings.

People, on the other hand, take this skill for granted. When we watch youngsters learn to speak, or when the system falls down in some way, we realize how complicated it all is. For example, we may miscommunication with someone owing to misunderstandings, or we may deal with someone who is a linguistic handicap as a result of birth, accident, or disease. When we try to connect with someone who grew up in a culture that speaks a language other than our own, we appreciate the complexities of language even more.

What is the significance of an anaphora problem to a computational linguistic?

It can help to interpret human language in a way that is valuable to humans as well able to understand multiple meaning of a sentence. When we correctly grasp what someone has said, it means we've identified the entities the speaker was referring to, as well as the relationships

between them, in a specific context. We can then act depending on our newfound understanding. The activity could be as basic as remembering a fact or as sophisticated drawing a sequence of conclusions that lead to a life-threatening circumstance being encountered. At the very least, a computational system must recognize the entities in a discourse and construct the links between those entities in a context.

The word "anaphora" is derived from the Greek word "to carry back." (Dawit, T., et.al. 2014) An anaphora is a language expression that cannot be understood by itself, according to computational linguistics. Its meaning, on the other hand, is determined by determining the meaning of another phrase in the discourse context. The antecedent is the second expression of an anaphora. As a result, understanding an anaphora in a sentence necessitates determining whatever statement in the preceding context, it relies on for meaning, and then interpreting that expression. Anaphora resolution is the process of determining the antecedent for an anaphora in the discourse or sentence.

For example, "Ram is going to school. He is a very clever boy."

Where, 'Ram' is an antecedent. 'He' is anaphora which is referring to an antecedent.

An antecedent and an anaphora present over there is called the discourse. The process of determining an anaphora to a correct antecedent is called Anaphora Resolution. As human being we can easily understand the 'He' is an anaphora and which is referring to an antecedent 'Ram', but for the machine it is a very difficult task. It is observed in our study that compare to English language very less work is done on anaphora resolution in Indian languages. The goal of this paper is to address numerous anaphora resolution challenges in Marathi and to investigate linguistic factors that aid in the resolution process.

Challenge: Who does each pronoun refer to?

Ex1: The fruits were given to the children because they were hungry.

Ex2: The fruits were given to the children because they were ripe.

Ex3: The fruits were given to the children because they were there.

In the examples, who does each 'they' are referring to? The overall meaning of the statement changes due to the diverse interpretations of "they" in each sentence.

In Ex1 – 'they' refer to 'children', Ex2- 'they' refer to 'fruits' and in EX3 – 'they' refer to 'children'. A human can easily understand this, but a machine cannot.

Anaphora appears frequently in both written and spoken words. The basic goal of natural language processing applications is to resolve anaphora, yet no theory or approach exists that can resolve all anaphora making it one of the challenging task. Anaphora resolution is required in almost all NLP applications.

Types of Anaphora

Anaphora is divided into subtypes based on the linguistic aspect. When a referring expression has a non-pronominal noun phrase as its antecedent, nominal anaphora occurs. Personal,

possessive, reflexive, and demonstrative pronouns are included in the pronominal anaphora. Syntactically, lexical noun-phrase anaphora takes the form of definite noun phrases, also known as definite descriptions, and proper names. The anaphoric relationship between a non-lexical act and the head noun or nominal group in a noun phrase is not to be confused with noun anaphora. A specific type of identity-of-sense anaphora is noun anaphora. The anaphoric relationship between the verb and its antecedent in the preceding clause determines the verb/adverb anaphora. The so-called zero anaphora or ellipsis is another sort of anaphora based on the form of anaphora. It's an anaphora that can't be seen. One anaphora is a sort of anaphora that refers to the antecedent before it. The pronominal anaphora is the most prevalent sort of anaphora of the ones listed above. The most challenging task in NLP is anaphora resolution, which necessitates not just knowledge, but also skill in language processing domains such as morphological, semantic, syntactic, real-world, and pragmatic importance.

Discourse Anaphora

The term "discourse" refers to how the sentences are connected. The initial initiative sentence is dependent on two or more sentences. To correctly resolve the anaphora, the researcher must first grasp how two or more sentences are dependent on each other, which requires processing at many different levels of complicated obstacles. It becomes more challenging in Marathi due to the intricate structure of the Marathi language. There are eleven different types of anaphora in discourse. The definite pronoun anaphora must refer to a noun (noun) that has previously been mentioned in the sentence. Definite Noun Phrase Anaphora is a type of anaphora in which the noun is referred to as an action-noun and refers to a verb phrase that represents an action, process, or event. One anaphora is an anaphoric noun phrase with the word one as the first word. 'Do –it' is a phrase that means 'do it.' The verb phrase do is an anaphora in linguistics; it refers to the left towards its antecedent. The verb phrase anaphoric process in which the string does so refer back to an antecedent verb phrase is known as 'do so' anaphora. As a result, anaphora is utilized to prove the internal structure of the verb phrase. When the antecedent is absent from the speech, Null Complement anaphora works similarly to deep anaphora. "Sentential it" anaphora has looked into sentential anaphora, which is defined as the presence of a sentential pro-form and does not allow pragmatic control. Sluicing is the word given to an ellipsis procedure that ensures the clause containing it is isomorphic in previous discourse. In a phrase or clause, the gapping anaphora is utilized similarly to the zero anaphora; it frequently refers back to a statement that provides the required information for comprehending the gap. Last but not least, the 'Such' anaphora is a straightforward and understandable anaphora. (Houser, M. 2010), (Webber, B. L.2016), (Cyrino, S. 2004), (Ander Bois, S. 2010), (Qiang, Y., et. al. 2013)

Types of Knowledge to Resolve Anaphora

There are five types of knowledge which is required the anaphora resolution to resolve the pronouns are described in Figure 1 and explanation is given below:

Morphological and lexical knowledge

It is the process by which a word is formed. It comprises the study of the structure of word

production by combining sounds into morphemes, which are the smallest separate units of meaning. Morphological knowledge is concerned with the way morphemes are combined to form words.

Syntactic knowledge

It's the process of mixing words to make phrases, phrases to build clauses, and clauses to make sentences. Syntactic analysis is concerned with determining how words can be placed together to produce valid sentences. It also establishes the structural significance of each word in the sentence.

Semantic knowledge

It's the process of figuring out what words and sentences mean. Semantic knowledge is the study of the context-independent meaning of sentences, regardless of their context. It's tough to define the meaning of a sentence because of the ambiguities.

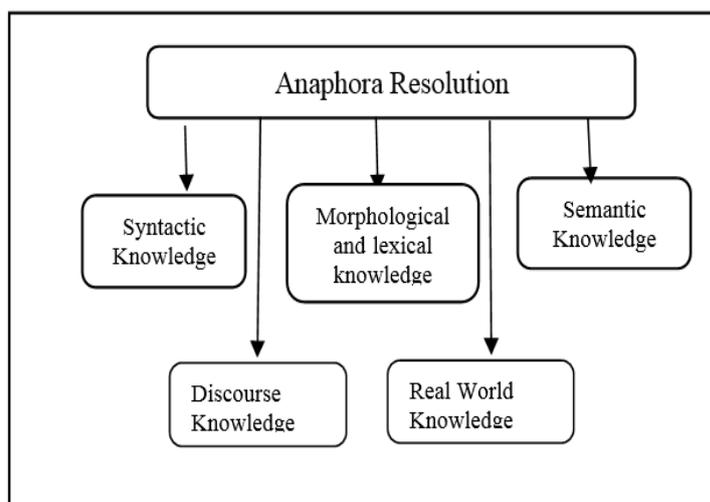
Discourse knowledge

It's the technique of identifying linked sentences that are longer than a single sentence. It is concerned with the intersentential and intrasentential connections that exist inside or between sentences.

Real-world knowledge

It is nothing more than common information about the world that all speakers share. It contains general knowledge about the world's structure as well as what each language user needs to know about the views and goals of the other. This greatly improves the language comprehension.

Figure 1: Required knowledge to resolve an anaphora



LITERATURE SURVEY

AR has been actively discussed and frequently written in English and Indian languages over the last three decades. For a full study of Anaphora Resolution (AR), a few of them have been referred to in this section.

In English, (Yu, J., Moosavi et.al.2021) the authors in their paper presented the challenge of resolving split-antecedent anaphora which is more difficult than resolving single-antecedent anaphora. The semantic relationship between each individual antecedent and the anaphora is element-of rather than identity, and the number of antecedents might vary. They had used the different types of evaluation measures like MUC, B3, CEAF-4, F1 score, LEA score. For anaphora recognition, full resolution (lenient and strict), and full resolution (lenient and strict), their models improved by up to 19%, 19.9%, and 14.7%, respectively, over baselines.

Kim et.al. (2021) presented various types of techniques to resolve an anaphora in Korean language like bidirectional encoder representation from transformers (BERT), FFNN feed-forward network, machine- learning SVM, and sequence-to-sequence deep-learning model. Their model greatly improved as a consequence of the experiments. The performance of the model in comparison to other models. Their model requires extremely minimal feature engineering for training and inference. Because BERT's pre-training is unsupervised, increasing the amount of unlabeled data could improve performance even further.

In Dutch language anaphora resolution by (Allein, L., et.al., 2020) they have used tokenized and parsed using the Dutch version of Tree Tagger, Bidirectional Long-Short Term Memory (BiLSTM), Recurrent Neural Network (RNN), Word2Vec Skip-gram model, Binary Cross-Entropy methods and they have achieved. The binary classification model achieves a promising accuracy of 84.56 %. The multitask classification model with a sentence and context encoder scores best on all assessment metrics, with an accuracy of 87.78%.

On the other hand, in Uyghur language anaphora resolution by (Yang, Q., 2020) with - Long Short-Term Memory (LSTM), independently recurrent neural network (IndRNN), capsule network and the result achieved by their four models is 96%. The authors faced lots of issues regarding the language.

The author has written a review of AR for Indian languages in the Indian language (Yadav et al. 2016). The authors divided AR techniques into four groups: rule-based, corpus-based, knowledge-poor, and discourse-based. (Mahato et al. 2019) published a survey of AR techniques and numerous aspects employed in these approaches in Hindi. The computational behavior of AR methods was also examined by the author.

On the other hand, for Arabic they have framed rules depend on the morphological, lexical, heuristic, syntactic, and the positional constraints. They have used the Arabic statistical tagger (ASPOST) and they have made the corpus from Quran of 6236 sentences and the system obtained the result 84.43%. (Abolohom, A., & Omar, N. 2017)

In Malayalam language the authors have framed rules for parsing the sentence and the system named VASISTH anaphora resolution developed for the morphological richness. The overall

result of this rule based system of Malayalam shows the success rate 82%. (Sobha, L., & Patnaik, B. N. 2002)

As mentioned above, an anaphora has been resolved various Foreign as well as Indian languages but for Marathi language an anaphora has not resolved up to the mark. Hence, we have to try to resolve anaphora in Marathi language using the rule based approach.

PROPOSED METHODOLOGY

The Indo-Aryan language family includes Marathi, which is spoken in India. Marathi is India's third most widely spoken language. Marathi is one of the oldest regional languages. There are 42 varieties of Marathi, the most notable of which are Ahirani, Khandeshi, Varhadi/Vaidarbhi, ZadiBoli, Konkani/Malvani, and Tanjour Marathi. Marathi is spoken fluently by around 90 million people around the world. The syntax and grammar of Marathi are derived from Prakrit and Pali. Marathi has a variable word order and a diversified morphology. The default word order in Marathi is Subject-Object-Verb (SOV). We'd like to investigate linguistic features that can be used to resolve anaphora, notably dependency structures as a source of syntactic and semantic information. Our studies revealed that promising performance can be achieved using simple rules based on dependence structure and agreement properties. The following categories can be used to categorize Marathi words (Naam, Sarvnam, Kriyapad, Visheshan, Shabdyogi Avyay, Kriya Visheshan Avyay, Ubhayanvayi Avyay, and Kevalprayogi Avyay).

There are six types of pronouns in Marathi such as personal मी, आम्ही, तुम्ही (Mee, Amhi, Tumhi), possessive माझा, माझी, तुझा (Maaza, Maazi, Tuza), demonstrative तो, ती, ते (To, Ti, Te), reflexive आपण, आम्ही, तुम्ही, तुम्हाला, स्वतः (Apan, Amhi, Tumhi, Tumhala, Swataah), reciprocal एकमेकांचा, एकमेकाला, आपल्याला (Ekmekancha, Ekmekala, Aaplyala) and interrogative काय, कसे, कोण, का (Kay, Kase, Kon, Ka) pronouns respectively.

Rule Based Approach

Rule-based approaches can make considerable use of language knowledge, such as syntactic rules. The proper antecedent is a crucial consideration when assigning a core pronoun anaphora for example, is the distance between anaphora and antecedent. As a result, the options for antecedents that are closer to the anaphora are more likely to be the true antecedent. Another factor is that when determining what is likely to be an antecedent, nouns are preferred over whole sentences. When Computational Linguistic is concerned, writing, grammar production for a language is a bit difficult because of different gender and number forms. The rule framed on the basis of the Marathi grammar. (Kulkarni, D. 2014).

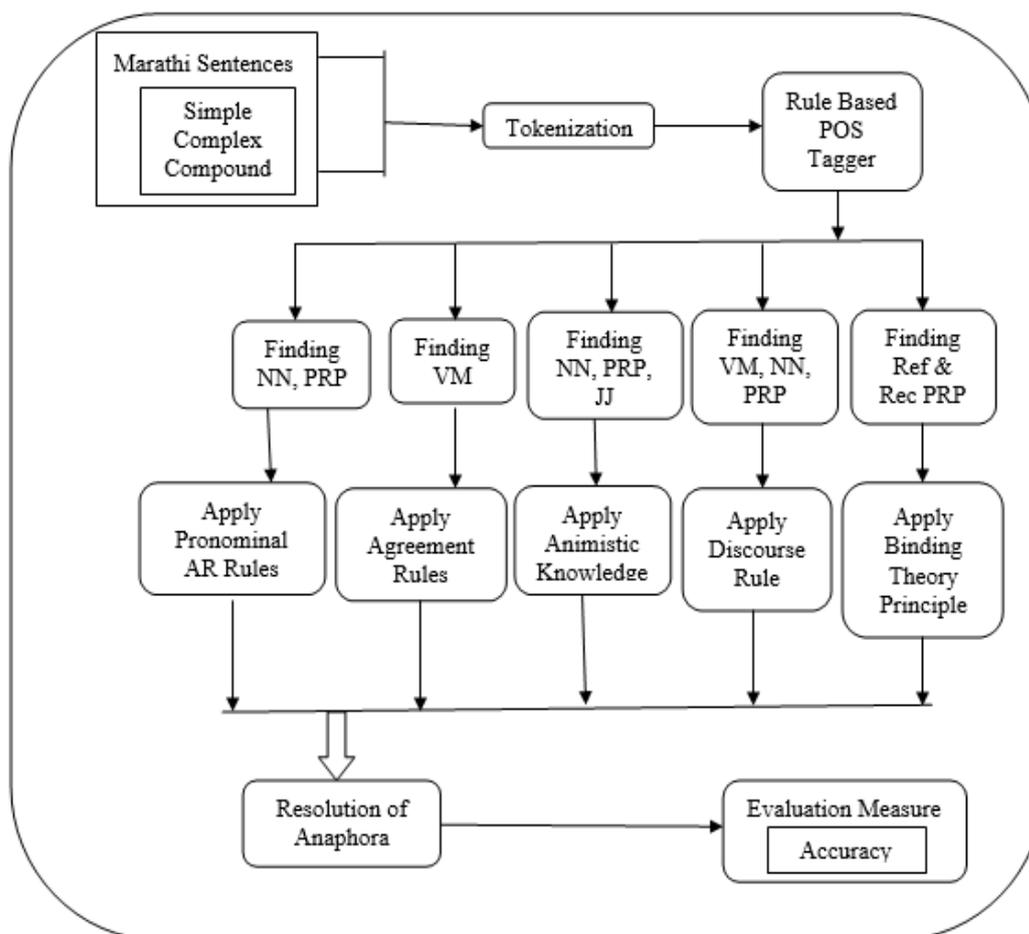
Hobb's algorithm uses noun phrases as antecedents to resolve personal and possessive pronouns. Hobbs' method additionally, takes into account split antecedents, coordinated noun phrases, and noun phrases as antecedents that must be reconstructed first if they are left out by verb phrases with do. They worked on the English language, which has a sentence structure (SVO), but the sentence structure in Marathi is far more intricate (SOV). In contrast to Hobbs' algorithm, our system takes into account all subtypes of central pronouns, including reflexive

and reciprocal pronouns.

Need of Anaphora Resolution in Marathi

Anaphora resolution is used in nearly every aspect of natural language processing. Anaphora resolution mechanisms exist in other Indian languages as well, such as Hindi, Tamil, Kannada, and many others, but the tools and methods for resolving an anaphora have not met the standards. Marathi is a morphologically rich and difficult to grasp language. As Marathi speakers, we can readily comprehend the meaning of the text; nevertheless, deciphering the meaning of the Marathi language for a machine is a challenging task. It means that, regardless of the usual format, the order in which sentences are formed is determined by the user's needs. Below figure shows the architecture of the Marathi Anaphora Resolution System is depicted in the diagram below (MARS).

Figure 2: Architecture of Marathi Anaphora Resolution System (MARS)



Input Text

There is no ready standard corpus available for Marathi thus we have developed our corpus

with 1350 sentences manually created which includes anaphora and antecedents. We have also taken 1000 discourses as input from 1st to 8th standard chapters from Marathi Balbharati Textbook for the discourse anaphora. Below table 1 depicts the types of sentences and example of each type.

Table 1: Types of Marathi sentence with example

Types of sentences	Example
Simple Sentences (One antecedent - one anaphora)	राम शाळेत जातो तो हुशार आहे. (Ram Shalet Jato To Hushar Ahe.)
Compound Sentences (Sentence is connected with conjunction “आणि”)	राधाने स्पर्धेत भाग घेतला आणि ती जिंकली. (Radhane Spardhet Bhag Ghetla Ani Ti Jinkali.)
Complex Sentences (Two same gender antecedents and only one anaphora)	रश्मीने राणीला बाजारात नेलं आणि ती खूप थकली. (Rashmine Ranila Bajarat Nele Ani Ti Khup Thakli.)
Discourse Sentences (Multiple sentences are linked together and multiple anaphora present over there)	राहुलने विजयला भेटायला बोलावले पण तो घरी नव्हता, म्हणून तो खूप नाराज झाला. त्याला त्याच्याकडे महत्वाचे काम होते. (Rahulne Vijayla Bhetayla Bolavle Pan To Ghari Navhta, Mahnun To Khup Naraj Zala. Tyala Tyachyakade Mahtwache Kam Hote.)

Tokenization

The processes of splitting each word from the input words have separated by the white space and the punctuation marks. By using this we can easily find out the tokens of the sentence.

Combined Rule Based Tagger

With the use of a rule-based approach, we designed a POS tagger that would assign POS to the word in a sentence provided as input to the POS tagger system. Marathi is a free-ordered language with a verb ending structure. We used the combination of the unigram and bigram tagger to train the data in order to develop a part-of-speech tagger for Marathi. As a result of combining both taggers, the result is superior. As specific corpus of Marathi text or dataset is not available, 2000+ sentences were manually created, with many variations to sense the difficulties while generating tags using the proposed POS tagger.

Extract Proper Noun (NNP), Pronoun(PRP) and Verb(VM)

In this we have extracted the proper noun, the pronoun and verb from the POS tagger for the resolution of pronominal discourse anaphora. Below table 2 shows the rules which are based on the Marathi transformational grammar which we have developed for the resolution of pronominal and discourse anaphora.

Table 2: Developed rules to resolve pronominal and discourse anaphora

Types of Anaphora	Rules to resolve anaphora in Marathi		Example
Pronominal Anaphora	Simple Sentence	<p>If PRP “तो”, ”ती”, ”ते” present in second sentence: Then pronoun refers to the proper antecedent Else pronoun is not found properly.</p>	<p>राम शाळेत जातो,तो हुशार आहे. (Ram Shalet Jato To Hushar Ahe.)</p>
	Complex Sentence	<p>If PRP “तो”, ”ती”, ”ते” present in second sentence: Then match successive word of proper noun (NNP) is noun (NN) and successive word of pronoun (PRP) is an adjective (JJ) If match found: Then an anaphora refers to the noun (NN) Else anaphora is not resolved properly</p>	<p>राधाने चकली खाल्ली पण ती तिखट होती. (Radhane Chakli Khalli Pan Ti Tikhat Hoti.)</p>
Discourse Anaphora	Possessive pronoun	<p>If the discourse is about three sentences: Then we found other sentences are linked together If (.) full stop present after the VM: Then the First word of the 2nd sentence and 3rd sentence is PRP If PRP is found in 2nd and 3rd sentence: Then back to the 1st sentence and check for the noun or proper noun and the antecedent is found Else: Pronoun is not found Anaphora resolved properly in discourse</p>	<p>आंबेडकरांचे भाषण ऐकण्यासाठी लोकांची गर्दी जमत असे .ते समाज प्रभोदन करायचे .त्यांना ऐकून लोकांच्या अंगात जणू वाघ संचारायचा. (Ambedkaranche Bhashan Aikanyasathi Lokanchi Gardi Jamat Ase. Te Samaj Prabodhan Karayche. Tyanna Aikun Lokanchya Angat Janu Vagh Sancharaycha.)</p>
	First Person Possessive and Third Person Plural Possessive Pronoun	<p>If the discourse is of two or three sentences: Then we found other sentences are linked together If PRP in between sentence and (.) full stop present after the VM: Then the First word of the 2nd sentence is PRP If ‘त्यांचा/त्यांची/त्यांचे’ is found in 2nd sentence: Then 3rd Person Possessive pronoun refers to antecedent NN or NNP Else: PRP is 1st Person Possessive Pronoun is not found Anaphora resolved properly as Possessive pronoun</p>	<p>सुजाताला भेटावे म्हणून तिच्याकडे गेले .पण ती गावाला गेली होती .तिचं भेट झाली नाही म्हणून मला वाईट वाटले. (Sujatala Bhetave Mhanun Tichyakade Gele Pan Ti Gavala Geli Hoti. Tichi Bhet Zali Nahi Mhanun Mala Vait Vatale.)</p>

--	--	--	--

Extract Verb (VM) and Extract Noun (NN), Pronoun (PRP), Adjective (JJ)

In this we have extracted the verb (VM) from the POS tagger. Below table shows the rules which we have framed to the resolution of anaphora using the gender and number agreements and we have also extracted noun (NN), pronoun (PRP) and adjective (JJ) from POS tagger to resolve an anaphora using animistic knowledge.

Table 3: Developed rules to resolve anaphora using agreements

Features /agreement	Rules to resolve anaphora in Marathi	Example
Gender Agreement	If compound or auxiliary verb ends with “ा” Then noun (NNP)and pronoun (PRP) refer as masculine. If compound or auxiliary verb ends with velanti “ी” Then noun (NNP)and pronoun (PRP) refer as feminine.	नितिन मिठाईच्या दुकानात गेला व तो रसमलाई खात होता. (Nitin Mithaichya Dukanat Gela V To Rasmalai Khat Hota.) Or अश्विनी मिठाईच्या दुकानात गेली व ती रसमलाई खात होती . (Ashwini Mithaichya Dukanat Geli V Ti Rasmalai Khat Hoti.)
Number Agreement	If the verb ends with Matra “े” Then noun or pronoun refers as plural If the verb ends with Kana “ा” and velanti “ी” Then noun or pronoun refers as singular.	प्रथमेश आणि आकाश मित्र आहेत ते दोघे सोबत सहलीला जाणार होते. (Prathamesh Ani Akash Mitra Ahet Te Doghe Sobat Sahlila Janar Hote.) or मृणाली स्पर्धा जिंकली कारण ती हुशार होती . (Mrunali Spardha Jinkali Karan Ti Hushar Hoti.) Or राजेश स्पर्धा जिंकला कारण तो हुशार होता. (Rajesh Spardha Jinkala Karan To Hushar Hota.)
Animistic Knowledge	If match the successive words of a proper noun (NNP) is the noun (NN) and the pronoun (PRP) is the adjective (JJ) Then the pronoun refers to the noun (NN)	मनीषने समोसा खाल्ला, तो तिखट होता. (Manishne Samosa Khalla, To Tikhat Hota.)

Extract Reflexive and Reciprocal Pronoun

We assessed the reflexive and reciprocal pronouns to resolve for the Marathi text since we noticed a large number of reflexive and reciprocal pronouns in the dialogue when researching for the database, necessitating the use of a binder to resolve anaphora. “Swatha” (स्वतः) fulfills the role of a reflexive pronoun.

The long-distance reflexive “Aapan” (आपण), which must be locally free within NPs and serves as the object of prepositions that assign their own, is another option.

The goal of binding theory is to establish the conditions of reference dependency between expressions in a given language. In the below sentence the reflexive pronoun “स्वतः” referring to the antecedent “विराज”. This type of sentence is easy to resolve because the pronoun is referring to which is present over there. “विराजला स्वतःचा राग आला.” (Virajla Swathacha Raag Aala.)But the sentence like, “तिला वाटते राधा चांगली मुलगी आहे.”(Tila Vatate Radha Changli Mulgi Ahe.)The pronoun “तिला” refers to a third person (other person in context). And when we talk about the sentence like, “राधाला वाटते ती चांगली मुलगी आहे.”(Radhala Vatate Ti Changli Mulgi Ahe.)

Here in the above sentence the pronoun or anaphora may refer to a third person or to self too. Such type of sentences makes things more complex.

In English the word order is SVO and we mostly concentrate on two reflexive gender pronoun like “himself” and “herself” but in Marathi the word order is SOV and it becomes more difficult to resolve the pronoun because in Marathi one sentence may carry multiple meaning like we saw in the above few examples. Hence, in Marathi two genders have come into the picture in four ways like “तो/ती”and “तिला/त्याला”.

Binding theory is a significant area of research in inquiries into the alliance between syntax and semantics. And have shown promising results in English language. As discussed above that both English and Marathi have different word order, and Marathi have complex forms of third person referring.

Still, we try to experiment with the Binding theory, to find whether it would helpful to bind the correct anaphora to the correct antecedent. We found that it gave better results with multiple sentences in discourse and with multiple anaphora and with one or two antecedence. The performance was seen to be 60% more accurate compared to rule based approach. The results are briefed in section 4.

Evaluation Measure

Accuracy is the primarily used metrics in natural language processing. This method helps to measure the performance of the system. It is the proportion between a true text selection and the total of all selections (all true + all false).

Figure 3: Evaluation measure

$$\text{Accuracy} = \frac{\text{No. of Correct Anaphora Resolved}}{\text{No. of Correct + Incorrect Anaphora Resolved}} * 100$$

RESULT

Anaphora resolution using Gender and Number agreement

In the database, we manually created 1350 sentences with antecedents and anaphora. On the basis of the number and gender agreement, 930 sentences are resolved. There are still 420 sentences that have not been resolved. Table 5 and 6 represents the examples of resolution of anaphora using gender and number agreement.

Table 4: Overall accuracy, using gender and number agreement

No. of sentences	Resolved Anaphora Correctly	Incorrect resolution of anaphora	Accuracy
1350	930	420	68.88%

Table 5: Example of anaphora resolution using gender agreement

Correctly Resolved by Gender Agreement	मधुरा दुकानातून कपडे खरेदी करत होती, ती फार खुश होती. Madhura Dukanatun Kapde Kharedi Karat Hoti, Ti Far Khush Hoti.
Incorrectly Resolved	अनिता व नीता जिवलग मैत्रिणी आहेत, ती तिला खूप आवडते. Anita V Nita Jivlag Mairini Ahet, Ti Tila Khup Avadte

Table 6: Example of anaphora resolution using number agreement

Correctly Resolved by Number Agreement	अथर्व दुकानातून कपडे खरेदी करत होता, तो फार खुश होता. Atharv Dukanatun Kapde Kharedi Karat Hota, To Faar Khush Hota.
Incorrectly Resolved	अब्दुल कलाम राष्ट्रपती होते आणि ते खूप प्रसिद्ध संशोधकही होते. Abdul Kalam Rashtrapati Hote Ani Te Khup Prasadha Sanshodhakhi Hote.

Anaphora resolution using Binding theory

We studied 1000 discourses from the Marathi Balbharati book of standard 1st to 8th for the resolution of the anaphora utilizing the binding theory. We found 674 discourse sentences that are linked together from this.

Table 7: Overall accuracy, using binding theory

Total Discourses	Correctly Bind	Incorrectly Bind	Accuracy
674	457	217	67.80%

Table 8: Example of anaphora resolution using binding theory

Correctly Resolved by Binding Theory	रमेशला वाटते की वर्षाला तो आवडतो. (Rameshla Vatate Ki Madhurala To Avadto.)
Incorrectly Resolved	सुजाता गावाला जाणार होती ; तिला बघायला मुलाकडची मंडळी येणार होती; तिला जरा भीतीच वाटत होती . (Sujata Gavala Janar Hoti; Tila Baghayla Mulakadchi Mandali Yenar Hoti; Tila Jara Bhitich Vat Hoti.)

Pronominal and discourse anaphora detected in Marathi text

For other languages like Hindi, Bengali and many more, the researcher has resolved the pronominal anaphora numerous times, but the pronominal anaphora for Marathi has yet to be resolved fully.

Total 1700 sentences or discourse was considered here to resolve anaphora. 1177 anaphora are detected in this text and correctly resolved with 733 probable antecedents correctly resolved, while he is unable to find 444 antecedents.

Table 9: Accuracy of Marathi Anaphora Resolution System

No. of sentences	Pronominal and discourse Anaphora detected	Anaphora with possible antecedents resolved	Anaphora without possible antecedents resolved	Accuracy
1708	1177	733	444	62.27%

CONCLUSION

This composition was inspired by the Marathi Anaphora Resolution. There hasn't been much progress, just like there hasn't been much advancement in Marathi. Because Marathi is a free-order language, we have had to cope with a number of linguistic structure issues.

Resolving the anaphora is one of the most challenging tasks for the researcher. The pronominal anaphora has been resolved in many languages and numerous efforts on the anaphora have

been made for Indian languages like Hindi, Bengali etc., but not for the Marathi text, in our research work we are striving to resolve the anaphora for the Marathi text. In order to resolve the anaphora, we focused on gender and number agreement, as well as animist knowledge. Concentrate on the Binding theory, which is a crucial aspect of discourse resolution. We developed the corpus which included self-made sentences, sentences from Marathi Balbharati Textbook lessons 1st to 8th class chapters. Our rules-based method to anaphora resolution has proven to be more effective than previous ways in resolving anaphora in Marathi literature. An antecedent and an anaphora pair from the sentence were discovered as part of the overall performance of the anaphora resolution system for Marathi text. The overall score is 68.88% in terms of gender and number agreement. In addition, we focus on the binding theory for resolving anaphora in Marathi text, and the system's accuracy is 67.80%. In comparison to other approaches for anaphora resolution, our rules-based approach has shown to be more effective in resolving anaphora in Marathi text.

For pronominal and discourse anaphora, the total performance of the anaphora resolution method for Marathi text was 62.27 %. In future we have to concentrate on the resolution of anaphora in paragraph in Marathi language.

Acknowledgement

We would like to thanks to the NLP research laboratory for supporting this work. Also thanks to Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India.

References

- Houser, M. The syntax and semantics of do so anaphora, 2010
- Webber, B. L. (Ed.). A formal approach to discourse anaphora Routledge, 2016
- Houser, M. J. On the anaphoric status of do so, URL http://linguistics.berkeley.edu/~mhouser/Papers/do_so_status.pdf. 2010
- Cyrino, S. M. L. "Null complement anaphora and null objects in Brazilian Portuguese. Buenos Aires: Coloquio de Morfosintaxis." Workshop on Morphosyntax. Vol. 112, 2004.
- Dawit, T., & Assabie, Y. Amharic Anaphora Resolution Using Knowledge-Poor Approach, In International Conference on Natural Language Processing. Springer, Cham. September. 2014, pp. 278-289.
- AnderBois, S. Sluicing as anaphora to issues. In Semantics and linguistic theory. August. 2010, Vol. 20, pp. 451-470.
- Qiang, Y., Ma, R., & Wei, W. Discourse Anaphora Resolution Strategy Based on Syntactic and Semantic Analysis, Information Technology Journal, 2013,12(24), pp.8204-8211.
- Yu, J., Moosavi, N. S., Paun, S., & Poesio, M. Stay Together: A System for Single and Split-antecedent Anaphora Resolution., arXiv preprint arXiv:2104.05320.2021
- Kim, Y., Ra, D., & Lim, S. Zero-anaphora resolution in Korean based on deep language representation model: BERT. ETRI Journal, 2021, 43(2), 299-312.

- Allein, L., Leeuwenberg, A., & Moens, M. F. Binary and multitask classification model for Dutch anaphora Resolution. Die/dat prediction. arXiv preprint arXiv:2001.02943.2020.
- Yang, Q., Yu, L., Tian, S., & Song, J. Multi-attention-based capsule network for Uyghur personal pronouns Resolution. IEEE Access, 8, 76832-76840.2020.
- Yadav DS, Dutta K, Singh P, Chandel P. Anaphora resolution for indian languages: the state of the art. Recent Innovations in Science and Engineering. 2016. 1(2):01–07
- Mahato S, Thomas A, Sahu N. A relative study of factors and approaches for Hindi Anaphora resolution. Int J Manag IT Eng., 2019. 7(12):176–188
- Abolohom, A., & Omar, N. A Computational Model for Resolving Arabic Anaphora using Linguistic Criteria. Indian Journal of Science and Technology, 2017. 10(3).
- Sobha, L., & Patnaik, B. N. Vasisth: An anaphora resolution system for Malayalam and Hindi. In Symposium on Translation Support Systems. March. 2002.
- Kulkarni, D. Specifying Context Free Grammar for Marathi Sentences. International Journal of Computer Applications, 99(2014), 38-41.