Rule Based Anaphora Resolution for ISL Generation System

Thesis submitted in partial fulfillment of the requirements
for the award of degree of

Master of Engineering

in

Computer Science and Engineering

Submitted By

Deepakshi Singla
(801532010)

Under the supervision of:

Dr. Parteek Kumar
Associate Professor

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

THAPAR UNIVERSITY

PATIALA - 147004

July 2017

Certificate

I hereby certify that the matter which is being presented in this thesis entitled as, "***Rule Based Anaphora Resolution for ISL Generation System***", in partial fulfillment of the requirements for the award of the degree of Master of Engineering in ***Computer Science and Engineering*** submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of ***Dr. Parteek Kumar*** and refers other researcher's work which is duly listed in the reference section.

The matter presented in this thesis has not been submitted for the award of any other degree of this or any other university.

Deepakshi

Deepakshi Singla
801532010

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

Parteek Kumar

Dr. Parteek Kumar
Associate Professor
Computer Science & Engineering Department
Thapar University, Patiala

Acknowledgement

This research work would be incomplete without acknowledging the people who supported and guided me for the successful completion of this task.

First of all I wish to acknowledge the benevolence of God who gave me courage and strength to face the challenges and to overcome the obstacles that occurred while working on this task.

This research work is the outcome of the excellent guidance and positive attitude of my guide **Dr. Parteek Kumar**, Associate Professor, Computer Science and Engineering Department, Thapar University. I am very thankful to him for his valuable guidance, constant encouragement, keen interest, valuable co-operation, generous attitude and above all his blessings. He has been a source of inspiration and motivation for me.

I am grateful to **Dr. Maninder Singh**, Head, Computer Science and Engineering Department, Thapar University for the motivation and inspiration for the completion of this thesis.

I wish to express my gratitude to **Patiala School for Deaf Blind, Village Saifidipur**, **Patiala** and **Bhagat Puran Singh School for Deaf, Pingalwara, Amritsar** for their support and guidance for the completion of this task.

I am thankful to my parents and my friends who encouraged and supported me with their whole hearted co-operation for completing this thesis.

Deepakshi Singla
(801532010)

Abstract

Sign language (SL) is the most well-known approach of articulation of thoughts and feelings for the deaf community. For transformation with one dialect then onto the next one needs itemized learning about the two dialects. There are relatively few specialists of gesture based communication. Machine interpretation of gesture based communication is an extreme assignment because of absence of phonetically commented on and all around recorded information on Indian Sign Language because of which inquire about on Indian Sign Language etymology and phonology is restricted.

Sign language translation is used to convert a simple sentence into Indian Sign Language. While doing so a number of problems have been faced. Among them, Anaphora resolution is one such problem that comes across while converting sentence into ISL. Anaphora is a discourse level phenomenon whereby the understanding of an event of one articulation relies upon the translation of an event of another. Anaphora occurs frequently in written texts and spoken languages.

The focus of this thesis is Entity Anaphora Resolution and pronominal forms. A web based system has been developed that works for Entity Anaphora using dependency information in Hindi. In this thesis, rule based approach which characterizes rules utilizing dependency structures and relations for referent determination have been developed. Rule based approach is the least difficult approach to take care of the issue of anaphora determination. Different tenets for various sorts of pronouns have been examined in this theory. This thesis discusses about the rule set and algorithms define and tested over a corpus of 300 sentences. Each pronominal form has an algorithm to be followed to find the referent corresponding to its anaphora. The algorithms are run against the type of anaphor to resolve it. The system produces the accuracy of 60%. The system resolves the anaphor for spatial pronouns more efficiently as compared to others. The system fails to intact the grammatically and syntactically meaning of the sentence.

.

Table of Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1
Introduction

1.1 Sign Language

Sign language (SL) is the most common approach of expression of thoughts and feelings for the deaf community. SLs are gestural languages which provide information by using hands and other body parts. These languages contain symbolic encoded messages for communication among deaf people without any speech channel. Many hearing impaired people learn sign language as their first language. Sign languages are unique in many ways from other languages. They cannot be written like other spoken language, a special mechanism is required for transcribing the signs. Most of the deaf people prefer to communicate by using sign language only as they don‘t have much knowledge about other spoken languages.

1.2 Indian Sign Language

In India, Indian Sign Language is used as a main mode of communication by deaf people. It is a gestural language which uses hands, arms, face and other body parts for communication. Sign Language being a complete natural language has its own syntax and grammar. Sign Language varies from region to region. It is used as the main mode of communication among deaf people in India and also in some other South Asian countries. It is used in various parts of Indian subcontinent including some regions of Pakistan across the Indian border, Nepal, Bangladesh and Sri Lanka [1]. In India various varieties of sign language are used in different regions. The various dialects of ISL have same grammar but these may differ in terms of lexical variations.

1.3 Challenges to Sign Language

Sign Language being a complete natural language has its own syntax and grammar. Sign Language varies from region to region. For transformation from one language to another one needs to have detailed knowledge about both languages. There are not many experts of sign language. Machine translation of sign language is a tough task because of absence of semantically commented on and very much archived information on Indian Sign Language due to which research on Indian Sign Language linguistics and phonology is limited [2].

1.4 Need of ISL Grammar

Indian Sign Language (ISL) Grammar is tied to the Deaf Community. ISL grammar specifies the rules that have been framed to remove the unnecessary parts of the sentence or that are not part of ISL grammar. Like modals, articles etc. are not used in ISL grammar [4]. These rules reduce the complexity of the sentence as it is always hard for

the people who are deaf and dumb to understand the language of other people.

1.5 Need of Anaphora Resolution for Sign Language Generation System

The hard hearing people understand only simple sentences. A project for Automated Conversion of Hindi Text into Sign Language for easy communication and education of Hearing Impaired People that has been discussed in section 2.1.5 is going on at Thapar University, Patiala. The proposed system is a part of this project. The anaphora resolution for ISL generation resolves the anaphors with its referent to make a sentence simple for the hard hearing people. They are not capable of resolving the anaphoric relations. Consider example (1.1).

एक प्यासा कौआ पानी की तलाश में इधर उधर भटक रहा था। तेज़ गर्मी से उसकी प्यास बढ़ने लगी। … (1.1)

In example (1.1), *उसकी* has been resolved in essence of proper Hindi Grammar as *कौए की.* This keeps the meaning intact. But proposed system resolves anaphors but fails to intact the meaning of the sentence.

To communicate with the people who are hard hearing is difficult if you are not aware of Sign Language. To interact with them in a better way we try to simply the system as much as possible by automating Sign Language Generation System [7]. One such problem we resolve while automating is Anaphora Resolution. It is possible that in a sentence the occurrence of one expression may depend on occurrence of another expression. Refer to example (1.2) which is as follows:

रमेश कल पटियाला गया था । वहाँ बहुत गर्मी थी। … (1.2)

In example (1.2), वहाँ cannot be signed while communicating. And deaf people might get confused as they may not able to identify whether वहाँ is referring to which Noun-phrase present in a sentence. Anaphora resolution solves this problem and replaces the anaphors or pronouns with its referent. Like in example (1.2), वहाँ is an anaphor which is replaced by पटियाला as shown in example (1.3).

रमेश कल पटियाला गया था । [पटियाला] बहुत गर्मी थी। … (1.3)

1.6 Anaphora Resolution

Anaphora is a discourse level phenomenon whereby the elucidation of an event of one articulation relies upon the translation of an event of another. It is used to derive the "Correct interpretation" of the text. The referring expression is called anaphor and

referred expression is called the antecedent. The procedure of distinguishing proof of referent is known as Anaphora Resolution [29]. Most of the times anaphora resolution is taken as co-reference resolution. However, co-reference task includes anaphora resolution and there is subtle difference between both the terms. Anaphora resolution is non-transitive. Anaphora resolution in Hindi is a difficult task. Anaphora resolution is of two types such as Entity anaphora resolution and Event anaphora resolution. Entity anaphora stands for those pronominal references which refer to a Concrete Entity such as person, place and other common nouns. Thus possible candidate referents for entity anaphora are noun phrases. Event anaphora stands for those pronominal references which refer to an abstract object, thus the possible referents are verbs, clauses and propositions.

1.7 Classification of references

In computational linguistics, terms related to anaphora resolution have been used quite interchangeably. Although, a particular term may be used to include broader reference relations, the emphasis of different tasks changes with the terminology. The related terminologies and their differences have been discussed later in this chapter.

1.7.1 Co-reference Vs Anaphora

The two terms and their difference can be understudied in following sections.

1.7.1.1 Co-reference

The term co-reference is often used for all types of reference relations including anaphoric and non-anaphoric references. In most of the research, co-reference resolution task also includes anaphora resolution.

1.7.1.2 Anaphora

Anaphora is the phenomenon whereby an event of an articulation has its referent provided by an event of some other articulation in the same or another sentence.

1.7.1.3 Difference between Co-reference and Anaphora

There is subtle difference between the two terms. There is co-reference between two expressions when both of them unambiguously refer to a unique physical or conceptual entity. On the other side, a reference is called anaphoric only when an expression refers to another expression called the antecedent and this antecedent is required for the correct interpretation of the anaphor. Considering example (1.4), the reference relation between '*वह*' and '*बैंक के निर्देशक*' is anaphoric, while the relation between '*क्षेत्रीय प्रबंधक*' and '*प्रबंधक*' in second sentence is anaphoric as well as co-referential.

बैंक के निर्देशक ने क्षेत्रीय प्रबंधक के खिलाफ जाँच के आदेश दिए । वह दावा करते हैं कि प्रबंधक लोन घोटाले में शामिल थे। … (1.4)

For simplicity, we consider all pronominal references to be anaphora and non-pronominal references to be co-referential. In this thesis, focus of our research is anaphora (pronominal) reference resolution.

1.7.2 Concrete Vs Abstract references

Reference relations can be arranged on the premise of linguistic category of the expression that an anaphor is referring to. This classification includes Concrete and Abstract anaphora.

1.7.2.1 Concrete or Entity reference

In solid reference an anaphor refers to a solid (singular) element like noun phrase (individual, place and so on), quantifiers and so forth. Consider example (1.5).

बैंक के निर्देशक ने क्षेत्रीय प्रबंधक के खिलाफ जाँच के आदेश दिए । वह दावा करते हैं कि प्रबंधक लोन घोटाले में शामिल थे। … (1.5)

Since, the pronoun '*वह*' refers to a concrete entity '*बैंक के निर्देशक*', this is an example of concrete reference.

1.7.2.2 Abstract reference

Abstract reference includes the situations where an anaphor refers to an abstract object such as event, preposition or clause. In abstract reference the referent is usually large textual expression. Consider example (1.6).

नेपाल में कल उच्च तीव्रता का भूपंक आया था। जिसके कारण जान माल का बहुत नुकसान हो गया। … (1.6)

In example (1.6), '*जिसके*' refers to an abstract entity '*उच्च तीव्रता का भूपंक*'. Specifically the verb represents the event being referred. Thus, this is an example of Abstract reference.

In many languages same lexical forms of pronoun can be used to refer to Concrete as well as Abstract referents depending on the context and for much pronominal form, it is not possible to decide only on the basis of lexical forms whether an anaphor is referring to a Concrete or an Abstract referent.

1.8 Different forms of Anaphora

In this section, the categorization of anaphora in Hindi has been discussed. There are three grammatical categories in Hindi which can be anaphoric that are discussed in section 1.9.1 to 1.9.3. Out of all the three grammatical categories, only pronominal form of anaphora has been handled by the proposed system.

1.8.1 Pronominal form of Anaphora

By pronoun, we here refer to words which have a grammatical category or POS-tag as 'Pronoun' (*PRP*) in a give context. The categorization of pronominal forms has been discussed from section 1.9.1.1 to 1.9.1.4.

1.8.1.1 Reflexives

A reflexive pronoun is a pronoun that is gone before/prevailing by preceded/succeeded by noun, adjective, qualifier or pronoun to which it refers (its precursor) inside a similar provision. In sentence structure, a reflexive pronoun is an anaphor that must be bound by its predecessor.

In Hindi there are two types of reflexives.

1.8.1.1.1 Possessive reflexives

Possessive reflexives are only used in possession relations within the same clause and are different from third person possessive pronouns. They are not inflected with the gender and number of the possessor, but with that of possesses. They include अपना, अपनी, अपने. Consider example (1.7).

रवि अपने परिवार को घुमाने ले गया। … (1.7)

In example (1.7), pronoun अपने is reflexive which refers to the subject of the clause रवि and shows the possession of the परिवार to the subject.

1.8.1.1.2 Non-possessive reflexives

Non-possessive reflexives can be used in any participant position, but mostly used in object position or to shows the emphasis. They include अपने आप, स्वयं, खुद representing 'self' for different persons. Consider example (1.8).

रानी खुद ही घर का काम करती है। … (1.8)

In example (1.8), reflexive pronoun खुद refers to the subject *रानी* and is used to emphasize the subject.

The table 1.1 contains the complete list of reflexive pronoun.

Table 1.1: Forms of reflexives

Form	**Root**	**Direct**	**Oblique**
Masculine(Possesive)	अपना	अपना	अपने
Feminine(Possesive)	अपनी	अपनी	अपनी
Plural Possesive	अपने	अपने	अपनों
Non-Possesive	अपने आप	अपने आप	अपने आप
Non-Possesive	स्वयं	स्वयं	स्वयं
Non-Possesive	खुद	खुद	खुद

1.8.1.2 Relative pronouns

A relative pronoun is a pronoun that denotes a relative statement inside a bigger sentence. It is known as a relative pronoun since it relates the relative (and henceforth subordinate) condition to the noun that it alters. A relative pronoun joins two conditions into a single complex provision. It is comparable in capacity to a subordinating conjunction. Dissimilar to a conjunction, be that as it may, a relative pronoun remains in place of a noun. Consider example (1.9), in Hindi जो, and its inflected forms are used as relative pronouns.

मोहन के पास दो बैग थे जिनमें वह दुकान का ज़रूरी सामान ले जाता था। … (1.9)

'Mohan had two bags in which he used to carry important items of the shop.'

In example (1.9), the relative clause *सामान ले जाता था* relativizes the Noun phrase *बैग* and is linked to the main clause via pronoun *जिनमें* which is an inflected form of जो (jo). In case of relative pronoun, its referent is the Noun phrase itself that it relativists or modifies. Thus in above example, the referent of the pronoun *जिनमें* is the NP *बैग.*

The following table 1.2 contains the complete list of relative pronoun.

Table 1.2: Forms of relative pronouns

Type	Root	Nomina-tive	Accusa-ive	Instrument -al/Ablative	Dative	Genitive	Locati-ve
Singular	जो	जिसने	जिसे	जिससे	जिसको	जिसका, जिसकी, जिसके	जिसमें
Plural	जो	जिन्होंने	जिन्हें	जिनसे	जिनको	जिनका, जिनकी, जिनके	जिनमें

The dative case is generally used to indicate the pronoun to which something is given. The genitive case is usually used to indicate possession over something. Ablative pronoun in general represents a motion away from something. Locative case indicates a location. Nominative case deals with the subject of the sentence. When you accomplish something, you'll utilize the accusative case.

1.8.1.3 Personal pronoun

Personal pronouns are associated with a particular grammatical person, that is, first, second or third person. It is important to note here that the term ‘Personal’ does not mean that it only applies in reference to ‘Persons’. Personal pronouns can refer to animals, objects and sometimes organizations or groups. Personal pronouns have different inflected forms for number and respect (honorific) and they also exhibit different forms for case marking.

1.8.1.3.1 First Person

First person pronoun include मैं, and हम and their different case forms such as मुझे, हमें etc.

1.8.1.3.2 Second Person

Hindi has three second persons, तू, तुम and आप, all meaning (you), where first two are singular forms for formal and informal usage and the third one is used for plural and for respect. They are also marked for different cases in Hindi.

1.8.1.3.3 Third Person

Third person pronouns are difficult to resolve in Hindi, as they are also the forms of demonstrative determiners. They are also marked for number and case. We categorize third person pronouns into proximal and distal forms. Proximal pronouns and their case forms are frequently used as abstract as well as concrete references.

The following table 1.3 contains the complete list of personal pronoun.

Table 1.3: Forms of personal pronoun

Type	**Root**	**Nomina -tive**	**Accusati -ve**	**Instrument- al/Ablative**	**Dative**	**Genitive**	**Locati -ve**
1st person (singular)	मैं	मैंने	मुझे	मुझसे	मुझको	मेरा, मेरी, मेरे	मुझमें
1st person (plural)	हम	हमने	हमें	हमसे	हमको	हमारा, हमारी, हमारे	हममें
2nd person (intimate)	तु	तूने	तुझे	तुझसे	तुझको	तेरी, तेरी, तेरे	तुझमें
2nd person (honorary)	आप	आपने	आपको	आपसे	आपको	आपका, आपकी, आपके	आपमें

Third person (distal)(singu -lar)	वह	उसने	उसको	उससे	उसको	उसका, उसकी, उसके	उसमें
Third person (distal) (plural/honor -ific)	वे	उन्होंने	उन्हें	उनसे	उनको	उनका, उनकी, उनके	उनमें
Third person (proximal) (singular)	यह	इसने	इसे	इससे	इसको	इसका, इसकी, इसके	इसमें
Third person (proximal) (plural/honor -ific)	ये	इन्होंने	इन्हें	इनसे	इनको	इनका, इनकी, इनके	इनमें

1.8.1.4 Spatial pronoun

Spatial pronouns refer to location or places. Spatial pronouns are also known as locative pronouns. They include यहाँ and वहाँ. Although, generally spatial pronouns are not classified separately and are considered a form of personal pronouns only, we consider 'locatives' as separate pronoun class. The reason for this separate classification is that in the dependency framework that we use in our work, separate labels are used to represent the locative case (the place where the action take place), thus it can help to identify the referents of the locative pronouns.

The following table 1.4 contains the complete list of locative pronoun.

Table 1.4: Forms of Spatial pronouns

Root
यहाँ
वहाँ

1.8.2 Demonstrative form of Anaphora

Besides pronouns, the other grammatical categories that can be anaphoric are demonstrative. Demonstratives are deictic words (they rely upon an outside edge of reference) that show which elements a speaker refers to and recognize those elements from others. Demonstratives do not individually refer to an entity, but they specify the entity referred by some other referring expression. Demonstratives are annotated with DEM pos-tag. Consider example (1.10).

नेपाल में कल से ही भूकंप के झटके महसूस हो रहे थे । परन्तु मौसम विभाग ने यह सूचना प्रशासन को नही दी । … (1.10)

In example (1.10), *यह* is a demonstrative because it does not directly refer to any entity, but specifies the Noun (‘information’) which in actual refers to the ‘information’ that is provided in the first sentence.

1.8.3 Gap and Ellipsis form of Anaphora

Besides the other two linguistic elements which can be termed as anaphoric are gap and ellipsis. Gaps are the instances where an element is omitted in a sentence to avoid redundancy. Sometimes gaps are anaphoric when a reference (not necessarily pronoun) is dropped. In such cases, the pronoun or the referring expression needs to be interpreted. Consider example (1.11)

रवि विराट का तो मित्र है लेकिन अभय का दुश्मन । …(1.11)

In the example (1.11), the possible pronoun वह is dropped between *लेकिन* and *अभय का.*

However, dropping in English makes it ungrammatical.

1.9 Our Contribution

The ISL from text named project is ongoing in Thapar University, Patiala under DST for Automation of ISL Generation System from Hindi Text for Communication and Education of Hearing Impaired People. The major objectives of the ISL from text project includes, building online multilingual multimedia Indian Sign Language dictionary, developing a parser to parse input language Hindi text for its ISL processing, developing HamNoSys generation system to create HamNoSys notation for ISL to express sign language phonological features, generating Indian Sign Language with avatar to animate ISL corresponding to input text using HamNoSys notation, developing an online web based and mobile based system where a user can input any simple sentence in Hindi language to get its corresponding Indian Sign Language animation through a virtual human character.

The proposed system in this thesis is part of developing a parser to parser input language Hindi text for its ISL processing. As the hearing impaired people understand simple sentences, the proposed system simplifies the sentence by resolving anaphors.

1.10 Preliminaries

The information about the tools and other related concepts used for the creation of proposed system are discussed in this section.

1.10.1 Hindi Dependency Parser

Hindi Dependency Parser that has been utilized is created by IIIT Hyderabad [3]. Dependency Parser uses dependency graphs to represent words and their relationships to syntactic modifiers using directed edges. Dependency Parser analyzes grammatical structure of sentence establishing relationship between “head” words and words which modify those heads.

1.10.2 Tags used in Hindi Dependency Parser

In grammar, a part of speech also known as word class, lexical class is a phonetic classification of words, which is for the most part characterized by the syntactic or morphological conduct of the lexical thing being referred to. Common linguistic categories include noun and verb. Parts of speech are one of the most important information required for any NLP task or application. Part of Speech tagging is the process of marking up a word in a text (corpus) as corresponding to a particular part of speech, based on both its definition, as well as its context is the way toward increasing a word in a content (corpus) as relating to a specific grammatical form, in light of the two its definition, and in addition its unique circumstance. Some of the important tags in dependency parser are given in table 1.5.

Table 1.5: Parts of Speech (POS) Tags

S. No.	Category	Tag Name
1	Noun	NN
2	NLoc	NST
3	Proper Noun	NNP
4	Pronoun	PRP
5	Demonstrative	DEM
6	Verb-finite	VM
7	Verb Auxiliary	VAUX
8	Adjective	JJ
9	Adverb	RB
10	Post Position	PSP
11	Particles	RP
12	Conjuncts	CC
13	Question Words	WQ
14	Quantifiers	QF
15	Cardinal	QC
16	Ordinal	QO
17	Classifier	CL
18	Intensifier	INTF
19	Interjection	INJ
20	Negation	NEG
21	Symbol	SYM

1.10.3 Information about Chunks

Chunks are group of closely related words, but unlike phrases they do not have syntactic behaviour. A typical chunk comprises of a single substance word encompassed by a constellation of function words. Chunks are typically taken to be an 'correlated group of words'. Chunks are important in dependency representation and parsing because usually in free word order languages, chunks are the minimal units which do not have internal

movement of its component words. Since, dependency structures do not follow order of words in a sentence, important dependency relations are realized and represented between chunks. Consider example (1.12).

((मोहन से)) ((दो बैग)) ((बरामद हुए)) ((जिनमें)) ((वह)) ((दुकान का)) ((ज़रूरी)) ((सामान)) ((ले जाता था)) । …(1.12)

'Two bags were found from Mohan in which he used to carry the important items of shop.'

1.10.4 Information about Syntax

In linguistics, syntax is "the study of the principles and processes by which sentences are constructed in particular languages". The term syntax is also used to refer to the rules which govern and describe the structure of Natural languages. Linguistic research in syntax is focused on analysis of languages in terms of such rules. There are different theoretical approaches and formalisms which are used to describe the syntactic structure of the sentences. These formalisms vary on their basic idea of how the complete structure and meaning of sentences is realized from their component elements in terms of the representation. The two such formalism which are most frequently used are discussed in section 1.10.5

1.10.5 Constituency grammar and Phrase structure syntax

The idea behind phrase structure syntax is that sentences consist of groups of words called constituents or phrases which belong to certain categories based on their linguistic behaviour, and in order to produce linguistically correct sentence for any particular language, these constituent can combine only in some specific ways which are govern by specific rules known as phrase structure grammar for that language. Moreover, these rules and combinations have recursive behaviour, that is, phrases are combined recursively to produce larger phrases which are finally combined to produce sentences. Consider example (1.13).

सुबह को रमेश ने खाना अपने मित्र के साथ खाया। …(1.13)

'Ramesh had (ate) his breakfast with his friend in the morning'

The Figure 1.1 shows the phrase structure of the example (1.13) based on given specific phrase structure grammar rules.

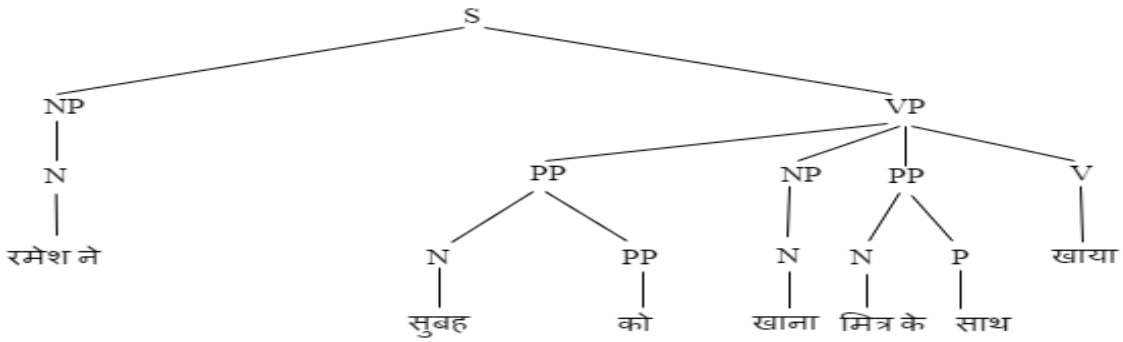

Figure 1.1: Phase structure Example

In a phrase structure tree, a terminal symbol representing the sentence node is the root of the tree. Each non-leaf node represents a non-terminal which is constituent of further non terminals or terminals as are described by the rules of the phrase structure grammar. The leaf nodes of the tree represent the words in the sentence.

1.10.6 Dependency grammar and structure

Dependency grammar is based on the idea that the relations between words in a sentence are one to one unlike phrase structure syntax. The term dependency stands for the dependence between the words in a sentence, that is, words in a sentence are dependent on other words forming a head-modified relationship where one word is head and some other words are its modifier. This head-modified relationship recursively result in dependency structures of larger lengths. The specific type of relationship may vary between different dependency frameworks but should essentially be modifier-modified relation.

Most dependency frameworks consider verb to be the root of the sentence and its arguments or participants of the action represented by the words attached under the verb. These arguments can further have modifiers. The following Figure 1.2 below shows the dependency structure of the example (1.13).

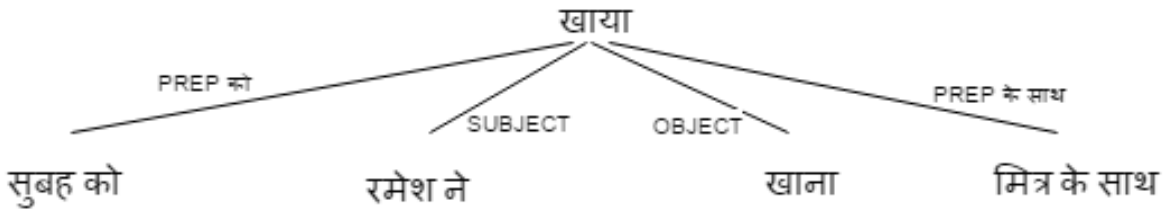

Figure 1.2: Dependency Structure Example

As seen from Figure 1.2, the head of the tree is the verb and all the arguments of the verb are represented as children of the root verb. The edges of the tree are labelled where the label represents the relationship between the words connected by the edge.

The dependency structure is used by the proposed system to resolve anaphors.

1.11 Thesis Outline

This thesis is divided into 6 chapters. Chapter 1 includes introduction about sign language, Indian sign language and the challenges to sign language. The problem of anaphora resolution and its need is described. Various terminologies required to understand the problem of anaphora resolution are introduced too. Chapter 2 describes the existing systems and techniques for Indian sign language and rule based anaphora resolution. Chapter 3 describes the problem statement, objectives and methodology used to develop a system for rule based anaphora resolution for sign language. Chapter 4 contains the detailed description of the proposed system and implementation of the system. Chapter 5 discusses the results for the sentences taken to test the system. Chapter 6 includes conclusion, limitations and future scope of the system proposed in this thesis.

Chapter Summary

In this chapter sign language is described along with the need to build a system that can be used for communication by deaf people. Also the need of ISL grammar has been discussed. The problem of anaphora has been discussed and need to resolve the anaphora for deaf people have been discussed.

Chapter 2
Literature Review

In this chapter, the existing systems for the sign language generation and the existing work on anaphora resolution have been discussed.

2.1 Existing Systems of Sign Language

With advancement in technology and growing research in field of NLP there has been significant improvement in Sign Language Systems. Researchers across the globe have been developing techniques to automate the process of text/speech to sign language.

2.1.1 Tessa

Tessa is an experimental system that was developed to convert input speech to British Sign Language. Tessa was developed in the year 1997 at the University of East Anglia; Norwich. The domain of this system was Post office where the developed system was tested to convert the words uttered by a clerk to sign language so that it was understood by deaf people. Designed using 115 phrases that covered 90% of daily business transactions, this system used Avatars for displaying the generated signs. The system was built on the Entropic Speech Recognizer that required a set of acoustic models and also a network. For complete sentences an accuracy of 61% was achieved and for sign units accuracy was 81%. On asking the participants about how acceptable the phrase was, an average acceptability rating was obtained as 2.2 ranging on the scale from 1.7 to 2.8. The architecture for the Tessa system is given by the Figure 2.1 [1].

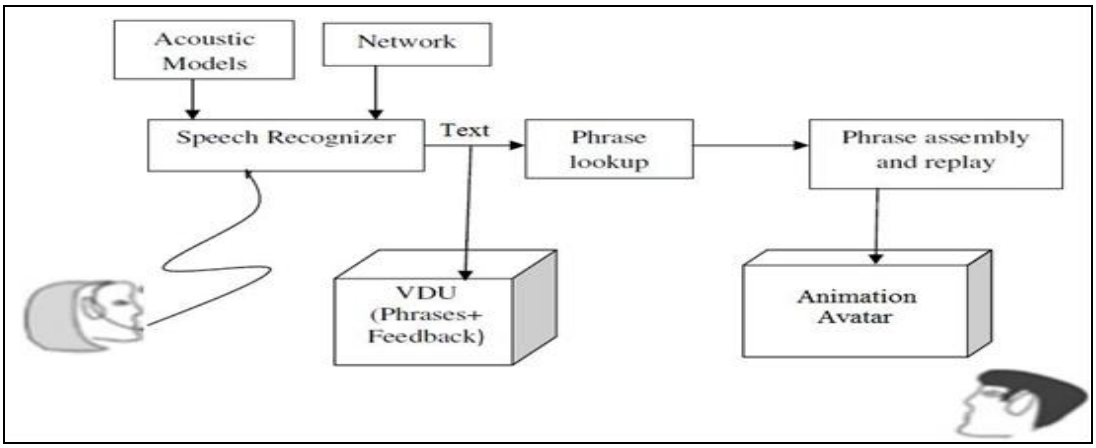

Figure 2.1: Architecture of Tessa System [11]

Tessa System has a speech recogniser which actively listens to phrases uttered by speaker. Tessa uses an entropic speech recogniser which has acoustic model that matches the input signal and a network that guides the speech recognizer during matching process.

By network system is able to constraint the speech recognizer to a finite number of predefined paths using available vocabulary. System software acts as an enabler between speech recogniser module and avatar module. System displays sign using avatars.

2.1.2 Arabic Text to Arabic Sign Langauge

System developed to convert Arabic text to Arabic Sign Language displays the Arabic signs on the screen using the saved videos .It is web based Java application that stores the video files corresponding to the signs and is played as and when demanded. Stored entries of the database are taken into the dictionary objects rendering fast retrieval and better performance. However this system does not do any pre-processing on input text and hence appropriate grammatical structure of sign language is not followed. Also the sign videos are stored on the Internet limiting its access to the end users. Figure 2.2 Shows Flow of Arabic Text To Arabic Sign Language System [12].

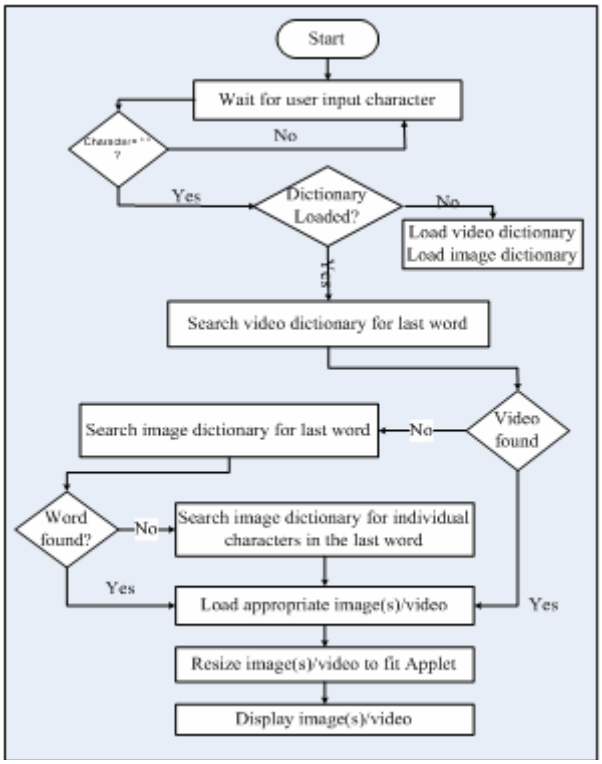

Figure 2.2: Flow of Arabic Text to Arabic Sign Language [12]

System has been developed using JSP as programming language and MS access as database. All the entries of the system are brought into dictionary object and are kept in memory to make search easy. When a word is entered and it matches an entry in memory then the location of file were word is present is accessed and subsequently is sent to browser. System works on three JSP files namely top.jsp, list.jsp and display.jsp.

2.1.3 INGIT

INGIT system is made keeping in mind Indian Railway Reservation System. INGIT is a cross modal translation system to convert Hindi String to Indian Sign Language. This uses semantically mediated formulaic framework for Hind-ISL mapping. It uses HamNoSys notation for displaying signs. The system architecture of the INGIT model is depicted in Figure 2.3. It adopts a construction grammar approach to handle standard contributions to terms of a development dictionary with single constituents and bigger expressions, with coordinate semantic mappings at each level [3].

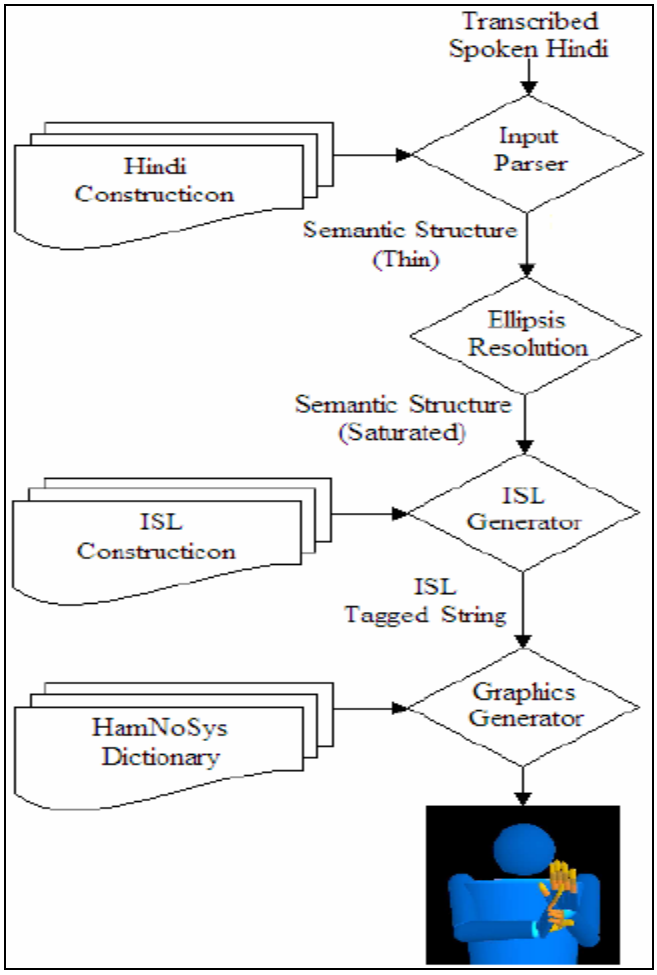

Figure 2.3: Architecture of INGIT System [13]

INGIT system works on formulaic approach which directly generates semantic structures wherever possible and compositional mode for others. It uses FCG framework for analyzing the input and generation of sign. If semantic structure of sentence is saturated it is passed to ISL generator otherwise it is passed to ellipsis resolution module. It handles elision of participants in ternary events and subjects in unary and binary events. INGIT uses HamNoSys for sign notation and avatars for displaying signs.

2.1.4 Zardoz

Zardoz is a multilingual sign interpretation framework was designed to translate textual or spoken language into various graphical animated sign-languages that include particularly Irish Sign Language (ISL), Japanese Sign Language (JSL) and American Sign Language (ASL). Zardoz was designed in the year 1991 at Trinity College, Dublin, Ireland. Being a form-driven approach, it works by processing the English input serially by using the morphological analysis, parsing by a unification grammar and idiomatic reduction. It considers the issues of broad syntactic and semantic categories of sign language, and employs spatial-dependency graphs to specify the output syntax for the sign languages in a robust and flexible way [4]. The systems described so far work on domain specific information and on a limited set of words. Figure 2.4 depicts architecture of Zardoz system [14].

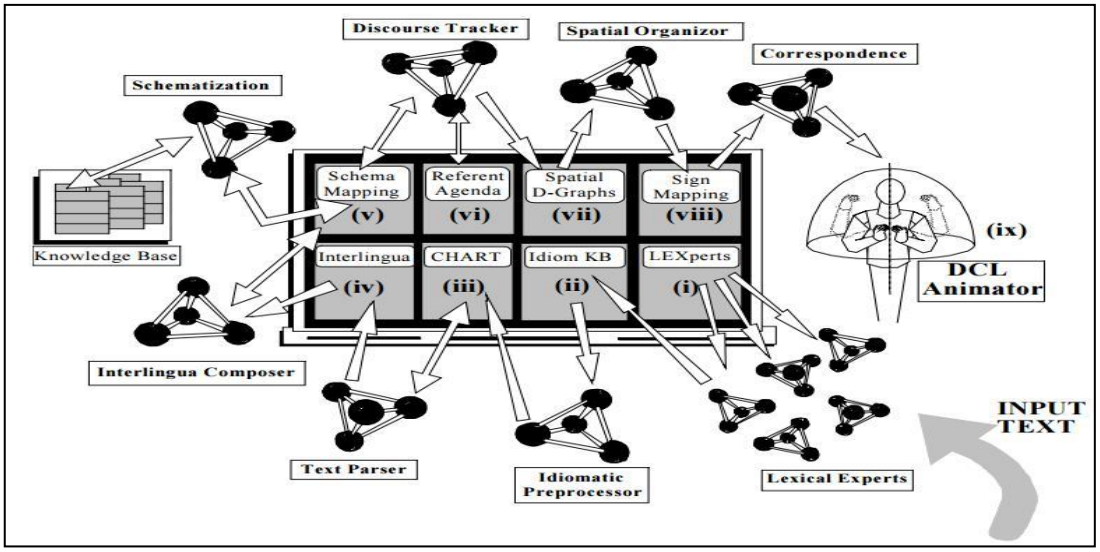

Figure 2.4: Architecture of Zardoz System [14]

It uses syntactic framework of spatial dependency graphs. SD-graph represents syntactic context and is collection of limitations for requesting the components of Interlingua outline structure. Whenever source and target language uses different anaphoric discrimination systems anaphoric resolution system is required for translation process. Fluid articulation of utterance is required to display an animated sequence.

2.1.5 ISL from text

It is an ongoing project of three duration started in August, 2016 at the Thapar University, Patiala, for the Automatic Generation of Sign Language from Hindi Text for Communication and Education of Hearing Impaired People sponsored by Science for Equity, Empowerment and Development (SEED) Division of DST.

For the system input Hindi sentence will be processed by shallow parser to identify the higher syntactic and useful data of the sentence. It will play out the tasks of tokenization, transform investigation, grammatical feature labelling and piecing for the handling of an

input sentence. The Hamburg sign language Notation System (HamNoSys) is a phonetic interpretation framework used to translate marking motions. It is a syntactic portrayal of a sign to encourage computer processing. The root words of input sentence will be mapped with HamNoSys notations with eSIGNEditor tool. eSIGNEditor will be used to pick signed sentences sign by sign from the lexicon and apply morphological changes to individual signs or strings of signs where necessary. After writing signs in HamNoSys, it will be converted into Signing Gesture Mark-up Language (SiGML). SiGML is a form of XML which defines an arrangement of XML labels for each phonetic symbol in HamNoSys. Generated SiGML file will finally be processed by Signing Avatar like "JA SiGML URL APP", which will play sign animation for the input text[15]. Figure 2.5 shows an interface of ISL from text conversion for automation.

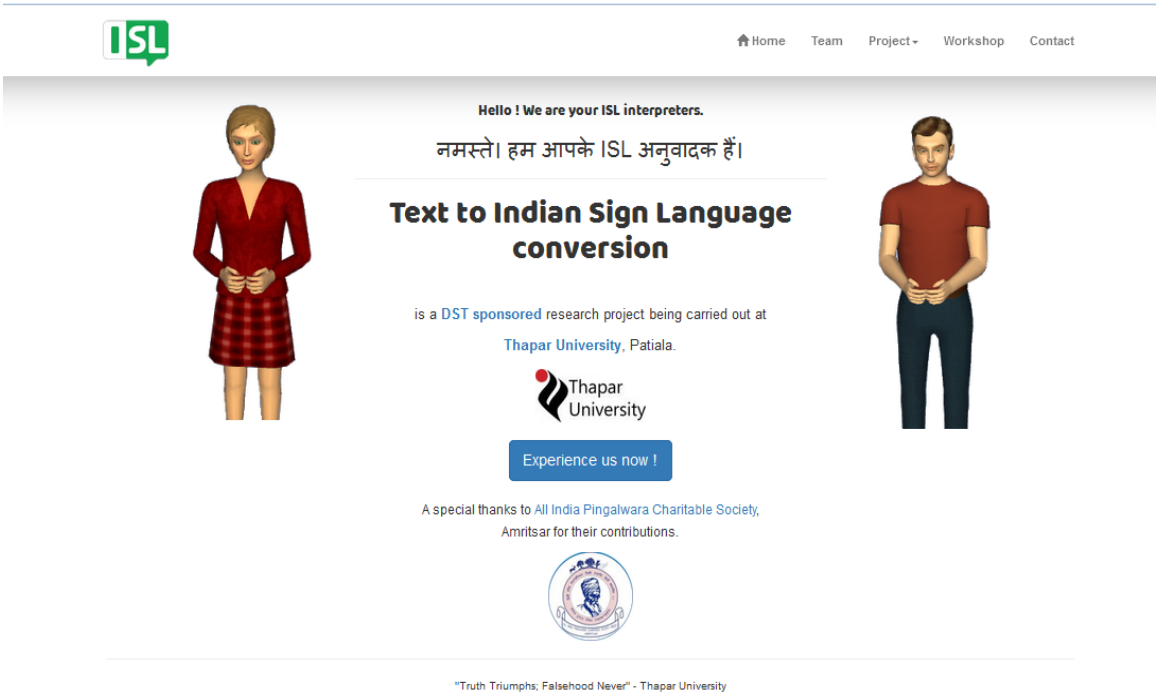

Figure 2.5: ISL from Hindi text [15]

2.2 Background and Existing Work on Anaphora Resolution

Hindi language is a free from the word order and it has no appropriate structure like English. So Pronoun in Hindi demonstrates a lot of vulnerability. To surely know the many-sided quality of anaphora determination one must be clear about grouping of anaphora and pronouns in Hindi. Distinctive methodologies used in determining anaphora resolution displays in session 2.2.1.

2.2.1 Different Approaches Used in Anaphora Resolution

Various approaches are used to solve the problem of anaphora resolution. One such approach is rule based approach.

2.2.1.1 Rule Based Approaches

Rule based methodologies combine information sources and different components that remove the things which are not required up to an arrangement of the attainable things are acquired. The requirements function as a channel to remove the undesirable applicants inside an arrangement of characterized rules. From that point, preference based elements are applied.

Various algorithms under the Rule based approach:

a) Tree search algorithm [16]
b) Shallow processing approach [17]
c) Multi-strategy approach [18]
d) Syntax based approach [19]
e) Combination of linguistic and statistical methods [20]

2.2.2 Related Work

In anaphora determination a great deal of the work is done in other Languages like English yet less measure of work is done in Hindi dialect. The scientists concentrated on different issues of anaphora determination for Hindi and utilized different existing or crossover ways to deal with settle these issues.

Sinha and Jain (2003) displayed an interpretation arrangement of English to Hindi Machine-Aided named as AnglaHindi. Anglabharti was a pseudo-interlingual rule based interpretation system. It additionally utilized case base and insights to get more exact interpretation for every now and again experienced thing and verb phrasals. It utilized semantics to determine the vast majority of the pronoun references. It had issue in settling on a decision of right reflexive pronouns .The framework produced around 90% satisfactory interpretation if there should arise of an occurrence of basic, compound and complex sentences up to a length of 20 words [21].

Pal *et al.* (2012) displayed techniques to deal with anaphora and utilized ellipsis and executed a model by utilizing a prototypel of natural language interface (NLI) to databases for Hindi – Matra2. It settled the issues of Reflexive Pronoun, Possessive Pronoun, and Demonstrative Pronoun where AnglaHindi had issue in settling on a decision of right reflexive pronouns. Matra2 was additionally contrast and the Google interpreter which had likewise issue in settling on a decision of right reflexive, possessive, expressive pronoun. It didn't separate pronouns on gender [22].

Lakhmani and Singh (2013) presented in paper the details regarding anaphora determination for Hindi dialect. The essential concentration was on the arrangement of pronominal anaphora. It likewise secured the issues identified with syntactic and semantic structure of Hindi. It played out a trial on various types of informational indexes and gave consequence of approx 71% however just on number agreement and animistic learning yet Gender understanding don't demonstrate any exactness [23].

Singh *et al.* (2014) exhibited the pronominal anaphora determination for Hindi Language and a computational model for anaphora determination in Hindi that depended on Gazetteer strategy which included numerous notable variables for settling anaphora. But, the proposed model settled anaphora by utilizing two factors that is Animistic and Recency. The investigation led on various informational indexes - informational index 1 - youngsters story – gave exactness of 65%, informational index 2 - news article - gave precision of 63%, informational collection 3 - biography from Wikipedia-gave exactness of 83% [24].

Chopra and Purohit (2013) displayed how Anaphora Resolution was helpful in performing computation linguistic task in various Natural dialects also in the Indian dialects and told about how the Anaphora Resolution was favourable in dealing with obscure words in Named Entity Recognition. Transliteration approach was utilized to settle name substance acknowledgment in different dialects [25].

आशिमा/PER बैडमिंटन खेल रही है … (2.1)

रामू/PER जम्मू/CITY में रहता है … (2.2)

In example (2.1) and (2.2), Named Entities in Hindi were transliterated into English as: Ashima, Ramu, and Jammu. Also, it had given approx 96% of result yet it didn't concentrate on different issues.

Dutta *et al.* (2008) exhibited the utilization of Hobbs calculation for pronominal determination in Hindi and tackled reflexive and possessive pronouns. Changed the Hobb''s calculation into hobb's naive algorithm for hindi and don't understand the gender agreement, number assention and so forth [26].

Uppalapu and Sharma (2009) exhibited a calculation which is in accordance with S-List to determine the Hindi third individual pronouns and demonstrated that there was a refinement of the S-List calculation in the execution by taking two records one was available and second was past rather than one. It additionally investigated how complex sentences could be broken into articulations as persuaded. The calculation additionally presented another calculation for settling the first and the second individual pronouns and it had given 61.11%, 77.45% of result on various informational indexes. In any case, creator did not concentrate on different issues [27].

Lalitha *et al.* (2014) introduced a framework for Indian dialects called a generic anaphora engine, which were poor recourses dialects. It had dissected the resemblance and unlikeness between the pronouns and their concurrence with precursors. The machine learning approach utilized the elements which could deal with significant Indian dialects. It took shallow parsed message as information and stamped Generic Engine and utilized CRFs, a direct graphical machine learning calculation to prepare the framework. It had tackled the sexual orientation issue in past [28].

Dakwale *et al.* (2013) drafted way to deal with Entity-pronoun references in Hindi called hybrid approach and utilized dependency structures as a wellspring of syntactic data. In

this approach, the dependency structure were utilized with a lead based gauge for settling the straightforward anaphoric references and a decision tree classifier was utilized to illuminate the more confusing sentences, utilizing linguistic and semantic components. The outcomes demonstrate that use of dependency structures that gives syntactic learning is utilized to determine some particular sorts of references. Semantic data, for example, animacy and Named Entity classifications additionally enhanced the determination exactness. It likewise settled the Reflexive, Locative, Relative and Personal pronouns. Furthermore, it had given 70% of exactness yet did not explain the sex understanding and different issues [29].

Lakhmani *et al.* (2014) displayed the Gazetteer strategy for pronominal anaphora determination for Hindi Language. It had built up a model that utilized Recency factor which was acted as the gauge and Animistic learning which frames the criteria of characterization of various nouns and pronouns and s for playing out the pronominal anaphora determination assignment for Hindi Language and gave approx 60 to 70% of result yet did not unravel different issues [30].

Dutta *et al.* (2011) exhibited the grouping of backhanded anaphora in Hindi corpus by utilizing machine learning approach. This was relying upon the learning of semantic structure given by the collocation of different examples and following pronouns were likewise penetrated out. It had given 12.44% outcome on roundabout anaphora in entire information however did not different issues [31].

The related work is summarised in table 2.1.

Table 2.1 Summary of related work

Author(s)	Proposed Approach	Strength	Weakness
Lakhmani et al., 2012	An approach has been proposed to determine the issue of pronominal anaphora on the premise of number agreement and animistic information and gender understanding.	An increased accuracy for number agreement and animistic knowledge. It produced result of approximately 71%.	In the proposed approach Gender agreement didn't show any accuracy. Moreover pronouns like reflexive, relative were not solved by the given approach.
Singh et al., 2014	The problem of pronominal anaphora	The experiments performed results	The proposed approach failed to

	resolution was solved using gazetteer method by using two factors, i.e., Animistic and Recency.	in the accuracy of 83%.	resolve pronouns like reflexive, spatial etc.
Dutta et al.,2008	The pronominal anaphora was sovled by using hobb's algorithm and focused on reflexive and possessive pronoun.	Successfully solve the gender agreement in past.	The algorithm works on limited set of sentences.
Uppalapu and Sharma, 2009	An adjusted calculation which was in accordance with S-List settling the Hindi third individual pronouns was utilized.	The algorithm produced 61.11%, 77.45% of result on different data sets considered. It handled first, second, third pronouns separately.	It didn't resolve other pronouns like, reflexive, relative etc. Gender agreement was not solved. It didn't solve pronouns in intersentential sentences.
Lakhmani et al., 2014	An approach was used to solve pronominal anaphora by using gazetteer method with the use of animistic knowledge and recency factor.	The results were fast for anaphora resolution system. It produced result of 60-70% and increases the accuracy.	It didn't focus on other factors like number, gender agreement. Intrasentential and intersential recency was not solved.

Lalitha et al., 2014	The generic algorithm was produced for Indian dialects by utilizing the machine learning approach for the similarity and varieties between the different pronouns and their concurrence with their antecedents.	The system developed was language independent as it used information from in-depth morphological analysis.	The system didn't focus on number agreement, recency. It had used very minimal resources.
Sinha and Jain, 2003	A system was developed, AnglaHindi: An English to Hindi Machine-Aided Translation System using semantic learning to determine a large portion of the intrasentence pronoun references.	The system produced result of 91% on simple, compound, complex sentences up to 20 words. It used example-base and statistics which lead to more accuracy.	It has problem in making a choice of correct reflexive pronouns.

In the table 2.1, the various algorithms and systems are considered with their strength and weakness for better understanding. It is observed that many attempts were made to handle pronominal form of anaphora in different ways and for different purposes. And it is observed that it is hard to handle pronominal form of anaphora considering every factor like, number, gender, recency etc.

Chapter Summary

In this chapter various systems have been discussed that worked on Sign Language generation. Also, various approaches and algorithms for anaphora resolution system have been discussed. The factors that play significant role in resolving pronominal form of anaphora have been discussed through various systems.

Chapter 3
Problem Statement

Sign Language is a way of communication for deaf and dumb. There are not many experts in India who can impart knowledge about Sign Language to deaf, thus resulting in poor Literacy Levels and Language skills among the group. Moreover in the absence of high end experts who can make these people well versed with their mode of talking to the outside world, there is a critical need to automate the Indian Sign Language System to ensure a better communication among deaf communities. It is recommended to keep the system as simple as possible. Hearing impaired people understand very straight forward language and very basic and simple sentences. Occurrence of anaphora in a sentence might confuse them as they might not understand the referred object/noun. So a system is required to have in place the resolution of pronominal forms of anaphora in Hindi Language. In order to resolve the Anaphors, some framework and algorithm has to be designed for handling various types of Pronominal Forms. These Pronominal anaphors are further replaced by their appropriate referents.

The proposed solution will make the life of hard hearing people a lot easier. Automated generation of Sign Language from sentences helps in better communication.

3.1 Objectives

The proposed system has been designed to achieve the following objectives:

a) To study the existing approaches and systems for anaphora resolution and ISL generation respectively.

b) To identify the different types of anaphora for Hindi Text for generation of Indian Sign Language.

c) To develop a framework for resolution of anaphora for ISL generation system.

d) To develop rules based system for resolving different types of anaphoras.

e) To test and validate the proposed anaphora resolution system for ISL generation for Hindi sentences.

3.2 Methodologies

To achieve the objective given in section 3.1 the following methodologies have been used.

a) The literature survey has been carried out to understand various approaches and techniques used for resolving anaphora resolution and for automatic generation of Indian Sign Language.

b) Hindi Grammar has been referred to identify all possible pronouns for each

category of the Pronominal Form which are; Reflexive, Spatial, Relative and First-Second Person Pronouns.

c) A web based system has been designed in Python that uses Hindi Dependency Parser and handles the anaphoras of an input sentence.

d) Rules have been designed and followed for the framework structured and formulated in Python Language to fine referents and further, replacing the anaphora with it.

e) A corpus of approximately 300 sentences has been designed to test and validate the working of proposed framework.

Chapter 4
Implementation

4.1 ISL Generation System

It is an ongoing project of three duration started in August, 2016 at Thapar University, Patiala, for the Automatic Generation of Sign Language from Hindi Text for Communication and Education of Hearing Impaired People sponsored by Science for Equity, Empowerment and Development (SEED) Division of DST.

For the system input Hindi sentence will be processed by shallow parser to identify the higher syntactic and functional information of the sentence. The Hamburg sign language Notation System (HamNoSys) is a phonetic transcription system used to transcribe signing gestures. It is a syntactic representation of a sign to facilitate computer processing. The root words of input sentence will be mapped with HamNoSys notations with eSIGNEditor tool. After writing signs in HamNoSys, it will be converted into Signing Gesture Mark-up Language (SiGML). SiGML is a form of XML which defines a set of XML tags for each phonetic symbol in HamNoSys. Generated SiGML file will finally be processed by Signing Avatar like "JA SiGML URL APP", which will play sign animation for the input text. The workflow of the system has been given in Figure 4.1

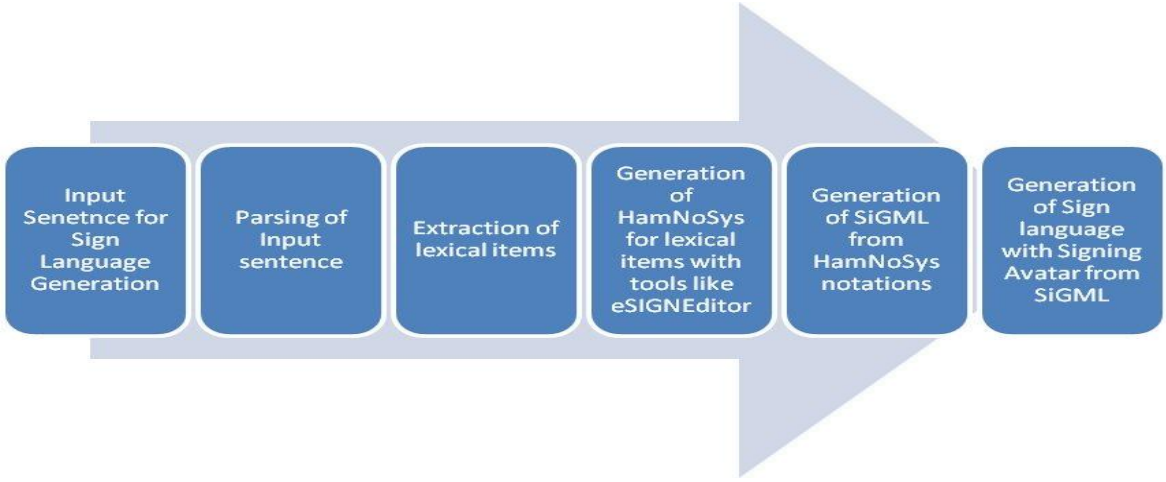

Figure 4.1: Workflow of ISL from Hindi Text

4.2 Architecture of proposed system

The architecture of Rule based Anaphora Resolution system has been shown in figure 4.2 is a part of the project ISL from Hindi text. It belongs to the module parsing of input sentence of the project ISL from Hindi text. The working of each module of the architecture has been explained further.

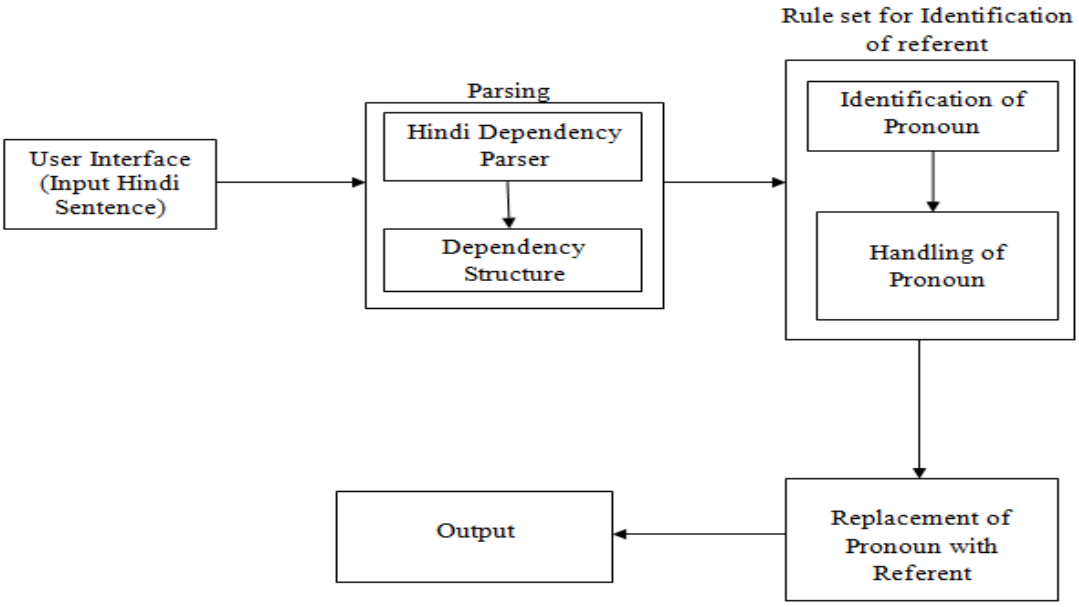

Figure 4.2: Proposed Architecture of Rule Based Anaphora Resolution in Hindi for ISL

4.2.1 User Interface

User interface is the front face of the system. The GUI of the system has been developed in Python. Through this user can provide input to the developed system which is then further processed by Dependency Parser discussed in section 4.1.2. The figure 4.3 shows the user interface.

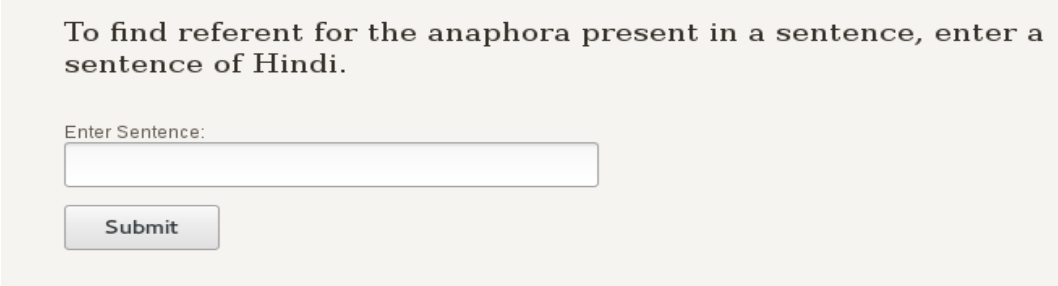

Figure 4.3: User Interface

4.2.2 Dependency Parser

Hindi Dependency Parser written in Python has been called as shown in figure 4.4 for the input sentence.

```
if request.method == 'POST':
        sent = request.form['sent']
        upload_dir = '/home/user1/flask/hindi-dependency-parser-2.0/'
        file_text=codecs.open(upload_dir+ "hindi.input.txt","w",encoding="utf8")
        file_text.write(sent)
        os.chdir("/home/user1/flask/hindi-dependency-parser-2.0/")
        file_text.close()
        make_process = subprocess.Popen("make hindi.output", shell=True,stdout=subprocess.PIPE, stderr=subprocess.STDOUT)
```

Figure 4.4: Code to call Dependency Parser

Once execution is done of parser the results will be saved in output file. The output file has same name as of input file for parsing. The file with *.output.txt* contains information of POS tag with all information of parsed sentence. The output file of Dependency Parser contains information in the following format:

<word_id> <word> <lemma> <pos_tag> <parent_id> <dependency_label>

The *word id* is the unique identification number given to each word present in the sentence. Next is the *word*, which is the basic word. *Lemma* represents the root word or conventional form of word which have the same meaning. The *POS tag* gives description of Parts of Speech that particular word belongs to and the possible tags of dependency parser have already been discussed in table 6. Parent identification number (*parent_id*) represents from which word has been derived. And the last information that is *dependency label* has been discussed later in section 4.2.1. Consider example (4.1).

विवेक ने अपनी कार की चाबी खो दी … (4.1)

The fig 4.5 shows the parsed output of Hindi sentence considered in example (4.1)

1	विवेक	विवेक	NN	7	k1
2	ने	ने	PSP:ने	1	lwg__psp
3	अपनी	अपनी	PRP	4	r6
4	कार	कार	NN	6	r6
5	की	का	PSP:का	4	lwg__psp
6	चाबी	चाबी	NN	7	k2
7	खो	खो	VM	0	main
8	दी	दे	VAUX	7	lwg__vaux

Figure 4.5: Parsed output for Hindi sentence

4.2.3 Dependency structure and relations

The dependency structure represents the structure obtained by representation of one-to-one correspondence between the words in a sentence which is based on the head-modifier relationship. The nodes in the sentence necessarily represent the words or chunks. In dependency structure, nodes need not to follow any specific word order. Edges in the dependency tree are usually labeled where the label represents the relation between the

two words. There are approximately 45 relations in total. Some of the relations with their meaning are shown in table 4.1.

Table 4.1: CPG relations and their meanings

Label	CPG relation	Meaning
k1	Karta	Most independent entity which carries out the action
k2	Karma	The entity on which action is carried out
k4/k4a	Sampradan	Experiencer/receiver
k7p	Apaadan	Location
r6	Sambandh	Genitive/possessive
Rh	Hetu	Purpose
Rs	Samaanadhikaran	Equivalence
Ccof	Conjunction	Conjunction

Considering example (4.2) the dependency structure has been shown in figure 4.6.

विवेक ने अपनी कार की चाबी खो दी … (4.2)

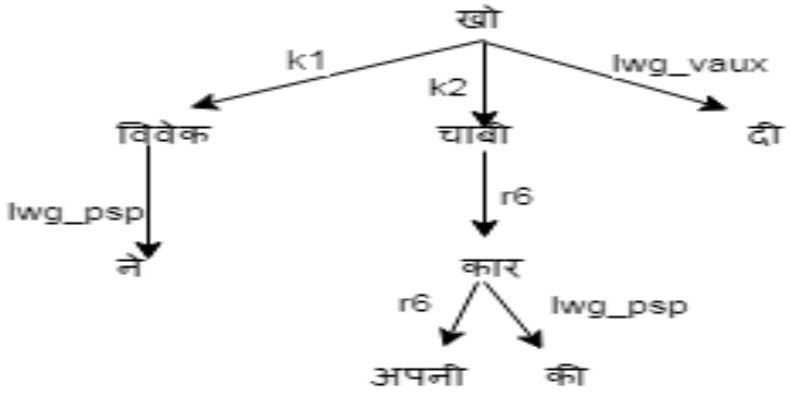

Figure 4.6: Dependency structure for example (4.2)

4.2.4 Identification and Handling of different types of anaphors

Rule sets have been framed to handle different types of anaphors that are written in python. Some categories of pronominal references such as Reflexives, Relatives, First and Second person can be easily resolved by formulating rules based on dependency structures. The rules have been discussed as follows for different forms of pronouns.

4.2.4.1 Reflexives

In Hindi, frequent types of reflexives are the possessive reflexives i.e. अपना, अपनी etc. The referent of the possessive reflexives is the possessor entity which is the 'SUBJECT' of the clause or sentence. The referent is frequently the Entity which has a role of '***karta***' ('***k1***'). The example (4.3) with dependency structure shown in figure 4.7 is given below:

महेश ने बताया कि रवि ने अशोक को अपनी कार दे दी है। … (4.3)

'Mahesh told that ravi gave his car to ashok.'

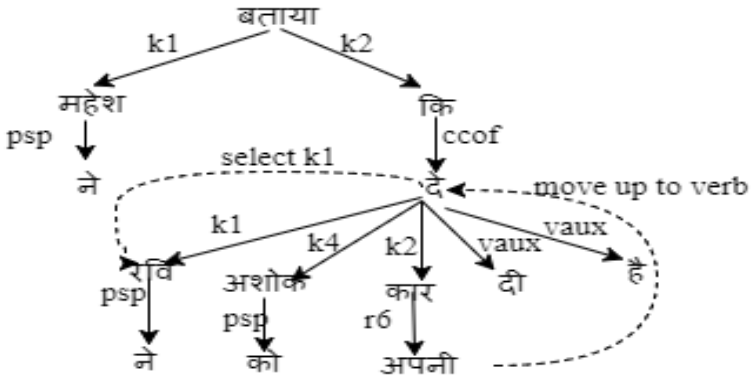

Figure 4.7: Dependency structure of example (4.3)

The algorithm 4.1 has been followed to resolve Reflexive Pronouns.

Algorithm 4.1: Resolution for Reflexive Pronouns

Input: Dependency Structure (*dep_str)*

Output: Referent

Ref_words= [अपना, अपनी, अपने] // Possible pronouns for reflexive

Ref_PPN = [ ] // Empty list to store pronoun

(a) ReferentIdent(dep_str):

1. **Search** for word *w* in *dep_str* where $w \in Ref\_words$ and $POS(w) == PRP$

   Add *w* to *ref_PPN***;**

2. **If** *(POS(par(w)) ==VM***), then**

   Search in child nodes *n* having dep_rel == 'r6'

   Else if $(POS(w) == NN \,||NNP)$

Search in child nodes having dep_rel == 'k1'

Else

w = par(*w*) and repeat step 2.

3. **Replace** the Referent found with *w* // Here, w belongs to Ref_PPN;

If a node, n having POS tag as PRP is found then store it in an empty list. Then check parent of pronoun if it is a verb node (VM or VAUX). If a verb node is encountered, then search for children nodes other n, if a children node with dependency label 'k1' is found, propose it as referent. Replace pronoun stored in a list with its referent.

And, if instead of verb node, parent node is noun (NN or NNP) then search for children nodes other than n, if a children node with dependency label 'r6' is found, propose it as referent. Replace pronoun with its referent. Consider example (4.4)

विवेक ने अपनी कार की चाबी खो दी। … (4.4)

Figure 4.8 given below shows example (4.4) as an input for the reflexive and figure 4.9 shows its corresponding output.

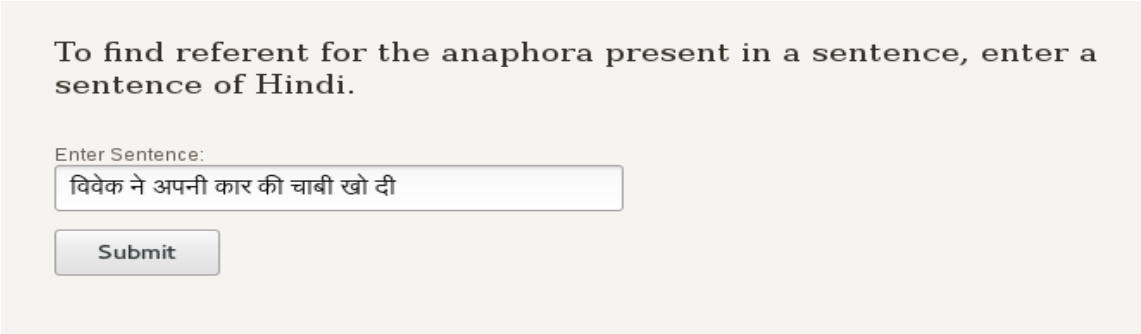

Figure 4.8: Input Sentence with Reflexive Pronoun

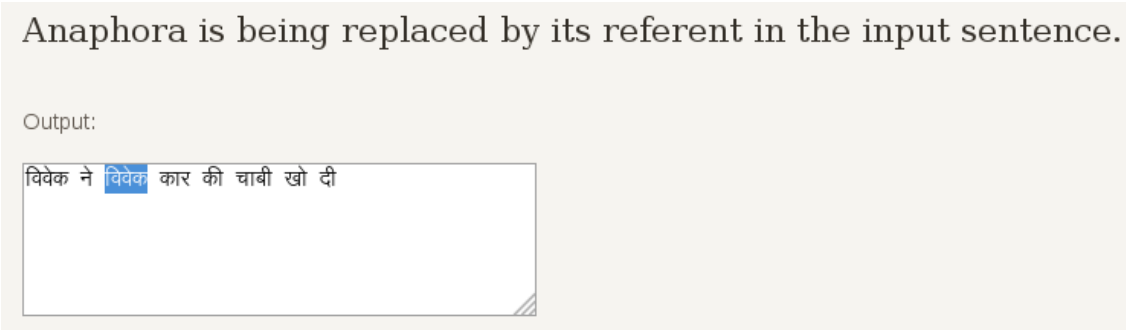

Figure 4.9: Output Sentence with referent for Reflexive Pronoun

4.2.4.2 Spatial Pronoun

It refers to places. Hindi has two spatial pronouns यहाँ and वहाँ. It identifies Noun phrases representing '*places*' and choose the most probable among them. A label '*k7p'* is used to annotate the entities which represent the location of the '*action taking place*' in the sentence. The Example (4.5) for spatial pronouns with dependency structure shown in figure 4.10 as below:

राम ने आगरा जाकर वहाँ का ताज-महल देखा। ... (4.5)

'Ram went to Agra to see the Taj Mahal there.'

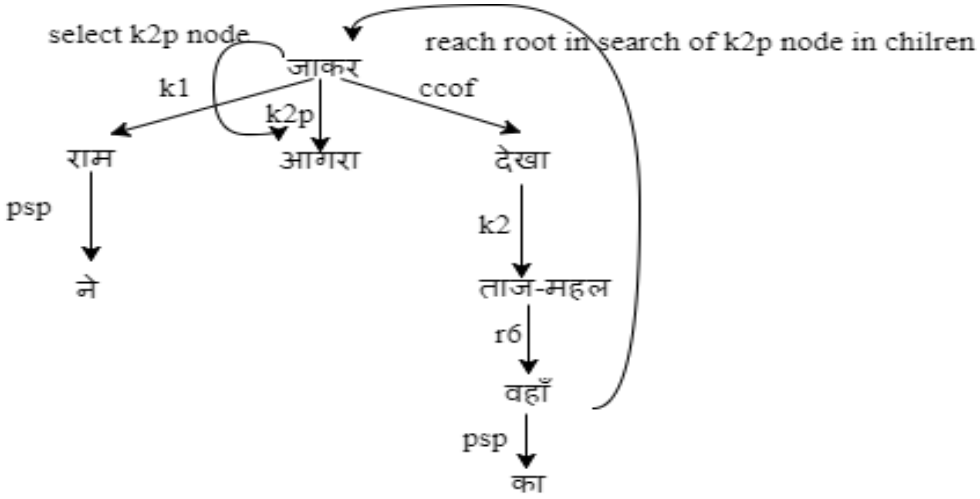

Figure 4.10: Dependency structure for example (4.5) sentence 2

No location is described in the second sentence, so search for the referent in previous sentence (sentence 1). The root verb of the sentence 1 has a child node with label '*k7p*'; hence it will be selected as the referent of the pronoun.

The algorithm 4.2 has been followed to resolve the spatial pronouns.

Algorithm 4.2: Resolution for Spatial Pronouns
Input: Dependency Structure (*dep_str)* Output: Referent *Spat_words*= [यहाँ, वहाँ] *Spat_PPN* = [ ] ***(b) SpatialIden(dep_str):***

1. **Search** for word *w* in dep_str where ($w \in Spat\_words$) and ($POS(w) == PRP$)

   Add *w* to *Spat_PPN***;** // pronoun stored in a list and say it, *p*

2. **If** (*POS(par(w)) ==VM*)**, then**

   Search in child nodes *n* having *dep_rel* == (*'k7p'*|| *'k2p'*|| *'vmod'*)

   Else

   w = *par*(*w*) and repeat step 2.

3. **If** *dep_rel*== (*'vmod'*) t**hen**

   Go to step 4.

   Else

   Go to step 5.

4. **If** (*POS(child(par(w))) == VM)* and *(dep_rel == 'vmod')*, **then**

   w = child*(w);*

   Search in child nodes n having *dep_rel == ('k2p')*

5. **Replace** the Referent found with w ∈ *Spat_PPN*;

Search for pronoun in dependency structure and identify it as spatial or place pronoun. Search in reverse order for a noun phrase and dependency label 'k7p' or 'k2p' or 'vmod'. In case of dependency label 'k2p' or 'k7p', if such a noun phrase is found then predict it as the referent. In case of 'vmod', search its child nodes. If a noun phrase having dependency label as 'k2p' is found, predict it as the referent. Else no referent is found (no solution). Consider example (4.6).

राहुल शिमला जाकर वहाँ की सुंदरता से प्रभावित हुआ … (4.6)

Figure 4.11 given below shows that example (4.6) is taken input for the reflexive and figure 4.12 shows its corresponding output.

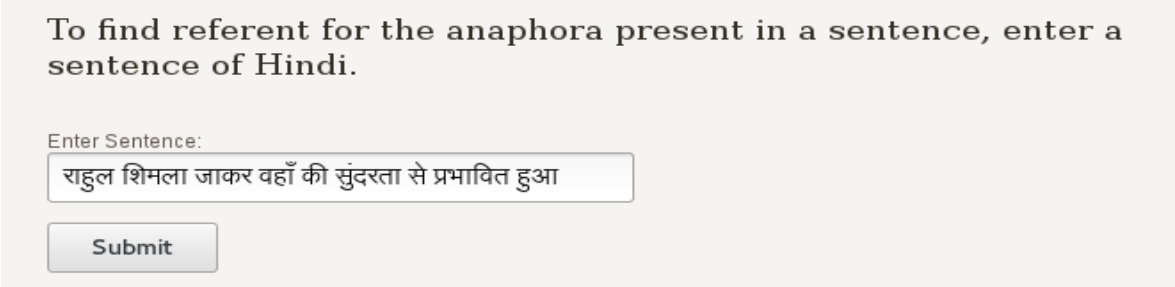

Figure 4.11: Input Sentence for Spatial Pronoun

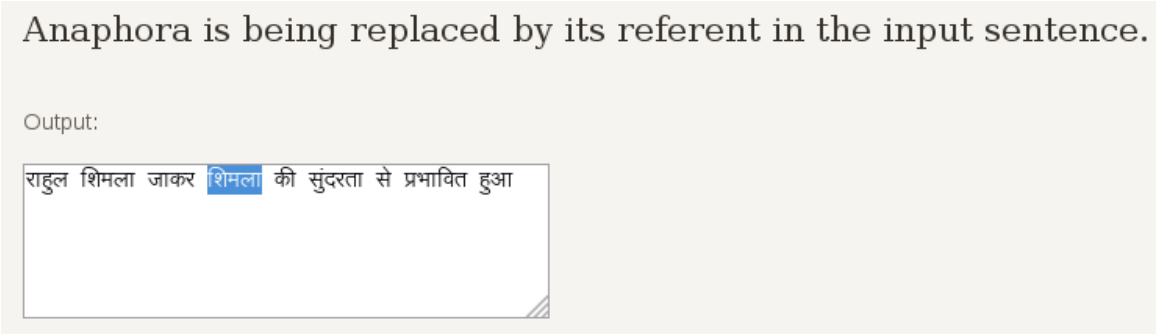

Figure 4.12: Output Sentence with referent for Spatial Pronoun

4.2.4.3 Relative Pronoun

A relative clause is a kind of subordinate clause which specifies an element, usually a Noun Phrase (NP), in the main clause. The referent of the relative pronoun is essentially that Noun phrase which is head of the root verb of the relative clause. The Example (4.7) with its Dependency structure is given in figure 4.13.

बदमाशों से बैग बरामद हुए जिनमें चोरी का सामान था । … (4.7)

'Bag was seized from the thugs in which there were looted items.'

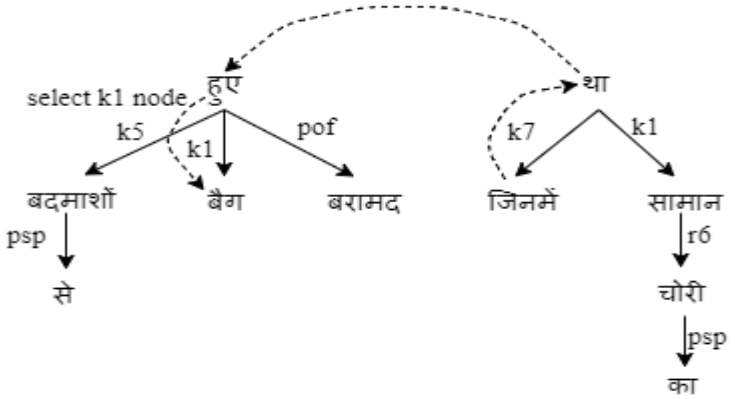

Figure 4.13: Dependency structure for example (4.7)

In the example, the relative pronoun is जिनमे and the head of the relative clause is the verb ले जाते थे which is attached below the NP node दस बैग with a relation ‘k1’ or ‘k2p’ or ‘k1s’.

The algorithm 4.3 has been followed to resolve relative pronouns.

Algorithm 4.3: Resolution for Relative Pronouns

Input: Dependency Structure (*dep_str)*

Output: Referent

Rel_words= [जिनमें, जिसमें]

Rel_PPN = [ ]

(c) RelativeIden(dep_str):

1. **Search** for word *w* in dep_str where ($w \in Rel\_words$) and ($POS(w) == PRP$)

   Add *w* to *Rel_PPN***;**

2. **If** (*POS(par(w)) ==VM || VAUX) and (Parent_id(par(w)) ==* 0**), then**

   Search in other nodes *p* having ($POS(p) == VM||VAUX$) and (*Parent_id == 0)*

   Else

   w = par(*w*) and repeat step 2.

3. **If** $(POS(p) == VM||VAUX)$ and (*Parent_id == 0),***then**

   Search in other child nodes n having *dep_rel == (‘k1’|| ‘k2p’ || ‘k1s’)* and *(POS(n) == ‘NN’);*

4. **Replace** the Referent found with w belongs to *Spat_PPN*;

Start with the pronoun node and, search up in the structure until a verb node (VGF or VGNF) is encountered. Search its children nodes n, if any noun phrase node having dependency relation ‘k1’ or ‘k2p’ or ‘k1s’ is found, then select it as referent. Else find for other verb node in the structure which is also a root. And repeat the same for it.

Consider example (4.8).

बदमाशों से बैग बरामद हुए जिनमें चोरी का सामान था … (4.8)

Figure 4.14 given shows the input for the relative and figure 4.15 shows its corresponding output.

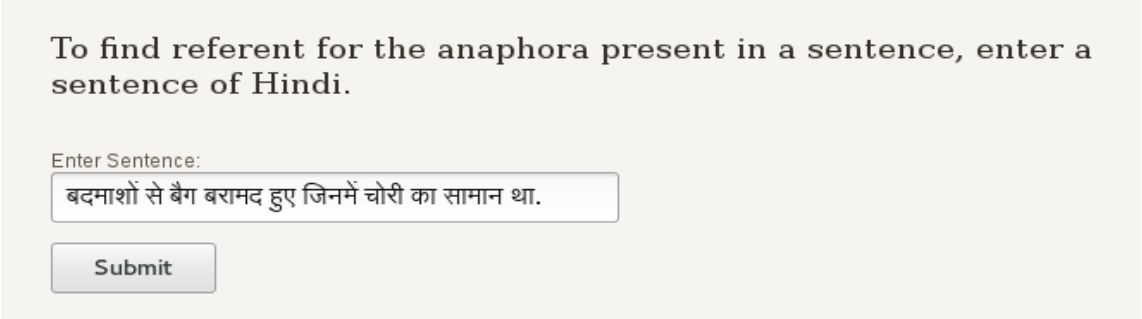

Figure 4.14: Input Sentence for Relative Pronoun

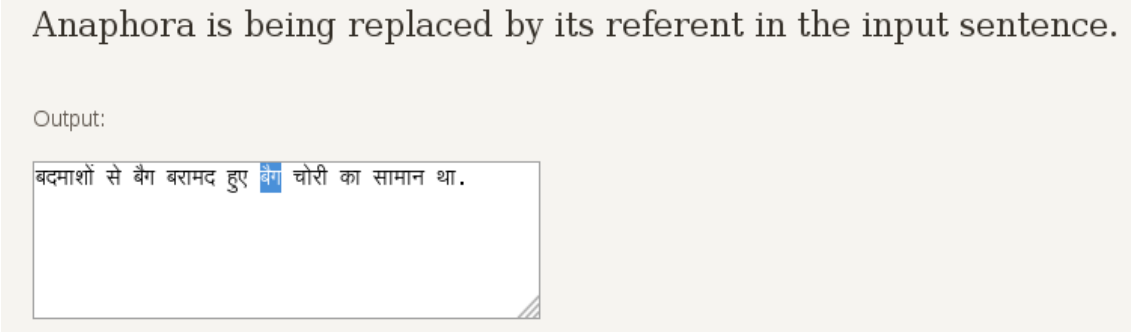

Figure 4.15: Output Sentence with referent for Relative Pronoun

4.2.4.4 First and Second Prounoun

First and second pronoun refers to speaker and listener of a communication. First pronouns include मैं, हम and their inflected forms and second person pronoun include तु, तुम, आप along with their inflected forms.

The Example (4.9) for first and second pronoun with dependency structure shown in figure 4.16 and figure 4.17 respectively is discussed below:

प्रधानमंत्री ने जनता से कहा कि आप मेरा मत्त स्वीकार करें। … (4.9)

'Prime Minister asked people to accept his viewpoint.'

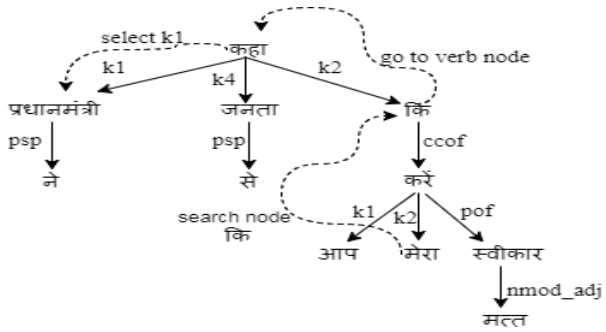

Figure 4.16: Dependency structure for मेरा

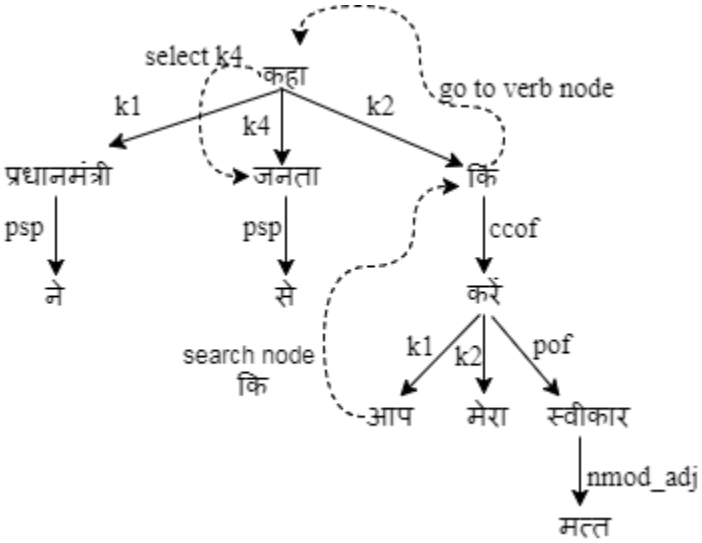

Figure 4.17: Dependency structure for आप

The dependency structure of example shows मेरा is the first person pronoun and आप is second person pronoun in the attribution clause rooted at स्वीकार.

The algorithm 4.4 has been followed to resolve first person pronoun.

Algorithm 4.4: Resolution for First Person Pronouns
Input: Dependency Structure (*dep_str*) Output: Referent *FirP_words*= [मैं, हम] // Possible First Pronouns *FirP_PPN* = [ ] *Comp* = [कि] // Complementizer node value ***(d) FirstIden(dep_str):***

1. **Search** for word *w* in *dep_str* ;
2. **If** ($w \in FirP\_words$) and ($POS(w) == PRP$) **then**

   Add *w* to *FirP_PPN***;**
3. **If (***POS(*par(*w)*) ==Comp**), then**

   w = par(*w*) and repeat step 3 until Comp is encountered.

   Else

   No referent.
4. **If** ($POS(par(w)) == VM$) **then**

   Search in child nodes having *dep_rel* == 'k1'

   Else

   w = par(*w*) and repeat step 3.
5. **Replace** the Referent found with *w* that belongs to *FirsP_PPN*;

Starting at the pronoun node, move up in the structure until a complementizer node कि ('that') is encountered. If no such node is found up-to the root, output 'no solution'. And if found then move up to the parent node of complementizer node. If this is a verb node, search in the children nodes of it, node with label 'k4'. Select node with 'k4' as the referent. If no such node is found, output 'no solution'.

The algorithm 4.5 has been followed to resolve second person pronoun

Algorithm 4.5: Resolution for Second Person Pronouns

Input: Dependency Structure (*dep_str*)

Output: Referent

SecP_words = [तू , तुम, आप]

SecP_PPN = []

Comp = [कि]

(e) SecIden(dep_str)*:*

1. **Search** for word *w* in dep_str where ($word \in Ref\_words$) and ($POS == PRP$)

   Add *w* to ref_PPN**;**

2. **If (**par(*w*) ==Comp)**, then**

   w = par(*w*) and repeat step 2 until Comp is encountered.

   Else

   No referent.

3. **If** $(par(w) == VM)$

   Search in child nodes having *dep_rel* == 'k4'

   Else

   No referent.

4. **Replace** the Referent found with w belongs to SecP_PPN;

Starting at the pronoun node, move up in the structure until a complementizer node कि ('that') is encountered. If no such node is found up-to the root, output 'no solution'. And if found then move up to the parent node of complementizer node. If this is a verb node, search in the children nodes of it, node with label 'k4'. Select node with 'k4' as the referent. If no such node is found, output 'no solution'.

Figure 4.18 shows input for the first and second person pronoun and figure 4.19 and figure 4.20 show the output for the first and second person for example (4.9) respectively.

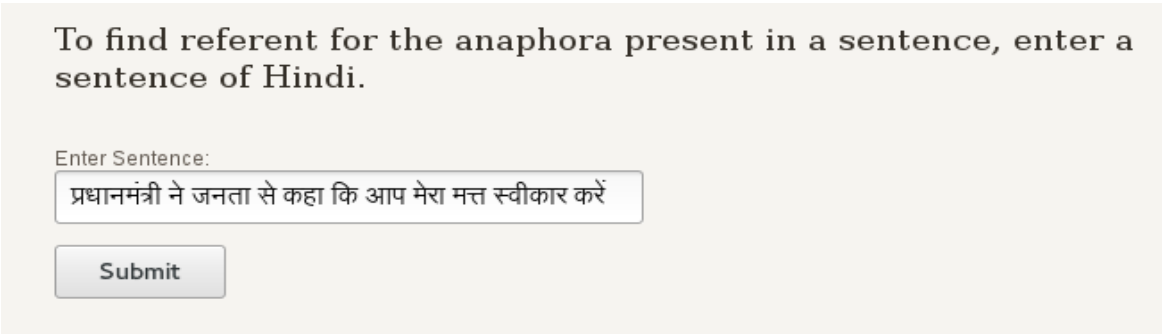

Figure 4.18: Input Sentence for First and Second Person Pronoun

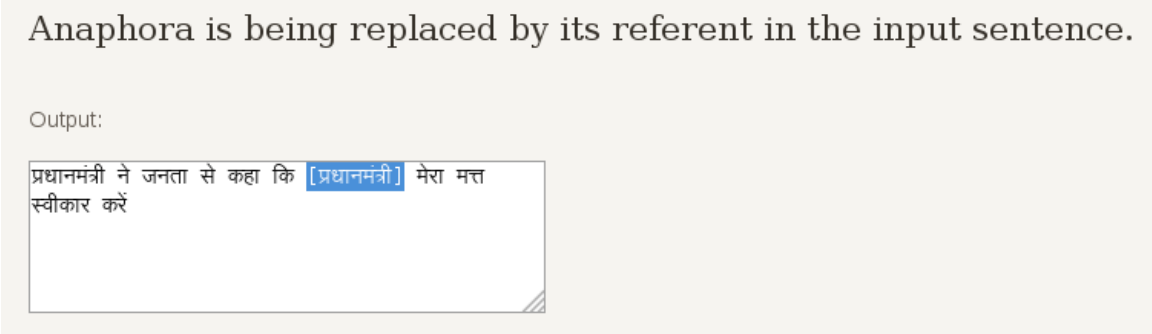

Figure 4.19: Output Sentence with referent for Second Person Pronoun

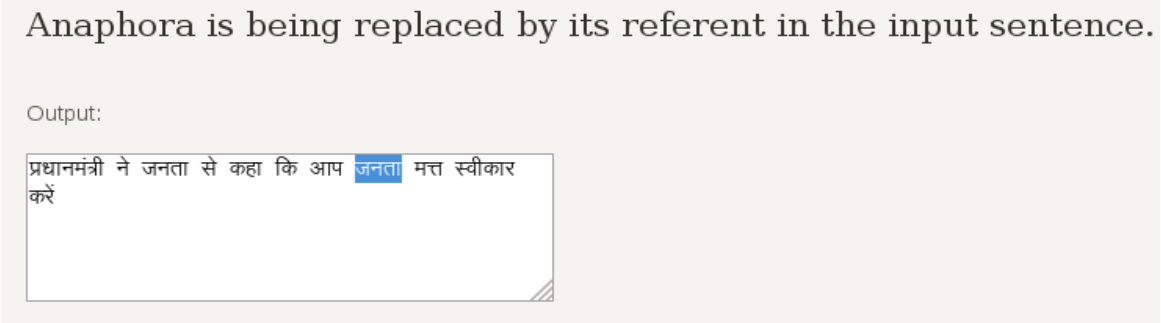

Figure 4.20: Output Sentence with referent for First Person Pronoun

Chapter Summary

In this chapter, different rules on the basis of types of pronoun and algorithms for them have been discussed. The dependency parser parses the sentence and provides the parsed output to the system. The system identifies the type of pronoun and resolves the anaphor which is then replaced by its referent.

Chapter 5
Results and Discussion

The proposed system has been tested over Hindi sentences. The system handles the four categories of pronominal anaphoras namely reflexive pronoun, relative pronoun, spatial pronoun, first and second person pronoun for which the developed system has been tested. A corpus of approximately 300 sentences has been used to test and validate the working of Rule based Anaphora Resolution system. A rating or a score has been given by the user physically to each sentence on the size of 5 which is portrayed in table 5.1. The sentences have been evaluated on the premise of precision of creating linguistically and syntactically correct result by the framework.

Table 5.1 Score Description

Rating/Score	Description
5	Excellent
4	Good
3	Average
2	Below Average
1	Poor

The table 5.2 shows the results for the sentences of reflexive, spatial, relative, first and second person pronoun. The score of each sentence has been given corresponding to each sentence which is used to calculate the accuracy later.

Table 5.2: Examples of Anaphora Resolution with sentence Score

Examples of Anaphora Resolution for Spatial Pronoun		
Sentence	**Resolution for sentence**	**Score**
राम ने आगरा जाकर वहाँ का ताज-महल देखा.	राम ने आगरा जाकर आगरा का ताज-महल देखा.	4
राहुल शिमला जाकर वहाँ की	राहुल शिमला जाकर शिमला की सुंदरता	4

सुंदरता से प्रभावित हुआ.	से प्रभावित हुआ.	
ओबामा ने इंडिया आकर यहाँ के लोगों से बात की.	ओबामा ने इंडिया आकर इंडिया के लोगों से बात की.	4
राजेश पंजाब आकर यहाँ के लोगों से घुल गया।	राजेश पंजाब आकर पंजाब के लोगों से घुल गया।	4
रामू ने पटियाला जाकर वहाँ दुकान किराए पर ली।	रामू ने पटियाला जाकर पटियाला दुकान किराए पर ली।	4
घर पर आए मेहमान आज यहाँ पर ही रुकने वाले हैं.	घर पर आए मेहमान आज घर पर ही रुकने वाले हैं.	4
राज विदेश जाकर वहाँ की सुंदरता से प्रभावित हुआ.	राज विदेश जाकर विदेश की सुंदरता से प्रभावित हुआ.	4
राज दिल्ली में वहाँ का लाल-किला देखने गया।	राज दिल्ली में दिल्ली का लाल-किला देखने गया।	4
राजेश जैपुर में यहाँ का हवा-महल देखने आया।	राजेश जैपुर में जैपुर का हवा-महल देखने आया।	4
लोगों ने ऋषिकेश जाकर वहाँ की सैर की.	लोगों ने ऋषिकेश जाकर ऋषिकेश की सैर की.	4
राम ने आगरा जाकर वहाँ का ताज-महल देखा.	राम ने आगरा जाकर आगरा का ताज-महल देखा.	4
राहुल शिमला जाकर वहाँ की सुंदरता से प्रभावित हुआ.	राहुल शिमला जाकर शिमला की सुंदरता से प्रभावित हुआ.	4

ओबामा ने इंडिया आकर यहाँ के लोगों से बात की.	ओबामा ने इंडिया आकर इंडिया के लोगों से बात की.	4
राजेश पंजाब आकर यहाँ के लोगों से घुल गया।	राजेश पंजाब आकर पंजाब के लोगों से घुल गया।	4
Examples of Anaphora Resolution for Relative pronoun		
परीक्षा आसान थी जिसमें अच्छे अंक हैं.	परीक्षा आसान थी परीक्षा अच्छे अंक हैं.	4
राम के पास गाड़ियाँ है जिनमें सामान जा रहा था.	राम के पास गाड़ियाँ है गाड़ियाँ सामान जा रहा था.	4
अभय ने मेज़ पर से थैला उठाया जिनमें सब्ज़ियाँ रखी.	अभय ने मेज़ पर से थैला उठाया थैला सब्ज़ियाँ रखी.	4
राम के पास थैलियाँ हैं जिनमें खाने का सामान है.	राम के पास थैलियाँ हैं थैलियाँ खाने का सामान है.	4
बदमाशों से बैग बरामद हुए जिनमें चोरी का सामान था.	बदमाशों से बैग बरामद हुए बैग चोरी का सामान था.	4
सामने आलीशान घर है जिसमें अभय रहता है.	सामने आलीशान घर है घर अभय रहता है.	4
शाम के घर के सामने मैदान है जिसमें बच्चे खेल रहे हैं.	शाम के घर के सामने मैदान है मैदान बच्चे खेल रहे हैं.	4
रमन ने बड़ा घर खरीदा जिसमें सुंदर बालकोनी है.	रमन ने बड़ा घर खरीदा घर सुंदर बालकोनी है.	4

राजू ने बैग लिया जिसमें समान डाल सके.	राजू ने बैग लिया बैग समान डाल सके.	4
जैसलमेर में एक बंगला है जिसमें भूतों का बसेरा है.	जैसलमेर में एक बंगला है बंगला भूतों का बसेरा है.	4
Examples of Anaphora Resolution for Reflexive Pronoun		
रमेश ने अपनी कार की चाबी खो दी है.	रमेश ने रमेश कार की चाबी खो दी है.	3
रमा ने रेखा को अपनी किताब दी.	रमा ने रेखा को रमा किताब दी.	3
राजेश अपने मित्रों से मिला.	राजेश राजेश मित्रों से मिला.	3
काजोल ने अपने किरदार को बाखूबी निभाया.	काजोल ने काजोल किरदार को बाखूबी निभाया.	3
राखी ने काम में अपनी माँ का हाथ बँटाया.	राखी ने काम में राखी माँ का हाथ बँटाया.	3
राज अपना काम कर गया.	राज राज काम कर गया.	3
ममता अपने बच्चों को खाना खिला रही है.	ममता ममता बच्चों को खाना खिला रही है.	3
कलाकार अपने हुनर से सब का दिल जीत रहे थे.	कलाकार कलाकार हुनर से सब का दिल जीत रहे थे.	3
रमा ने अपने भाई को पढ़ाने की ज़िम्मेदारी ली.	रमा ने रमा भाई को पढ़ाने की ज़िम्मेदारी ली.	3
रमा ने अपना काम पूरा किया.	रमा ने रमा काम पूरा किया.	3

Examples of Anaphora Resolution for First and Second Pronoun		
प्रधानमंत्री ने जनता से कहा कि आप मेरा मत्त स्वीकार करें.	प्रधानमंत्री ने जनता से कहा कि जनता प्रधानमंत्री मत्त स्वीकार करें.	2
अभय ने राज से कहा कि तुम मेरे साथ चलो.	अभय ने राज से कहा कि राज अभय साथ चलो.	3
राजेश ने लोगों से कहा कि आप मेरा चुनाव करें.	राजेश ने लोगों से कहा कि लोगों राजेश चुनाव करें.	2
रमा ने रानी से कहा कि तुम मेरी प्रिय सखी हो.	रमा ने रानी से कहा कि रानी रमा प्रिय सखी हो.	3
रामू ने मालिक से कहा कि आप मेरी तनख़ाह दे दो.	रामू ने मालिक से कहा कि मालिक रामू तनख़ाह दे दो.	3
उमा ने आडवाणी से कहा कि आप मेरा इस्तीफ़ा स्वीकार करें.	उमा ने आडवाणी से कहा कि आडवाणी उमा इस्तीफ़ा स्वीकार करें.	3
रजनी ने रीमा से कहा कि तुम मेरे घर खाने पर आना.	रजनी ने रीमा से कहा कि रीमा रजनी घर खाने पर आना.	3
राधा ने रमा से कहा कि आप मेरा तोहफा स्वीकार करें.	राधा ने रमा से कहा कि रमा राधा तोहफा स्वीकार करें.	3
राज ने रमेश से पूछा कि तुम मेरे साथ फिल्म देखने चलोगे.	राज ने रमेश से पूछा कि रमेश राज साथ फिल्म देखने चलोगे.	3
बच्चे ने माँ से कहा कि आप मेरा खाना बना दें.	बच्चे ने माँ से कहा कि माँ बच्चे खाना बना दें.	3

The score is summed up for the illustrations considered to ascertain the precision of the framework. For the 40 sentences considered in table the framework creates an accuracy of 70%. The system resolves anaphors but fails to keep them the meaning of the sentence intact. The system works more efficiently for spatial pronouns. The resolution of anaphors of first and second person pronoun shows that there is no need to resolve them as the entire meaning of sentence changes.

Chapter 6
Conclusion and Future Scope

6.1 Conclusion

Anaphora resolution is an identification process of referent whereby interpretation of an occurrence of one expression depends on the interpretation of occurrence of another. The hearings impaired understand simple sentences. They might not be able to identify the referent of anaphor. So, a rule based anaphora resolution system has been proposed. In this system the parts of speech are used as a feature to extract the pronoun in a sentence. The dependency structure plays an important role. It represents one-to- one correspondence between the words in a sentence which is based on the head-modifier relationship. On this basis algorithms have been written to handle four pronominal anaphora forms which are; Reflexive, Spatial, Relative, and First-Second person pronouns. The system has been tested for approximately 300 sentences. The score has been given manually out of 5 according to their accuracy and correctness on the basis of Grammar and syntax. Depending upon the score generated (out of 5) the system results in accuracy of 70%.

6.2 Limitations

Some of the limitations of the system are as follows:

a) Among three sorts of Anaphoras, just a single is taken care of in the created framework that is pronominal type of anaphora.

b) In the event that Dependency parser brings about various dependency relations, the framework neglects to deliver yield that is seen from it.

c) The framework is not equipped for dealing with complex and compound sentences.

d) The framework is not dynamic for the utilization of others.

e) There is just a limited set of sentences for which framework has been tried.

f) The sentences have been made physically for the rule based anaphora resolution system and are exceptionally basic.

g) The system resolves the anaphors but don’t produce results grammatically correct.

6.3 Future Scope

The following can be achieved in the future for the framework.

a) The framework can be prepared to deal with different sorts of anaphora like Demonstratives, Gaps and so on as well.

b) The Rule based anaphora determination framework can be made dynamic for the use of others.

c) More calculations for the framework can be composed to deal with complex and compound sentences.

d) More rule sets can be shaped to deal with more dependency relations come about by the Hindi Dependency Parser.

e) No. of sentences can be expanded for testing reason.

References

[1] Zeshan U., "Indo-Pakistan Sign Language Grammar:A Typological Outline", Sign Language Studies, pp. 157-212, 2003

[2] Zeshan U., "Sign language in Indo-Pakistan: A description of a signed language", John Benjamins Publishing, 2000

[3] "Hindi Dependency Parser" [Online] Available: http://sivareddy.in/downloads [Accessed May 2017]

[4] Johnson, Jane E., and Russell J. Johnson., "Assessment of regional language varieties in Indian Sign Language", Report submitted to Work Papers of the Summer Institute of Linguistics, University of North Dakota, pp. 90-101,2005

[5] Zeshan U., Vasishta M. N. and Sethna M., "Implementation of Indian Sign Language in educational settings", Asia Pacific Disability Rehabilitation Journal, pp. 116-134, 2014

[6] World Federation of the Deaf: Sign Language, [Online] Available: http://wfdeaf.org/human-rights/crpd/sign-language [Accessed February 2017]

[7] T. Dasgupta and A. Basu, "Prototype machine translation system from text-to-Indian sign language", Proceedings of the 13th international conference on intelligent user interfaces, pp. 313-316, 2008

[8] T. Dasgupta, S. Shukla, S. Kumar, S. Diwakar, and A. Basu, "A Multilingual Multimedia Indian Sign Language Dictionary Tool", Proceedings of International Joint Conference on Natural Language Processing, Hyderabad, India, pp.57-64, 2008

[9] S. Samar Sinha, "A skeletal grammar of Indian Sign Language", PhD thesis under supervision of A. Kidwai, 2003

[10] A. K. Sahoo, G. S. Mishra and K. K. Ravulakollu, "Sign Language Recoginition: State of the Art", ARPN Journal of Engineering and Applied Sciences, vol. 9, no. 2, pp. 116-134, 2014

[11] Cox et al., "Tessa, a system to aid communication with deaf people", proceedings of the fifth international ACM conference on Assistive technologies. ACM, 2002

[12] M. Mohandes, "Automatic translation of Arabic text to Arabic sign language", AIML Journal 6.4, pp. 15-19, 2006

[13] P. Kar et al., "Ingit: Limited domain formulaic translation from hindi strings to indian sign language", ICON, 2007

[14] T. Veale et al., "The challenges of cross-modal translation: English-to-Sign-Language translation in the Zardoz system", Machine Translation 13.1 pp: 81-106, 1998

[15] “ISL text from Hindi”: http://islfromtext.in/working.php [Accessed June 2017]

[16] Dutta K, Prakash N and Kaushik S., "Resolving pronominal anaphora in hindi using hobbs algorithm", Web Journal of Formal Computation and Cognitive Linguistics. 2008 Jan; 1(10):5607

[17] Hobbs J., "Resolving pronoun references", Lingua, 1978 Jan; 44:311–38

[18] Carter DM, "A shallow processing approach to anaphor resolution", PhD thesis, University of Cambridge, 1987

[19] Carbonell JG and Brown RD., "Anaphora Resolution: A Multi- Strategy Approach", proceedings 12th International Conference on Computational Linguistics, Budapest. 1988. p. 96–101

[20] Mitkov R., "Robust anaphora with limited Knowledge", COLING98, Canada. 1994

[21] Mitkov R., "Anaphor Resolution: a combination Of linguistic and statistical approaches.", proceedings Discourse Anaphora and Anaphor Resolution, UK. 1996

[22] Sinha RM and Jain A. Angla, "Hindi: an English to Hindi machine-aided translation system", MT Summit IX, New Orleans, USA. 2003 Sep 23; 494–7

[23] Pal TL, Dutta K and Singh P., "Anaphora Resolution in Hindi: Issues and Challenges", International Journal of Computer Applications. 2012 Mar; 42(18)

[24] Lakhmani P and Singh S., "Anaphora Resolution in Hindi Language", International Journal of Information and Computation Technology. 2013; 3:609–16

[25] Singh S, Lakhmani P, Mathur P and Morwal S., "Anaphora resolution in hindi language using gazetteer method", International Journal on Computational Sciences and Applications IJCSA. 2014 Jun; 4:567–9

[26] Chopra D and Purohit GN, "Handling ambiguities and unknown words in named entity recognition using anaphora resolution", International Journal on Computational Sciences and Applications IJCSA. 2013 Oct; 3:456–63

[27] Uppalapu B and Sharma DM, "Pronoun Resolution for Hindi", proceedings DAARC2009. 2009 Apr 22; 5847

[28] Lalitha Devi S, Sundar Ram V and Rao PRK, "A Generic Anaphora Resolution Engine for Indian Languages", proceedings 25th International Conference on Computational Linguistics, Coling. 2014. p. 67–84

[29] Dakwale P, Mujadia V and Sharma DM, "A Hybrid Approach for Anaphora Resolution in Hindi Praveen", proceeding 6th International Joint Conference on Natural Language Processing, IJCNLP, Nagoya, Japan. 2013 Oct 14-18:80–6

[30] Duttaa K, Kaushikb S and Prakash N., "Machine Learning Approach for the Classification of Demonstrative Pronouns for Indirect Anaphora in Hindi News Items", The Prague Bulletin of Mathematical Linguistics. 2011 Apr; 95:33–50

[31] Lakhmani P, Singh S and Mathur P., "Gazetteer Method for Resolving Pronominal Anaphora in Hindi Language", International Journal of Advances in Computer Science and Technology. 2014 Mar; 3

List of Publications

Research Paper

Deepakshi Singla and Parteek Kumar, "Rule based Anaphora Resolution in Hindi" *in International conference on Computational Intelligence in Data Science*, ICCIDS 2017, Chennai, India

[Accepted]

Video URL

A Video has been uploaded on YouTube to describe the working of the system named "Rule Based Anaphora Resolution for ISL Generation System". The URL of the video is as follows:

https://youtu.be/DO6g97yDcR4

Plag Report

ORIGINALITY REPORT

%10	%9	%2	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.i-scholar.in Internet Source	%7
2	anilkumarsingh.me Internet Source	%1
3	jovianlin.com Internet Source	<%1
4	Vandan Mujadia, Darshan Agarwal, Radhika Mamidi, Dipti Misra Sharma. "Paninian grammar based hindi dialogue anaphora resolution", 2015 International Conference on Asian Language Processing (IALP), 2015 Publication	<%1
5	Smith, Robert, and Brian Nolan. "Manual evaluation of synthesised sign language avatars", Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS 13, 2013. Publication	<%1
6	Lecture Notes in Computer Science, 2010. Publication	<%1