

Structural Ambiguity in Hindi

Madhu Priya

Central Institute of Indian Languages, Mysuru 570006, Karnataka, India
madhupriyapathak@gmail.com

Abstract

Structural ambiguity resolution has been always the most important problem in Natural Language Processing (NLP) because parser often fail to handle these types of problems. Structural ambiguity occurs when a phrase or sentence has more than one underlying structure. In this research article I discuss the structural ambiguity and its resolution in Hindi by using hybrid approach. The approach combines both linguistics and probabilistic approach.

Keywords: Natural language processing, structural ambiguity, Hindi, phrases, sentences.

1. Introduction

We live in the world where we constantly face ambiguous information. A word, phrase, or sentence is ambiguous if it has more than one meaning. Ambiguity resolution has always been the most important testing ground in linguistics for parsing models. There are some ways of classifying ambiguity. Hurford and Hesley [1] group ambiguity into two, **lexical** and **structural ambiguity**. Structural ambiguity occurs when the meaning of the component words can be combined in more than one way [2]. One of the central issues in natural language analysis is structural ambiguity resolution. Correct phrase attachment determines the quality of structural analysis and as a consequence the semantic analysis of the natural language sentence. Contextual information, linguistics knowledge and world knowledge are the important requirements for successful resolution of structural ambiguity.

Hindi is the third most spoken language of the world. There are many ambiguous expressions found in Hindi at the structural level. For example,

maine ne gali se aate hue ladke ko dekhaa.

I ne. Erg street by.postposition coming. IMP.PTCP.M boy.M DAT see.PERF.M.SG
I saw a boy coming from the street.

This is an ambiguous sentence because it has more than one meaning-

a. I am coming from the street. b. Boy is coming from the street

A human being can easily analyze the ambiguous sentence. This is because they, using contextual information and world knowledge from their lifetime experience. Automatic systems lack such common sense. To make the ambiguous sentences unambiguous and grammatical, it is

necessary to have some sort of linguistics and world knowledge which help the machine recognize the sentence structure.

In this paper I describe different types and structural ambiguity in Hindi. Also, I will discuss about the process of disambiguation. Then, the findings are discussed. Finally, the paper ends with a conclusion and suggestions for further research.

2. Related Works

A number of works has been done in this area which are described below

- I. Ratnaparkhi, Reynar and Roukos: PP-attachment ambiguity resolution, Maximum Entropy model for English with 82 % accuracy [3]
- II. Nakov and Hearst: PP-attachment ambiguity resolution, Web as a training set for English with 83.82% accuracy [4]
- III. Zhao and Lin: PP-attachment ambiguity resolution, Nearest Neighbor Method for English with 86.5% accuracy [5]
- IV. Jignashu Parikh, Jagadish Khot, Shachi Dave and Pushpak Bhattacharyya: Predicate preserving parsing UNL for English with 78% accuracy [6]
- V. Patrick Pantel and Dekang Lin, An unsupervised approach to prepositional phrase attachment using contextually similar words unified framework for English with 84% accuracy [7]
- VI. Eric Brill and Philip Resnik: A rule-based approach to prepositional phrase attachment disambiguation, transformation-based error driven learning to learn the transformation rule for English with 81.1 % accuracy. [8]
- VII. Ashish Almeida: Prepositional Phrase Attachment and Generation of Semantic Relations UNL, Knowledge based approach for English with 84% accuracy. [9]
- VIII. Medimi Srinivas and Pushpak Bhattacharyya: Prepositional Phrase Attachment through Semantic Association using Connectionist Approach, UNL and Corpus based approach for English with 79.1% accuracy [10]
- IX. Muhammad Bilal, Mohammad Abid Khan, Rahman Ali and Rashid Ahmed: An adaptive approach of syntactic ambiguity resolution in Pashto, Knowledge based adaptive approach with 90% accuracy.[11]
- X. Niladri Chatterjee, Shailly Goyal, and Anjali Naithani: Resolving Pattern Ambiguity for English to Hindi Machine Translation Using WordNet Rule based approach English-Hindi with 75% accuracy [12].

In these works, the accuracy in ambiguity resolution is one of the major issues due to improper knowledge of linguistics in NLP.

3. Structural Ambiguity in Hindi

Structural ambiguities arises when a sequence of words has more than one syntactic interpretation. There are mainly two types of structural ambiguity 1. Coordination ambiguity and 2. Attachment ambiguity.

3.1. Coordination Ambiguity Coordination ambiguity arises by coordinate conjunction *aur*. There are three different structures in coordination ambiguity.

Structure-1: Noun1 +Conjunction+Noun2+ Noun3

geeta aur sudha ki sahelieyan ja rahin hain.

Geeta.F Conj Sudha.F.Sg GEN Friend.F.PL go.PROG be

Geeta and Sudha's friends are going.

It has two different possible interpretations: a) Geeta and Sudha's friends are going. b) Friends of both Geeta and Sudha are going.

Structure-2: Adjective + Noun1 + Conjunction + Noun2

bujurg mahilayen aur purush ja rahe hain.

old women.F.PL Conj men.M.PL go.PROG be

Old women and men are going.

It has two possible interpretations: a. ((Old women and men) are going.) b. ((Old women) and men are going.)

Structure-3: Noun1 + Noun2 + Conjunction + Noun3

Steel ki katori aur plate

Noun GEN bowl Conj Plate

steel bowl and plate

Two possible interpretations a) ((steel bowl) and plate.) and b) (steel (bowl and plate))

3.2. Attachment Ambiguity

Attachment ambiguity arises from uncertainty of attaching a phrase or clause to a part of sentence. Attachment problems are mostly problems of modifier placement.

3.2.1. Noun Attachment

Structure-1: Noun1 + Noun2

haveli prasad kha raha hai.

Haveli.M pro/common noun eat PROG be

Haveli Prasad is eating.

There are two different possible interpretations. a) Haveli Prasad is eating. b) Haveli is eating Prasad (Prasad is a religious offering in Hinduism. Most often Prasad is vegetarian food especially cooked for devotees after praise and thanks giving to the Lord.).

Structure-2: Noun1 + Adjective +Noun2

mohan lal shirt pahana hai.

Mohan.M prop N/Adj Shirt wear be

Mohan is wearing a red shirt.

There are two different possible interpretations: a) Mohan is wearing a red shirt. b) Mohan lal is wearing a shirt.

Structure-3: Pronoun + Adjective + Noun

wo pagal ladki ko dekh raha hai.

He.M mad girl. F GEN see PROG be

He is looking at the mad girl.

There are two different possible interpretation a) He is looking at the mad girl. b) That mad boy is looking at the girl.

3.2.2. Adjective Attachment

Structure-1:

Adjective + Noun + GEN (ka/ki) + Noun

purani kitab ki dukan

old books GEN shop

Old book shop.

There are two different possible interpretation: a) Books are old. b) Shop is old.

Structure- 2:

Adjective + Adjective + Noun

atyadhik vaastvik vivaran

Most real data

3.2.3. Adverb Attachment

Structure-1: Noun + adverb + Noun + VP

seet par baithe hue kutte ko bandhiye.

Seat on.Postposition sit. IMP.PTCP. dog.M. DAT tie. PRESENT TENSE. M.SG

Tied the dog sitting on the seat.

a. Tied when you are sitting on the seat. b. Tied when dog is sitting on the seat.

4. Structural Ambiguity Resolution

Structural ambiguity resolution is based on hybrid approach which is combination of linguistics and probabilistic approach.

4.1. Linguistics Approach

Sometimes structures of sentence are same as ambiguous sentences but actually meaning is not ambiguous. Ambiguity occurs in some specific conditions. These conditions are based on semantic: similarity, selectional preference and other linguistic cues.

4.1.1. Semantic Similarity

In coordination ambiguity sentence become ambiguous if and only if semantic features of words are same.

For example,

a) *geeta (noun1) aur sudha (noun2) ki saheliyean (noun3) ja rahin hain*

Geeta.F Conj Sudha.F.Sg GEN Friend.F.PL go.PROG be

Geeta and Sudha's friends are going.

This is an ambiguous sentence because noun1, noun2, noun3 have same semantic features.

b) *mujhe chaaval aur chane kee daal pasand hai* (I like rice and chickpeas.)

pronoun noun1 conj noun2 pp noun3 vp

This sentence is not ambiguous because here semantic features of noun1 (-animate,+grain), noun2 (-animate, +type of grain, subset of noun3) and noun3 (-animate, +grain) are different.

In semantic similarity conjoined heads appear to play an important role:

- a) **kachche aam aur papite**
raw. modifier mango.noun1 and.conj papaya.noun2
raw mangoes and papayas

This sentence is ambiguous because modifier modify noun1 and noun2

- b) **Kachche aam aur pudina**
raw. modifier mango. noun1 and. conj mint.noun2
raw mango and mint

This sentence is unambiguous because modifier modify noun1 only.(mint leaves does not have the property of being raw/ripe.)

4.1.2. The Appropriateness of Noun-Noun Modification

- c) **aam aur pudine ki chatni**
Mango.noun1 and. conj mint.noun2 GEN sauce.noun3
Mango and Mint Chutney

(ambiguous because noun1, noun2 belongs to noun3)

- d) **haldiraam ki bhujiaa aur mithaai**
Haldiram.noun1 GEN noun2.fries and.conj sweets. noun3
Haldiram's Bhujia and Sweets

(ambiguous because noun1, noun2 belong to noun3)

- e) **thode se rupaye aur chandi ke gahanon se maan jaayegi**
modifier. little money.noun1 and. conj noun2.silver GEN jewellery.noun3 agree.vp
Will agree with a little money and silver jewelry

This sentence is unambiguous because modifier modify noun1 only.

4.1.3. Lexical Preferences play an important role in disambiguation. For example,

- g) **raam ne daudhate hue ghode ko dekhaa.** (ambiguous)

ram.M erg run.IMP.PTCP.M horse.M DAT see.PERF.M.SG
Ram saw the running horse.

a. Ram saw the running horse. b. Ram saw the horse while ram is running.

- h) **raam ne daudhate hue ball ko fekaa.** (unambiguous)

ram.M erg run.IMP.PTCP.M ball.M DAT throw.PERF.M.SG
Ram threw the ball while running.

(unambiguous because the ball does not have the property of running.)

4.1.4. Grammatical Conditions

1. In attachment ambiguity if postposition comes after noun, then sentence is not ambiguous. For example,

- i) **haveli prasad ne khana khaya.** (unambiguous)

Haveli Prasad.M erg food. ate. PERF.M.SG
Haveli Prasad ate food.

This sentence is not an ambiguous sentence. Here 'ne' postposition comes after 'prasad'

2. If another surname or last name comes after noun, sentence is not ambiguous. For exam-

Hanuman Prasad Shukla vishwavidyalaya ja rhe hain.
Hanuman Prasad Shukla.M university go. PROG be PROG be
Hanuman Prasad Shukla is going to university.

3. In adverbial phrase attachment. if agreement of adverb has masculine gender, then sentence is ambiguous. otherwise not.

j) raam ne daudhate hue ghode ko dekhaa. (ambiguous)

ram.M erg run.IMP.PTCP.M horse.M DAT see.PERF.M.SG

Ram saw the running horse.

a. Ram saw the running horse. b. Ram saw the horse while ram is running

k) raam ne daudhati hui bachchi ko dekhaa.

ram.M erg run.IMP.PTCP.M girl.M DAT see.PERF.M.SG

Ram saw the running girl.

(unambiguous because there is agreement between adverb 'daudhati' and 'bachchi')

4.2. Probabilistic Approach

I have give questionnaire of 100 ambiguous sentences with its two possible interpretation to 45 students of Mahatma Gandhi Antarrashtriya Hindi Vishwavidyalaya, Wardha Maharashtra.

On the basis of their selection of meaning I have calculated percentage of probability of most selected meaning. These are the results -

- Probability of Noun+VP (56%) is higher than Noun+Noun (36%).
- Probability of attachment of adjective + noun (37%) is higher than probability of adjective (Noun + noun) (28.3%).
- Probability of attachment of adverb+ noun2 (19%) is higher than noun1+ adverb (16%).

In group of 45 students it has been observed that maximum number of students interpreted the given sentences in only one possible way. The second interpretation was identified by a very few students. A few questions in the handout were left not attempted by some students.

Results: Evaluation of linguistics rule on LDC-IL, CIIL, Hindi corpus, overall 42.53 % ambiguous sentence resolved by linguistics knowledge. 57.47% resolved by probabilistic theory.

Conclusion: Structural ambiguities resolution is a challenging issue in natural language processing. In Hindi, structural ambiguity is found in noun-noun attachment, noun adjective attachment, adjective-noun attachment, and noun-adverb attachment. On the basis of linguistic approach, learners disambiguate 42.53% ambiguous sentences and 57.47% resolve it by probabilistic theory.

References

1. Hurford, James R. 1983. *Semantics: A Coursebook*. Cambridge: Cambridge University Press.
 2. O’Grady, William, Michael Dobrovolsky, and Mark Aronof. 1997. *Contemporary Linguistics: An Introduction*. 3rd Edition. New York: St.Martin’s Press.
 3. Adwait Ratnaparkhi, Jeff Reynar, and Salim Roukos. 1994. A maximum entropy model for prepositional phrase attachment. In *Proceedings of the ARPA Workshop on Human Language Technology*, pages 250–255.
 4. Nakov, P. and Hearst, M. 2005. Using the web as an implicit training set: Application to structural ambiguity resolution. In *Proceedings of HLTNAACL*.
 5. Zhao, S and Lin, D. 2004. A Nearest-Neighbor Method for Resolving PPAttachment Ambiguity. In *Proceedings of IJCNLP-04*
 6. Jignashu Parikh, Jagadish Khot, Shachi Dave and Pushpak Bhattacharyya .2003 . Predicate preserving parsing. In *Proceeding*.
 7. Patrick Pantel and Dekang Lin. 2000. An unsupervised approach to prepositional phrase attachment using contextually similar words. In *Proceedings of ACL*.
 8. Eric Brill and Philip Resnik. 1994. A rule-based approach to prepositional phrase attachment disambiguation. In *Proceedings of COLING*.
 9. Ashish Almeida. *Prepositional Phrase Attachment & Generation of Semantic Relation*. Master Thesis.
 10. Medimi Srinivas and Pushpak Bhattacharyya. 2006. Unsupervised pp attachment disambiguation using semantics Prepositional Phrase Attachment through Semantic Association using Connectionist Approach. In *Proceedings, The Third International WordNet conference (GWC-06)*.
 11. Muhammad Bilal, Mohammad Abid Khan, Rahman Ali and Rashid Ahmed. 2011. An adaptive approach of syntactic ambiguity resolution in Pashto, *African Journal of Mathematics and Computer Science Research* Vol.4(2), pp. 55-63.
 12. Niladri Chatterjee, Shailly Goyal and Anjali Naithani: Pattern Ambiguity and its Resolution in English to Hindi Translation, in the *Proceedings of International Conference " Recent Advances in Natural Language Processing-2005"*, ISBN: 954-91743-3-6, Borovets, Bulgaria, 2005, pp152 – 156.
-