



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

Ambiguities in Natural Language Processing

Anjali M K¹, Babu Anto P²

Department of Information Technology, Kannur University, Kerala, India^{1,2}

ABSTRACT: Ambiguity can be referred as the ability of having more than one meaning or being understood in more than one way. Natural languages are ambiguous, so computers are not able to understand language the way people do. Natural Language Processing (NLP) is concerned with the development of computational models of aspects of human language processing. Ambiguity can occur at various levels of NLP. Ambiguity could be Lexical, Syntactic, Semantic, Pragmatic etc. This paper presents a study about different types of ambiguities that comes under Natural Language Processing.

KEYWORDS: Ambiguity, Natural Language Processing, Lexical, Syntactic, Semantic, Anaphora, Pragmatic.

I. INTRODUCTION

Natural Language Processing (NLP) is an area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things [1]. The Text based NLP has been regarded as consisting of various levels. They are:

Lexical Analysis:- Analysis of word forms

Syntactic Analysis:-Structure processing

Semantic Analysis:- Meaning representation

Discourse Analysis:- Processing of interrelated sentences

Pragmatic Analysis:-The purposeful use of sentences in situations.

Ambiguity can occur at all these levels. It is a property of linguistic expressions. If an expression (word/phrase/sentence) has more than one interpretation we can refer it as ambiguous.

For eg: Consider the sentence,

The chicken is ready to eat

The interpretations in the above phrase can be, The chicken (bird) is ready to be fed or The chicken (food) is ready to be eaten.

Consider another sentence,

There was not a single man at the party

The interpretations in this case can be Lack of bachelors at the party or Lack of men altogether.

II. DIFFERENT TYPES OF AMBIGUITY

There are different types of ambiguities

2.1 Lexical Ambiguity: is the ambiguity of a single word. A word can be ambiguous with respect to its syntactic class. Eg: book, study.

For eg: The word silver can be used as a noun, an adjective, or a verb.[2]

She bagged two silver medals.

She made a silver speech.

His worries had silvered his hair.

Lexical ambiguity can be resolved by Lexical category disambiguation i.e, parts-of-speech tagging. As many words may belong to more than one lexical category part-of-speech tagging is the process of assigning a part-of-speech or lexical category such as a noun, verb, pronoun, preposition, adverb, adjective etc. to each word in a sentence.

2.1.1 Lexical Semantic Ambiguity: The type of lexical ambiguity, which occurs when a single word is associated with multiple senses. Eg: bank, pen, fast, bat, cricket etc.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

For eg:

The tank was full of water.

I saw a military tank.

The occurrence of tank in both sentences corresponds to the syntactic category noun, but their meanings are different.

Lexical Semantic ambiguity resolved using word sense disambiguation (WSD) techniques, where WSD aims at automatically assigning the meaning of the word in the context in a computational manner.

2.2 Syntactic Ambiguity: The structural ambiguities were syntactic ambiguities.

Structural ambiguity is of two kinds: Scope Ambiguity and Attachment Ambiguity.

2.2.1 Scope Ambiguity: Scope ambiguity involves operators and quantifiers.

Consider the example:

Old men and women were taken to safe locations.[3]

The scope of the adjective (i.e., the amount of text it qualifies) is ambiguous. That is, whether the structure (old men and women) or ((old men) and women)?

The scope of quantifiers is often not clear and creates ambiguity.

Every man loves a woman.[7]

The interpretations can be, For every man there is a woman and also it can be there is one particular woman who is loved by every man.

2.2.2 Attachment Ambiguity

A sentence has attachment ambiguity if a constituent fits more than one position in a parse tree. Attachment ambiguity arises from uncertainty of attaching a phrase or clause to a part of a sentence.[3]

Consider the example:

The man saw the girl with the telescope.[2]

It is ambiguous whether the man saw a girl carrying a telescope, or he saw her through his telescope.

The meaning is dependent on whether the preposition 'with' is attached to the girl or the man.

Consider the example:

Buy books for children

Preposition Phrase 'for children' can be either adverbial and attach to the verb buy or adjectival and attach to the object noun books.

2.3 Semantic Ambiguity: This occurs when the meaning of the words themselves can be misinterpreted. Even after the syntax and the meanings of the individual words have been resolved, there are two ways of reading the sentence.

Consider the example,

Seema loves her mother and Sriya does too.[2]

The interpretations can be Sriya loves Seema's mother or Sriya likes her own mother.

Semantic ambiguities born from the fact that generally a computer is not in a position to distinguishing what is logical from what is not.

Consider the example:

The car hit the pole while it was moving.

The interpretations can be The car, while moving, hit the pole and The car hit the pole while the pole was moving. The first interpretation is preferred to the second one because we have a model of the world that helps us to distinguish what is logical (or possible) from what is not. To supply to a computer a model of the world is not so easy.[4]

Consider the example:

We saw his duck

Duck can refer to the person's bird or to a motion he made.

Semantic ambiguity happens when a sentence contains an ambiguous word or phrase.

2.4 Discourse: Discourse level processing needs a shared world or shared knowledge and the interpretation is carried out using this context. Anaphoric ambiguity comes under discourse level.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

2.4.1 Anaphoric Ambiguity: Anaphoras are the entities that have been previously introduced into the discourse. Consider the example,
The horse ran up the hill. It was very steep. It soon got tired.
The anaphoric reference of 'it' in the two situations cause ambiguity.
Steep applies to surface hence 'it' can be hill. Tired applies to animate object hence 'it' can be horse.

2.5 Pragmatic Ambiguity: Pragmatic ambiguity refers to a situation where the context of a phrase gives it multiple interpretation [2]. One of the hardest tasks in NLP. The problem involves processing user intention, sentiment, belief world, modals etc.- all of which are highly complex tasks.[3]

Consider the example,

Tourist (checking out of the hotel): Waiter, go upstairs to my room and see if my sandals are there; do not be late; I have to catch the train in 15 minutes.

Waiter (running upstairs and coming back panting): Yes sir, they are there.

Clearly, the waiter is falling short of the expectation of the tourist, since he does not understand the pragmatics of the situation.[3]

Pragmatic ambiguity arises when the statement is not specific, and the context does not provide the information needed to clarify the statement. Information is missing, and must be inferred.[5]

Consider the example

I love you too.[6]

This can be interpreted as

I love you (just like you love me)

I love you (just like someone else does)

I love you (and I love someone else)

I love you (as well as liking you)

III. CONCLUSION

This paper made a detailed study about Ambiguity which occurs across all levels of NLP. It is highly complex task to resolve these kinds of ambiguities, especially in upper levels of NLP. The meaning of a word, phrase, or sentence cannot be understood in isolation and contextual knowledge is needed to interpret the meaning, pragmatic and world knowledge is required in higher levels. It is not easy to create a world model for disambiguation tasks. Linguistic tools and lexical resources are needed for the development of disambiguation techniques. Resourceless languages are lagging behind in these field compared to resourceful languages in implementation of these techniques. Rulebased methods are language specific where as stochastic or statistical methods are language independent. Automatic resolution of all these ambiguities contains several long standing problems but again we can put a positive thinking towards the development of full-fledged disambiguation techniques which take care of all the ambiguities, because it is very much necessary for the accurate working of the NLP applications such as Machine Translation, Information Retrieval, Question Answering etc.

REFERENCES

1. Gobinda G Chowdhury, 'Natural Language Processing', http://www.cis.strath.ac.uk/cis/research/publications/papers/strath_cis_publication_320.pdf.
2. Tanveer Siddiqui, U S Tiwary, 'Natural Language Processing and Information Retrieval', Oxford University Press.
3. Pushpak Bhattacharyya, 'Natural Language Processing: A Perspective from Computation in Presence of Ambiguity, Resource Constraint and Multilinguality', CSI Journal of Computing | Vol. 1 • No. 2, 2012.
4. <http://swdevelopers.tripod.com/english/language/chap4.html>
5. <http://michealaxelsen.com/blog/?p=348>
6. plato.stanford.edu/entries/ambiguity/
7. <http://www.coli.uni-saarland.de/projects/milca/courses/comsem/html/node92.html>