Semantic Analysis

4

OBJECTIVES

After reading this chapter, the student will be able to understand :

- Lexical semantics
- Attachment for fragment of English
- Relations among lexemes and their senses
- WordNet
- Robust Word Sense Disambiguation (WSD), Dictionary based approach

SEMANTIC ANALYSIS

The process whereby meaning representations are composed and assigned to linguistic inputs.

Semantics involves figure out the meaning of linguistic input (construct meaning representations) and process language to produce common-sense knowledge about the world (extract data and construct models of the world).

Semantics is associated with the meaning of language. General idea of semantic interpretation is to take natural language sentences or utterances and map the monitor some representation off meaning semantic analysis is concerned with creating representations for the meaning of linguistic inputs this chapter deals with the meaning of written text we can divide semantics into two parts as follows:

- The study of meaning of individual words (lexical semantics) and
- the study of how individual words combine to give meaning to a sentence (or larger units).

Once we have the meaning of the words we need to combine them into meaning of the whole sentence the principle of semantic compositionality sometimes called freg's principle states that the meaning of the whole sentence is comprised of the meaning of its parts that is the meaning of the sentence can be composed from the meaning of its constituent words. Natural languages do not obey this principle of compositionality. Often, the meaning of the whole is only partially dependent on the meaning of its constituents (as in collocates) and sometimes entirely different from the meanings of individuals (e.g., in idioms).

Hence, this decomposition of semantics may not be realistic. Word meaning is just one component of semantics. The relationships that exist between words, the domain, the word order, the semantic structure, the underlying context and the real-world knowledge, all contribute to the meaning of a sentence. Still, compositional and lexical word semantics remain the dominant approach to semantics in compositional linguistics. There are theories supporting the idea that linguistic knowledge is knowledge about words. This means that the lexicon contains all the knowledge about language.

These theories completely dispense with grammar as an independent entity which has been accepted for so many years. The idea that Syntax and lexicon cannot be described adequately without reference to each other is gaining increasing support from the results of Corpus analysis which attempts to investigate the role of contexts in syntax and semantics.

Lexical semantics has been the starting point for all the early theories of semantics. The most common paradigm involves decomposing lexical meanings in terms of semantic primitives or atomic units of meaning. However, this theory turns out to be inadequate in handling compositional semantics. This has led to the development of model theoretic semantics which, along with the structural semantics, is the dominant approach to semantics within linguistics. Model theoretic semantics is inspired by the semantics that logicians used for formal logical languages. Logicians focus on logical words (and, or, not, if, all, some, only, etc.) and do not pay attention to non-logical words (most nouns, verbs and adjectives). The primary advantage of this theory is its ability to explain compositional semantics - how the meaning of a sentence is determined from the meaning of its parts. One important aspect off meaning is that it related sentences to the outside world. However, we do not know the world is. A model theoretic approach to semantics attempt to create a model of the world and determine the truth of a sentence using this model. In this theory, the truth of a sentence does not mean that sentence is actually true; it simply means that the sentence is true in the world being modelled. Model theoretic semantics is effective in studying pragmatic as well as semantics.

Among the source of knowledge typically used are

- the meanings of words,
- the meanings associated with grammatical structures,
- knowledge about the structure of the discourse,
- Knowledge about the context in which the discourse is occurring, and
- Common-sense knowledge about the topic at hand

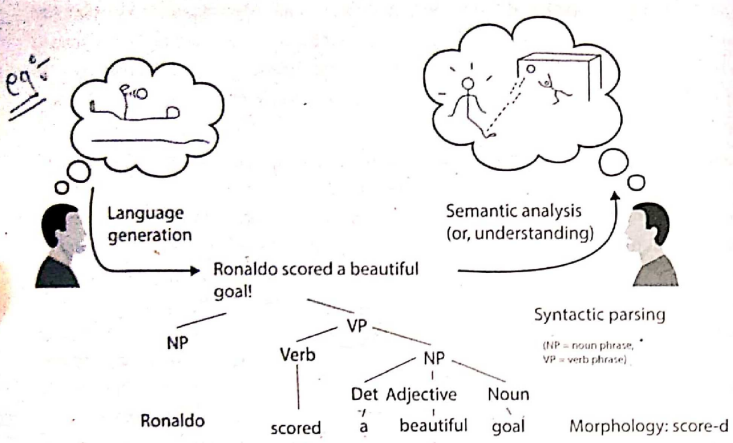

Figure 1: Understanding Semantics

Syntax-driven semantic analysis

Assigning meaning representations to input based solely on static knowledge from the lexicon and the grammar.

1. Lexical semantics

So far, we have focused on the creation of meaning representation of whole sentence. The approach we took supports the view that words contribute to the meaning of a sentence but do not themselves have meaning. It is perhaps this view that has led some to consider the lexicon as a simple list of words. However, this is narrow concept, too far from reality. Words have meanings; they have internal structure, and are involved in different relationship with other words. All this information can be captured in a systematic structure by lexicon. In this section, we focus on this and other issues related to words. More precisely, we focus on lexical semantics, which is concerned with the linguistic study of systematic, meaning related structure of words or lexemes (the minimal unit and lexicon). It involves meanings of component words

Compositional semantics

It involves how words combine to form larger meanings. So, if knowledge of each word's meaning isn't enough for comprehension of a sentence or phrase as it is meant to be understood, how is the overall compositional meaning derived? Obviously the overall meaning must have something to do with the meanings of those words contained within the phrase. "I don't eat" and "I don't drink", for example, express different ideas because of the difference in lexical meaning between "eat" and "drink." However, syntax, or the way in which the sentence is constructed , plays a role as well, as demonstrated below: 1. I like you. 2. You like me. Both sentences express totally different propositions, however they have the same words and each word has a clearly understood meaning. How is it that the meaning has changed? In sentence 1, "I" = subject, "you" = object, while in the second sentence these are reversed. This means that overall meaning relies not only on the meaning of each part, but additionally on syntactic composition. This premise is known as the principle of compositionality. All languages contain an infinite number of word combinations, so memorization of each separate phrasal meaning is impossible. This means that in order to understand the meanings of new phrases, one must rely on individual word meanings combined with the specific syntactic structure. The roles played by lexical and compositional semantics are equally necessary with regard to total understanding of a phrase or sentence. Knowledge of one without the other will invariably lead to miscommunication, and an understanding of denotation, connotation, and syntactical structure is necessary for compositional understanding of the whole. Understanding the difference between these terms and how the ideas they represent interact is paramount to understanding the meaning of a phrase or sentence.

What is language understanding

What exactly is meant by "understanding" an utterance?

- Extracting knowledge about the world from the utterance–in the case of concrete, physical things: ultimately translating linguistic input to physical and geometrical terms... uh huh
- "Understanding language means knowing how to use it.
- To master a technique. At matriculation examination: the student is presented with a portion of text or speech and then asked questions about it

- It is difficult to give one generic definition
- Thinking of a concrete application of semantic analysis is probably the best way of defining the problem

Semantic analysis vs. other areas of natural language processing

Phonetics: the study of linguistic sounds

Morphology: the study of the meaning components of words

scored = score –d = Verb score + past tense

employ, employee, employment, ...

Syntax: a study of the structural relationship between words

Semantics: the study of meaning

Pragmatics: the study of how language is used to accomplish goals; discourse conventions (turn taking, politeness, etc.); relation between language and context-of-use

Semantic analysis often requires syntactic parsing, pragmatics etc. too (in some form; not necessarily formal linguistics)

Approaches to semantic analysis

Predicate logic

The sentence "a restaurant that serves Chinese food near TUT" corresponds to the meaning representation $\exists x$ Restaurant(x) $\wedge$ Serves(x, ChineseFood) $\wedge$ Near(LocationOf(x), LocationOf(TUT) semantic analysis is equivalent to creating meaning representations from language.

scalability problem (large vocabulary or unrestricted domain)

Statistical approach

statistical machine translation (as an example)

find a bilingual database (e.g. parliamentary proceedings in two languages)

learn an alignment: words and phrases that correspond to each other

learn word order in the target language (probabilities of target word strings)

translated by matching source fragments against a database of real examples, identifying the corresponding translation fragments, and then recombining these to give the target text

Information retrieval

Google solves a certain part of the problem in a statistical way: answers to "trivial" kind of questions can be located using a web search engine assumes a database (Internet) and a clever page ranking system

Domain knowledge driven analysis

expect certain "slots" of information to be filled in football example in the beginning: hearer is aware of missing details and may expect to hear them another example: booking a flight restricting to a certain domain allows the use of specific patterns, rules, expectations, etc. customer at a restaurant buying train tickets

Applications of semantic analysis

Information extraction : extract small amounts of pertinent information from large bodies of text to find an answer to a question for example eg:- document searching

Text summarization : Information retrieval (cf. Google) and document classification

Machine translation

Human-computer interaction : conversational agents: book plane tickets, query for a restaurant

Expert systems : free help: "please show me how to widen the margins of my document"

Surveillance

Several well-defined problems and applications yet unsolved

List on the right is from Jim Gray s Turing talk: "What next? A few remaining problems in IT"

Why is semantic analysis difficult?

Ambiguity of language

"I made her duck", for example, could mean

I cooked waterfowl for her.

I created the (plaster?) duck she owns.

I caused her to quickly lower her head or body.

I waved my magic wand and turned her into undifferentiated waterfowl.

Commonsense knowledge is typically omitted from social communications example: "Laura hid George's car keys. He was drunk."

Language understanding often requires unsound inference abduction ((A □ B) and B) € infer A (which is not sound logic)

Language is dynamic: allows defining new terms, allegory, etc.

Why is semantic analysis important?

Power of language: transfer thoughts from a head to another transfer between brains and a computer as well?

Language is a very generic representation (the most generic?)

- words can describe almost anything
- ability to reason with language ability to reason about almost anything (assuming the ability to construct a model of the world, too)

2. Attachment for Fragment of English

Words in a sentence are not tied together as a sequence of part of speech. Language puts constraints on word order. For example, certain words both together with each other more than with others, and to behave as a unit. The fundamental idea of Syntax is that words grouped together to form constituents (often termed phrases), each of which acts as a single unit. They combine with other constituents to form larger constituents, and eventually, a sentence. *The bird, The rain, The Wimbledon Court, The beautiful garden* are all noun phrases that can occur in the same syntactic context. For example, they can all function as the subject or the object of a verb.

These constituents combine with others to form a sentence constituent. For example, the noun phrase *the bird* can combine with the verb phrase *flies* to form the sentence *the bird flies*. Different types of phrases have different internal structures. In this section, we discuss some of the major phrase types and try to build phrase structure rules to identify them.

Phrase level Constructions

As discussed earlier, a fundamental notion in natural language is that certain groups of words behave as constituents. These constituents are identified by their ability to occur in similar context. One of the simplest ways to decide whether a group of words is a phrase, is to see if it can be substituted with some other group of words without changing the meaning. If such a substitution is possible in the set of words forms a face. This is called the substitution test. Consider sentence below where we can substitute a number of phrases:

Henna reads a book.

Henna reads storybook.

Those girls read a book.

She read the comic book.

We can easily identify the constituents that can be replaced for each other in the sentences. These are *Henna*, *she*, and *Those girls* and *a book*, *a story book*, and *a comic book*. These are the words that form a phrase. In linguistics, such constituents represent a paradigmatic relationship. Elements that can substitute each other in certain syntactic positions are said to be members of one paradigm.

Phrase types are named after their head, which is a lexical category that determines the properties of the phrase. Thus, if the head is a noun, the phrase is called a noun phrase, if head is a verb, then the phrase is a verb phrase and so on for other lexical categories such as adjective and prepositions. Figure below shows the sentences with the noun phrase verb phrase and preposition phrase.

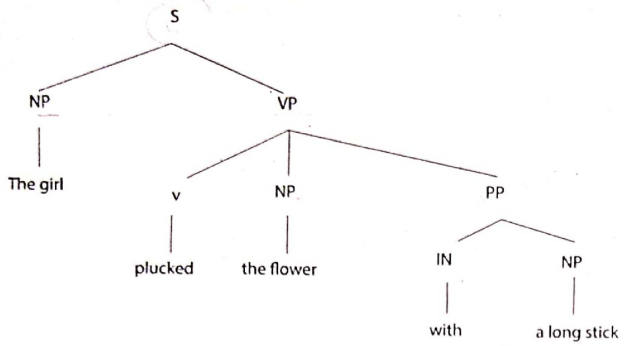

Figure 2: A sentence with NP, VP and PP

Noun phrase

A noun phrase is a phrase whose head is a noun or a pronoun, optionally accompanied by a set of modifiers. It can function as a subject, object or compliment. The modifiers of a noun phrase can we determine hours or adjective phrase. The obligatory constituent of a noun phrase is the noun head-all other constituents are optional. These structures can be represented using the phrase structure rule. As discussed earlier, phrase structure rules are of the form A →BC, which states that constituent A can be rewritten as two constitutes B and C. These rules specify the element can occur in a phrase and in what order. Using this notation, we can represent the phrase structure rules for a noun phrase as follows.

NP → Pronoun

NP → Det Noun

NP → Noun

NP → Adj Noun

NP → Det Adj Noun

We can combine all these rule in a single phrase structure rule as follows:

NP → (Det) (Adj) Noun|Pronoun

NP → (Det) (AP) Noun (PP)

NP → (Det) (AP) Nom (PP) Nom → Noun | Noun Nom

The constituents in parenthesis are optional. This rule states that a noun phrases consists of a noun, possibly preceded by determiner and adjective (in that order). This rule does not cover all possible NPs. A noun phrase may include post modifiers and more than one at adjective. For example, it may include a prepositional phrase (PP). More than one adjective is handled by allowing an adjective phrase (AP) for the adjective in the rule. After incorporating PP and AP in the phrase structure rule, we get the following

NP → (Det) (AP) Noun (PP)

The following are a few examples of noun phrases

They .... (1)

The foggy morning .... (2)

Chilled water .... (3)

A beautiful lake in Kashmir .... (4)

Cold banana shake .... (5)

Let's see how the above phrases can be generated using phrase structure rules. The (1) phrase consists only of a pronoun. The (2) phrase consists of a determiner, an adjective (foggy) that stands for an entire adjective phrase, and a noun. The (3) phrase comprises and adjective phrase and noun, (4) phrase consists of a determiner (the), an adjective phrase (beautiful), a noun (lake) and a prepositional phrase (in Kashmir) and (5) phrase consists of an adjective followed by a sequence of nouns. A noun sentence is termed as nominal. None of the phrase structure rules discussed so far are able to handle nominals. So, we modify our rules to cover this situation.

NP → (Det) (AP) Nom (PP)

Nom → Noun | Noun Nom

A noun phrase can act as a subject, an object or a predicate. The following sentences demonstrate each of these uses.

The foggy damped weather disturbed the match. .... (1)

I would like a nice cold banana shake. .... (2)

Kula botanical garden is a beautiful location. .... (3)

In (1), the noun phrase acts as subject. In (2), its acts as an object, and in (3), it is a predicate.

Verb Phrase

Analogous to the noun phrase is the verb phrase, which is headed by of verb. There is a fairly wide range of reasons that can modify of verb. This makes verb phrase a bit more Complex. The verb phrase organises various elements of the sentence that depends syntactically on the verb find the following are some examples of phrases:

Khushbu slept. …. (1)

The boy kicked the ball. …. (2)

Khushbu slept in the garden. …. (3)

The boy gave the girl a book. …. (4)

The Boy gave the girl a book with blue cover. …. (5)

As you can see from these examples a verb phrase can have a verb {VP → Verb in (1)}, a verb followed by an NP {VP → Verb NP in (2)}, a verb followed by a PP {VP → Verb PP in (3)}, a verb followed by two NPs {VP → Verb NP NP in (4)}, or a verb followed by two NPs and a PP {VP → Verb NP NP PP in (5)}. In general, the number of NPs in BP is limited to two, whereas it is possible to add more than two PPs.

VP → Verb (NP) (NP) (PP)*

Things are further complicated by the fact that objects may also be entire clauses as in the sentence, *I know that Taj is one of the seven wonders*. Hence, we must also allow for an alternative phrase statement rule, in with NP is replaced by S.

VP → Verb S

Prepositional Phrase

Prepositional phrases are headed by a preposition. They consist of a preposition, possibly followed by some other constituent, usually a noun phrase.

We played volleyball on the beach.

We can have a preposition phrase that consists of just a preposition.

John went outside.

The phrase structure rule that captures the above eventualities is as follows.

PP → Prep (NP)

Adjective Phrase

The head of an adjective phrase (AP) is an adjective. AP consists of an adjective, which may be preceded by an adverb and followed by a PP. Here are some examples.

Ashish is love.

The train is very late.

My sister is fond of animals.

The phrase structure rule for adjective phrase is

AP → (adv) Adj (PP)

Adverb Phrase

An adverb phrase consists of an adverb, possibly preceded by a degree adverb. Here is an example.

Time passes very quickly.

AdvP → (Intens) Adv

Sentences

Having discussed phrase structures, sentences. A sentence can have bearing structure commonly known structures are declarative structure, imperative structure, yes-no question structure and WH question structure.

Sentences with a declarative structure have a subject followed by a predicate. The predicate of a declarative sentence is a noun phrase and predicate is a verb phrase, e.g., *I like horse riding*. The phrase structure rule for declarative sentences are sentences is

S → NP VP

Sentences with an imperative structure usually begin with the verb phrase and lack subject. The subject of these types of sentences is implicit and is understood to be "you". These types of sentences are used for commands and suggestions, and hence are called imperative. The grammar rule for this kind of sentence structure is

S → VP

Examples of this kind of sentences are as follows:

Look at the door.

Give me the book.

Stop talking.

Show me the latest design.

Sentences with the yes-no question structure ask questions which can be answered using yes or no. These sentences begin with an auxiliary verb, followed by a subject NP, followed by a VP. Here are some examples

Do you have a Red pen?

Is awake and cotter?

Is the game over?

Can you show me your album?

We expand our grammar by adding another rule for the expansion of as, as follows:

S → Aux NP VP

Sentences with WH question structure are more complex. These sentences begin with WH words (who, which, where, what why and how). A WH question may have a WH phrase as a subject or may include another subject. Consider the following WH questions:

Which team won the match?

This sentence is similar to a declarative sentence except it contains a WH word. How simple rule to handle this type of sentence structure is

S → Wh-NP VP

Another type of WH questions structure is one that involves more than one NP. In this type of questions, the auxiliary verb comes before the subject NP, just as in yes-no questions structure.

Which cameras can you show me in your shop?

The rule for this type of WH questions is

- S → Wh-NP Aux NP VP

3. Relations among lexemes & their senses – Homonymy, Polysemy, Synonymy, Hyponymy

One way to approach lexical semantics is to study the relationship among lexemes (an abstract representation of a "word", the lexical entry in a dictionary). Semantics of a lexeme can be understood by analysing the relationship of lexemes with other lexemes. Lexical semantics information is useful for wide variety of NLP applications. This section discusses a variety of relationship that holds among lexemes and their senses.

Homonym

The first relationship that we discuss is homonymy which is perhaps the simplest relationship that exists among lexemes. Homonyms are words that have the same form but have different, unrelated meanings. A classic example of homonymy is Bank (river bank or financial institution). A related idea is that of homophones that refers to words that are pronounced in the same way but different meaning or spelling of both (e.g., be and bee, bear and bare).

Polysemy

Many words have more than one meaning of sense. Unlike homonyms, polysemes are words with related meanings. This linguistic phenomenon is called polysemy or lexical ambiguity. Words that have several senses are ambiguous and called polysemous. For example, the word "chair" can refer to a piece of furniture, a person, the act of presiding over a discussion etc. The word "employ" is a polysemy as its two meaning- to hire (employ a person) and to accept (employ an idea) are related. In a particular use, only one of these meanings is correct.

Hyponymy

The hypernym is a word with the more general sense. The word automobile is a hypernym for a car and a truck. The hyponym is a word with the most specific meaning. In the relationship between car and automobile, car is a hyponym of automobile. Antonym is a semantic relationship that holds between words that express opposite meanings. The word Good is an antonym of Bad, and White is an antonym of Black.

Synonym

The word synonym defines the relationship between different words that have a similar meaning. A simple way to decide whether two words are synonymous is to check for substitutability. Two words are synonyms in a context if they can be substituted for each other without changing the meaning of the sentence.

These relationships are useful in organising words in lexical databases one widely known lexical database is WordNet discussed in next topic.

4. WordNet

WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. The resulting network of meaningfully related words and concepts can be navigated with the browser. WordNet is also freely and publicly available for download. WordNet's structure makes it a useful tool for computational linguistics and natural language processing.

WordNet superficially resembles a thesaurus, in that it groups words together based on their meanings. However, there are some important distinctions. First, WordNet interlinks not just word forms—strings of letters—but specific senses of words. As a result, words that are found in close proximity to one another in the network are semantically disambiguated. Second, WordNet labels the semantic relations among words, whereas the groupings of words in a thesaurus does not follow any explicit pattern other than meaning similarity.

Structure

The main relation among words in WordNet is synonymy, as between the words shut and close or car and automobile. Synonyms--words that denote the same concept and are interchangeable in many contexts--are grouped into unordered sets (synsets). Each of WordNet's 117 000 synsets is linked to other synsets by means of a small number of "conceptual relations." Additionally, a synset contains a brief definition ("gloss") and, in most cases, one or more short sentences illustrating the use of the synset members. Word forms with several distinct meanings are represented in as many distinct synsets. Thus, each form-meaning pair in WordNet is unique.

Relations

The most frequently encoded relation among synsets is the super-subordinate relation (also called hyperonymy, hyponymy or ISA relation). It links more general synsets like {furniture, piece_of_furniture} to increasingly specific ones like {bed} and {bunkbed}. Thus, WordNet states that the category furniture includes bed, which in turn includes bunkbed; conversely, concepts like bed and bunkbed make up the category furniture. All noun hierarchies ultimately go up the root node {entity}. Hyponymy relation is transitive: if an armchair is a kind of chair, and if a chair is a kind of furniture, then an armchair is a kind of furniture. WordNet distinguishes among Types (common nouns) and Instances (specific persons, countries and geographic entities). Thus, armchair is a type of chair; Barack Obama is an instance of a president. Instances are always leaf (terminal) nodes in their hierarchies.

Meronymy, the part-whole relation holds between synsets like {chair} and {back, backrest}, {seat} and {leg}. Parts are inherited from their super ordinates: if a chair has legs, then an armchair has legs as well. Parts are not inherited "upward" as they may be characteristic only of specific kinds of things rather than the class as a whole: chairs and kinds of chairs have legs, but not all kinds of furniture have legs.

Verb synsets are arranged into hierarchies as well; verbs towards the bottom of the trees (troponyms) express increasingly specific manners characterizing an event, as in {communicate}-{talk}-{whisper}. The specific manner expressed depends on the semantic field; volume (as in the example above) is just one dimension along which verbs can be elaborated. Others are speed (move-jog-run) or intensity of emotion (like-love-idolize). Verbs describing events that necessarily and unidirectionally entail one another are linked: {buy}-{pay}, {succeed}-{try}, {show}-{see}, etc.

Adjectives are organized in terms of antonymy. Pairs of "direct" antonyms like wet-dry and young-old reflect the strong semantic contract of their members. Each of these polar adjectives in turn is linked to a number of "semantically similar" ones: dry is linked to parched, arid, dessicated and bone-dry and wet to soggy, waterlogged, etc. Semantically similar adjectives are "indirect antonyms" of the contral member of the opposite pole. Relational adjectives ("pertainyms") point to the nouns they are derived from (criminal-crime).

There are only few adverbs in WordNet (hardly, mostly, really, etc.) as the majority of English adverbs are straightforwardly derived from adjectives via morphological affixation (surprisingly, strangely, etc.)

Applications of WordNet

WordNet has found numerous applications in problems related with IR and NLP. Some of these are given below:

Concept Identification in Natural Language: WordNet can be used to identify concepts pertaining to a term, to suit them to the full semantic richness and complexity of a given information need.

Word Sense Disambiguation: WordNet combines features of a number of the other resources commonly used in disambiguation work. It offers sense definitions of words, identifies synsets of synonyms, defines a number of semantics relations and is freely available. This makes it the (currently) best known and most utilized resource for word sense disambiguation.

Automatic Query Expansion: WordNet semantic relations can be used to expand queries so that the search for a document is not confined to the pattern-matching of query terms, but also covers synonyms.

Document Summarization: WordNet has found useful application in text summarization. Few approaches utilize information from WordNet to compute lexical chains.

5. Robust Word Sense Disambiguation (WSD) - Dictionary based approach

Most of us find it difficult to understand by computers are not able to understand language the way people are. This is because we do not realise how weight and ambiguous natural languages are. Ambiguity that is having more than one meaning, can be result of syntax or semantics. Vagueness is not the same as ambiguity. In vagueness, words or phrases have only one meaning but they lack clarity, which makes it difficult to arrive at a precise meaning.

In many cases, a single word in a language corresponds to more than one thought, for example the noun Bank (financial institution of bank of a river) and the verb run (to move fast or to direct and manage). But this does not create a problem for us. We hardly give any thought to understanding that what constitutes the correct meaning of a word or phrase, but we generally arrive at the correct one. The process is almost effortless. People are good at resolving ambiguity by considering the context of the written text or coconut princess. Except for jokes and puns, where it is intended, ambiguity is not perceived as such. However, ambiguities existing at different levels are one of the major challenges in computational linguistics.

Let us try to understand why ambiguity is difficult. This is because it increases the range of possible interpretations of natural language. Suppose each word in an 8-word sentence is ambiguous and has three possible interpretations. The total number of interpretations of the whole sentence is $3^8 = 3561$. Further, synthetic and pragmatic ambiguities make the actual number of interpretations even larger. Resolving all these interpretations in a reasonable amount of time is difficult. There are words with a much larger number of sensors that then considered in this example. This gives a clear picture of the difficulty involved in the automatic interpretation of natural languages.

Ambiguity is the property of linguistic expressions. Ambiguity means capable of being understood in more than one way of having more than one meaning. It refers to a situation where an expression can have more than one partition. Ambiguity can occur at four different levels:

Lexical

Lexical ambiguity is ambiguity of a single word. A word can be ambiguous with respect to its internal structure or to its syntactic class. For example in the sentence, *look at the screen*, look is verb where as in *she gave me a stern look*, it is a noun. However this type of ambiguity is viewed as part of speech tagging in NLP and is considered to have been solved with reasonable accuracy.

Syntactic

There are different ways in which sequence of words can be grammatically structured. Each structuring leads to a different interpretation. For example, *the man saw the girl with the telescope*. It is unclear & ambiguous whether the Man saw a girl carrying a telescope, or he saw her through the telescope. It is the syntax, not the meaning of the word which is unclear. The meaning is dependent on whether the preposition "with" is attached to the girl or the man.

Semantic

Semantic ambiguity occurs when the meaning of the word themselves can be misinterpreted. For example, the meaning of words in the phrase can be combined in different ways, leading to different interpretations.

Iraqi head seeks arms

The homograph "head" can be interpreted as a noun meaning either chief or anatomical head of a body. Likewise the homograph arms can be interpreted as a plural noun meaning either weapons or body parts.

Pragmatic

Pragmatic ambiguity refers to a situation where the context of a phrase gives it multiple interpretations.

For example, *Give it to the kids*. Here "it" may refer to many things depending on the context. Consider a larger context.

Cake is on the table. I have prepared some snacks. Give it to kids.

It is not clear whether it refers to cake or snacks for both. Perhaps a larger context may help us.

Cake is on the table. I have prepared some stacks. Give it to kids. Kids enjoyed cake and snacks

Now it is clear that it refers to both snacks and the cake. Resolving these type of ambiguous required Discourse Processing

We understand that words have different meanings based on the context of its usage in the sentence. If we talk about human languages, then they are ambiguous too because many words can be interpreted in multiple ways depending upon the context of their occurrence.

Word sense disambiguation in natural language processing (NLP), may be defined as the ability to determine which meaning of word is activated by the use of words in a particular context. Lexical ambiguity, syntactic or semantic, is one of the very first problems that any NLP system faces. Part-of-speech (POS) taggers with high level of accuracy can solve Word's syntactic ambiguity. On the other hand, the problem of resolving semantic ambiguity is called WSD (word sense disambiguation). Resolving semantic ambiguity is harder than resolving syntactic ambiguity.

For example, consider the two examples of the distinct sense that exist for the word *"bank"* –

*The **bank** will not be accepting cash on Saturdays.*

*The river overflowed the **bank.***

The occurrence of the word **bank** clearly denotes the distinct meaning. In the first sentence, it means commercial (finance) banks, while in the second sentence; it refers to the river bank. Hence, if it would be disambiguated by WSD then the correct meaning to the above sentences can be assigned as follows –

*The **bank/financial institution** will not be accepting cash on Saturdays.*

*The river overflowed the **bank/riverfront.***

The evaluation of WSD requires the following two inputs:

A Dictionary: The very first input for evaluation of WSD is dictionary, which is used to specify the senses to be disambiguated.

Test Corpus: Another input required by WSD is the high-annotated test corpus that has the target or correct-senses. The test corpora can be of two types:

- **Lexical sample** - This kind of corpora is used in the system, where it is required to disambiguate a small sample of words.
- **All-words** - This kind of corpora is used in the system, where it is expected to disambiguate all the words in a piece of running text.

Dictionary-based or Knowledge-based Approach (WSD)

As the name suggests, for disambiguation, these methods primarily rely on dictionaries, treasures and lexical knowledge base. They do not use corpora evidence for disambiguation. The Lesk method is the seminal dictionary-based method introduced by Michael Lesk in 1986. The Lesk definition, on which the Lesk algorithm is based, is **"measure overlap between sense definitions for all words in context"**. However, in 2000, Kilgarriff and Rosensweig gave the simplified Lesk definition as **"measure overlap between sense definitions of word and current context"**, which further means identify the correct sense for one word at a time. Here the current context is the set of words in the surrounding sentence or paragraph.

The Lesk algorithm is based on the assumption that words in a given "neighbourhood" (section of text) will tend to share a common topic. A simplified version of the Lesk algorithm is to compare the dictionary definition of an ambiguous word with the terms contained in its neighbourhood.

Versions have been adapted to use WordNet. An implementation might look like this:

1. for every sense of the word being disambiguated one should count the amount of words that are in both neighbourhood of that word and in the dictionary definition of that sense
2. the sense that is to be chosen is the sense which has the biggest number of this count

A frequently used example illustrating this algorithm is for the context "pine cone". The following dictionary definitions are used:

Pine:

- *Kinds of evergreen tree with needle-shaped leaves*
- *Waste away through sorrow or illness*

Cone:

- *Solid body which narrows to a point*
- *Something of this shape whether solid or hollow*
- *Fruit of certain evergreen trees*

As can be seen,

Pine#1 Ç Cone#1 = 0

Pine#2 Ç Cone#1 = 0

Pine#1 Ç Cone#2 = 1

Pine#2 Ç Cone#2 = 0

Pine#1 Ç Cone#3 = 2

Pine#2 Ç Cone#3 = 0

the best intersection is Pine #1 Ç Cone #3 = 2.

Expected Questions

1. What is semantic analysis? why semantic analysis difficult? Explain various approaches to semantic analysis.
2. Explain with suitable examples following relationships between word meanings: Homonymy, Polysemy, Synonymy, Antonymy, Hypernomy, Hyponomy, Meronomy
3. What is semantic analysis? Discuss different semantic relationships between the words.
4. What is WordNet? How is "sense" defined in WordNet? Explain with example.
5. What do you mean by word sense disambiguation (WSD)? discuss dictionary based approach for WSD.
6. What do you mean by word sense disambiguation (WSD)? Discuss knowledge based WSD
7. What do you mean by word sense disambiguation (WSD)? discuss machine learning based(Navie based) approach for WSD.
8. Explain how a supervised learning algorithm can be applied for word sense disambiguation.