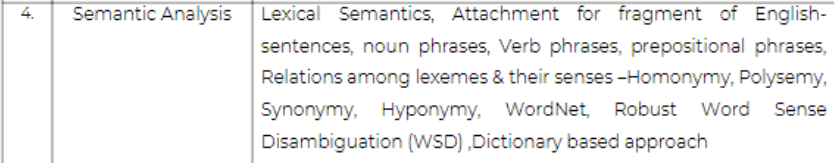
Module 04 : Semantic Analysis.



4.1 What are the different components in Lexical Semantic Analysis ?

Lexical Semantic : It is the study of the meanings of the words. It includes the study of how words structure their meaning, how they act in grammar and compositionality, and the relationships between the distinct senses and uses of a word.

Lexical Semantic Analysis : It is the process of identifying the meaning of each word in a sentence.

The following are the elements that should be taken into account while processing Natural Language.

* Hyponymy :

Hyponymys refers to a term that is an instance of a generic term. They can be understood by taking class-object as an analogy. For example : ‘Colour’ is a hypernymy while ‘grey’, ‘blue’, ‘red’, etc, are its hyponyms. It is usually a transitive relation, if A is a hyponym of B and B is a hyponym of C, then A is a hyponym of C. For example, the word ‘Purple’ hyponym is ‘Lavender’ and ‘Colour’ hyponym is ‘Purple’ therefore ‘Colour’ hyponym is ‘Lavender’.

* Homonymy :

It refers to two or more words that have same spellings but completely different meaning. For example, the word ‘bank’ means different in ‘river bank’ and different in ‘savings bank’.

* Synonymy :

When two or more words that are spelled differently but have the same meaning are called Synonymy. For example : (*large, big), (job, occupation).*

* Antonymy :

When two words have opposite meaning to each other or are in contrast to each other is called Antonymy. Two senses can be antonyms if they define a binary opposition. For example : *(light, dark)*

* Polysemy :

It refers to terms that have the same spelling but multiple closely related meanings. For example, *‘man’* may mean *‘the human species’* or *‘a male human’* or *‘an adult male human’* – since all these different meanings bear a close association, the lexical term ‘man’ is a polysemy.

* Meronomy :

It refers to a relationship wherein one lexical term is a constituent of some larger entity.

For example :

The terms ‘root’, ‘branch’,’leaf’ are all meronym of ‘Tree’

4.2 What is WordNet? Explain with an example.

WordNet is a lexical database of words in more than 200 languages. It is a database of lexical relations. It consists of three separate databases.

* Nouns
* Verbs
* Adjectives and Adverbs

The database consists of a set of lemmas, each one annotated with a set of senses. WordNet 3.0 release has

* 117,097 Nouns
* 11,488 Verbs
* 22,141 Adjectives
* 4,601 Adverbs

WordNet can be accessed via the web or it can be downloaded and used locally.

Example :

The noun “bass” has 8 senses in WordNet.

1. bass – (the lower part of musical range)

2. bass part (lowest part in polyphonic music)

3. basso (adult male singer with lowest voice)

4. sea bass (the lean flesh of a saltwater fish of the family Serranidae)

5. freshwater bass (any of various North American freshwater fish with lean flesh)

6. bass voice1, basso2 (the lowest adult male singing voice)

7. bass (member with the lowest range of a family of musical instruments)

8. bass (nontechnical name for any edible marine fish)

The adjective “bass” has 1 sense in WordNet.

1. bass1, deep (having a low vocal range)

*“a deep voice”, “a bass voice is lower than a baritone voice”, “a bass clarinet”*

There are

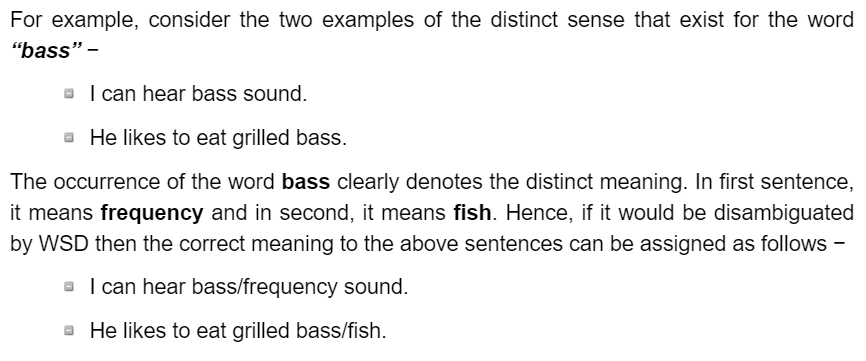
* 8 Senses for Noun
* 1 Sense for Adjective
* Each of which has a gloss (a dictionary-style definition)
* A list of synonyms for the sense (called a synset)
* Usage examples (shown for adjective sense)

4.3 What is Synset? Explain with an example.

The set of near-synonyms for a WordNet sense is called a synset. It is an important primitive in WordNet. It actually constitute the senses associated with WordNet entries. It is synsets and not wordforms/lemmas/individual senses that participate in most of the lexical sense relations in WordNet. Each synset is related to its immediately more general and more specific synsets via direct hypernym and hyponym relations.

4.4 Explain Different Approaches to Word Sense Disambiguation.

Word Sense Disambiguation is defined as the ability to determine which meaning of word is activated by the use of word in a particular context. The problem of resolving semantic ambiguity is called WSD. Resolving semantic ambiguity is harder than resolving syntactic ambiguity.



Approaches/Methods to Word Sense Disambiguation.

* Dictionary-Based/Knowledge Based Methods

It relies on dictionaries, treasures and lexical knowledge base. They do not use corpora evidences. E.g. Lesk Method.

Lesk Definition :

“measure overap between sense definitions for all words in context”.

* Supervised Methods

It makes use of sense-annotated corpora to train. It assumes that context can provide enough evidence to disambiguate the sense. The context is represented as a set of “features” of words. It relies on substantial amount of manually sense-tagged corpora.

SVM and Memory-Based Learning are the most popular algos used.

* Semi-Supervised Methods

Due to the lack of training corpus, most of the WSD algos used semi supervised methods. It is because semi-supervised methods use both labelled and unlabelled data.

* Unsupervised Methods

It assumes that similar senses occur in similar context. That is why the senses can be induced from text by clustering word occurrences by using some measure of similarity of the context. This task is called word sense induction/discrimination.

Applications of WSD

* Machine Translation
* Information Retrieval

Module 5 :

5.1 What is Reference Resolution? What are the components in Reference Resolution?

Reference is defined as the linguistic expression to denote and entity or individual.

In the following sentence *“Ram, the manager of ABC bank, saw his friend Shyam at a shop. He went to meet him”* The linguistic expressions like Ram, His, He are reference.

Reference Resolution is defined as the task of determining what entities are referred to by which linguistic expression.

Terminology:

- Referring Expression: The natural language expression that is used to perform reference is called a referring expression.

- Referent: It is the entity that is referred. For example, Ram is the referent.

- Corefer: When two expressions are used to refer to the same entity, they are called corefers. For eg, *Ram* and *he* are corefers.

- Antecedent: Ram is the antecedent of the reference he.

- Anaphora: The reference to an entity that has been previously introduced into the sentence.

Example :

1. *Guiliani left Bloomberg to be mayor of a city with a big budget problem. It is unclear how he will be able to handle it during his term.*

Referring Expressions : he, it, his.

Referents : Guiliani, Big Budget Problem

Co-referring Expressions : he, his (both referring to Bloomberg)

Antecedent : Bloomberg, Big Budget Problem

Anaphora : he, his, it

2. *Captain Farragut was a good seaman, worthy of the frigate he commanded. His vessel and he were one. He was the soul of it.*

Referring Expressions : he, his, it

Referents : Captain Farragut, worthy of the frigate

Co-referring Expressions :

Antecedents : Captain Farragut is the antecedent of he.

Anaphora : he

3. Five types of referring expression ?

- Indefinite Noun Phrases : Such kind of reference represents the entities that are new to the hearer into the discourse context. For example : *Ram had gone around one day to bring him some food –* some is an indefinite reference.

- Definite Noun Phrases : Opposite to above, such kind of reference represents the entities that are not new or identifiable to the hearer into the discourse context. For example, in the sentence - I used to read The Times of India – The Times of India is a definite reference.

- Pronouns : It is a form of definite reference. For example, Ram laughed as loud as he could. The word he represents pronoun referring expression.

- Demonstratives : These demonstrate and behave differently than simple definite pronouns. For example, this and that are demonstrative pronouns.

- One-anaphora : It blends properties of definite and indefinite reference.

4. Explain the three types of referents that complicate the reference resolution problem.

- Inferrables

A referring expression that does not refer to any entity that has been explicitly evoked in the text, but instead one that is inferentially related to an evoked entity. Such referents are called *inferrables.* Consider the expressions *a door* and *the engine* in sentence.

*I almost bought an Acura Integra today, but* ***a door*** *had a dent and* ***the engine*** *seemed noisy.*

The indefinite noun phrase a door would normally introduce a new door into the discourse context, but in this case the hearer is to infer something more: that it is not just any door, but one of the doors of the Integra. Similarly, the use of the definite noun phrase the engine normally presumes that an engine has been previously evoked or is otherwise uniquely identifiable. Here, no engine has been explicitly mentioned, but the hearer infers that the referent is the engine of the previously mentioned Integra.

- Discontinuous Sets

In some cases, references using plural referring expressions like *they* and *them* refers to sets of entities that are evoked together, for instance, using another plural expression (*their Acuras)* or a cojoined noun phrase (*John and Mary) :*

*John and Mary love their Acuras.* ***They*** *drive* ***them*** *all the time.*

However, plural references may also refer to sets of entities that have been evoked by discontinuous phrases in the text :

*John has an Acura, and Mary has a Mazda.* ***They*** *drive* ***them*** *all the time.*

Here, they refers to John and Mary, and likewise them refers to the Acura and the Mazda. Note also that the second sentence in this case will generally receive what is called a pairwise or respectively reading, in which John drives the Acura and Mary drives the Mazda, as opposed to the reading in which they both drive both cars.

* Generics :

Making the reference problem even more complicated is the existence of generic reference.

Consider example :

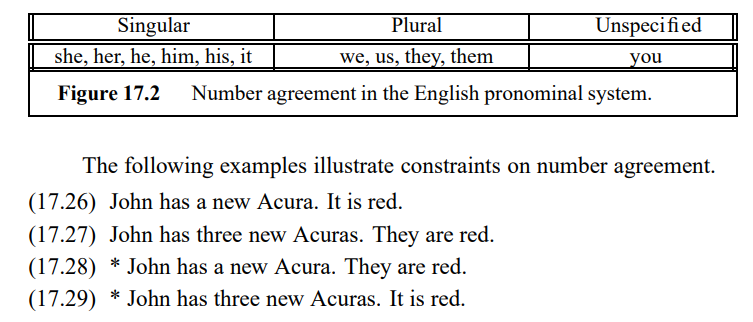
I saw no less than 6 Acura Integras today. *They* are the coolest cars.

Here, the most natural reading is not the one in which they refers to the particular 6 Integras mentioned in the first sentence, but instead to the class of Integras in general.

5. Explain the Syntactic and Semantic Constraints on Coreference

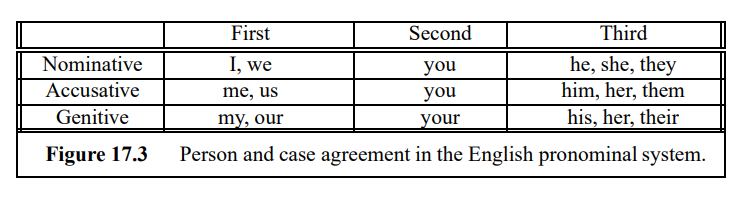
* Number Agreement

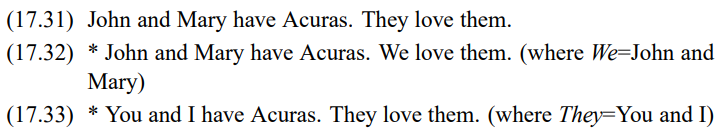
Referring Expressions and their referents must agree in number; for English this means distinguishing between singular and plural references. A categorization of pronouns with respect to number is shown



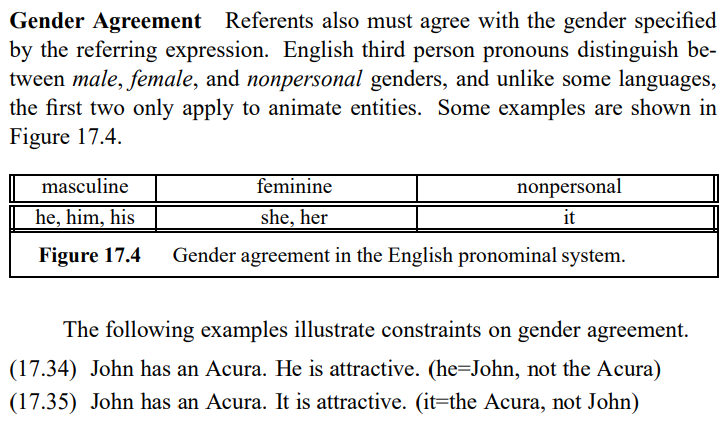
* Person and Case Agreement

English distinguishes between three forms of person : first, second and third.





* Gender Agreement



* Syntactic Constraints

Reference relations may also be constrained by the syntactic relationships between a referential expression and a possible antecedent noun phrase when both occur in the same sentence.