# NLP U3
# Semantic Analysis

8/6/23

[Introduction, Meaning Representation, Lexical Semantics, Ambiguity, Word sense Disambiguation
Discourse processing : Introduction, Cohesion, Reference Resolution, Discourse Coherence & Structure]

## Q # Introduction :

- Semantic analysis subfield of NLP that attempts to understand ~~Not~~ meaning of natural language.
- Understanding human language for humans is a straightforward process but for machines it is quite difficult task.
- Captures meaning of given text in concern with context, logical structuring of sentences & grammer rules.

### a) Parts of semantic analysis :

① Lexical semantic analysis ⇒ understands meaning of each word of text individually.
- fetches dictionary meaning of word.

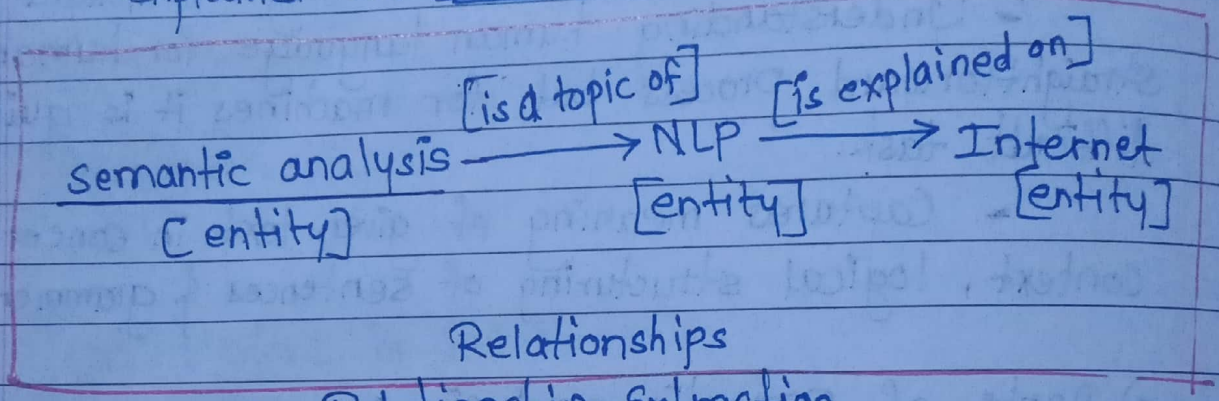
② Compositional semantic analysis ⇒ Understands meaning of text which formed by combination of individual words.
- Example ⇒
  1) students love NLP.
  2) NLP loves students.

In above two sentences, both have same set of root words but they convey entirely different meanings.

b) Tasks of semantic analysis :

① Word Sense Disambiguation
② Relationship extraction ⇒ first identifies entities present in sentence & then extracts relationships bet$^n$ those entities.

Ex. ⇒
Semantic analysis is a topic of NLP which is explained on Internet.




ex. Relationship Extraction

c) Elements of Semantic analysis :

① Hyponomy ⇒
- Instance of generic term
- 'color' is hypernymy & 'red', 'black', 'blue' are its hyponyms.

② Homonymy ⇒ same spelling but different meanings.
- 'Rose' (past form of 'rise' or flower).

③ Synonomy ⇒ diff. spellings but same meanings.
- (Job, occupation) (large, big)

④ Antonomy ⇒ pair of terms with contrast meanings.
- (Hot, cold)

⑤ Meronomy ⇒ Relationship where one term is a part of larger entity.
- 'Wheel' is meronym of 'Automobile'.

(6) Polysemy ⇒ Term have same spelling with multiple closely related meanings.
- 'Man' meanings ⇒ 'the human species', 'a male human', 'an adult male human'.

d) Applications of semantic analysis :

(1) Machine Translation ⇒ translate text from one language to another
(2) Information Retrieval ⇒ find relevant information from large collection of text.
(3) Question Answering ⇒ Used to answer the questions about a text.
(4) Chatbots ⇒ Used to create chatbots which can have natural conversations with humans.

e) challenges :

(1) Ambiguity
(2) Polysemy
(3) World knowledge (requires knowledge of world)

## 9.# Meaning Representation :

- Representation of meaning of text in natural language.
- It is easy to understand meaning of text for human but not for machines. So machines tends to represent text in specific format to extract its meaning.
- This is called meaning representation.
- Appl$^{ns}$ ⇒ Machine translation, question answering & NLP.

### a) Basic units of meaning representation :

① Entity ⇒ particular unit or individual in specific text (i.e. person or location).
eg. Australia, Pune, Riyan

② Concept ⇒ understood as generalization of entities.
eg. City, students, etc.

③ Relations ⇒ Establish relationships bet$^{n}$ entities & concepts.
eg. Pune is a city.

④ predicate ⇒ represent verb structure of sentences.

### b) Ways of meaning representation :

① Symbolic representation ⇒ represent meaning of natural language expressions using formal languages.
- More expressive but difficult to develop & maintain

(2) Statistical representation ⇒ Uses statistical methods to represent meaning of natural language expressions.
- Less expressive but easy to develop & maintain.

c) Popular approaches to meaning representation :

1. First order predicate Logic (FOPL)
2. Case grammer
3. Semantic nets
4. Conceptual dependency (CD)
5. Conceptual graphs
6. Frames.

d) Examples of meaning representation :

(1) Logic based representation ⇒ Uses formal language like first order logic to represent meaning of natural language expressions.
(2) Frame based representation ⇒ Uses frames like data structures.
(3) Semantic networks ⇒ Uses network of nodes & edges.

Q. # Lexical Semantics :

- Study of how words & phrases relate to each other & to world.
- Essential for NLP & AI.
- Helps machines to understand meaning & context of human language.
- By using lexical semantics, machines can perform :

① machine translation ⇒ translate text from one language to another.
② Information extraction ⇒ extract relevant facts & events from texts or speech.
③ Question answering ⇒ finds & provide answers to natural language queries.
④ dialogue system ⇒ helps to engage in natural & interactive conversations with humans.
⑤ text summarization ⇒ converts long text into shorter one.

a) Some key components :

['elements of semantic analysis' write here]

b) Challenges :

① Ambiguity
② Variability
③ creativity
④ Evolution of language.

(1) Ambiguity ⇒ lexical items have multiple meanings. depends on context.

(2) Creativity ⇒ lexical items can be invented or modified to create new meanings of effects.

(3) Variability ⇒ lexical items can have different forms or expressions.

(4) Evolution ⇒ lexical items can change or ~~mod~~ disappear due to cultural, social or technological factors.

c) ~~Cha~~ Opportunities :

(1) Semantic representation ⇒ lexical items can be encoded in a way that captures their meanings & relations.

(2) Semantic similarity ⇒ lexical items can be compared or measured in terms of closeness or distance in their meaning.

(3) Semantic inference ⇒ lexical items can be used to derive new information.

(4) Semantic evaluation ⇒ lexical items can be assessed or validated in terms of their accuracy or quality

# NLU - Natural Language Understanding

## # Ambiguity :

- Stylish programming technique to confuse conscious mind.
- Arises when words or phrases have multiple meanings.
- Leads to uncertainty in choosing right meaning of sentence.
- challenging task in NLU & can be handled by disambiguation process.
- Name entity recognition (NER) eliminates ambiguities.
- Types :
  - (a) Lexical
  - (b) Syntactic
  - (c) Semantic

### a) Lexical ambiguity : (words)

- Causes by a word when it have multiple senses (meanings).
- eg (1) 'Metal' have two meanings 'a substance' & 'a music genre'.
  (2) 'bat' ⇒ 'an animal' & 'sports bat'.
- Refered as homonymy.

### b) Syntactic ambiguity : (sentence)

- Also refered as structural or grammatical ambiguity.
- occurs when sentence structure have two or more possible meanings.
- eg. (1) 'I invited the person with the microphone'

Two meanings: I spoke using the microphone to invite the person.
I invited the person who has the microphone.

c) Semantic ambiguity :
- occurs when there are more than one way of reading a sentence.
- Word with diff. meanings in text.
- Eg. The old car needed constant attention.
The troops stood at attention.

d) Pragmatic ambiguity
e) discourse ambiguity

- In ambiguity, syntactic ambiguity resolves by using POS tagging & Semantic ambiguity resolves using WSD.

# Word Sense Disambiguation (WSD):

- Basically solves ambiguities arises in determining meaning of same word used in diff. situations.
- Subfield of NLP that determines intended (actual) meaning of words from given text.

## a) Evaluation of WSD:

- Two inputs are required for evaluation:
  1. Dictionary
  2. Text corpus

1. Dictionary ⇒ specify words that needs to be disambiguated.
2. Text corpus ⇒ another input to WSD.
   - Disambiguate small sample of words or all words in given text.

## b) Approaches & Methods:

1. Dictionary-based or knowledge-based methods:
   - method primarily depends on dictionaries, & knowledge base.
   - Uses dictionary to lookup diff. senses of word.
   - The correct sense is then chosen based on context.

2. Statistical approaches:
   - Uses statistical methods to find probability of word having particular meaning.
   - correct sense is chosen based on highest probability.

(3) Hybrid approaches :
combination of dictionary-based & statistical approaches.

(4) supervised methods
(5) Unsupervised methods
(6) Semisupervised methods

C) Applications :

(1) Lexicography ⇒ with lexicography, WSD provides statistically significant contextual indicators of sense.

(2) Text mining & Information extraction ⇒
- Used for accurate analysis of text.
- WSD provide intelligent gathering system for flagging of correct words.
- Eg. In medical, might need 'illegal drugs' than 'medical drugs'.

(3) Information retrieval ⇒
WSD used to resolve ambiguities of queries provided by IR system.

(4) Machine translation ⇒
most obvious application of WSD.

## d) Difficulties in WSD :

① Word sense discreteness → words cannot be easily divided into discrete submeanings.
② Inter-judge variance ⇒ WSD systems generally tested by their results of task ~~as~~ compared against ~~task~~ of human beings.
③ Differences betⁿ dictionaries.
④ Different algorithms for different applications.

→ training part

## b) Approaches & methods :

④ Supervised methods ⇒ Uses sense-oriented corpora to train.
⑤ Semi-supervised methods ⇒ Uses secondary source of knowledge like small ~~annoted~~ annotated corpora as seed data in bootstrapping.
⑥ Unsupervised methods ⇒ Uses directly from raw unannotated corpora.

Linguistics ⇒ grammatical
Utterances ⇒ wordings/phrasing/statements

## Q # Discourse Processing :

- There are no. of major problems in NLP, one is Discourse Processing.
- It is building of theories & models of how utterances stick together to form coherent discourse.
- Language always consists of allocated, structured & coherent group of sentences, these groups are 'discourse'.

### a) Coherence :

- Sequence of sentences is 'text' when there is some dependence betⁿ sentences.
- Task of textual analysis is to identify features caused by this dependence.
- These features are 'cohesion' & 'coherence'.

**Cohesion ⇒**

- linguistic features that link sentences together & are easy to identify.
- While framing sentences & phrases, cohesion is how you put yout ideas & views together.
- Measurable property ⇒ measured using grammer rules.
- Quantitative property.

- Coherence concerns to concepts & relation which are mutually accessible & relevant.

Coherence ⟹

- Evaluates output quality of natural language generation system along with good text property.
- Superset of cohesion which includes cohesive sentences, properties like consistency & understandability of content.
- Qualitative property.
- Abstract which deals with ideas.

b) Discourse Structure :

- Imp question ⟹ regarding discourse ⟹ what kind of structure it must have.
- Answer ⟹ depends on segmentation applied to its discourse.
- Discourse segmentations ⟹ determines types of structures for large discourse.
- Discourse segmentation ⟹ quite difficult to implement but imp for IR, info extraction & text summarization.
- Human discourse exhibits structures that ~~intended~~ indicates common experiences & respond to them.

- Discourse Segmentation :
  - Documents are automatically partitioned into fragments i.e. passages, these are discourse segments.
  - Techniques to separate doc into passages :

① Rule based systems based on clue words & phrases.
② probabilistic techniques to separate fragments
③ Text filling algorithm uses cohesion to identify segments.

- Algorithms for discourse segmentation :

① Unsupervised discourse segmentation ⇒
- 'linear segmentation.'
- Depends on cohesion which means use of linguistic features to bind textual units together.

② Supervised discourse segmentation ⇒
- Needs boundary lebeled training data.
- Discourse marker or cue words plays an imp role.
- Discourse marker or cue word ⇒ word or phrase that functions to signal discourse structure.
- Discourse markers are domain specific.

c) Reference Resolution :

- Reference ⇒ linguistic expression to denote an entity or individual.
- Ex. Ram, manager of Kotak bank,
  saw his friend Shyam.
  He went to meet him.

Here, Ram, his, he are reference.

- Reference resolution ⇒ task of determining what entities referred to by which linguistic expression.

[ Discourse processing intro part,
a) coherence part &
b) discourse structure part
in short you can write here. ].