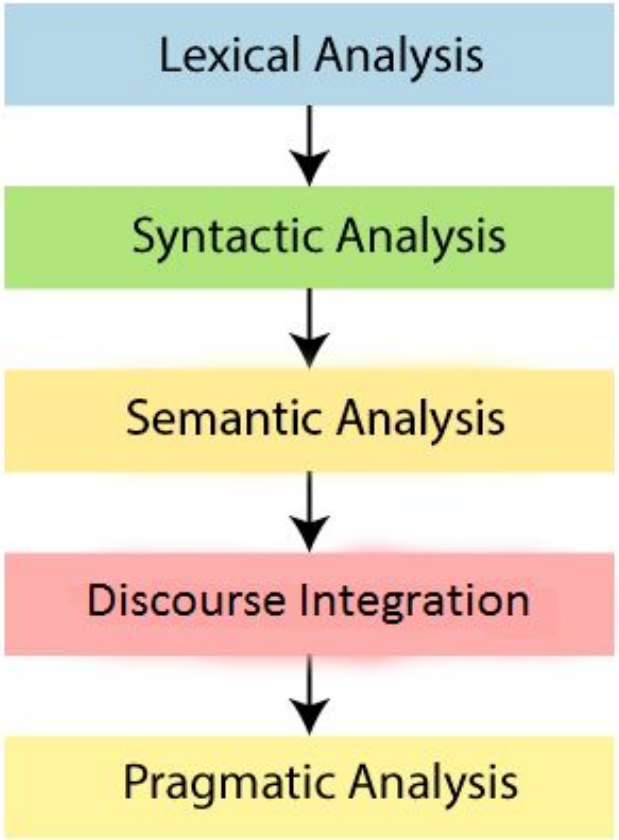
Natural Language Processing

Introduction -NLP

- Language is a method of communication with the help of which we can speak, read and write.
- Can human beings communicate with computers in their natural language?
- It is a challenging to develop NLP applications because computers need structured data, but human speech is unstructured and often ambiguous in nature.
- Natural Language Processing (NLP) is the sub-field of AI that is concerned about enabling computers to understand and process human language.
-

NLP Phases


Lexical Analysis
Syntactic Analysis
Semantic Analysis
Discourse Integration
Pragmatic Analysis

Lexical Analysis

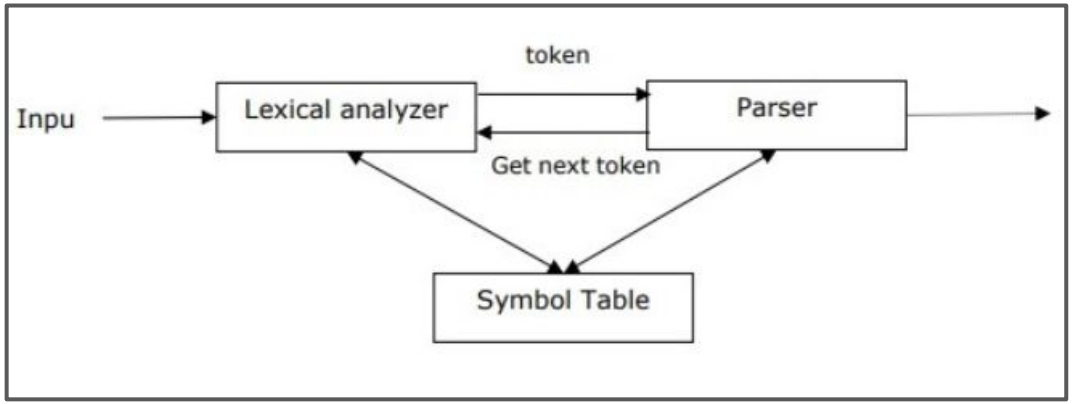
- The first phase of NLP is the Lexical Analysis.
- This phase scans the source code as a stream of characters and converts it into meaningful lexemes.
- It divides the whole text into paragraphs, sentences, and words.

Syntactic Analysis

- Syntactic analysis (parsing or syntax analysis) draw exact structure of sentence or arrangement of tokens.
- Syntax analysis checks the text for meaningfulness comparing to the rules of formal grammar.
    - E.g. sentence like “playing and are ram shyam” would be rejected in this phase.
    - Statement like “2)+2(print” would be rejected in this phase.
    - Statement c = a + b would be accepted.
-

Parser

- It’s software component, used to implement the task of parsing.
- It takes input data (text) and giving structural representation of the input after checking for correct syntax as per formal grammar.
- It also builds a data structure generally in the form of parse tree or abstract syntax tree or other hierarchical structure.



Role of Parser

The main roles of the parse include −

- To report any syntax error.
- To recover from commonly occurring error so that the processing of the remainder of program can be continued.
- To create parse tree.
- To create symbol table.
- To produce intermediate representations .

Types of Parsing

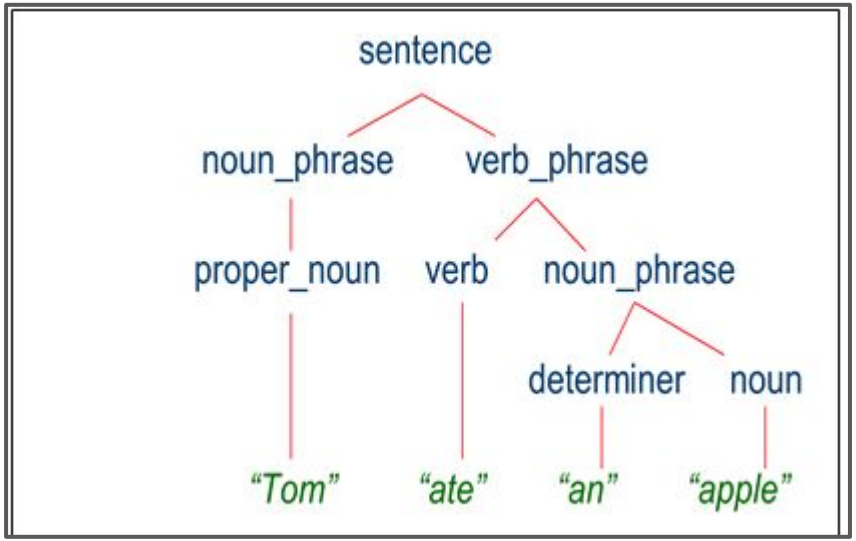
- Top-down Parsing
  - The parser starts constructing the parse tree from the start symbol and then tries to transform the start symbol to the input.
- Bottom-up Parsing
  - The parser starts with the input symbol and tries to construct the parser tree up to the start symbol.
- Derivation
  - Derivation is a set of production rules.
  - the non-terminal, which is to be replaced along with deciding the production rule.
  - Two Types of Derivation
    - Left-most Derivation :- An input is scanned and replaced from the left to the right.
    - Right-most Derivation:- An input is scanned and replaced from right to left.

Parsing Example

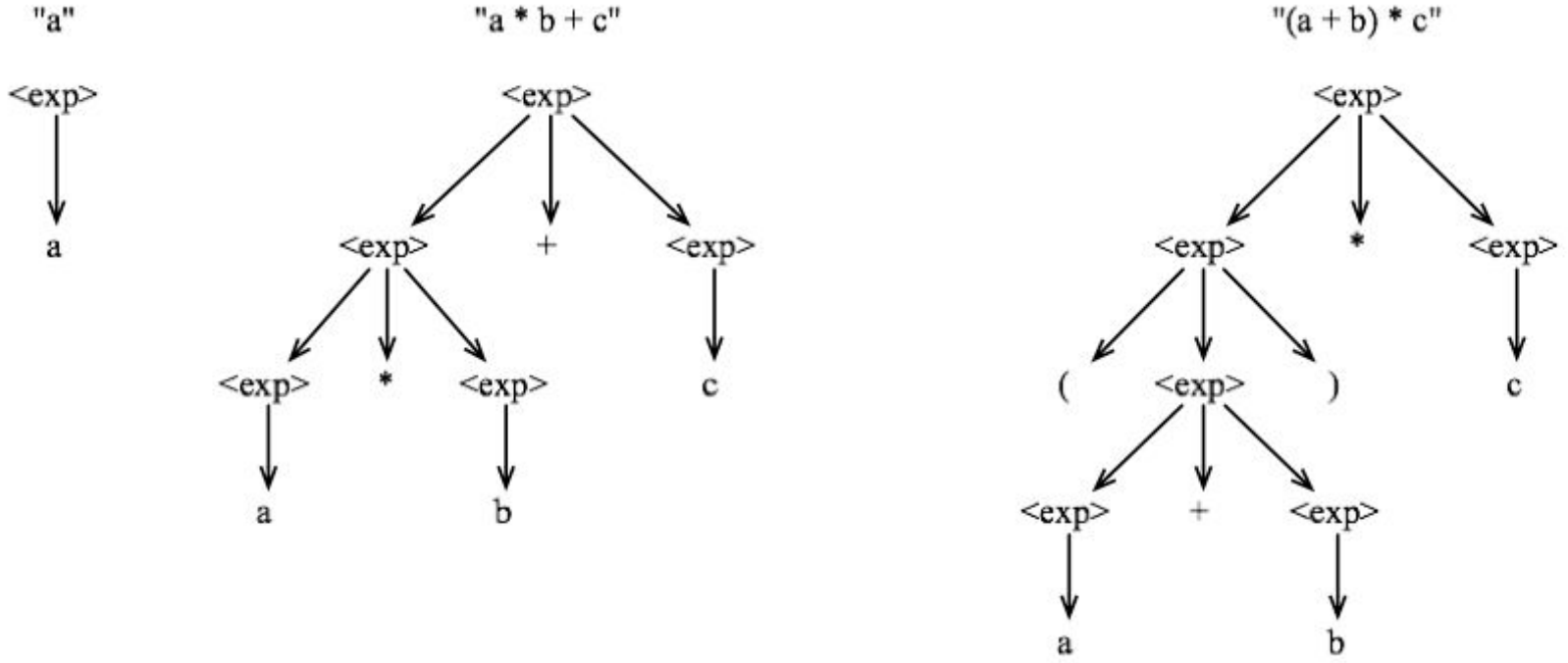
Grammer

```
sentence -> noun_phrase, verb_phrase
noun_phrase -> proper_noun
noun_phrase -> determiner, noun
verb_phrase -> verb, noun_phrase
proper_noun -> [Tom]
noun -> [apple]
verb -> [ate]
determiner -> [an]
```

Parse Tree



Parsing Example



Semantic Analyser

- Semantic analysis is to draw exact meaning of text .
- E.g. sentence like “hot ice-cream” would be rejected by semantic analyzer.

Building Blocks of Semantic System

- **Entities** − It represents the individual such as a particular person, location etc. For example, Haryana. India, Ram all are entities.
- **Concepts** − It represents the general category of the individuals such as a person, city, etc.
- **Relations** − It represents the relationship between entities and concept. For example, Ram is brother of Laxman.
- **Predicates** − It represents the verb structures.

Approaches to Meaning Representations

Semantic analysis uses the following approaches for the representation of meaning

- First order predicate logic (FOPL)
- Semantic Nets
- Frames
- Conceptual dependency (CD)
- Rule-based architecture
- Case Grammar
- Conceptual Graphs

Tasks involved in Semantic Analysis

To understand the meaning of a sentence, the following are the major processes involved in Semantic Analysis:

- Word Sense Disambiguation
  - In Natural Language, the meaning of a word may vary as per its usage in sentences and the context of the text.
  - Rock - Types of Music / BIg Stone
  - Board - Verb / noun
  - accurate meaning of the word is highly dependent upon its context and usage in the text.
  - The ability of a machine to overcome the ambiguity involved in identifying the meaning of a word based on its usage and context is called **Word Sense Disambiguation.**

Tasks involved in Semantic Analysis

- Relationship Extraction
  - It involves firstly identifying various entities present in the sentence and then extracting the relationships between those entities.
  - **Mohan** and **Suresh** are **friends.**

Elements of Semantic Analysis

- Hyponymy
  - refers to a term that is an instance of a generic term
  - They can be understood by taking class-object as an analogy.
  - Color - red,blue,black
- Homonymy
  - refers to two or more lexical terms with the same spellings but completely distinct in meaning.
  - rose - past form of rise
  - rose - flower
- Synonymy
  - When two or more lexical terms that might be spelt distinctly have the same or similar meaning.
  - (Job, Occupation), (Large, Big), (Stop, Halt).

Elements of Semantic Analysis

- **Antonymy**
  - refers to a pair of lexical terms that have contrasting meanings
  - (Day, Night), (Hot, Cold), (Large, Small).
- **Polysemy**
  - refers to lexical terms that have the same spelling but multiple closely related meanings.
  - It differs from homonymy because the meanings of the terms need not be closely related in the case of homonymy.
  - E.g: ‘man‘ may mean ‘the human species‘ or ‘a male human‘ or ‘an adult male human‘.
- **Meronomy**
  - refers to a relationship wherein one lexical term is a constituent of some larger entity.
  - E.g.: ‘Wheel‘ is a meronym of ‘Automobile‘

Semantic Analysis Techniques

Two of the most common Semantic Analysis techniques are:

- Text Classification
- Text Extraction

Text Classification

- In-Text Classification, label the text according to the insights to gain from the textual data.
  - **Sentiment Analysis** :
    - Label the text with the prominent emotion they convey.
    - It is highly beneficial when analyzing customer reviews for improvement.
  - **Classification**
    - Categories text into some predefined categories.
    - For example: Identifying whether a research paper is of Physics, Chemistry or Maths
  - **Classification**
    - Determine the intent behind a text message.
    - For example: Identifying whether an e-mail received at customer care service is a **query**, **complaint** or **request.**
-

Text Extraction

In-Text Extraction, our aim is to obtain specific information from our text.

- Keyword Extraction
  - obtain the essential words that define the entire document.
- Entity Extraction
  - obtain all the entities involved in a document.

Discourse Analysis

- Discourse analysis is a research method.
- Discourse parsing is used for distinguishing the connectedness and specific talk relations among various units in a content.
- It helps us to understand how a language can be used in our daily life situations.
- Discourse analysis is used for understanding written or spoken language and its relation to social context.
- Language always consists of collocated, structured and coherent groups of sentences

Discourse

- Discourse is a coherent structured group o.f textual units (e.g., sentences)
- Discourse comprises a sequences that must be interpreted with respect to the context.
- For a set of sentences to make sense, it must consist of sentences that are related to each other.
- The structure is needed to interpret the text- is called discourse structure
- Collection of interrelated sentences is called discourse group of sentences.
- E.g :- The Tin woodman went to the Emerald city to see the Wizards of Oz and ask for a heart.After **he** asked for **it**, the woodman waited for the Wizard’s response.
- The goal of deciding what pronouns refer to is cohesion resolution work at discourse level.
- Discourse is Monologues
    - Speaker/writer + hearer/reader
- Dialogues
    - Human-human
    - Human-computer

features of discourse

- Position: opening sentence , Ending sentence.
- Order: different orders lead to various events/meaning
  - I said the magic words, and a genie appeared. **vs.**
  - A genie appeared, and I said the magic words.
- Adjacency: attributed material and contrasts are visible through sentences nearby
- Context : intended meaning can only be conveyed when understood in context.

Coherence

- John hid Bill’s car keys. He was drunk.
- John hid Bill’s car keys. He likes spinach
- Which one is more coherent??
- Coherence as the main characteristic of discourse
- How do you recognize discourse?
  - It makes sense!
  - It is relevant!
  - It ‘hangs together’
  - "It is coherent!! !
- Ram got caught in the rain. He fell ill.

Textual Coherence

- John went to his favorite music store to buy a piano.
- He had frequented the store for many years.
- He was excited that he could finally buy a piano.
- He arrived just as the store was closing for the day.

- ❏ John went to his favorite music store to buy a piano.
- ❏ It was a store John had frequented for many years.
- ❏ He was excited that he could finally buy a piano.
- ❏ It was closing just as John arrived.

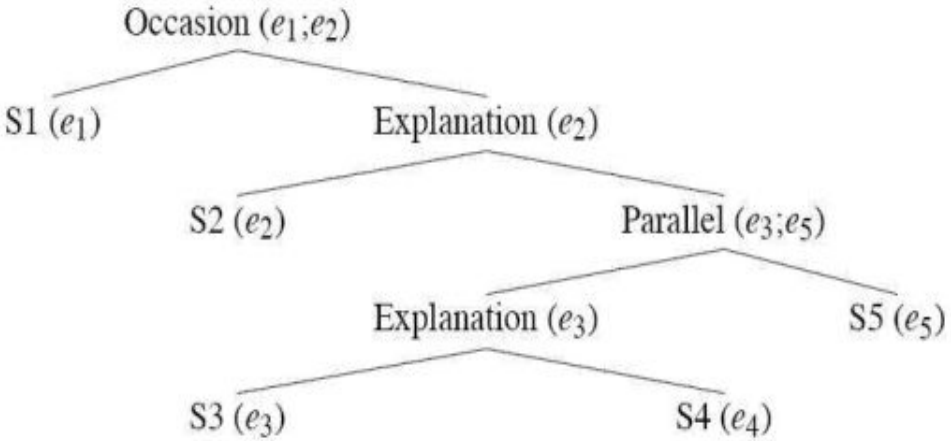
Two entities --- John and the store: Depending on the sentence structure, the focus differs

Discourse Relation

- Discourse relations (Coherence relations) specify the relations between sentences or clauses. Due to these relations, two adjacent sentences can look coherent.
- Explanation
  - John hid Bill’s car keys. He was drunk.
- Elaboration
  - Dorothy was from Kansas. She lived on the Kansas prairies.
- Result
  - The tin woodman was caught in the rain. His joints rusted.
- Parallel
  - The scarecrow wanted some brains. The tin woodsman wanted a heart.
- Occasion
  - Dorothy picked up the oil-can. She oiled the Tin Woodman’s joints.

Discourse parsing

- John went to the bank to deposit the paycheck. (e1)
- He then took a train to Bill's car dealership. (e2)
- He needed to buy a car. (e3)
- The company he works for now isn't near any public transportation. (e4)
- He also wanted to talk to Bill about their softball league. (e5)



Pragmatic Analysis

- Pragmatics is a study of language that is not directly spoken, instead speaker hints at suggest at meaning and Listener assume the meaning.
- Pragmatic Analysis deals with the overall communicative and social content and its effect on interpretation.
- It means abstracting or deriving the meaningful use of language in situations.
- In this analysis, the main focus always on what was said in reinterpreted on what is meant.
- E.g., “close the window?” should be interpreted as a request instead of an order
- E.g. Will you crack open the door? I am getting hot.

5 aspects of Pragmatic Analysis

1. Deixies
    - It refers to the words or phrases such as me, here that cant not be fully explaind without additional information.
    - E.g meet me here
2. Implicature
    - More being communicated than said
        i. Conversational Implicature

            A : i am out of gas

            B : There is a gas station around the corner.

        ii. Conventional Implicature

            Abhishek is rich but sad

3.Presupposition

Something speaker assume prior to making utterance.

Manan’s brother wants three superbikes.

4.Speech act

Sentence convey the action rather than something saying.

I smashed potatoes.

5.Conversational structure

Analysis of sequential and Anti sequential nature of conversation.