**Need of NLP**

* **Understanding Human Language**: NLP enables computers to understand and process human language in a meaningful way, allowing them to interact with users more naturally through text or speech.
* **Information Extraction**: NLP techniques extract structured information from unstructured text data, such as extracting named entities (e.g., people, organizations) or relationships between entities from news articles or social media posts.
* **Text Analytics**: NLP provides tools for analyzing and summarizing large volumes of text data, allowing organizations to gain insights from customer feedback, social media discussions, news articles, and other sources.
* **Sentiment Analysis**: NLP techniques can analyze the sentiment expressed in text data, helping businesses understand customer opinions, monitor brand reputation, and gauge public opinion on social and political issues.
* **Language Translation**: NLP powers machine translation systems that automatically translate text from one language to another, facilitating communication and collaboration across linguistic barriers.
* **Question Answering**: NLP systems can answer questions posed in natural language by extracting relevant information from textual sources, such as FAQs, knowledge bases, or documents.
* **Virtual Assistants and Chatbots**: NLP is essential for building virtual assistants and chatbots that can understand user queries, engage in conversational interactions, and provide personalized assistance or information retrieval services.
* **Information Retrieval**: NLP techniques improve search engines' ability to understand user queries and retrieve relevant documents or web pages based on the user's intent.
* **Content Generation**: NLP enables the generation of human-like text, such as generating product descriptions, news articles, or personalized recommendations.
* **Healthcare and Life Sciences**: In healthcare, NLP is used for analyzing medical records, extracting clinical information, and identifying patterns in patient data to support diagnosis, treatment planning, and biomedical research.

**Challenges in ambiguity**

Ambiguity, generally used in natural language processing, can be referred as the ability of being understood in more than one way.

Natural language is very ambiguous. NLP has the following types of ambiguities −

1. **Lexical Ambiguity**

* The ambiguity of a single word is called lexical ambiguity. For example, treating the word **silver** as a noun, an adjective, or a verb.
* There are various types of lexical ambiguity, including:
  + **Polysemy:** A single word having multiple related meanings. For example, the word "bank" can refer to a financial institution or the side of a river.
  + **Homonymy:** Different words with the same spelling or pronunciation but different meanings. For instance, "bat" can refer to a flying mammal or a piece of sports equipment.
  + **Heteronymy:** Words with the same spelling but different pronunciations and meanings. An example is "tear," which can mean to rip or a drop of liquid from the eye.

1. **Syntactic Ambiguity**

* This kind of ambiguity occurs when a sentence is parsed in different ways.
* types of syntactic ambiguity:
  + **Structural Ambiguity:** This occurs when the syntactic structure of a sentence allows for multiple interpretations. For example:
  + "I saw the man with the telescope." Is the man holding the telescope or did the speaker use the telescope to see the man?
  + **Attachment Ambiguity:** In sentences with multiple clauses, the attachment of a phrase or clause to one part of the sentence over another can lead to ambiguity. For instance:
  + "I told my friend I would help." Did the speaker tell their friend that they would help, or did they tell their friend about their intention to help someone else?
  + **Modifier Attachment Ambiguity:** This occurs when a modifier, such as an adjective or adverb, can be associated with more than one word in a sentence. For example:
  + "She almost told him everything." Did she almost tell him everything, or did she tell him almost everything?
  + **Coordination Ambiguity:** Ambiguity can arise in sentences with coordination, where multiple phrases are joined by conjunctions. For example:
  + "She likes chocolate and vanilla ice cream." Does she like both flavors, or does she like chocolate ice cream and vanilla ice cream separately?

1. **Semantic Ambiguity**

* This kind of ambiguity occurs when the meaning of the words themselves can be misinterpreted.
* In other words, semantic ambiguity happens when a sentence contains an ambiguous word or phrase.
* For example, the sentence “The car hit the pole while it was moving” is having semantic ambiguity because the interpretations can be “The car, while moving, hit the pole” and “The car hit the pole while the pole was moving”.

1. **Anaphoric Ambiguity**

* This kind of ambiguity arises due to the use of anaphora entities in discourse. For example, the horse ran up the hill. It was very steep. It soon got tired. Here, the anaphoric reference of “it” in two situations cause ambiguity.

1. **Pragmatic ambiguity**

* Such kind of ambiguity refers to the situation where the context of a phrase gives it multiple interpretations.
* In simple words, we can say that 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.
* For example, the sentence “I like you too” can have multiple interpretations like I like you (just like you like me), I like you (just like someone else dose).

**Level of NLP**

* The analysis of NL is broken down into various broad levels such as phonological, morphological, lexical, syntactic, semantic, pragmatic and discourse analysis.
* For example some applications require the first 3 levels only. Also, the levels could be applied in a different order independent of their granularity

1. **Phonology**

* This level is applied *only*if the input is a speech. Input is acoustic waveform and output is string of words.
* It deals with speech recognition and generation. That is interpretation of speech sounds within and across words
* Speech sound might give a big hint about the meaning of a word or a sentence.
* The area of computational linguistic that deals with speech analysis is computational phonology.

1. **Morphological Analysis**

* Morphology is study of internal structure of words. Given a particular word in a language, what are the different meaningful units it is made up of and each small unit is called a morpheme.
* cat : stem cats: cat + s unhappy happily unhappily
* cats : N +PL cat: N+SG sort: V+SG sorts: V+PL sort: N+PL
* **Computational tool to perform morphological parsing is finite state transducer.**

1. **Lexical Analysis**

* It deals with understanding of everything about distinct words according to their position in the speech, their meanings and their relation to other words.
* It identify and analyze the structure of words with respect to their lexical meaning and part-of-speech.
* Lexicon is a dictionary. Lexicon of a language means the collection of words and phrases in a language.
* Validity of word is checked according to lexicon.

1. **Syntactic Analysis (Parsing)** –

* It involves analysis of words in the sentence for grammar and arranging words in a manner that shows the relationship among the words.
* The sentence such as “The school goes to boy” is rejected by English syntactic analyzer.
* Syntax refers to the study of formal relationships between words of sentences. Validity of a sentence is checked according to rules of grammar.
* To perform syntactic analysis, the knowledge of grammar and parsing techniques is required.
* Grammar is formal specification of rules allowable in the language.
* Parsing is a method of analysing a sentence to determine its structure according to the grammar. CFG is used for syntactic analysis.
* Two basic parsing techniques are top down parsing and bottom-up parsing.
* **I eat banana.**
* **I banana eat.**

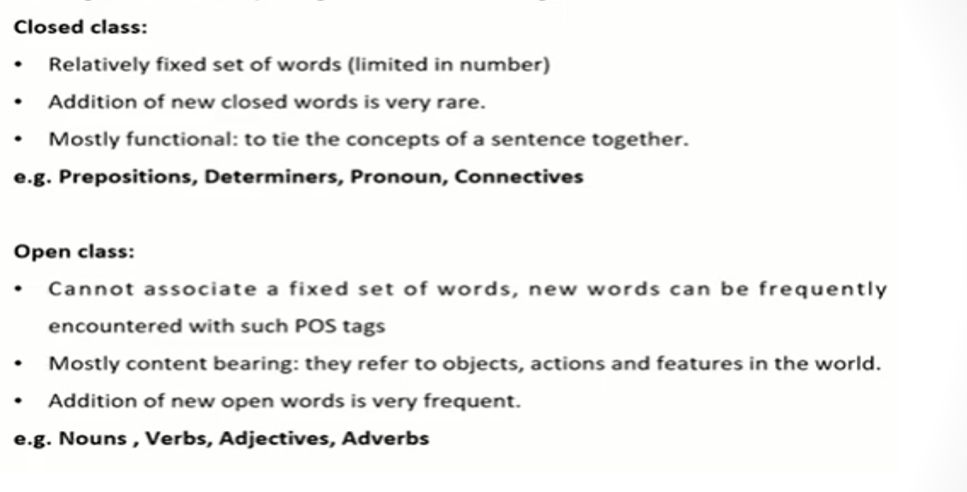
1. **Semantic Analysis**

* Semantics deals with the meaning of natural language sentences. Meaning of the sentences is understood in this phase.
* It draws the exact meaning or the dictionary meaning from the text.
* The text is checked for meaningfulness. It is done by mapping syntactic structures and objects in the task domain. The semantic analyzer disregards sentence such as “hot ice-cream”.
* e.g.
* She eats banana
* Machine eats banana.

1. **Pragmatic :**

* During this, what was said is re-interpreted on what it actually meant. It involves deriving those aspects of language which require real world knowledge.
* Explains how extra meaning is read into texts without actually being encoded in them. This requires much world knowledge, including the understanding of intentions, plans, and goals.
* Consider the following 2 sentences:
* The city counsellors refused the demonstrators a permit because they feared violence.
* The city counsellors refused the demonstrators a permit because they advocated revolution.
* The meaning of “they” in the 2 sentences is different. In order to figure out the difference, world knowledge in knowledge bases and inferencing modules should be utilized.

**Difference btn open and close class**



|  |  |  |
| --- | --- | --- |
| **Aspect** | **Rule-Based POS Tagging** | **Stochastic-Based POS Tagging** |
| Definition | Rules are manually crafted based on linguistic principles. | Probability theory is used to model uncertainty and make predictions. |
| Approach | Relies on predefined rules to assign POS tags. | Utilizes statistical models or machine learning algorithms. |
| Linguistic Knowledge | Requires linguistic expertise to design rules. | Relies less on linguistic expertise but more on annotated data. |
| Adaptability | May not adapt well to new domains or languages. | Can adapt to different domains and languages with sufficient training data. |
| Handling Ambiguity | May struggle with handling ambiguity and exceptions. | Can handle ambiguity by assigning probabilities to POS tags. |
| Performance | Generally has lower performance compared to statistical models. | Can achieve higher accuracy with large annotated datasets and advanced algorithms. |
| Examples | Regular expression-based taggers, handcrafted grammars. | Hidden Markov Models (HMMs), Maximum Entropy Markov Models (MEMMs), Conditional Random Fields (CRFs). |