

## **Module 1 : Introduction**

- Introduction - 2
- History of NLP - 3
- Steps in Natural Language Processing- 4
- Knowledge requires in NLP - 7
- Ambiguity in Natural Language Processing - 9
- Phases in NLP - 11
- Application of NLP - 13

[www.Lastmomenttuitions.com](http://www.Lastmomenttuitions.com)

# Natural Language Processing

- Language is a method of communication with the help of which we can speak, read and write. Natural Language Processing (NLP) is a subfield of Computer Science that deals with Artificial Intelligence (AI), which enables computers to understand and process human language.
- Natural Language Processing is the technology used to aid computers to understand the human's natural language.

(ek machine ka insano ki language samjhna fir uspe answer back karna ya jo action kaha hai usko perform karna is natural Language processing )

## Components of NLP

There are two components of NLP as given –

### Natural Language Understanding (NLU)

Understanding involves the following tasks –

- Mapping the given input in natural language into useful representations.
- Analyzing different aspects of the language.

### Natural Language Generation (NLG)

It is the process of producing meaningful phrases and sentences in the form of natural language from some internal representation.

It involves –

- Text planning – It includes retrieving the relevant content from the knowledge base.
- Sentence planning – It includes choosing required words, forming meaningful phrases, setting tone of the sentence.
- Text Realization – It is mapping sentence plan into sentence structure.

## History of Natural Language Processing

- We have divided the history of NLP into four phases. The phases have distinctive concerns and styles.

### First Phase (Machine Translation Phase) - Late 1940s to late 1960s

- The work done in this phase focused mainly on machine translation (MT). This phase was a period of enthusiasm and optimism

Let us now see all that the first phase had in it –

- The research on NLP started in early 1950s after Booth & Richens' investigation and Weaver's memorandum on machine translation in 1949.
- 1954 was the year when a limited experiment on automatic translation from Russian to English demonstrated in the Georgetown-IBM experiment.
- In the same year, the publication of the journal MT (Machine Translation) started.
- The first international conference on Machine Translation (MT) was held in 1952 and second was held in 1956.
- In 1961, the work presented in Teddington International Conference on Machine Translation of Languages and Applied Language analysis was the high point of this phase.

### Second Phase (AI Influenced Phase) – Late 1960s to late 1970s

In this phase, the work done was majorly related to world knowledge and on its role in the construction and manipulation of meaning representations. That is why, this phase is also called AI-flavored phase.

The phase had in it, the following –

- In early 1961, the work began on the problems of addressing and constructing data or knowledge base. This work was influenced by AI.
- In the same year, a BASEBALL question-answering system was also developed. The input to this system was restricted and the language processing involved was a simple one.

- A much advanced system was described in Minsky (1968). This system, when compared to the BASEBALL question-answering system, was recognized and provided for the need of inference on the knowledge base in interpreting and responding to language input.

### Third Phase (Grammatico-logical Phase) – Late 1970s to late 1980s

This phase can be described as the grammatico-logical phase. Due to the failure of practical system building in the last phase, the researchers moved towards the use of logic for knowledge representation and reasoning in AI.

The third phase had the following in it –

- The grammatico-logical approach, towards the end of decade, helped us with powerful general-purpose sentence processors like SRI's Core Language Engine and Discourse Representation Theory, which offered a means of tackling more extended discourse.
- In this phase we got some practical resources & tools like parsers, e.g. Alvey Natural Language Tools along with more operational and commercial systems, e.g. for database query.
- The work on lexicon in the 1980s also pointed in the direction of grammatico-logical approach.

### Fourth Phase (Lexical & Corpus Phase) – The 1990s

We can describe this as a lexical & corpus phase. The phase had a lexicalized approach to grammar that appeared in late 1980s and became an increasing influence. There was a revolution in natural language processing in this decade with the introduction of machine learning algorithms for language processing.

## Steps in Natural language processing

- 1)Tokenizations
- 2)Stemming
- 3)Lemmatization
- 4)POS tags
- 5)Named Entity Recognition

## Tokenization

- Cutting the big sentence into small tokens
- Example : Welcome to Last moment tuitions will be divided into tokens  
[welcome] [to] [last] [moment] [tuitions]

## Stemming

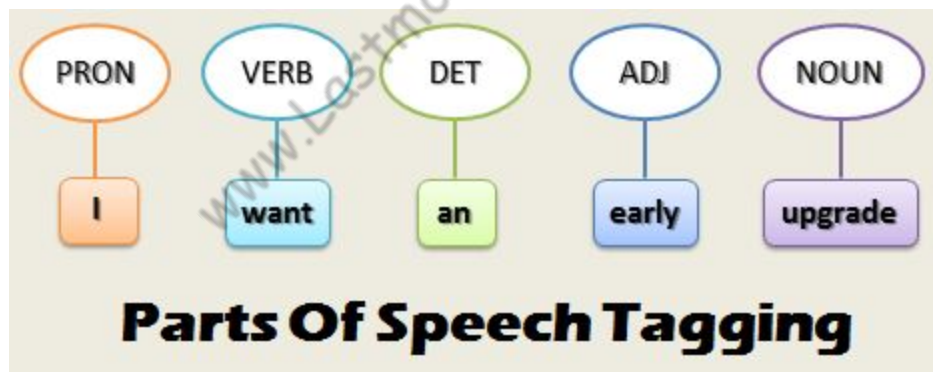
- Normalize words into its base or root forms
- Waits, waited, waiting -> wait ( In this example the root meaning or word is wait so in stemming we cut remaining part and find the root word)

## Lemmatization

- Group together different inflected forms of a word called Lemmatization
- Somehow similar to stemming, as it maps several words into one common root
- Output of lemmatization is a proper word
- Example : Gone, going and went -> Go

## POS Tags

- POS stands for Parts of speech tags
- It indicated how a word function in meaning as well as grammatically within a sentence.



- But problem in POS tag is sometime one word can have different meaning example “ google something on internet” now we know google is a proper noun but here it is use as verb
- To solve this problem we use Named entity recognition

## Named Entity Recognition

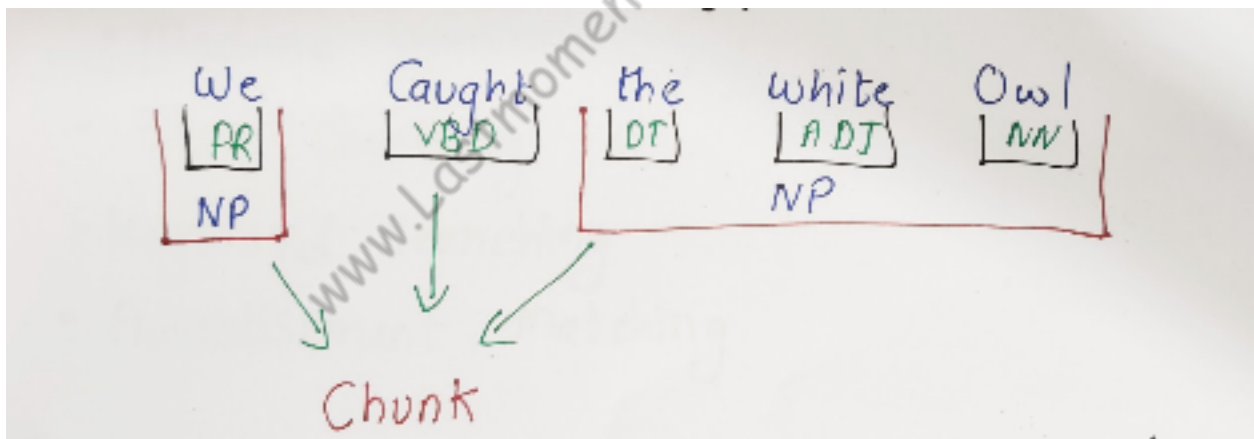
- It is a process of name entity such as: person , organization , location , quantities ,etc.



- Here Apple is an Organization , Tim cook is a person and cupertino flint center is location. In NER we get exact information like apple is a organization not a fruit.

## Chunking

- Picking individual pieces of information and grouping them into bigger pieces.



- This helps getting insight and meaningful information from the text.

# Knowledge Requires in NLP

A natural language understanding system must have knowledge about what the words mean, how words combine to form sentences, how word meanings combine to form sentence meanings and so on. The different forms of knowledge required for natural language understanding are given below.

- **Phonetic And Phonological Knowledge**
- **Morphological Knowledge**
- **Syntactic Knowledge**
- **Semantic Knowledge**
- **Pragmatic Knowledge**
- **Discourse Knowledge**
- **World Knowledge**

## ❑ **Phonetic And Phonological Knowledge**

1. -Phonetics is the study of language at the level of sounds while phonology is the study of combination of sounds into organized units of speech
2. -Phonetic and phonological knowledge are essential for speech based systems as they deal with how words are related to the sounds that realize them.

## ❑ **Morphological Knowledge**

1. Morphology concerns word formation.
2. It is a study of the patterns of formation of words by the combination of sounds into minimal distinctive units of meaning called morphemes
3. Morphological knowledge concerns how words are constructed from morphemes.

## ❑ **SYNTACTIC KNOWLEDGE**

1. Syntax is the level at which we study how words combine to form phrases, phrases combine to form clauses and clauses join to make sentences.
2. Syntactic analysis concerns sentence formation.
3. It deals with how words can be put together to form correct sentences

### ❑ **Semantic Knowledge**

1. It concerns the meanings of the words and sentences.
2. Defining the meaning of a sentence is very difficult due to the ambiguities involved.

### ❑ **Pragmatic Knowledge**

1. Pragmatics is the extension of the meaning or semantics.
2. Pragmatics deals with the contextual aspects of meaning in particular situations
3. It concerns how sentences are used in different situations

### ❑ **Discourse Knowledge**

1. Discourse concerns connected sentences. It is a study of chunks of language which are bigger than a single sentence.
2. Discourse language concerns inter-sentential links that is how the immediately preceding sentences affect the interpretation of the next sentence.
3. Discourse knowledge is important for interpreting pronouns and temporal aspects of the information conveyed.

### ❑ **World Knowledge**

1. Word knowledge is nothing but everyday knowledge that all speakers share about the world.
2. It includes the general knowledge about the structure of the world and what each language user must know about the other user's beliefs and goals.
3. This is essential to make the language understanding much better.



## Ambiguity in Natural Language Processing

- Ambiguity, generally used in natural language processing, can be referred to as the ability of being understood in more than one way.
- In simple terms, we can say that ambiguity is the capability of being understood in more than one way.
- Natural language is very ambiguous.

NLP has the following types of ambiguities –

### 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.

- She won two silver medals
- She made a silver speech
- His worries has silvered his hair

### Syntactic Ambiguity

- This kind of ambiguity occurs when a sentence is parsed in different ways.
- For example, the sentence “The man saw the girl with the telescope”. It is ambiguous whether the man saw the girl carrying a telescope or he saw her through his telescope.

### Semantic Ambiguity

- This kind of ambiguity occurs when the meaning of the words themselves can be misinterpreted even after syntax and the meaning of individual word have been resolved.
- In other words, semantic ambiguity happens when a sentence contains an ambiguous word or phrase.

### Example 1

- “Seema loves her mother and shreya does too.”
- Here there are two meaning “seema loves her mother and shreya loves her own mother” and “seema loves her mother and shreya also loves seema mother”

## Example 2

“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”.

## Anaphoric Ambiguity

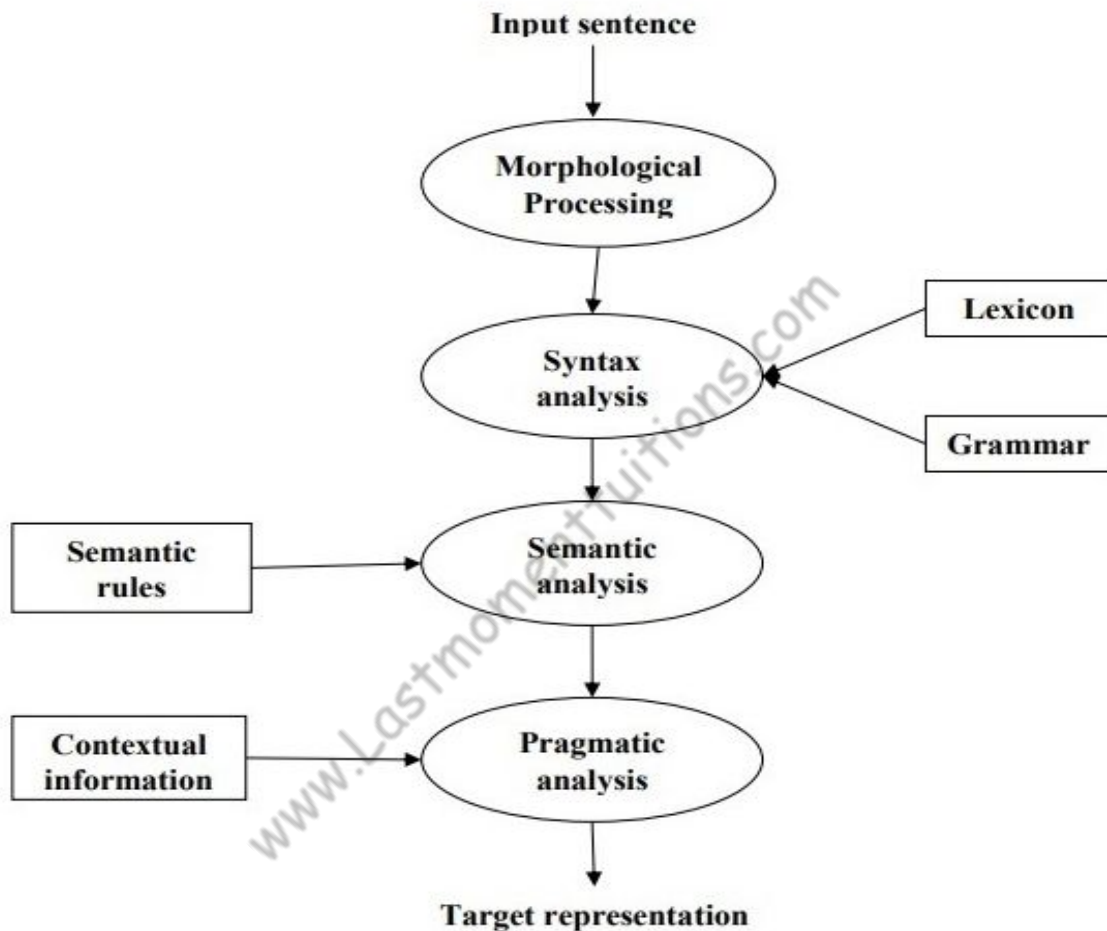
- This kind of ambiguity arises due to the use of anaphora entities in discourse.
- Anaphora : when same beginning of sentence is repeated several times
- Example of what anaphora is “my mother liked the house very much but she couldn’t buy it” here we are not repeating mother again and again we replaced it by she so here she is anaphora
- Now lets come back to anaphoric ambiguity and understand it with the help of example
- 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 can either be horse or hill which cause anaphoric ambiguity.

## Pragmatic ambiguity

- It occurs when sentence gives it multiple interpretations or it is not specific.
- 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 does).

## Natural Language Processing Phases

Following diagram shows the phases or logical steps in natural language processing



### Morphological Processing

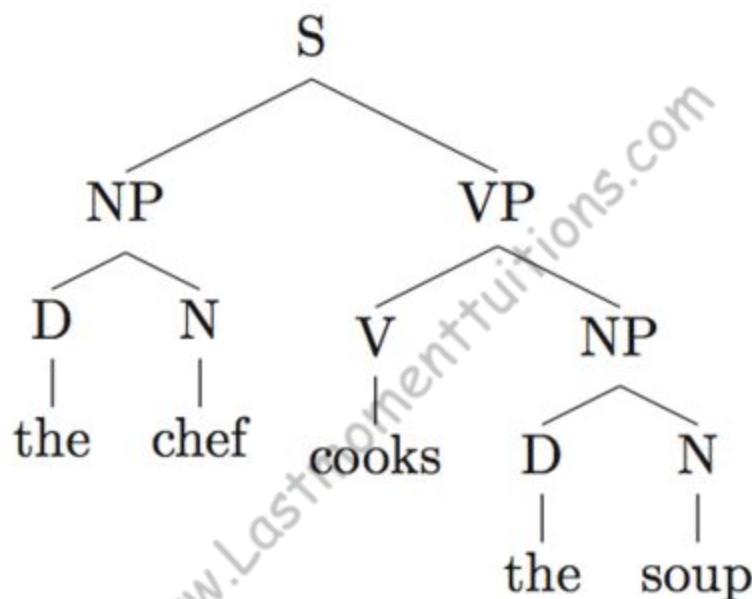
- It is the first phase of NLP.
- The purpose of this phase is to break chunks of language input into sets of tokens corresponding to paragraphs, sentences and words.
- For example, a word like “uneasy” can be broken into two sub-word tokens as “un-easy”.

## Syntax Analysis

It is the second phase of NLP.

The purpose of this phase is two folds:

1. To check that a sentence is well formed or not
  - Eg: “The school goes to the boy” would be rejected by syntax analyzer or parser.
2. Taking input data and giving structural representation of input data.
  - Example : “ The chef cooks the soup”



## Semantic Analysis

- It is the third phase of NLP.
- The purpose of this phase is to draw exact meaning, or you can say dictionary meaning from the text.
- The text is checked for meaningfulness.
- For example, semantic analyzer would reject a sentence like “Hot ice-cream”.

## Pragmatic Analysis

- It is the fourth phase of NLP.
- Pragmatic analysis simply fits the actual objects/events, which exist in a given context with object references obtained during the last phase (semantic analysis).

- For example, the sentence “The car hit the pole while it was moving” can have two semantic interpretations and pragmatic analyzer will choose between these two possibilities.

## **Application of Natural Language Processing**

- Machine Translation
- Sentimental Analysis
- Automatic Summarization
- Question answering
- Speech recognition

### **Machine Translation**

- Machine translation (MT), the process of translating one source language or text into another language, is one of the most important applications of NLP.
- There are different types of machine translation systems. Let us see what the different types are.
- Bilingual MT systems produce translations between two particular languages.
- Multilingual MT systems produce translations between any pair of languages. They may be either unidirectional or bi-directional in nature.
- Example : Google translator

### **Sentiment Analysis**

- Another important application of natural language processing (NLP) is sentiment analysis.
- As the name suggests, sentiment analysis is used to identify the sentiments among several posts.
- It is also used to identify the sentiment where the emotions are not expressed explicitly.
- Companies are using sentiment analysis, an application of natural language processing (NLP) to identify the opinion and sentiment of their customers online.
- It will help companies to understand what their customers think about the products and services.

- Companies can judge their overall reputation from customer posts with the help of sentiment analysis.
- In this way, we can say that beyond determining simple polarity, sentiment analysis understands sentiments in context to help us better understand what is behind the expressed opinion.

### **Automatic Summarization**

- In this digital era, the most valuable thing is data, or you can say information.
- However, do we really get useful as well as the required amount of information? The answer is 'NO' because the information is overloaded and our access to knowledge and information far exceeds our capacity to understand it.
- We are in a serious need of automatic text summarization and information because the flood of information over the internet is not going to stop.
- Text summarization may be defined as the technique to create short, accurate summary of longer text documents.
- Automatic text summarization will help us with relevant information in less time. Natural language processing (NLP) plays an important role in developing an automatic text summarization.

### **Question-answering**

- Another main application of natural language processing (NLP) is question-answering. Search engines put the information of the world at our fingertips, but they are still lacking when it comes to answering the questions posted by human beings in their natural language.
- We have big tech companies like Google are also working in this direction.
- Question-answering is a Computer Science discipline within the fields of AI and NLP.
- It focuses on building systems that automatically answer questions posted by human beings in their natural language.
- A computer system that understands the natural language has the capability of a program system to translate the sentences written by humans into an internal representation so that the valid answers can be generated by the system.
- The exact answers can be generated by doing syntax and semantic analysis of the questions. Lexical gap, ambiguity and multilingualism are some of the challenges for NLP in building a good question answering system.

## **Speech recognition**

- Speech recognition (enables computers to recognize and transform spoken language into text – dictation – and, if programmed, act upon that recognition – e.g. in case of assistants like Google Assistant Cortana or Apple's Siri)
- Speech recognition is simply the ability of a software to recognise speech.
- Anything that a person says, in a language of their choice, must be recognised by the software.
- Speech recognition technology can be used to perform an action based on the instructions defined by the human.
- Humans need to train the speech recognition system by storing speech patterns and vocabulary of their language into the system.
- By doing so, they can essentially train the system to understand them when they speak.
- Speech recognition and Natural Language processing are usually used together in Automatic Speech Recognition engines, Voice Assistants and Speech analytics tools.