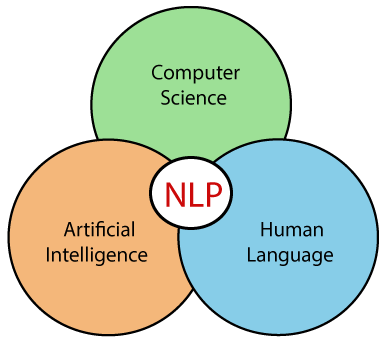
What is NLP?

NLP stands for **Natural Language Processing**, which is a part of **Computer Science, Human language,** and **Artificial Intelligence**. It is the technology that is used by machines to understand, analyse, manipulate, and interpret human's languages. It helps developers to organize knowledge for performing tasks such as **translation, automatic summarization, Named Entity Recognition (NER), speech recognition, relationship extraction,** and **topic segmentation**.



**Case Grammar**

Case Grammar was developed by **Linguist Charles J. Fillmore** in the year 1968. Case Grammar uses languages such as English to express the relationship between nouns and verbs by using the preposition.

In Case Grammar, case roles can be defined to link certain kinds of verbs and objects.

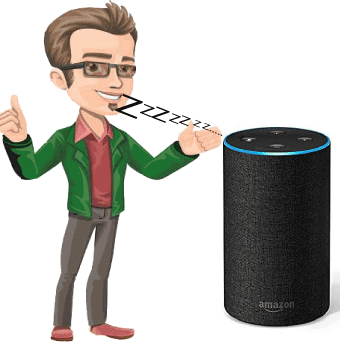
**For example:** "Neha broke the mirror with the hammer". In this example case grammar identify Neha as an agent, mirror as a theme, and hammer as an instrument.

## Applications of NLP

There are the following applications of NLP -

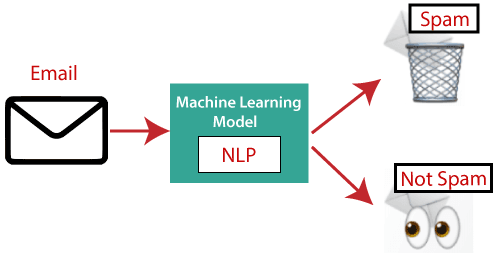
**1. Question Answering**

Question Answering focuses on building systems that automatically answer the questions asked by humans in a natural language.



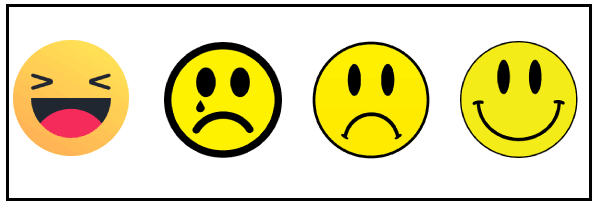
**2. Spam Detection**

Spam detection is used to detect unwanted e-mails getting to a user's inbox.



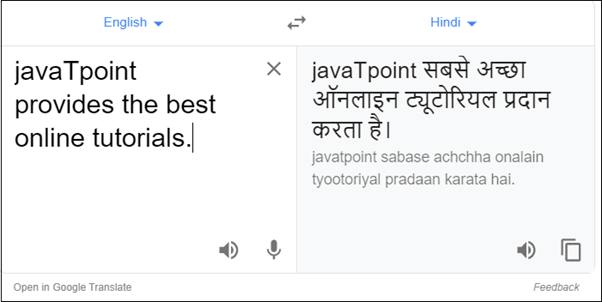
**3. Sentiment Analysis**

Sentiment Analysis is also known as **opinion mining**. It is used on the web to analyse the attitude, behaviour, and emotional state of the sender. This application is implemented through a combination of NLP (Natural Language Processing) and statistics by assigning the values to the text (positive, negative, or natural), identify the mood of the context (happy, sad, angry, etc.)



**4. Machine Translation**

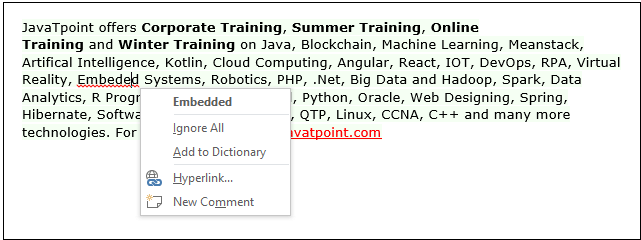
Machine translation is used to translate text or speech from one natural language to another natural language.



**Example:** Google Translator

**5. Spelling correction**

Microsoft Corporation provides word processor software like MS-word, PowerPoint for the spelling correction.

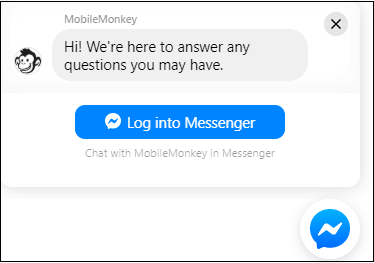


**6. Speech Recognition**

Speech recognition is used for converting spoken words into text. It is used in applications, such as mobile, home automation, video recovery, dictating to Microsoft Word, voice biometrics, voice user interface, and so on.

**7. Chatbot**

Implementing the Chatbot is one of the important applications of NLP. It is used by many companies to provide the customer's chat services.



**8. Information extraction**

Information extraction is one of the most important applications of NLP. It is used for extracting structured information from unstructured or semi-structured machine-readable documents.

**9. Natural Language Understanding (NLU)**

It converts a large set of text into more formal representations such as first-order logic structures that are easier for the computer programs to manipulate notations of the natural language processing.

## How to build an NLP pipeline

There are the following steps to build an NLP pipeline -

**Step1: Sentence Segmentation**

Sentence Segment is the first step for building the NLP pipeline. It breaks the paragraph into separate sentences.

**Example:** Consider the following paragraph -

**Independence Day is one of the important festivals for every Indian citizen. It is celebrated on the 15th of August each year ever since India got independence from the British rule. The day celebrates independence in the true sense.**

**Sentence Segment produces the following result:**

1. "Independence Day is one of the important festivals for every Indian citizen."
2. "It is celebrated on the 15th of August each year ever since India got independence from the British rule."
3. "This day celebrates independence in the true sense."

**Step2: Word Tokenization**

Word Tokenizer is used to break the sentence into separate words or tokens.

**Example:**

JavaTpoint offers Corporate Training, Summer Training, Online Training, and Winter Training.

Word Tokenizer generates the following result:

"JavaTpoint", "offers", "Corporate", "Training", "Summer", "Training", "Online", "Training", "and", "Winter", "Training", "."

**Step3: Stemming**

Stemming is used to normalize words into its base form or root form. For example, celebrates, celebrated and celebrating, all these words are originated with a single root word "celebrate." The big problem with stemming is that sometimes it produces the root word which may not have any meaning.

**For Example,** intelligence, intelligent, and intelligently, all these words are originated with a single root word "intelligen." In English, the word "intelligen" do not have any meaning.

**Step 4: Lemmatization**

Lemmatization is quite similar to the Stamming. It is used to group different inflected forms of the word, called Lemma. The main difference between Stemming and lemmatization is that it produces the root word, which has a meaning.

**For example:** In lemmatization, the words intelligence, intelligent, and intelligently has a root word intelligent, which has a meaning.

**Step 5: Identifying Stop Words**

In English, there are a lot of words that appear very frequently like "is", "and", "the", and "a". NLP pipelines will flag these words as stop words. **Stop words** might be filtered out before doing any statistical analysis.

**Example:** He **is a** good boy.

#### Note: When you are building a rock band search engine, then you do not ignore the word "The."

**Step 6: Dependency Parsing**

Dependency Parsing is used to find that how all the words in the sentence are related to each other.

**Step 7: POS tags**

POS stands for parts of speech, which includes Noun, verb, adverb, and Adjective. It indicates that how a word functions with its meaning as well as grammatically within the sentences. A word has one or more parts of speech based on the context in which it is used.

**Example: "Google"** something on the Internet.

In the above example, Google is used as a verb, although it is a proper noun.

**Step 8: Named Entity Recognition (NER)**

Named Entity Recognition (NER) is the process of detecting the named entity such as person name, movie name, organization name, or location.

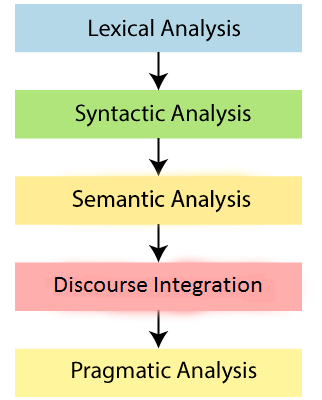
**Example: Steve Jobs** introduced iPhone at the Macworld Conference in San Francisco, California.

**Step 9: Chunking**

Chunking is used to collect the individual piece of information and grouping them into bigger pieces of sentences.

## Phases of NLP

There are the following five phases of NLP:



**1. Lexical Analysis and Morphological**

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.

**2. Syntactic Analysis (Parsing)**

Syntactic Analysis is used to check grammar, word arrangements, and shows the relationship among the words.

**Example:** Agra goes to the Poonam

In the real world, Agra goes to the Poonam, does not make any sense, so this sentence is rejected by the Syntactic analyzer.

**3. Semantic Analysis**

Semantic analysis is concerned with the meaning representation. It mainly focuses on the literal meaning of words, phrases, and sentences.

**4. Discourse Integration**

Discourse Integration depends upon the sentences that proceeds it and also invokes the meaning of the sentences that follow it.

**5. Pragmatic Analysis**

Pragmatic is the fifth and last phase of NLP. It helps you to discover the intended effect by applying a set of rules that characterize cooperative dialogues.

**For Example:** "Open the door" is interpreted as a request instead of an order.

## Why NLP is difficult?

NLP is difficult because Ambiguity and Uncertainty exist in the language.

**Ambiguity**

There are the following three ambiguity -

* **Lexical Ambiguity**

Lexical Ambiguity exists in the presence of two or more possible meanings of the sentence within a single word.

**Example:**

Manya is looking for a **match**.

In the above example, the word match refers to that either Manya is looking for a partner or Manya is looking for a match. (Cricket or other match)

* **Syntactic Ambiguity**

Syntactic Ambiguity exists in the presence of two or more possible meanings within the sentence.

**Example:**

I saw the girl with the binocular.

In the above example, did I have the binoculars? Or did the girl have the binoculars?

* **Referential Ambiguity**

Referential Ambiguity exists when you are referring to something using the pronoun.

**Example:** Kiran went to Sunita. She said, "I am hungry."

In the above sentence, you do not know that who is hungry, either Kiran or Sunita.