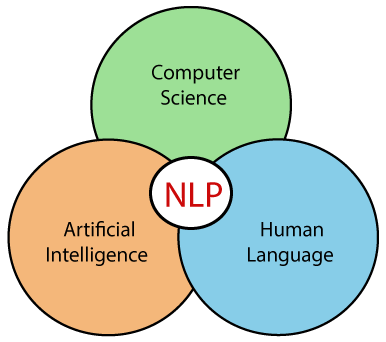
**NLP: THE BASIC PHILOSOPHY**

Natural Language Processing (NLP) is the phenomenon of processing the textual information in natural language form (commonly ‘english’ language), which usually involves one of the two types of processing.

***Note:*** NLP is also referred to as computational linguistics since it combines computation- intensive computer science with the processing of human (natural) language, formally referred to as linguistics.



***Fig: Interrelationship amongst AI, CS, NLP and Natural Language***

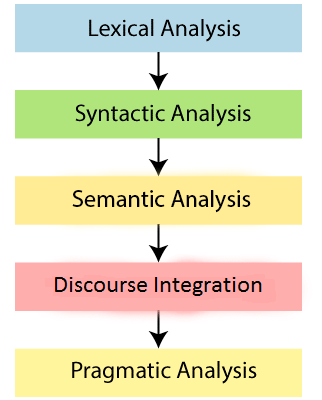
**I.Natural Language Generation(NLG):**

In this phenomenon, the ma- chine generates a response to the environment in natural language form. A concept called sentiment analysis where the intent or the underlying sentiment of the given statement is analyzed, is where we can see a practical example of natural language understanding phenomenon. In applications like Question- Answer systems, the solution model in that case shows the phenomenon of natural language generation, where answer-text is generated in response to the question- text.

**II.Natural Language Understanding (NLU):**

it’s the phenomenon where the machine understands the text that is present in its environment in natural language form. The process of understanding, interpreting and analyzing is multi- staged and detailed in a scientific and systematic way.

Next, we briefly discuss components of the NLU pipeline. Consider the following diagram for reference:



***Fig: The Various Stages of the NLU Pipeline***

**1.Lexical Analysis:**

It is the analysis of the textual information at the level of individual words or lexicons. Lexicons can have ambiguously taggable POS ( part of speech), which can be resolved by POS- parse trees . Also, the lexicons with equivalent meanings can be connected through their root- words known as morphemes.

**2.Syntactic Analysis:**

It is the analysis of the sentences in terms of their structure. Here, we see words in terms of their POS- tags and functions they have in sentences be- cause of their specific POS tags.

**3. Semantic Analysis:**

It is the analysis of the words and sentences in terms of the meanings they possess , while being present in that specific part of the sentence. Words are given distinctive representations according to the meanings they possess. They are very important in understanding the concepts of word- embeddings from the depth.

**4. Disclosure Integration:**

It is the component that represents the fact that there is actually a semantic relationship or connection amongst consecutive sentences. It relies on the long- term dependencies between words and or phrases across sentences. So, it finds its conceptual use in algorithms involv- ing long-term dependencies like RNN, LSTM, B-LSTM , etc.

**5.Pragmatic Analysis:**

It is the analysis of the given words and sentences in terms of contextual meaning that they possess out of many ambiguously possible alternative possibilities.

**NLTK (Natural Language Toolkit)**

NLTK is a toolkit built for working with NLP in Python. It provides various text processing libraries with a lot of test datasets. A variety of tasks can be performed using NLTK such as tokenizing, parse tree visualization, etc, many of which will be discussed later in detail.

**Advantages:**

1.Nltk provides very efficient ways to deal with NLP problems that would take much more number of lines if other common libraries were to be used.

2. The libraries of NLTK provide very intuitive and easy-to-understand classes that let us think in more abstract and high- level ways.

**Regular Expressions**

A regular expression (shortened as regex or regexp) ,sometimes referred to as rational expression, is a sequence of characters that specifies a search pattern in text. Usually such patterns are used by string-searching algorithms for "find" or "find and replace" operations on strings, or for input validation. Regular expression techniques are developed in theoretical computer science and formal language theory.

Regular expressions, together with NLTK, help the programmers to solve even the complicated problems of computational linguistics in a few easy steps and in a very intuitive way.