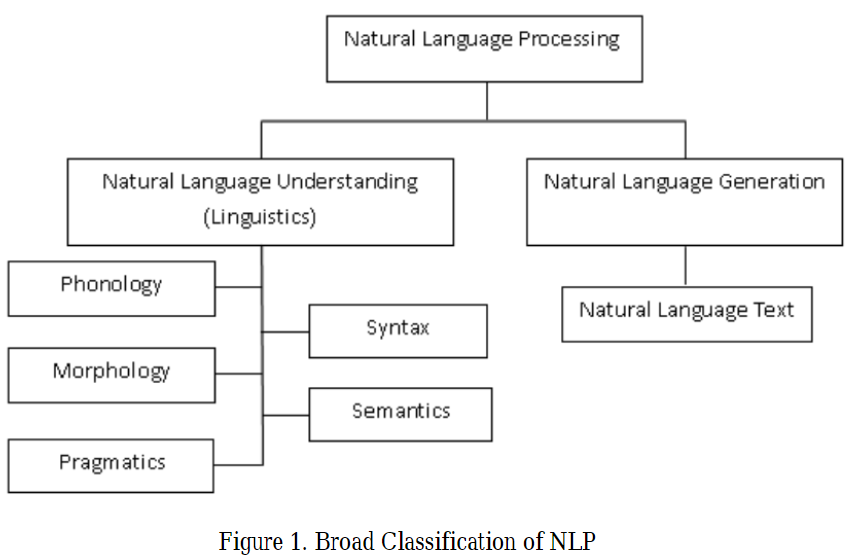
Natural Language Processing belongs to an area of AI and linguistics, it deals with making the computer to understand the statements or words written in human language. It helps humans to communicate with the computer in natural language. It deals with understanding text or speech by any software or machine. Usually, humans interrelate with others, know each other views, and react. In NLP, this interaction, understanding, and response are made by a computer instead of a human.

A language a set of rules or set of a symbol. A symbol is used for carrying information or to broadcasting the information.

NLP has two main parts,

Natural Language Understanding

Natural Language Generation (a process of understanding and generating text).



Linguistics is a study of language, it includes Phonology which deals with sound, Morphology is a word formation, Syntax which gives a structure of the sentence, Pragmatics and Semantics syntax deals with understanding. Basically, linguistics has two levels, Higher Level and lower level. The tasks concerned in NLP are Machine Translation, Part Of Speech classification, Morphological Segmentation, Optical Character acknowledgment, and Named Entity acknowledgment, Automatic Summarization, etc. The real world applications tasks among these are Machine translation, Optical character recognition, Named entity recognition, etc.

Levels are methods for demonstrating the NLP which generates NLP text by, Sentence setting up, realizing Content setting up and Surface Realization phase.

* Phonology

Phonology is the division of Linguistics which deal with the orderly understanding of sound. The term phono- means resonance, and the –logy refers to language. Phonology makes use of noise to encrypt meaning of human verbal communication.

* Morphology

Morphological segmentation is a process of breaking the word into segments called morphemes and identify the class of each. For example, the word prerequisite can be morphologically divided into three morphemes: the prefix pre, the root requires, and the suffix -site. The understanding of every morpheme remains similar for all words. It is the same as humans break the unidentified word into morphemes. The words that cannot be divided but have meaning are called Lexical morpheme for example table, chair, etc. The words such as -full,–ed, -est, -ing, ly, etc. can be combined with the lexical morpheme which is called Grammatical morphemes like Consulting, Likely, Smallest, Worked.

* Lexical

If the phrase tagged has an additional one or more part-of-speech, they assign with mainly part-of-speech tag based on the environment in which they are there. At this level, Semantic representations are replaced by the language that has one sense.

* Syntactic

This stage significantly stresses on the words in a sentence to bring out the grammatical arrangement of the sentence. Both parser and grammar are required in this. The output of this level is an illustration of the sentence that brings out the structural reliance association between the words.

* Semantic

Semantic processing finds the likely denotation of a sentence by pointing on the interactions amongst word-level meanings in the sentence. It merges the semantic disambiguation of words with multiple senses.

It is the development of generating sentences, phrases, and paragraphs that are significant from an interior illustration. This process is in the four phases of NLP which involves identifying the goals and planning how they can be achieved by using available communicative source and realizing plan as text.

Machine Translation

Text categorization

Spam filtering

Information extraction

Summarization

Medicine

[**Natural Language Toolkit**](http://www.nltk.org/)**:** It is a Python library which provides units for classification, analyzes grammatical structure of the sentence, stemming, tokenizing, categorizing things, dealing out text, and more.

**Gensim:** it is a library for unverified [natural language process](https://en.wikipedia.org/wiki/Natural_language_processing), and [topic modeling](https://en.wikipedia.org/wiki/Topic_model) using modern statistical [machine learning](https://en.wikipedia.org/wiki/Machine_learning).it is implemented in [Python](https://en.wikipedia.org/wiki/Python_(programming_language)) and [Cython](https://en.wikipedia.org/wiki/Cython). It can handle big text collection by means of incremental online algorithm and data streaming.

[**Apache OpenNLP**](https://opennlp.apache.org/)**:** It is a machine learning toolkit which provides sentence segmentation, tokenizes,  [named entity extraction](https://algorithmia.com/algorithms/ApacheOpenNLP/NamedEntityRecognition), [part-of-speech tagging](https://algorithmia.com/algorithms/ApacheOpenNLP/POSTagger), [parsing](https://algorithmia.com/algorithms/ApacheOpenNLP/ParseBySentence).

[**Standford NLP**](http://stanfordnlp.github.io/CoreNLP/)**:** provides the [sentiment analysis](https://algorithmia.com/algorithms/StanfordNLP/SentimentAnalysis), [named entity recognizer](https://algorithmia.com/algorithms/StanfordNLP/NamedEntityRecognition), [part-of-speech tagging](https://algorithmia.com/algorithms/StanfordNLP/PartofspeechTagger), [coreference resolution](https://algorithmia.com/algorithms/StanfordNLP/DeterministicCoreferenceResolution) system,  and more.

[**MALLET**](http://mallet.cs.umass.edu/index.php)**:** It is a Java package which provides bunching, information mining, topic modeling, [Latent Dirichlet Allocation](https://algorithmia.com/algorithms/mallet/LDA), document classification, and more.

NLTK means Natural Language Toolkit. It is the most powerful NLK libraries which hold packages to make

equipment recognize human spoken communication and respond to it with a suitable reply. NLTK is intended to hold up study and education in [NLP](https://en.wikipedia.org/wiki/Natural_language_processing) or closely linked areas, including experiential [linguistics](https://en.wikipedia.org/wiki/Linguistics), [cognitive science](https://en.wikipedia.org/wiki/Cognitive_science), [artificial intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence), [information retrieval](https://en.wikipedia.org/wiki/Information_retrieval), and [machine learning](https://en.wikipedia.org/wiki/Machine_learning). NLTK has been used productively as an education tool, as an individual study tool, and as a stage for prototyping and construction research systems.

Gensim is a [free](https://radimrehurek.com/gensim/intro.html#availability) Python library planned to mechanically take out semantic topics from papers, as strongly and painlessly as possible.

Gensim is planned to process raw, unformed digital text ("plain text”).

The algorithms in Gensim, such as [Word2Vec](https://radimrehurek.com/gensim/models/word2vec.html#gensim.models.word2vec.Word2Vec), [FastText](https://radimrehurek.com/gensim/models/fasttext.html#gensim.models.fasttext.FastText), etc, routinely find out the semantic arrangement of documents by investigative statistical co-occurrence patterns within a number of training documents. These algorithms are **unconfirmed**, which means no human input is essential – you need a body of plain text documents.

Once this statistical outline is found, any plain text documents (sentence, phrase, word…) can be in a few words be expressed in the new, semantic illustration and query based for topical resemblance alongside other documents (words, phrases…).