**How does NLP work?**

Before we learn how NLP works, let's understand how humans use language-

Every day, we say thousand of a word that other people interpret to do countless things. We, consider it as a simple communication, but we all know that words run much deeper than that. There is always some context that we derive from what we say and how we say it., NLP never focuses on voice modulation; it does draw on contextual patterns.

Example:

Man is to woman as king is to \_\_\_\_\_\_\_\_\_\_?

Meaning (king) – meaning (man) + meaning ( woman)=?

The answer is- queen

Here, we can easily co-relate because man is male gender and woman is female gender. In the same way, the king is masculine gender, and its female gender is queen.

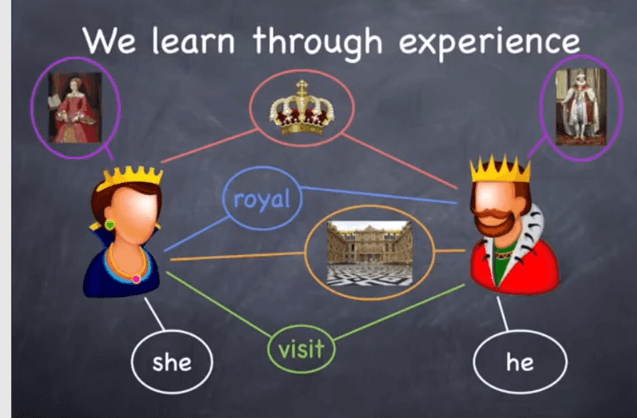
Example:

Is King to kings as the queen is to\_\_\_\_\_\_\_?

The answer is--- queens

Here, we can see two words kings and kings where one is singular and other is plural. Therefore, when the world queen comes, it automatically co-relates with queens again singular plural.

Here, the biggest question is that how do we know what words mean? Let's, say who will call it queen?

[](https://www.guru99.com/images/1/122118_0534_NLPNaturalL1.png)

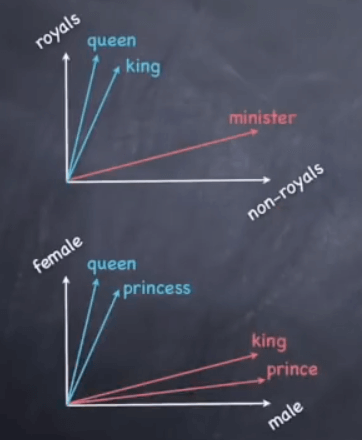
The answer is we learn this thinks through experience. However, here the main question is that how computer know about the same?

We need to provide enough data for Machines to learn through experience. We can feed details like

* Her Majesty the Queen.
* The Queen's speech during the State visit
* The crown of Queen Elizabeth
* The Queens's Mother
* The queen is generous.

With above examples the machine understands the entity Queen.

The machine creates word vectors as below. A word vector is built using surrounding words.

[](https://www.guru99.com/images/1/122118_0534_NLPNaturalL2.png)

The machine creates these vectors

* As it learns from multiple datasets
* Use Machine learning (e.g., Deep Learning algorithms)
* A word vector is built using surrounding words.

Here is the formula:

Meaning (king) – meaning (man) + meaning (woman)=?

This amounts to performing simple algebraic operations on word vectors:

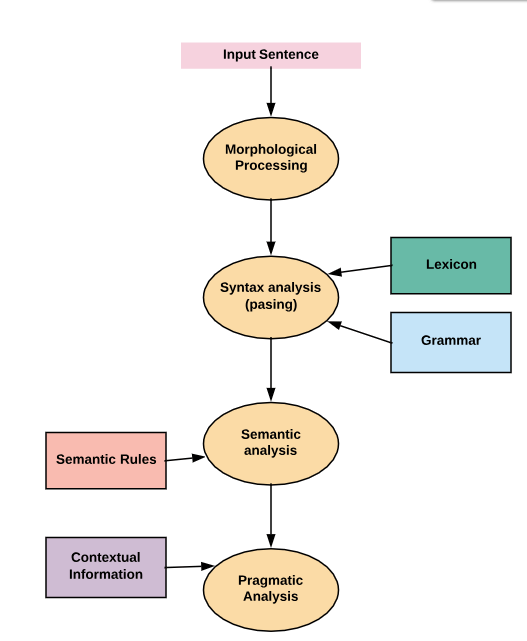
Vector ( king) – vector (man) + vector (woman)= vector(?)

To which the machine answers queen.

**Components of NLP**

Five main Component of Natural Language processing are:

* Morphological and Lexical Analysis
* Syntactic Analysis
* Semantic Analysis
* Discourse Integration
* Pragmatic Analysis

[](https://www.guru99.com/images/1/122118_0534_NLPNaturalL3.png)

**Morphological and Lexical Analysis**

Lexical analysis is a vocabulary that includes its words and expressions. It depicts analyzing, identifying and description of the structure of words. It includes dividing a text into paragraphs, words and the sentences

Individual words are analyzed into their components, and nonword tokens such as punctuations are separated from the words.

**Semantic Analysis**

Semantic Analysis is a structure created by the syntactic analyzer which assigns meanings. This component transfers linear sequences of words into structures. It shows how the words are associated with each other.

Semantics focuses only on the literal meaning of words, phrases, and sentences. This only abstracts the dictionary meaning or the real meaning from the given context. The structures assigned by the syntactic analyzer always have assigned meaning

E.g.. "colorless green idea." This would be rejected by the Symantec analysis as colorless Here; green doesn't make any sense.

**Pragmatic Analysis**

Pragmatic Analysis deals with the overall communicative and social content and its effect on interpretation. It means abstracting or deriving the meaningful use of language in situations. In this analysis, the main focus always on what was said in reinterpreted on what is meant.

Pragmatic analysis helps users to discover this intended effect by applying a set of rules that characterize cooperative dialogues.

E.g., "close the window?" should be interpreted as a request instead of an order.

**Syntax analysis**

The words are commonly accepted as being the smallest units of syntax. The syntax refers to the principles and rules that govern the sentence structure of any individual languages.

Syntax focus about the proper ordering of words which can affect its meaning. This involves analysis of the words in a sentence by following the grammatical structure of the sentence. The words are transformed into the structure to show hows the word are related to each other.

**Discourse Integration**

It means a sense of the context. The meaning of any single sentence which depends upon that sentences. It also considers the meaning of the following sentence.

For example, the word "that" in the sentence "He wanted that" depends upon the prior discourse context.

**NLP and writing systems**

The kind of writing system used for a language is one of the deciding factors in determining the best approach for text pre-processing. Writing systems can be

1. Logographic: a Large number of individual symbols represent words. Example Japanese, Mandarin
2. Syllabic: Individual symbols represent syllables
3. Alphabetic: Individual symbols represent sound

Majority of the writing systems use the Syllabic or Alphabetic system. Even English, with its relatively simple writing system based on the Roman alphabet, utilizes logographic symbols which include Arabic numerals, Currency symbols (S, £), and other special symbols.

This pose following challenges

* Extracting meaning(semantics) from a text is a challenge
* NLP is dependent on the quality of the corpus. If the domain is vast, it's difficult to understand context.
* There is a dependence on the character set and language

**How to implement NLP**

Below, given are popular methods used for Natural Learning Process:

**Machine learning:**The learning nlp procedures used during machine learning. It automatically focuses on the most common cases. So when we write rules by hand, it is often not correct at all concerned about human errors.

**Statistical inference:**NLP can make use of statistical inference algorithms. It helps you to produce models that are robust. e.g., containing words or structures which are known to everyone.

**NLP Examples**

Today, Natual process learning technology is widely used technology.

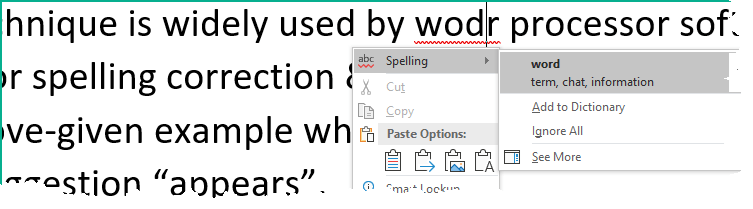
Here, are common Application' of NLP:

**Information retrieval & Web Search**

Google, Yahoo, Bing, and other search engines base their machine translation technology on NLP deep learning models. It allows algorithms to read text on a webpage, interpret its meaning and translate it to another language.

**Grammar Correction:**

NLP technique is widely used by word processor software like MS-word for spelling correction & grammar check.

[](https://www.guru99.com/images/1/122118_0534_NLPNaturalL4.png)

**Question Answering**

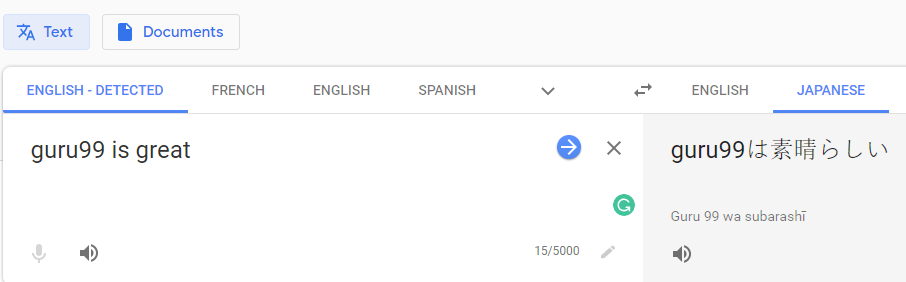
Type in keywords to ask Questions in Natural Language.

**Text Summarization**

The process of summarising important information from a source to produce a shortened version

**Machine Translation**

Use of computer applications to translate text or speech from one natural language to another.

[](https://www.guru99.com/images/1/122118_0534_NLPNaturalL5.png)

**Sentiment analysis**

NLP helps companies to analyze a large number of reviews on a product. It also allows their customers to give a review of the particular product.

**Future of NLP**

* Human readable natural language processing is the biggest Al- problem. It is all most same as solving the central artificial intelligence problem and making computers as intelligent as people.
* Future computers or machines with the help of NLP will able to learn from the information online and apply that in the real world, however, lots of work need to on this regard.
* Naturla language toolkit or nltk become more effective
* Combined with natural language generation, computers will become more capable of receiving and giving useful and resourceful information or data.

**Natural language vs. Computer Language**

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Natural Language** | **Computer Languages** |
| Ambiguous | They are ambiguous in nature. | They are designed to unambiguous. |
| Redundancy | Natural languages employ lots of redundancy. | Formal languages are less redundant. |
| Literalness | Natural languages are made of idiom & metaphor | Formal languages mean exactly what they want to say |

**Advantages of NLP**

* Users can ask questions about any subject and get a direct response within seconds.
* NLP system provides answers to the questions in natural language
* NLP system offers exact answers to the questions, no unnecessary or unwanted information
* The accuracy of the answers increases with the amount of relevant information provided in the question.
* NLP process helps computers communicate with humans in their language and scales other language-related tasks
* Allows you to perform more language-based data compares to a human being without fatigue and in an unbiased and consistent way.
* Structuring a highly unstructured data source

**Disadvantages of NLP**

* Complex Query Language- the system may not be able to provide the correct answer it the question that is poorly worded or ambiguous.
* The system is built for a single and specific task only; it is unable to adapt to new domains and problems because of limited functions.
* NLP system doesn't have a user interface which lacks features that allow users to further interact with the system

**Summary**

* Natural Language Processing is a branch of AI which helps computers to understand, interpret and manipulate human language
* NLP started when Alan Turing published an article called "Machine and Intelligence".
* NLP never focuses on voice modulation; it does draw on contextual patterns
* Five essential components of Natural Language processing are 1) Morphological and Lexical Analysis 2)Syntactic Analysis 3) Semantic Analysis 4) Discourse Integration 5) Pragmatic Analysis
* Three types of the Natural process writing system are 1)Logographic 2) Syllabic 3) Alphabetic
* Machine learning and Statistical inference are two methods to implementation of Natural Process Learning
* Essential Applications of NLP are Information retrieval & Web Search, Grammar Correction Question Answering, , Text Summarization, Machine Translation, etc.
* Future computers or machines with the help of NLP and Data Science will able to learn from the information online and apply that in the real world, however, lots of work need to on this regard
* NLP is are ambiguous while open source computer language is designed to unambiguous
* The biggest advantage of the NLP system is that it offers exact answers to the questions, no unnecessary or unwanted information
* The biggest draw back of the NLP system is built for a single and specific task only so it is unable to adapt to new domains and problems because of limited functions