**1. HISTORY:**

This article is about language processing by computers. For the processing of language by the human brain, see [Language processing in the brain](https://en.wikipedia.org/wiki/Language_processing_in_the_brain).

**Natural language processing** (**NLP**) is a field of [computer science](https://en.wikipedia.org/wiki/Computer_science), [artificial intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence), and [computational](https://en.wikipedia.org/wiki/Computational_linguistics) language concerned with the interactions between [computers](https://en.wikipedia.org/wiki/Computer) and [human (natural) languages](https://en.wikipedia.org/wiki/Natural_language) and, in particular, concerned with programming computers to process large [natural language corpora](https://en.wikipedia.org/wiki/Corpus_linguistics).

Challenges in natural language processing frequently involve [natural language understanding](https://en.wikipedia.org/wiki/Natural_language_understanding), [natural language generation](https://en.wikipedia.org/wiki/Natural_language_generation) frequently from [formal, machine-readable logical](https://en.wikipedia.org/wiki/Formal_language) forms  [connecting language and machine perception](https://en.wikipedia.org/wiki/Symbol_grounding_problem), [managing human-computer dialog systems](https://en.wikipedia.org/wiki/Dialog_system), or some combination thereof.

**Natural Language Programming** (NLP) is an ontology-assisted way of programming in terms of natural language sentences like English language.

A structured document with Content, sections and subsections for explanations of sentences forms a NLP document, which is actually a computer program. Natural languages and natural language user interfaces include a natural programming language for making interactive function.

The history of NLP generally starts in the 1950s, although work can be found from earlier periods. In 1950, Alan Turing published an article titled "Computing Machinery and Intelligence" which proposed what is now called the Turing test as a criterion of intelligence.

Many of the early successes occurred in the field of machine translation, due especially to work at IBM Research, where successively more complicated statistical models were developed.

These systems were able to take advantage of existing multilingual textual corpora that had been produced by the of Canada and the European Union as a result of laws calling for the translation of all governmental proceedings into all official languages of the corresponding systems of government.

However, most other systems depended on corpora specifically developed for the tasks implemented by these systems, which was (and often continues to be) a major limitation in the success of these systems. As a result, a great deal of research has gone into methods of more effectively learning from limited amounts of data.

In 1950, Alan Turing published an article titled "Machine and Intelligence" which advertised what is now called the turning test as a subfield of intelligence. Some beneficial and successful Natural language systems were developed in the 1960s were SHRDLU, a natural language system working in restricted "blocks worlds" with restricted vocabularies was written between 1964 to 1966.

**2. INTRODUCTION:**

Natural language processing (NLP) is a field of computer science and linguistics concerned with the interactions between computers and human (natural) languages. NLP is the branch of computer science focused on developing systems that allow computers to communicate with people using everyday language.

Natural Language processing refers to the way humans use words to communicate ideas and feelings, and how such communications are processed and understood. Thus, it is how the brain creates and understands language. Most recent theories consider that this process is carried out entirely by and inside the brain.

Humans can process natural languages, but for us the question is whether digital computers can or ever will process natural languages. This article is about language processing by computers. For the processing of language by the human brain, see Language processing in the brain.

Natural languages are those languages that are spoken by the people. A natural language processing girdle everything a computer needs to understand natural language and also generates natural language. Natural Language Processing is a subfield of Artificial Intelligence and linguistic, devoted to make computers understand the statements or words written in human languages.

A Natural language also known as ordinary language that is spoken or written by people (humans) for general purpose communication. A language is a system, a set of rules or set of symbols.

* Symbols are combined and used for conveying information or broadcasting the information.
* Rules tyrannize handling of symbols.NLP Besets anything a computer or machine needs to understand typed or spoken (natural language).

A better human-computer interface that could convert from a natural language into a computer language and vice versa. A natural language system could be the interface to a database system, such as for a travel agent to use in making reservations. Blind people could use natural language system (with speech recognition) to interact with computers, and Steven Hawking uses one to generate speech from his typed text.

Even if programs that translate between human languages are not perfect, they would still be useful in that they could do the rudimentary translation first, with their work checks and corrected by a human translator. Programs that could check for grammar and writing techniques in a word processing document.

NLP has significant overlap with the field of computational linguistics, and is often considered a sub-field of artificial intelligence. Modern NLP algorithms are grounded in machine learning, especially statistical machine learning. Research into modern statistical NLP algorithms requires an understanding of a number of disparate fields, including linguistics, computer science, statistics (particularly Bayesian statistics), linear algebra and optimization theory.

**3. DEFINITION OF NLP**

**(Natural Language Processing)**

"Natural language processing” here refers to the use and ability of systems to process sentences in a natural language such as English, rather than in a specialized artificial computer language such as C++.

The systems of real interest here are digital computers of the type we think of as personal computers and mainframes (and not digital computers in the sense in which "we are all digital computers," if this is even true).

Natural language came into existence because when user wishes to communicate with the computer we can’t force the users to learn machine specific language so this basically caters to managers or children’s who do not have enough time to learn new specific languages or get skilled in them. Languages can be any like Hindi, French, English, and Chinese etc…

**4. DETAILS Of NLP:**

This section provides a brief history of NLP, introduces some of the main problems involved in extracting meaning from human languages and examines the kind of activities performed by NLP systems.

**4.1 Background:**

Natural language processing systems take strings of words (sentences) as their input and produce structured representations capturing the meaning of those strings as their output. The nature of this output depends heavily on the task at hand. A natural language understanding system serving as an interface to a database might accept questions in English which relate to the kind of data held by the database.

In this case the meaning of the input (the output of the system) might be expressed in terms of structured SQL queries which can be directly submitted to the database. The first use of computers to manipulate natural languages was in the 1950s with attempts to automate translation between Russian and English [Locke & Booth].

These systems were spectacularly unsuccessful requiring human Russian-English translators to pre-edit the Russian and post-edit the English. Based on World War-II code breaking techniques, they took individual words in isolation and checked their definition in a dictionary. They were of little practical use. Popular tales about these systems cite many miss-translations including the phrase "hydraulic ram" translated as "water goat".

In the 1960s natural language processing systems started to examine sentence structure but often in an ad hoc manner. These systems were based on pattern matching and few derived representations of meaning. The most well known NLP System is *Eliza*, though this system was not the most impressive in terms of its ability to extract meaning from language.

Serious developments in natural language processing took place in the early & mid 1970s as systems started to use more general approaches and attempt to formally describe the rules of the language they worked with. LUNAR [Woods 1973] provided an English interface to a database holding details of moon rock samples.

Since that time there has been parallel development of ideas and technologies that provide the basis for modern natural language processing systems. Research in computer linguistics has provided greater knowledge of grammar construction and Artificial Intelligence researchers have produced more effective mechanisms for parsing natural languages and for representing meanings [Allen]. Natural language processing systems now build on a solid base of linguistic study and use highly developed semantic representations

Recently (during the 1990s) natural language systems have either focused on specific, limited domains with some success or attempted to provide general purpose language understanding ability with less success. A major goal in contemporary language processing research is to produce systems which work with complete threads of discourse (with human like abilities) rather than only with isolated sentences. Successes in this area are currently limited.

**4.2 Problems:**

Two problems in particular make the processing of natural languages difficult and cause different techniques to be used than those associated with the construction of compilers etc for processing artificial languages. These problems are:

1. The level of ambiguity that exists in natural languages
2. The complexity of semantic information contained in even simple sentences.

Typically, language processors deal with large numbers of words, many of which have alternative uses, and large grammars which allow different phrase types to be formed from the same string of words. Language processors are made more complex because of the irregularity of language and the different kinds of ambiguity which can occur.

For example:

1. Cats play with string.

\* Cat play with string.

1. I saw the racing pigeons flying to Paris.

I saw the Eiffel Tower flying to Paris.

**4.3 NLP using machine learning**

Modern approaches to natural language processing (NLP) are grounded in machine learning. The paradigm of machine learning is different from that of most prior attempts at language processing. Prior implementations of language -processing tasks typically involved the direct hand coding of large sets of rules.

The machine-learning paradigm calls instead for using general learning algorithm, although not always, grounded in statistical inference — to automatically learn such rules through the analysis of large corpora of typical real-world examples.

Example:

Consider the task of part of speech tagging, i.e. determining the correct part of speech of each word in a given sentence, typically one that has never been seen before. A typical machine-learning-based implementation of a part of speech tagger proceeds in two steps, a training step and an evaluation step.

Systems based on machine-learning algorithms have many advantages over hand-produced rules:

1. The learning procedures used during machine learning automatically focus on the most common cases, whereas when writing rules by hand it is often not obvious at all where the effort should be directed.
2. Automatic learning procedures can make use of statistical inference algorithms to produce models that are robust to unfamiliar input (e.g. containing words or structures that have not been seen before) and to erroneous input (e.g. with misspelled words or words accidentally omitted).
3. Systems based on automatically learning the rules can be made more accurate simply by supplying more input data. However, systems based on hand-written rules can only be made more accurate by increasing the complexity of the rules, which is a much more difficult task.

**4.4 Statistical NLP**

Statistical natural-language processing uses probabilistic and statistical methods to resolve some of the difficulties discussed above, especially those which arise because longer sentences are highly ambiguous when processed with realistic grammars, thousands or millions of possible analyses.

Methods for disambiguation often involve the use of corpora and Markov models. Statistical NLP comprises all quantitative approaches to automated language processing, including probabilistic modeling, information theory, and linear binomial. The technology for statistical NLP comes mainly from machine learning and data mining, both of which are fields of artificial intelligence that involve learning from data.

**4.5 Evaluation of natural language processing**

The goal of NLP evaluation is to measure one or more qualities of an algorithm or a system, in order to determine whether (or to what extent) the system answers the goals of its designers, or meets the needs of its users. Research in NLP evaluation has received considerable attention, because the definition of proper evaluation criteria is one way to specify precisely an NLP problem, going thus beyond the vagueness of tasks defined only as language understanding or language generation. A precise set of evaluation criteria, which includes mainly evaluation data and evaluation metrics, enables several teams to compare their solutions to a given NLP problem.

Depending on the evaluation procedures, a number of distinctions are traditionally made in NLP evaluation:

* Intrinsic vs. extrinsic evaluation
* Black-box vs. glass-box evaluation
* Automatic vs. manual evaluation

**4.6 Selection of NLP Tasks**

1. **Word segmentation:**

Separate a chunk of continuous text into separate words. For a language like English, this is fairly trivial, since words are usually separated by spaces. However, some written languages. Like - Chinese, Japanese etc…

1. **Topic segmentation and recognition:**

Given a chunk of text, separate it into segments each of which is devoted to a topic, and identify the topic of the segment.

1. **Part-of-speech tagging:**

Given a sentence, determine the part of speech for each word. Many words, especially common ones, can serve as multiple parts of speech. For example, “book” can be a noun (“the book on the table”) or

Verb (“to book a flight”); “set” can be a noun, verb or adjective; and “out” can be any of at least five different parts of speech.

1. **Word sense disambiguation:**

Many words have more than one meaning, we have to select the meaning which makes the most sense in context. For this problem, we are typically given a list of words and associated word senses, e.g. from a dictionary of from an online resource such as WordNet.

1. **Named entity recognition (NER):**

Given a stream of text, determine which items in the text map to proper names, such as people or places, and what the type of each such name is (e.g. person, location, organization).

1. **Parsing:**

Determine the parse tree (grammatical analysis) of a given sentence. The grammar for natural languages is ambiguous and typical sentences have multiple possible analyses.

**Where does it fit in the Classification?**

**Computer**

**Databases**

**Networking**

**Algorithms**

**Artificial Intelligence**

**Natural Language Processing (NLP)**

**Search**

**Robotics**

**Information Retrieval**

**Language Analysis**

**Machine Translation**

**Parsing**

**Semantics**

### H****ow does natural language processing work?****

To understand how NLP works, we have to take a look at the two main components of it, NLU and NLG. These two parts of NLP are very different from each other and are achieved by using different methods.

**Why NLP ??**

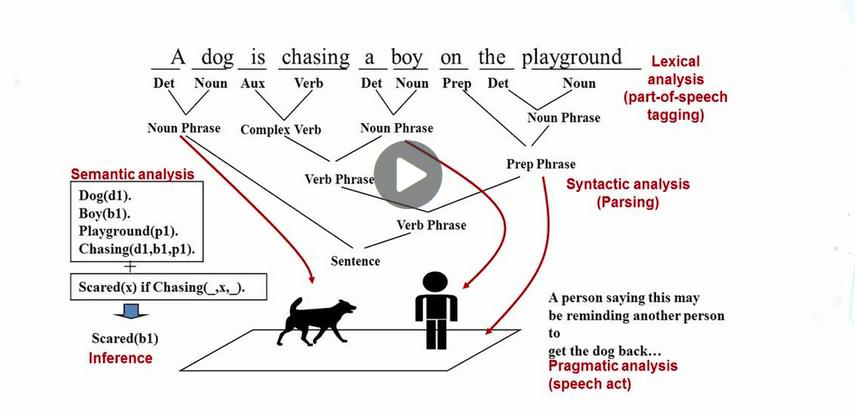
Huge amounts of data

Internet = at least 20 billions pages

Applications for processing large amounts of texts require NLP expertise

**5. EXMPLES OF NLP:**

**Examples of Natural Language Processing are** Based on artificial intelligence algorithms and driven by an increased need to manage unstructured enterprise information along with structured data, **Natural Language Processing  (NLP) is influencing a rapid acceptance of more intelligent solutions** in various end‐use applications. In this post, we’ll look at a few natural language processing techniques.



**Natural Language Processing - the tasks involved**

A simplified view of Natural Language Processing emphasizes four distinct stages [Fig 2.1].In real systems these stages rarely all occur as separated, sequential processes. In the overview that follows it is assumed that syntactic analysis and semantic analysis will be dealt with by the same mechanism - the parser. The rest of this section examines processes shown in the diagram.

**Input Sentence**

**Lexicon**

**Grammar**

### 

**Semantic Rules**

### 

### 

**Contextual Information**

### 

**Target**

**Representation**

### 

### 

### Figure: 1The logical steps in Natural language Processing

### 6. COMPONENTS OF NLP:

There are two components of NLP as given –

1. Natural Language Understanding(NLU) ,
2. Natural Language Generation(NLG)

**1)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.

### 2) 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 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.

### ****7. APPLICATION OF NLP:****

There are seven applications of NLP as follows –

### Text Processing

### Morph Analyzer

### POS Tagger

### Parsing

### Machine Translation

### Speech Processing

### Text to Speech (TTS)

### Text Processing:

### 

### Text processing lets you process (natural) language texts. You can detect the text’s language, the quality of the writing, find entity mentions, tag part-of-speech, extract dates, extract locations, or determine the sentiment of the text.

### Morph Analyzer:

### Morphology is a part of linguistics that deals with the study of words, their internal structure and partially their meanings .Morph Analyzer is a program for analyzing the morphology of an input word. It detects the morphemes of any text.

### Example:-

### Phrase level Morph Analyzer, Word level Morph Analyzer.

### 

### POS Tagger:

### A Part-Of-Speech Tagger (POS Tagger) is a piece of software that reads text in some language and assigns part of speech to each word ,such as noun verb ,adjective , etc ., although generally computational applications use more fine-grained POS tags like ‘noun-plural’.

### Parsing:

### In computer technology , a parser is a program , usually part of compiler , that receives input in the form of sequential source program instructions ,interactive online commands ,markup tags ,or some other defined interface and breaks them into parts (for example ,the nouns (objects , verbs (methods), and their attributes or options) that can then be managed by other programming (for example other components in a compiler).

### A parser may also check to see that all input has been provided that is necessary.

### Machine Translation:

### 

### Machine translation is a translation of a text by a computer with no human involvement. Machine translation system use combination of language and grammar rules plus dictionary for common words.

### Speech Processing:

### 

### Speech processing is the study of speech signals and the processing methods of these signals .The signals are usually processed in digital representation, so speech processing can be regarded as a special case of digital signal processing, applied to speech signal.

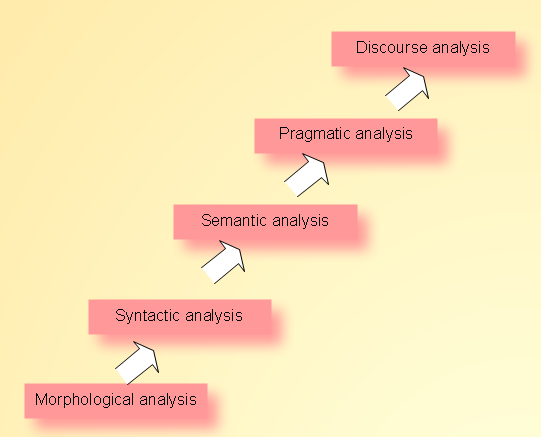
### Text to Speech (TTS):

### 

### A TTS converts normal language text into speech. Synthesized speech can be created by concatenating pieces of recorded speech that are stored in database.

### ****8. STEPS OF NLP:****

There are five steps of NLP as follows –



1. **Morphological analysis:**

A morpheme is the smallest part of a word that can carry a discrete meaning. Morphological analysis works with words at this level.

1. **Syntactic analysis:**

At this level, natural-language processors focus on structural information and relationships.

1. **Semantic analysis:**

Natural-language processors derive an absolute (dictionary definition) meaning from context.

1. **Pragmatic analysis:**

Natural-language processors derive knowledge from external commonsense information.

1. **Discourse analysis:**

Resolving references between sentences.

### ****9. NLP Related Diagrams:****

### pqr.jpg

### a.png

### ****10. ADVANTAGES:****

### 1. Better results all the way around

### Far and above any keyword matching or text-driven search, semantic search provides results that are true to form: exactly what your customers are looking for.

### 2. Search processing deciphers what your customers really mean

### Your customers are human, which means they’re fallible. They make spelling errors, confuse brands with products and forget details — it’s up to your on-site search to bridge the gap when these errors occur.

### 3. More data mined means more data for growth

### Measuring what your customers are searching is key in improving your business. This data can be applied across numerous facets of your business, from merchandising to SEO, marketing campaigns to sales and promotions and beyond.

### 4. Complex search capabilities eliminate ineffective results

### Being able to process numerous variables in a single search means providing a cumulative result that’s indicative of your customer’s end requirements.

### 5. Contextual understanding delivers answers

### Today’s search engines are slowly becoming Q&A boxes — customers ask questions and expect answers.

### 

### ****DISADVANTAGES:****

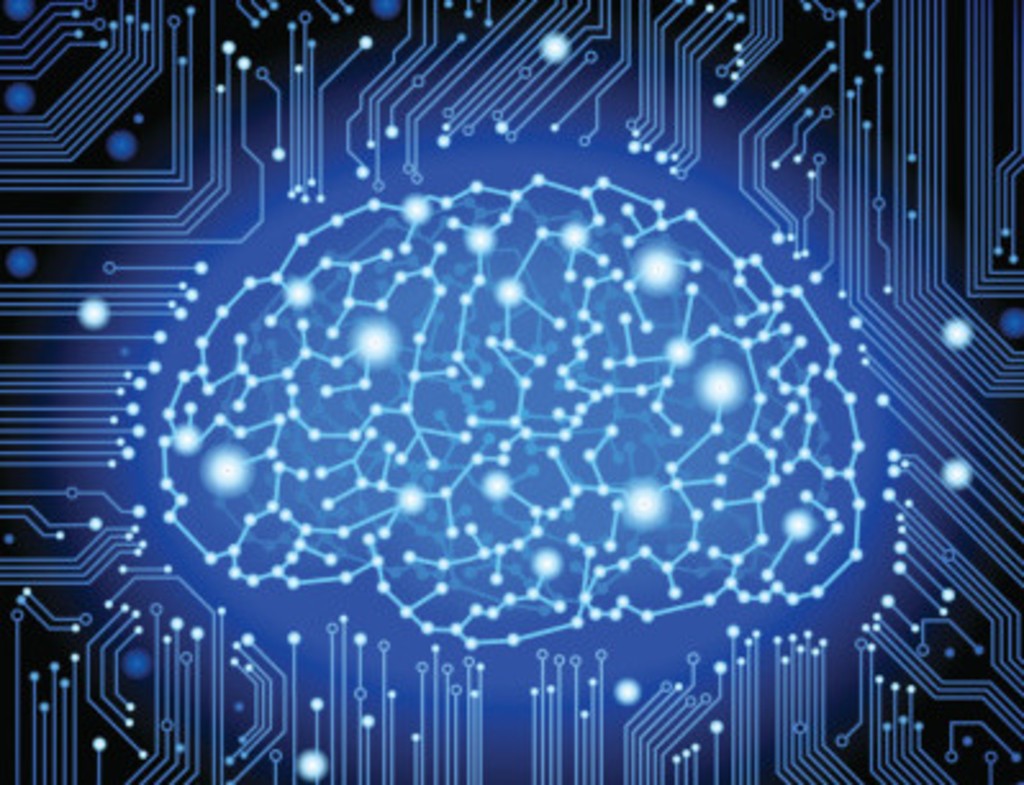
### Requires clarification dialogue

### May require more keystrokes

### May not show context

### Is unpredictable

### ****11. FUTURE OF NLP / Enhancement:****

****

We’re already seeing new ways and developing even better systems. Companies like Google are experimenting with Deep Neural Networks (DNNs) to push the limits of NLP and make it possible for humans-to-machine interactions to feel just like human-to-human interactions.

You can read more about how DNNs can significantly improve text-to-speech technology in this article. While we’re still a ways away from DNN-based text-to-speech engines from hitting the market, the potential for this technology is exciting!

**12. CONCLUSION:**

Natural language processing is a form of artificial intelligence that helps computers read and responds by simulating the human ability to understand everyday language. Many organizations use NLP techniques to optimize customer support, improve the efficiency of text analytics by easily finding the information they need, and enhance social media monitoring. For example, banks might implement NLP algorithms to optimize customer support; a large consumer products brand might combine natural language processing and semantic analysis to improve their knowledge management strategies and social media monitoring.

**13. REFERENCES:**

http://en.Wikipedia.org

http://www.Slideshare.net

<http://www.Doc.Ic.Ac.Uk>

<http://googlesystem.blogspot.in>

http://en.wikipedia.org/wiki/Natural\_language\_processing