# Brief History of Natural Language Processor (NLP)

**Research** · September 2023



**1 author:**

Hammad Hassan
COMSATS University Islamabad

# Brief History of NLP

**Hammad Hassan** **[Comsats University Islamabad, Lahore| HammadHassan611@outlook.com]**

**Abstract.** This article provides a comprehensive overview of the history of Natural Language Processing (NLP), tracing its evolution from its early roots in linguistics to its current prominence in the field of artificial intelligence. Beginning with Ferdinand de Saussure's groundbreaking ideas on language meaning, the article explores key milestones such as the IBM-Georgetown Demonstration, Noam Chomsky's generative grammar, and the development of semantic networks. It delves into the emergence of early chatbots like ELIZA and the challenges faced during the early years of NLP research.

Moving forward, the article discusses the paradigm shift from rule-based systems to machine learning algorithms in the late 1980s, propelled by increased computational power and the acceptance of corpus linguistics. Recurrent neural networks (RNNs), Long Short-Term Memory (LSTM) networks, and Boltzmann machines are introduced as pivotal developments. It highlights the transition to transformer architectures and the significance of large language models like GPT-3, GPT-2, and Microsoft's Turing Natural Language Generation.

The article also touches upon real-world applications such as Google Translate and the role of the Internet in data collection and distribution. It reflects on the history of AI's hype cycles, known as AI winters, and the renewed interest driven by deep learning. While acknowledging the current AI hype, it emphasizes the need for realistic expectations and the potential of AI to address a wide array of challenges in the future.

## 1 Introduction

Natural language processing (NLP), a combination of linguistics and computer science, is an aspect of Artificial intelligence that is trained to communicate with humans in human nature. This communication involves voice recognition, speech-to-text, generative text, and various more. According to Charles Rao NLP is:

> *"A branch of artificial intelligence that extracts interesting patterns in textual data, using a unique set of techniques."*

## 2 Early History

The History of NLP started back in early 1916 when Swiss professor Ferdinand de Saussure's book was published by his colleagues. He argued that the meaning of a sentence in a language is in the relations and differences between parts of a sentence and context. He proposed that the meaning of a word is not inherent in the word itself but is created by its relationship to other words in the language. This means that there is no natural or necessary connection between the two. The meaning of a word is not just determined by its relationship to other words, but also by its difference from other words. Saussure's theory of meaning has become the basis of the modern NLP. This theory led us to the very basic and simple computer Russian-to-English translator. This experiment was exhibited in the IBM-Georgetown Demonstration of 1954 and was very limited.

Later in 1957, the idea of generative grammar was introduced by Chomsky in the famous book Syntactic Structures, which helped researchers to better understand how machine translation could work. Chomsky attempts to create a set of rules able to correctly predict whether a text is grammatically (concerning a specific language) correct or not. It's a system of explicit (and usually recursive) rules that can be used to generate all the possible sentences in a language.

Later in 1956, Richard H. Richens first implemented semantic nets for computers as an interlingua for machine translation of natural languages. A semantic network is a graph-based knowledge representation system that represents knowledge as a network of nodes and edges. The nodes represent concepts, and the edges represent the relationships between concepts. Semantic networks can be used to represent a wide variety of knowledge, including facts, concepts, and relationships. However, the importance of this work was not fully recognized until later. Robert F. Simmons and Sheldon Klein independently implemented semantic nets in 1965 using first-order predicate calculus. They were inspired by a demonstration by Victor Yngve, who is considered the founder of this line of research. In 1960, Yngve published descriptions of algorithms for using phrase structure grammar to generate syntactically well-formed nonsense sentences.

Charles J. Fillmore, an American linguist, created the case grammar system in 1968 as part of transformational grammar. This theory analyzes the surface syntactic structure of sentences by examining the combination of deep cases (semantic roles, such as Agent, Object, Beneficiary, Location, or Instrument) required by a specific verb. For example, the verb "give" in English requires an Agent (A), an Object (O), and a Beneficiary (B). The sentence "Ali gave money to the hospital" illustrates this, with Ali (A) giving money (O) to the hospital (B).

In 1964 at MIT by Joseph Weizenbaum, Eliza a computer chatbot was developed. It was designed to explore communication between humans and machines by simulating conversation. ELIZA used a pattern matching and substitution methodology to give users the illusion of understanding, but they did not understand what was being said. The ELIZA program itself was written in MAD-SLIP, but the pattern-matching directives that contained most of its language capability were provided in separate "scripts". The script, DOCTOR, simulated a psychotherapist of the Rogerian school. This school of therapy often involves the therapist reflecting on the patient's words to the patient.

ELIZA used rules in the script to respond with non-directional questions to user inputs. As such, ELIZA was one of the first chatterbots (or chatbots) and one of the first programs capable of attempting the Turing test. The Turing test is a test of a machine's ability to simulate human conversation in such a way that a human interrogator cannot reliably tell whether they are talking to a machine or a human.

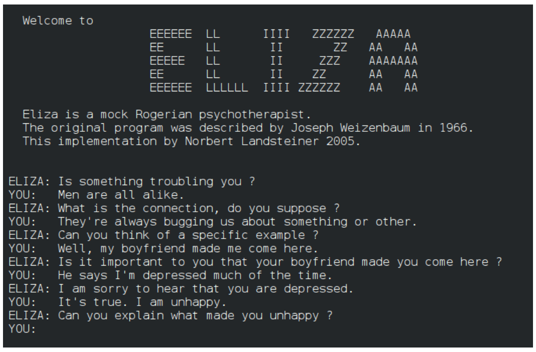


Image from https://en.wikipedia.org/wiki/ELIZA.

ELIZA was not without its limitations. It was unable to understand the meaning of words or phrases, and it could only respond to a limited range of prompts and questions. However, it was a significant step forward in the development of natural language processing and artificial intelligence. Although it was a very weak system, it created hype about artificial intelligence among researchers and the public and is called the first AI winter (hype of Artificial intelligence).

In the same year 1964, the U.S. The National Research Council (NRC) established the Automated Language Processing Advisory Committee, or ALPAC for short for the development of NLP and Artificial Intelligence (AI). This committee was tasked with reviewing developments in natural language research. In 1966, the NRC and Alpac launched their first AI-NLP break and stopped funding research in natural language processing and machine translation After 12 years of research, and $20 million, machine translation was still more expensive than human manual translation, and computers could and there was no basic conversation. In 1966, many (though not all) viewed artificial and natural intelligence language processing (NLP) as a formidable strategy because of the lack of resources, not enough storage space, and the slow speed of computers. This was a big shock for the researchers.

Roger Schank introduced conceptual dependency theory in 1969 at Stanford University, in the early days of artificial intelligence. Schank developed the model to represent knowledge for natural language input into computers. His goal was to make the meaning of a sentence independent of the words used in the input so that two sentences with the same meaning would have the same representation. The system was also intended to draw logical inferences.

Despite having very rare resources, no internet facilities, and low storage capacity of computers, there was much research on the linguistics and modern study of languages. The research was conducted internationally across borders including the USSR, USA, Europe, and Japan. Although it does not seem much, it was a progressive step toward modern and advanced AI models and chatbots.

# 3 History From 1980s to 2000s

Early natural language processing (NLP) systems were based on complex sets of hand-written rules. These systems were limited in their capabilities and could not handle the nuances of natural language. In the late 1980s, there was a revolution in NLP with the introduction of machine learning algorithms. Machine learning algorithms can learn from data, and this allowed NLP systems to become more powerful and capable. The rise of machine learning in NLP was due to the steady increase in computational power, which allowed machine learning algorithms to be trained on larger and larger datasets and the gradual lessening of the dominance of Chomskyan theories of linguistics. Chomskyan theories of linguistics focus on the formal rules of language, and they discouraged the use of corpus linguistics, which is the study of large collections of text. Corpus linguistics is essential for machine learning algorithms, as it provides the data that the algorithms need to learn from. Today, machine learning is the dominant approach to NLP. Machine learning algorithms are used for a variety of NLP tasks, including Machine translation, Text classification, Sentiment analysis, named entity recognition and Question answering.

The first recurrent neural networks (RNNs) and Long short-term memory (LSTM) networks were trained in these years. Recurrent neural networks (RNNs) are a type of artificial neural network that can process sequential data. They are characterized by the fact that the output from one layer can be fed back into the same layer, allowing the network to maintain a state or memory. This makes them well-suited for tasks such as speech recognition and natural language processing, where the context of previous inputs is important for understanding the current input. RNNs can be divided into two types: finite impulse response (FIR) and infinite impulse response (IIR). FIR RNNs have a finite number of states, and their output can be calculated by unrolling the network into a feedforward neural network. IIR RNNs have an infinite number of states, and their output cannot be calculated by unrolling the network. In addition to the basic RNN architecture, several variants have been developed to address specific challenges. For example, long short-term memory (LSTM) networks and gated recurrent units (GRUs) are designed to address the problem of vanishing gradients, which can occur in RNNs when the sequence length is long. They have been used successfully in a variety of applications, including speech recognition, natural language processing, machine translation, and music generation.

In the 1980s, Geoffrey Hinton developed the Boltzmann machine, a type of neural network that can be used for language modeling. A language model is a statistical model that predicts the probability of a sequence of words. It is trained on a corpus of text, which is a collection of text documents. The language model learns the statistical relationships between words in the corpus. Pure statistical models, such as word n-gram language models, were the first type of language models. They are based on the idea that the probability of a word depends on the words that have come before it. However, pure statistical models are limited in their ability to learn long-range dependencies between words. Boltzmann machines can learn the statistical relationships between words in a corpus of text.

In the 1990s, Yoshua Bengio proposed the use of feed-forward neural networks for language modeling. Feedforward neural networks are a type of neural network that can learn complex relationships between data. Transformers are a type of neural network that is particularly good at learning long-range dependencies between words. Large language models are the most advanced form of language models. They are trained on massive datasets of text and code. They can generate text, translate languages, write different kinds of creative content, and answer your questions in an informative way. Large language models are a combination of feedforward neural networks and transformers.

In the 2000s, Tomas Mikolov developed the word2vec model, a type of neural network that can be used to learn word embeddings. Word embeddings are a way of representing words as vectors in a high-dimensional space. This allows language models to learn the semantic and syntactic relationships between words.

The development of personal computers, the internet, and networking has revolutionized the field of NLP. These technologies have made it possible to collect and store massive amounts of data, which is essential for training machine learning algorithms. They have also made it possible to distribute NLP systems to a wider audience. As a result of these advances, NLP has made significant progress in recent years. Machine learning algorithms are now able to perform tasks such as machine translation, text summarization, and question answering with a high degree of accuracy. These developments are a major step towards the goal of creating artificial intelligence that can understand and generate human language.

## 3 History From 2000s to 2023

In April 2006, Google launched a wonderful and evolutionary product “Google Translate”. It is a free-to-user web-based translation service. It can translate text, phrases, and web pages in multiple languages. Originally, Google Translate used statistical machine translation (SMT) to translate text. SMT is a technique that uses predictive algorithms to translate text from one language to another. However, SMT can have poor grammatical accuracy, especially for languages with complex grammar. Today, Google Translate is one of the most popular translation services in the world. It is used by millions of people every day to translate text, phrases, and web pages in over 100 languages.

Improvements in machine learning have led to rapid progress in natural language processing (NLP). These improvements have been driven by advances in algorithms, computing power, and the availability of large datasets of text. One important development in NLP is the use of transformer architectures. Transformers are a type of neural network that is particularly well-suited for tasks such as machine translation and text summarization.

In 2018, OpenAI researchers introduced the first generative pre-trained transformer (GPT). GPT is a type of language model that is trained on a massive dataset of text. It can then be fine-tuned to perform specific tasks, such as generating text or answering questions. GPT has been followed by several other large language models, including GPT-2 and Microsoft's Turing Natural Language Generation (T-NLG). These models are trained on even larger datasets and have achieved even greater performance. The development of large language models is a major step forward in NLP. These models have the potential to revolutionize the way we interact with computers. They can be used to create chatbots that can hold natural conversations, generate realistic text, and answer our questions in an informative way. They can learn long-range dependencies between words, which is essential for tasks such as machine translation and text summarization. They are more efficient than previous neural network architectures, which makes it possible to train on massive datasets of text. They have achieved state-of-the-art results on a variety of NLP tasks. The development of large language models is a rapidly evolving field. As these models become more powerful, they will be able to perform even more complex tasks. Large language models will likely play a major role in the future of NLP.

# Ending Notes

The field of natural language processing (NLP) has seen a significant shift in these years towards statistical methods data-driven approaches, machine learning, and neural networks. This is due in part to the increasing availability of large corpora of machine-readable text, which can be used to train statistical models of language. Another factor driving this trend is the success of statistical methods in other areas of artificial intelligence, such as speech recognition. As a result of these trends, NLP research is now focused on several key areas. In this era, a lot of contribution is made by many scientists, and different techniques are developed. So, we cannot name it to a single person.

The history of artificial intelligence (AI) has been marked by a series of hype cycles, followed by disappointment, funding cuts, and a period of little research (known as AI winters), followed by renewed interest and hype. The first cycle began in the 1950s with the development of new AI techniques, such as language processing and machine learning. This led to a great deal of excitement and optimism about the potential of AI, and funding for AI research increased significantly. However, by the mid-1960s, it became clear that AI was not yet living up to its promises. Some of the early AI systems were found to be brittle and unable to generalize to new situations. This led to a period of disappointment and funding cuts, known as the "AI winter" of the 1970s.

The second AI winter ended in the 1980s, with the development of new AI techniques, such as expert systems and neural networks. This led to a renewed interest in AI, and funding for research increased again. However, the second AI winter also ended in disappointment, as it became clear that AI could still not solve some of the most challenging problems, such as natural language understanding and commonsense reasoning.

The third AI winter ended in the 2000s, with the development of new AI techniques, such as deep learning and GPT. Deep learning has enabled AI systems to achieve superhuman performance on a variety of tasks, such as image recognition and natural language processing. This has led to a new wave of excitement and optimism about the potential of AI, and funding for AI research is at an all-time high.

It is still too early to say whether the current AI hype cycle will end in disappointment or not. However, the history of AI suggests that it is important to be realistic about the capabilities of AI systems. AI systems are still far from being as intelligent as humans and will likely continue to make mistakes. However, AI is a powerful tool that can be used to solve a variety of problems, and it is likely to play an increasingly important role in our lives in the years to come. As we stand on the cusp of the AI revolution, it is essential to maintain a realistic perspective on AI's capabilities and potential limitations. While AI has made remarkable progress, it is not yet on par with human intelligence. Mistakes are inevitable, but the power of AI to solve complex problems and improve our lives is undeniable.

# Bibliography

Charles Roe. Identify Data Patterns with Natural Language Processing and Machine Learning - DATAVERSITY". Www.Dataversity.Net, 2023, https://www.dataversity.net/identify-data-patterns-natural-language-processing-machine-learning/. Accessed 2 Sep 2023."

"Keith D. Foote. A Brief History of Natural Language Processing - DATAVERSITY". Www.Dataversity.Net, 2023, https://www.dataversity.net/a-brief-history-of-natural-language-processing-nlp/. Accessed 2 Sep 2023.

"Hodgkin and Huxley: Superheroes". Cheever.Domains.Swarthmore.Edu, 2023, https://cheever.domains.swarthmore.edu/Ref/HH/HHmain.htm. Accessed 6 Sep 2023.

"Natural Language Processing: A Historical Review" Karen Sparck Jones Computer Laboratory, University of Cambridge

"*Cours de Linguistique Générale*" published in 1916, written by Saussure and compiled by Albert Sechehaye and Charles Bally

Joseph, Sethunya & Sedimo, Kutlwano & Kaniwa, Freeson & Hlomani, Hlomani & Letsholo, Keletso. (2016). Natural Language Processing: A Review. Natural Language Processing: A Review. 6. 207-210.

Hutchins, W. J., and H. L. Somers. An Introduction to Machine Translation. Academic Press, 1992.

Text-Based Intelligent Systems. Lawrence Erlbaum Associates, 1992.

Joshi, A. K., et al. (Eds.) Elements of Discourse Understanding. Cambridge University Press, 1981.

Kay, M., et al. Verbmobil: A Translation System for Face-to-Face Dialogue. Stanford University, 1991.

Nagao, M. 'A Japanese View of Machine Translation in Light of the Considerations and Recommendations Reported by ALPAC, USA'. Japan Electronic Industry Development Association, 1989.

Plath, W. "Multiple Path Analysis and Automatic Translation", in Booth 1967, 267315.

Reifler, E. "Chinese-English Machine Translation, Its Lexicographic and Linguistic Problems", in Booth 1967, 317–428.

Rumelhart, D. E. 'McClelland, J.L. and the PDP Research Group, Parallel Distributed Processing, 2 Vols, Cambridge'. MA: MIT Press, 1986.

Natural Language Processing. Algorithmics Press, 1973.

Schank, R. 'C., Conceptual Information Processing, Amsterdam'. North Holland, 1975.

Schank, R.C. "Language and Memory", 1980; Reprinted in Grosz et al. 1986, 171–191.

Verbmobil: Mobiles Dolmetschgerat; Studie. Siemens AG, 1991.

'K."Natural Language Processing: An Overview", International Encyclopedia of Linguistics (Ed W. Bright)'. New York: Oxford University Press, vol. 3, 1992.

Sparck Jones, K. "Thesaurus", Encyclopedia of Artificial Intelligence (Ed Shapiro). Wiley, 1992.

Walker, D.E. "SAFARI: An online Text-Processing System", Proceedings of the American Documentation Institute Annual Meeting, 1967, 144–147.

Winograd, T. "A Procedural Model of Language Understanding", 1973; Reprinted in Grosz et al. 1986, 249–266.

Woods, W.A. “Semantics and Quantification in Natural Language Question Answering”, 1978; Reprinted in Grosz et al. 1986, 205–248.

Yngve, V.H. “MT at MIT”, in Booth 1967, 451–523.

‘The MITRE Syntactic Analysis Procedure for Transformational Grammars, Proceedings of the Fall Joint Computer Conference, 1965; A FIPS Conference Proceedings Vol. 27’. Part, vol. 1, 1965.

"ELIZA". En.Wikipedia.Org, 2023, https://en.wikipedia.org/wiki/ELIZA. Accessed 6 Sep 2023.

"Recurrent neural network". En.Wikipedia.Org, 2023, https://en.wikipedia.org/wiki/Recurrent_neural_network. Accessed 7 Sep 2023.

"Google Translate". En.Wikipedia.Org, 2023, https://en.wikipedia.org/wiki/Google_Translate. Accessed 7 Sep 2023.

"Transfer learning". En.Wikipedia.Org, 2023, https://en.wikipedia.org/wiki/Transfer_learning. Accessed 7 Sep 2023.