

Title: History of natural language processing

URL: [https://en.wikipedia.org/wiki/History\\_of\\_natural\\_language\\_processing](https://en.wikipedia.org/wiki/History_of_natural_language_processing)

PageID: 27837170

Categories: Category:History of artificial intelligence, Category:History of linguistics, Category:History of software, Category:Natural language processing

Source: Wikipedia (CC BY-SA 4.0). Content may require attribution.

-----

The history of natural language processing describes the advances of natural language processing . There is some overlap with the history of machine translation , the history of speech recognition , and the history of artificial intelligence .

#### Early history

The history of machine translation dates back to the seventeenth century, when philosophers such as Leibniz and Descartes put forward proposals for codes which would relate words between languages. All of these proposals remained theoretical, and none resulted in the development of an actual machine.

The first patents for "translating machines" were applied for in the mid-1930s. One proposal, by Georges Artsrouni , was simply an automatic bilingual dictionary using paper tape . The other proposal, by Peter Troyanskii , a Russian, was more detailed. Troyanski proposal included both the bilingual dictionary, and a method for dealing with grammatical roles between languages, based on Esperanto .

#### Logical period

In 1950, Alan Turing published his famous article " Computing Machinery and Intelligence " which proposed what is now called the Turing test as a criterion of intelligence. This criterion depends on the ability of a computer program to impersonate a human in a real-time written conversation with a human judge, sufficiently well that the judge is unable to distinguish reliably — on the basis of the conversational content alone — between the program and a real human.

In 1957, Noam Chomsky 's Syntactic Structures revolutionized Linguistics with ' universal grammar ', a rule-based system of syntactic structures.

The Georgetown experiment in 1954 involved fully automatic translation of more than sixty Russian sentences into English. The authors claimed that within three or five years, machine translation would be a solved problem. However, real progress was much slower, and after the ALPAC report in 1966, which found that ten years long research had failed to fulfill the expectations, funding for machine translation was dramatically reduced. Little further research in machine translation was conducted until the late 1980s, when the first statistical machine translation systems were developed.

Some notably successful NLP systems developed in the 1960s were SHRDLU , a natural language system working in restricted " blocks worlds " with restricted vocabularies.

In 1969 Roger Schank introduced the conceptual dependency theory for natural language understanding. This model, partially influenced by the work of Sydney Lamb , was extensively used by Schank's students at Yale University , such as Robert Wilensky, Wendy Lehnert, and Janet Kolodner .

In 1970, William A. Woods introduced the augmented transition network (ATN) to represent natural language input. Instead of phrase structure rules ATNs used an equivalent set of finite-state automata that were called recursively. ATNs and their more general format called "generalized ATNs" continued to be used for a number of years. During the 1970s many programmers began to write 'conceptual ontologies', which structured real-world information into computer-understandable data. Examples are MARGIE (Schank, 1975), SAM (Cullingford, 1978), PAM (Wilensky, 1978), TaleSpin (Meehan, 1976), QUALM (Lehnert, 1977), Politics (Carbonell, 1979), and Plot Units

(Lehnert 1981). During this time, many chatterbots were written including PARRY , Racter , and Jabberwacky .

### Statistical period

Up to the 1980s, most NLP systems were based on complex sets of hand-written rules. Starting in the late 1980s, however, there was a revolution in NLP with the introduction of machine learning algorithms for language processing. This was due both to the steady increase in computational power resulting from Moore's law and the gradual lessening of the dominance of Chomskyan theories of linguistics (e.g. transformational grammar ), whose theoretical underpinnings discouraged the sort of corpus linguistics that underlies the machine-learning approach to language processing. Some of the earliest-used machine learning algorithms, such as decision trees , produced systems of hard if-then rules similar to existing hand-written rules. Increasingly, however, research has focused on statistical models , which make soft, probabilistic decisions based on attaching real-valued weights to the features making up the input data. The cache language models upon which many speech recognition systems now rely are examples of such statistical models. Such models are generally more robust when given unfamiliar input, especially input that contains errors (as is very common for real-world data), and produce more reliable results when integrated into a larger system comprising multiple subtasks.

### Datasets

The emergence of statistical approaches was aided by both increase in computing power and the availability of large datasets. At that time, large multilingual corpora were starting to emerge. Notably, some were produced by the Parliament 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.

Many of the notable early successes occurred in the field of machine translation . In 1993, the IBM alignment models were used for statistical machine translation . Compared to previous machine translation systems, which were symbolic systems manually coded by computational linguists, these systems were statistical, which allowed them to automatically learn from large textual corpora . Though these systems do not work well in situations where only small corpora is available, so data-efficient methods continue to be an area of research and development.

In 2001, a one-billion-word large text corpus, scraped from the Internet, referred to as "very very large" at the time, was used for word disambiguation .

To take advantage of large, unlabelled datasets, algorithms were developed for unsupervised and self-supervised learning . Generally, this task is much more difficult than supervised learning , and typically produces less accurate results for a given amount of input data. However, there is an enormous amount of non-annotated data available (including, among other things, the entire content of the World Wide Web ), which can often make up for the inferior results.

### Neural period

Neural language models were developed in 1990s. In 1990, the Elman network , using a recurrent neural network , encoded each word in a training set as a vector, called a word embedding , and the whole vocabulary as a vector database , allowing it to perform such tasks as sequence-predictions that are beyond the power of a simple multilayer perceptron . A shortcoming of the static embeddings was that they didn't differentiate between multiple meanings of homonyms .

Yoshua Bengio developed the first neural probabilistic language model in 2000

In recent years, advancements in deep learning and large language models have significantly enhanced the capabilities of natural language processing, leading to widespread applications in areas such as healthcare, customer service, and content generation.

### Software

### References

### Bibliography

Crevier, Daniel (1993). *AI: The Tumultuous Search for Artificial Intelligence* . New York, NY: BasicBooks. ISBN 0-465-02997-3 .

McCorduck, Pamela (2004), *Machines Who Think* (2nd ed.), Natick, MA: A. K. Peters, Ltd., ISBN 978-1-56881-205-2 , OCLC 52197627 .

Russell, Stuart J. ; Norvig, Peter (2003), *Artificial Intelligence: A Modern Approach* (2nd ed.), Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13-790395-2 .