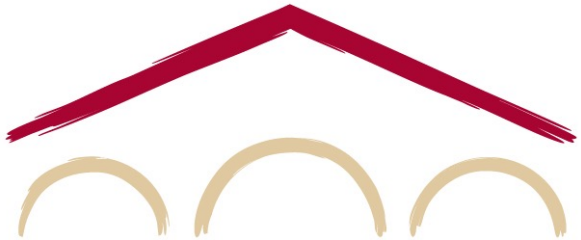
History of Natural Language Processing

CS 324H



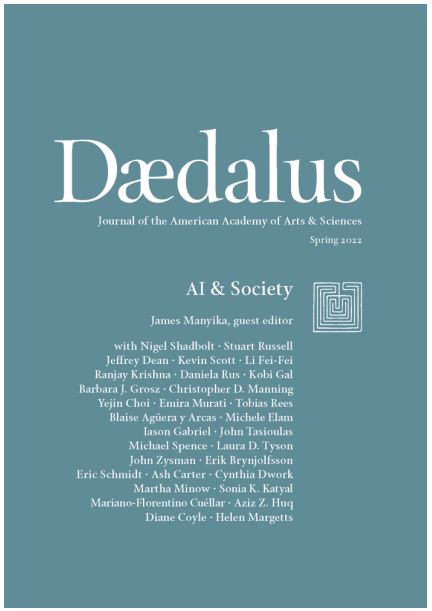
Dan Jurafsky and Christopher Manning

Lecture 1

Christopher D. Manning: Human Language Understanding & Reasoning

Four eras of NLP

- 1940–1969
  Early Explorations
- 1970–1992
  Hand-built demonstration NLP systems, of increasing formalization
- 1993–2012
  Statistical or Probabilistic NLP and then more general Supervised ML for NLP
- 2013–now
  Deep Learning or Artificial Neural Networks for NLP. Unsupervised or Self-Supervised NLP. Reinforcement Learning



Dædalus 151(2): 127–138
https://www.amacad.org/publication/human-language-understanding-reasoning

Early Explorations

1940–1969

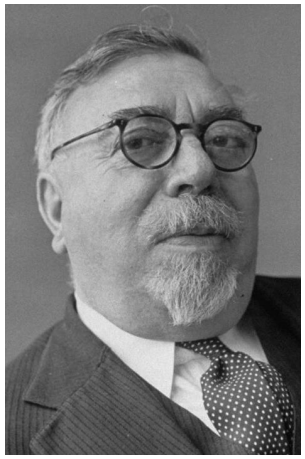
Machine Translation: Just a Code?

"Also knowing nothing official about, but having guessed and inferred considerable about, the powerful new mechanized methods in cryptography—methods which I believe succeed even when one does not know what language has been coded—one naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.' "

– Warren Weaver (1955:18, quoting a letter he wrote in 1947)

"When I look at an article in Russian, I say: 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.'" – Warren Weaver, March 1947

Weaver was a mathematician & engineer known for his work as a science funder at the Rockefeller Foundation and OSR&D (US Govt WWII science funder) and for coauthoring an approachable Info Theory intro with Shannon



"… as to the problem of mechanical translation, I frankly am afraid that the [semantic] boundaries of words in different languages are too vague … to make any quasi-mechanical translation scheme very hopeful."

– Norbert Wiener, April 1947

Wiener: MIT originator of **cybernetics**, which sought to tie together communication, control, and feedback in living things and computers. Nerd note: "cybernetics" draws from the same Greek word as Kubernetes

"All the News That's Fit to Print"

The New York Times.

LATE CITY EDITION
Fair and cold today. Increasing cloudiness and cold tomorrow.
Temperature Range Today-Max., 35; Min., 22
Temperatures Yesterday—Max., 36; Min., 26
Full U. S. Weather Bureau Report, Page 41

VOL. CIII..No. 35,048. Entered as Second-Class Matter. Post Office, New York, N. Y. NEW YORK, FRIDAY, JANUARY 8, 1954. Times Square, New York 36, N. Y. Telephone LAckawanna 4-1000 FIVE CENTS

PIER STRIKE CURB SOUGHT IN COURTS AS TIE-UPS SPREAD

Walkout on Another Brooklyn Dock Stirs Fear of 'Flash' Stoppage Throughout Port

VOTE DISPUTE CONTINUES

New Questions Delay Sending Poll Report to Capital—City Tightens Its Precautions

By A. H. RASKIN

Shipping employers sought court help yesterday to halt the spread of Brooklyn pier strikes before they exploded into a port-wide shutdown.

The injunction moves came after a second dock had been closed by a work stoppage. The walkout by seventy-five members of the old International Longshoremen's Association stopped loading of the British freighter City of Barcelona at Pier 2, Bush Terminal.

It occurred on the sixth day of a strike by 150 American Federation of Labor dock workers, which has prevented removal of cargo from two Isbrandtsen Line freighters at Pier 29, Brooklyn. Although the walkouts had no apparent connection with one another, shipping executives expressed fear that they might provide the spark for a general tie-up of the strife-ridden harbor.

City Widens Precautions

Mayor Wagner took new steps to perfect the city's preparations for possible violence in the event of a waterfront strike on a broad

White Miners Veto Rhodesia Color Bar

By ALBION ROSS
Special to The New York Times.

JOHANNESBURG, South Africa, Jan. 7—White miners at the American-controlled Roan Antelope Mine in Northern Rhodesia have shaken the mining and industrial world of southern Africa by voting for removal of the color bar.

The mine's white workers, representing roughly one-fourth of the white persons in the Northern Rhodesian copper industry, voted to strike out the first clause of their agreement with the copper producers. The clause limits employment in the higher paid and more skilled jobs to white men.

In the new goldfields of Odendaalsrus, Orange Free State, Gert Lombaard, speaking for the white South African Gold Miners Union, now under Nationalist leadership, called on white miners to resist this

Continued on Page 5, Column 5

MRS. LUCE TO RUSH TO ITALY IN CRISIS

Ambassador Cuts Visit to U.S. by Week to Report Events —Vatican Backs Pella

Special to The New York Times.

WASHINGTON, Jan. 7—Mrs. Clare Boothe Luce, United States Ambassador to Rome, will return to her post ahead of schedule because of the Italian Government crisis.

Mrs. Luce, who returned to Washington shortly before Jan. 1, plans to leave New York for

MAYOR ASKS STUDY OF STATE TAX SHIFT TO AID CITY INCOME

Urges Dewey to Join Move for 'Proper' Sharing of Levies as Way to End Problems

Text of address by the Mayor is printed on Page 14.

By PAUL CROWELL

Mayor Wagner offered last night to join with Governor Dewey in the appointment of a commission to consider the advisability of giving the city a "proper share" of the unlimited taxing powers now exercised by the state.

He said the purpose of such a commission would be to "study all state aid with a view to abolishing it and returning to us the taxing powers from which the money is now obtained."

The Mayor's offer was made in an address on city finances at the annual public-service-awards dinner of the Citizens Budget Commission at the Plaza Hotel. He declared that his proposal looked to a long-range solution of the city-state fiscal problem and could not, in all probability, bear any fruit in time to help the city prepare its 1954-55 budget.

"I said almost a year ago," the Mayor declared, "that I was tired of the city going to Albany for a handout. I said that I agreed with the Governor—and this was in the winter of 1952-53—that the state should get out of the collection agency business. I added that as a substitute for it the state should return to the localities their proper share of the taxing powers themselves; that it should give us a decent share

PRESIDENT WOULD CUT SPENDING, KEEP SECURITY AND PROSPERITY, END CITIZENSHIP OF SUBVERSIVES

'AKIN TO TREASON'

Eisenhower Maps Plan to Deal With Reds Guilty of Plots

By W. H. LAWRENCE
Special to The New York Times.

WASHINGTON, Jan. 7—President Eisenhower proposed today to strip United States citizenship from Communists convicted in the future of conspiring to advocate the overthrow of Government by force and violence.

His advocacy of a modern-day "men without a country" status for native-born or naturalized American conspirators surprised and delighted Congress.

The Communist conspiracy, the President said, is "akin to treason" and should be dealt with as such. Any person convicted of conspiring armed overthrowal of Government, he proposed, should be "treated as having, by such act, renounced his allegiance to the United States and forfeited his United States citizenship."

This goes well beyond the present deprivation of civil rights suffered by those convicted of a felony. Under Federal law, those convicted of violating the Smith Act ban on conspiracy to advocate overthrowal of the Government cannot vote, hold public

Associated Press Wirephoto

THE PRESIDENT REPORTS TO CONGRESS: President Eisenhower as he delivered his annual State of the Union message yesterday at a joint session of Congress. Behind him on the dais are Vice President Richard M. Nixon, left, and Speaker Joseph W. Martin Jr.

MESSAGE ON UNION

Eisenhower Proposes Amendment to Give 18-Year-Olds Vote

Text of State of Union Message is printed on Page 10.

By JAMES RESTON
Special to The New York Times.

WASHINGTON, Jan. 7—President Eisenhower, in his second State of the Union Message, today asked the almost evenly divided Eighty-third Congress to moderate but sustain the foreign and welfare policies of the New Deal era.

He proposed to cut the military budget, to cut taxes in relation to Government expenditures, and to reduce the guarantees to farmers very slowly. The cuts, however, were designed to reconcile security and solvency without jeopardizing the collective security or prosperity policies of the past.

The message contained only two surprises.

One was an appeal for a law that would take away the citizenship of any American hereafter convicted in the courts of "conspiring to advocate the overthrow of this Government by force or violence."

Russian Is Turned Into English By a Fast Electronic Translator

By ROBERT K. PLUMB

A public demonstration of what is believed to be the first successful use of a machine to translate meaningful texts from one language to another took place here yesterday afternoon.

This may be the culmination of centuries of search by scholars for "a mechanical translator." So far the system has a vocabulary of only 250 words. But there are no foreseeable limits to the number of words that the device can store or the number of languages it can be directed to translate.

Scholars and scientists who worked on it believe that within a few years the system may greatly increase communication, particularly in technical subjects, by making translation quick, accurate and easy.

The demonstration was at the headquarters of the International Business Machines Corporation, 590 Madison Avenue. It is the result of cooperative research by scientists of the corporation and scholars of the Georgetown University Institute of Languages and Linguistics in Washington.

The "mechanical" part of the translation system, which is mostly electronic, is a standard commercial model of the largest International Business Machines "stock" computer. This device, called the IBM Type 701 Electronic Data Processing Machine, was put on the market last April. Since then twelve of the machines have been sold to commercial, military and university computation laboratories.

The "literary" part of the system is a mechanical model of language devised at Georgetown by Prof. Leon Dostert and Dr. Paul Garvin. The corporation's share in the project was conducted by Dr. Cuthbert C. Hurd, director of its Division of Applied Science.

In the demonstration, a girl

Continued on Page 5, Column 2

Russian Is Turned Into English By a Fast Electronic Translator

By ROBERT K. PLUMB

A public demonstration of what is believed to be the first successful use of a machine to translate meaningful texts from one language to another took place here yesterday afternoon.

This may be the culmination of centuries of search by scholars for "a mechanical translator." So far the system has a vocabulary of only 250 words. But there are no foreseeable limits to the number of words that the device can store or the number of languages it can be directed to translate.

Scholars and scientists who worked on it believe that within a few years the system may greatly increase communication, particularly in technical subjects, by making translation quick, accurate and easy.

The demonstration was at the headquarters of the International Business Machines Corporation, 590 Madison Avenue. It is the result of cooperative research by scientists of the corporation and scholars of the Georgetown University Institute of Languages and Linguistics in Washington.

The "mechanical" part of the translation system, which is mostly electronic, is a standard commercial model of the largest International Business Machines "stock" computer. This device, called the IBM Type 701 Electronic Data Processing Machine, was put on the market last April. Since then twelve of the machines have been sold to commercial, military and university computation laboratories.

The "literary" part of the system is a mechanical model of language devised at Georgetown by Prof. Leon Dostert and Dr. Paul Garvin. The corporation's share in the project was conducted by Dr. Cuthbert C. Hurd, director of its Division of Applied Science.

In the demonstration, a girl

Continued on Page 5, Column 2

LANGUAGE DEVICE TRANSLATES FAST

Continued From Page 1

operator typed out on a keyboard the following Russian text in English characters:

"Mi pyeryedayem mislyi posryedstvom ryechyi."

The machine printed a translation almost simultaneously:

"We transmit thoughts by means of speech."

The operator did not know Russian. Again she typed out the meaningless (to her) Russian words:

"Vyelyichyina ugla opryedyelyayetsya otnoshyenyiyem dlyini dugi k radyiusu."

And the machine translated it to:

"Magnitude of angle is determined by the relation of length of arc to radius."

Guided by Language Code

Several short messages, within the 250-word range of the device, were tried. Included were brief statements in Russian about politics, law, mathematics, chemistry, metallurgy, communications and military affairs. The sentences were turned into good English without human intervention.

The heart of the system is the mechanical model of language devised at Georgetown. There the scholars first assembled a 250-word vocabulary in Russian covering the seven broad fields. Then they determined the rules of syntax required for a meaningful statement and reduced them to six instructions for the data-processing calculator.

These instructions are introduced into the calculator's short-term electrostatic "memory" with punch cards. The cards tell the machine how to cope with syntax.

In translating, for instance, a word "A" which precedes a word "B" in Russian, may be reversed in some cases in English. Each of the 250 words is coded for this inversion. Sometimes words must be inserted in the English text, sometimes they must be omitted, following code instructions.

When there are several possible English meanings for a Russian word, the instructions tell the machine to pick out the meaning that best fits the context.

Foreign words are typed on a keyboard that punches out I.B.M. cards. These are fed into the calculator, where they encounter the vocabulary, also punched on cards. On a standard printer meaningful English texts emerge.

According to Dr. Hurd, the calculator is a general-purpose data processing machine not designed specifically for translating. Nevertheless, it has a memory capable of storing roughly 1,000,000 five-letter words. There are 600,000 entries in the latest Webster's unabridged New International dictionary.

Dr. Hurd said that the corporation would now design a machine particularly fit for translating rather than for general computing utility. Such a device should be ready within three to five years. when the Georgetown scholars believe they can complete the "literary" end of the system.

Dr. Dostert and Dr. Garvin said they chose Russian for their first experiments because it was a difficult language and a system that could translate it could handle anything.

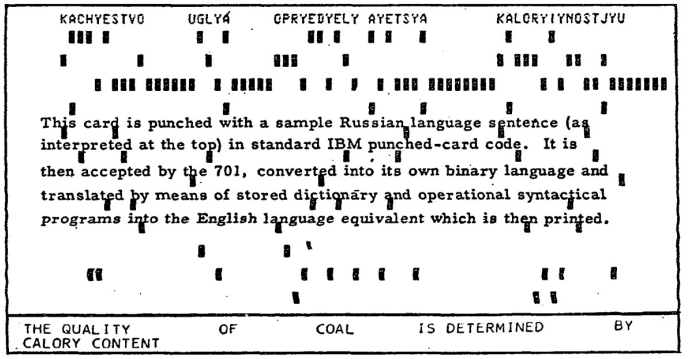
The machine will not accept incoherent statements, Dr. Dostert said. If they are introduced for "translation" the machine balks, and rings a bell. And it will ring the bell when it encounters a misprint. It now prints eighty letters in two seconds.

As soon as cards for Russian are completed, sets will be made for German and French. Then other Slavic, Germanic and Romance languages can be set up at will.

Calculator Takes on a New Job: Language Translation

The New York Times

An electronic calculator produced by the International Business Machines Corporation was demonstrated yesterday in a new role: translating Russian phrases into English. The device has a "vocabulary" of 250 Russian words and can be adapted to other languages. Miss Marilyn Polle "types" Russian phrases on I. B. M. punch cards that are fed into the machine.

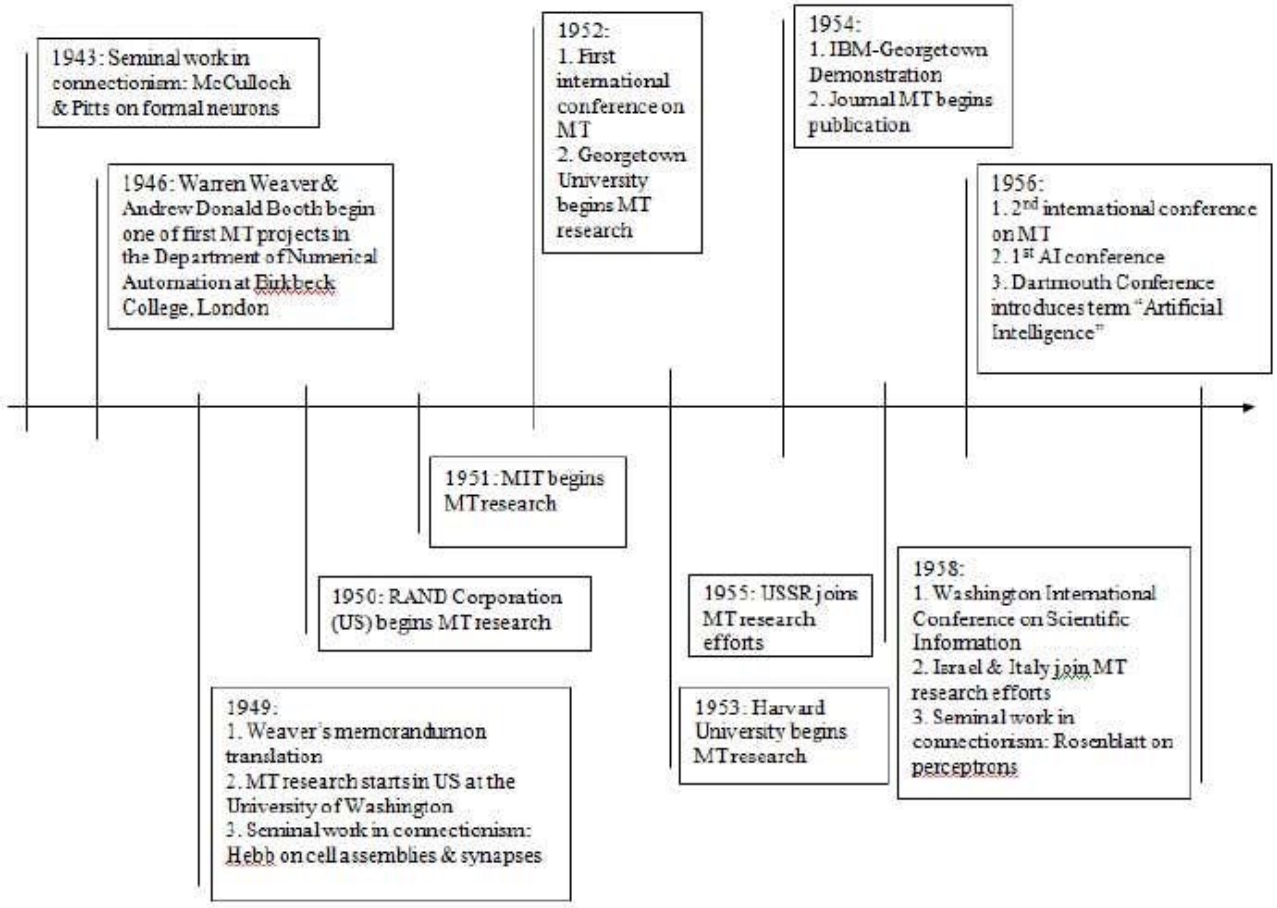


This card is punched with a sample Russian language sentence (as interpreted at the top) in standard IBM punched-card code. It is then accepted by the 701, converted into its own binary language and translated by means of stored dictionary and operational syntactical programs into the English language equivalent which is then printed.

Above, specimen punch card and below a strip with translation, typed almost simultaneously

The early history of MT: 1950s

Machine Translation: The origin of NLP/Computational Linguistics



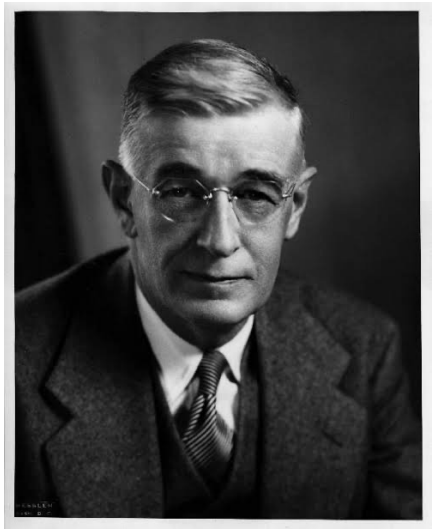
I grabbed these timelines from Ruth Camburn's "A Short History of Computational Linguistics". She was a CSU Fresno Linguistics grad student around 2013.

Information Retrieval: Vannevar Bush

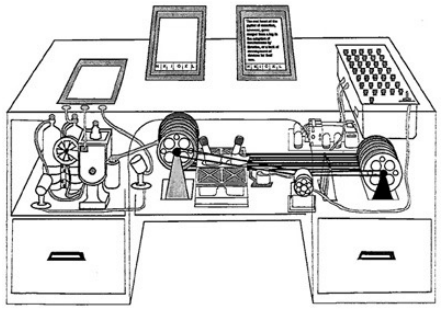
Bush (1945): As We May Think

"Consider a future device for individual use, which is a sort of mechanized private file and library. It needs a name, and, to coin one at random, "memex" will do. A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory.

https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/



Bush went from being Dean of Engineering at MIT and president of the Carnegie Institution of Washington to chairman of the National Defense Research Committee in WWII (which had huge funding, directing all wartime science, including the Manhattan Project)



Calvin Mooers

- Coined the term information retrieval in 1948/1950

- Zator Company (1947) doing Zatocoding as descriptor codes for IR.

Cyril Cleverdon

- Cranfield tests (1957–1967)
- Defined the idea of benchmark tests for language, with a document collection, queries, and correct answers.
- He had exhaustive answers over a small corpus!

HP (Hans Peter) Luhn

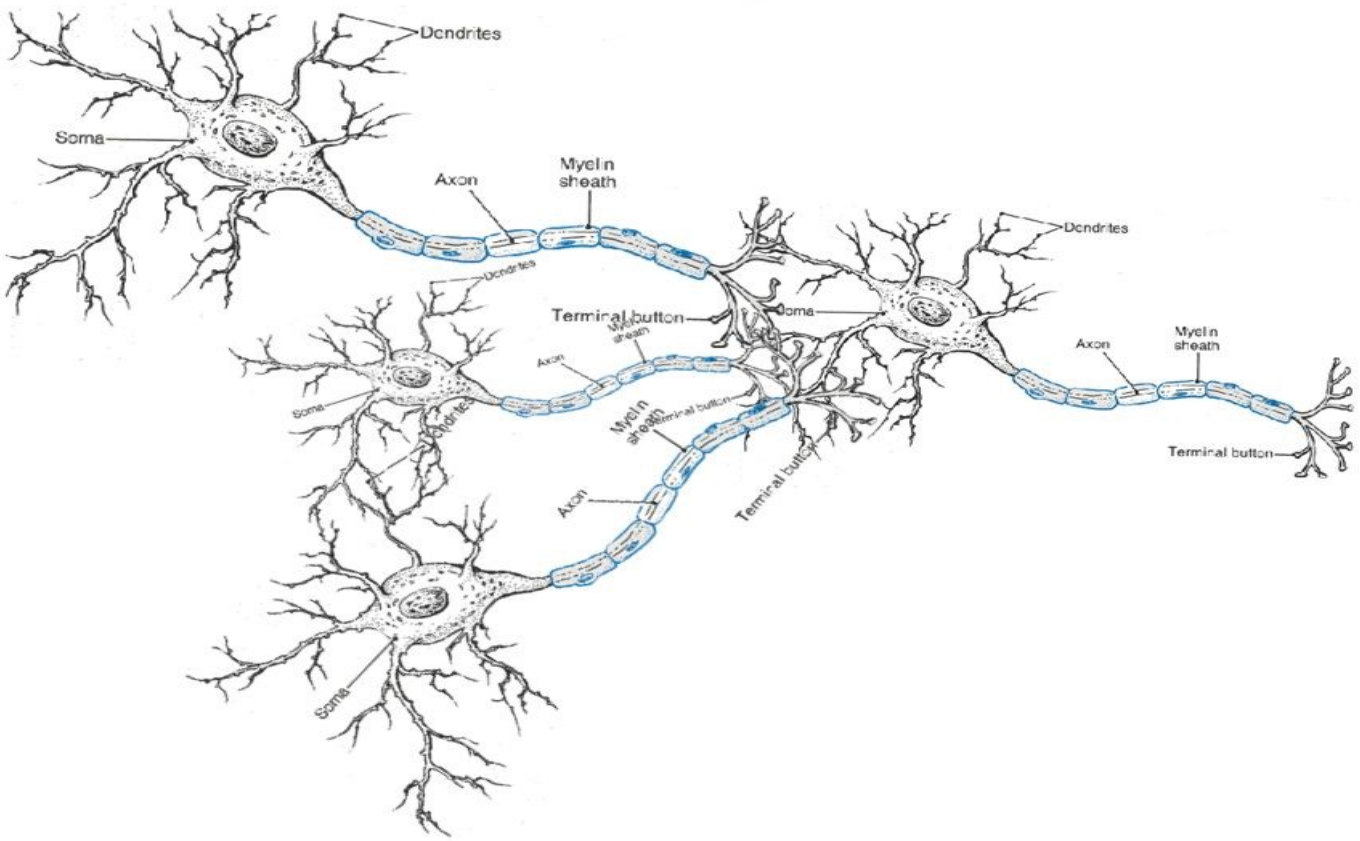
- IBM Information Retrieval Group manager from 1941
- Pioneered full-text processing, hash codes, keyword-in-context (KWIC) indexing, and the term “Business Intelligence”

Gerald Salton (Cornell)

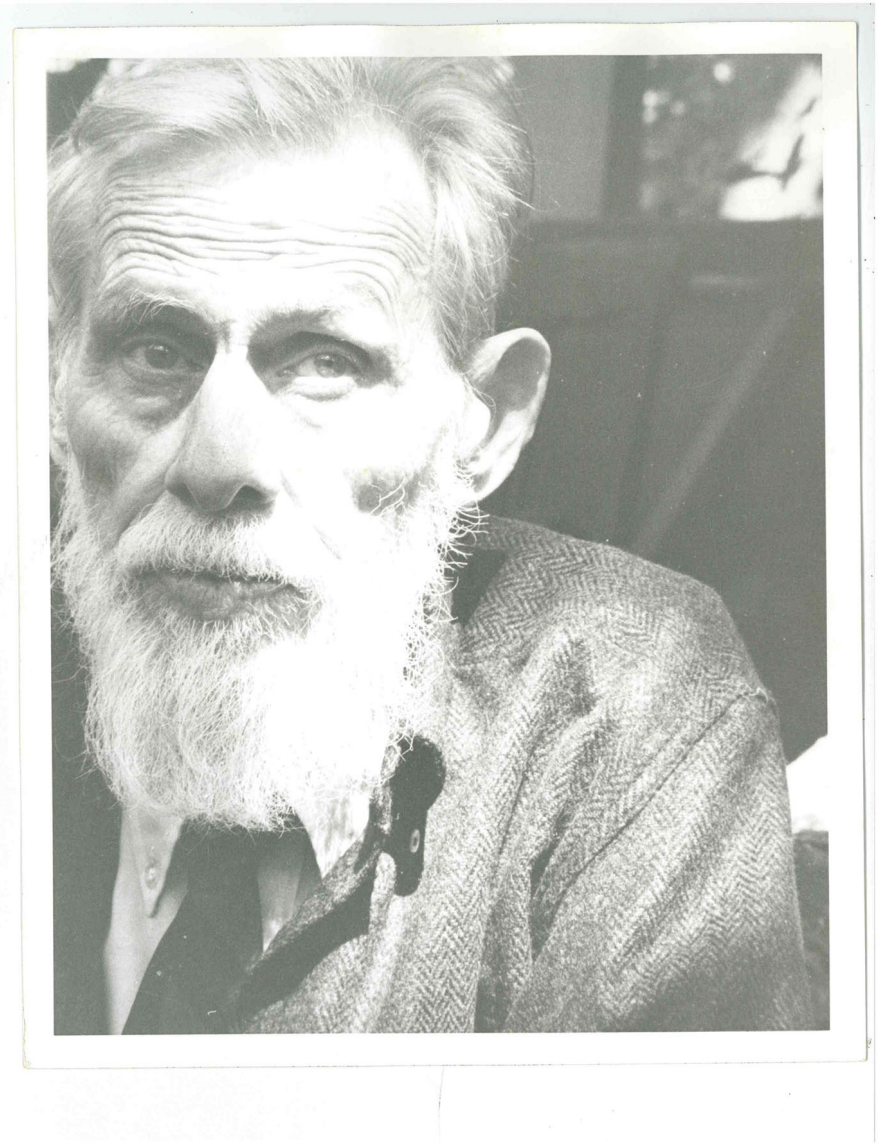
- Dominated U.S. information retrieval for years, just like Fisher or Chomsky in their fields

- Direct link from his group to Google web search through Amit Singhal**

Neural Networks: The inspiration from knowledge of neurons



Warren McCulloch



Walter Pitts

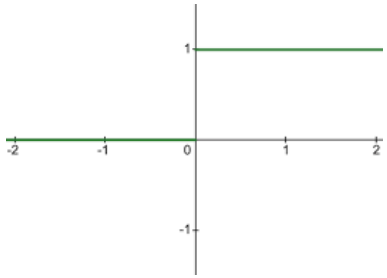
Logical Calculus of the Ideas immanent in Nervous Activity
1943.
The “McCulloch-Pitts neuron”

Original McCulloch & Pitts 1943 threshold unit:

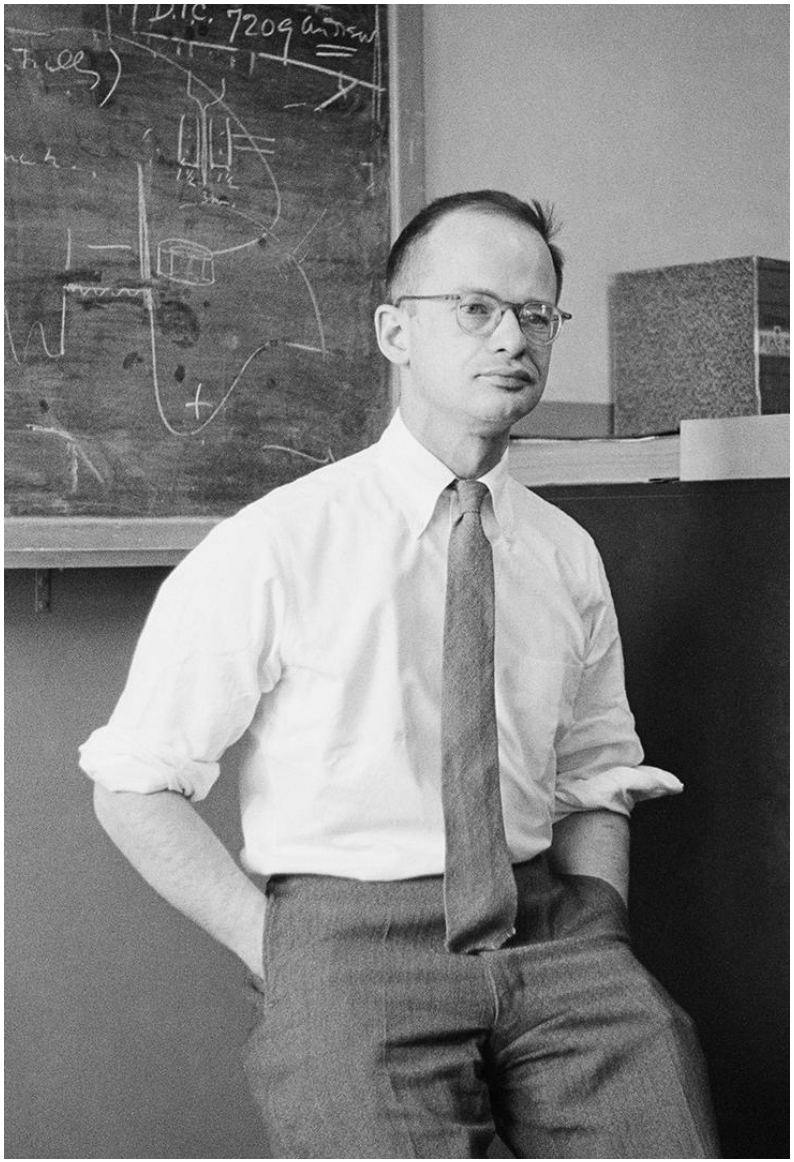
$\mathbf{1}(Wx > \theta)$

$= \mathbf{1}(Wx - \theta > 0)$

This function has no slope, so, no gradient-based learning



https://nautil.us/the-man-who-tried-to-redeem-the-world-with-logic-235253/



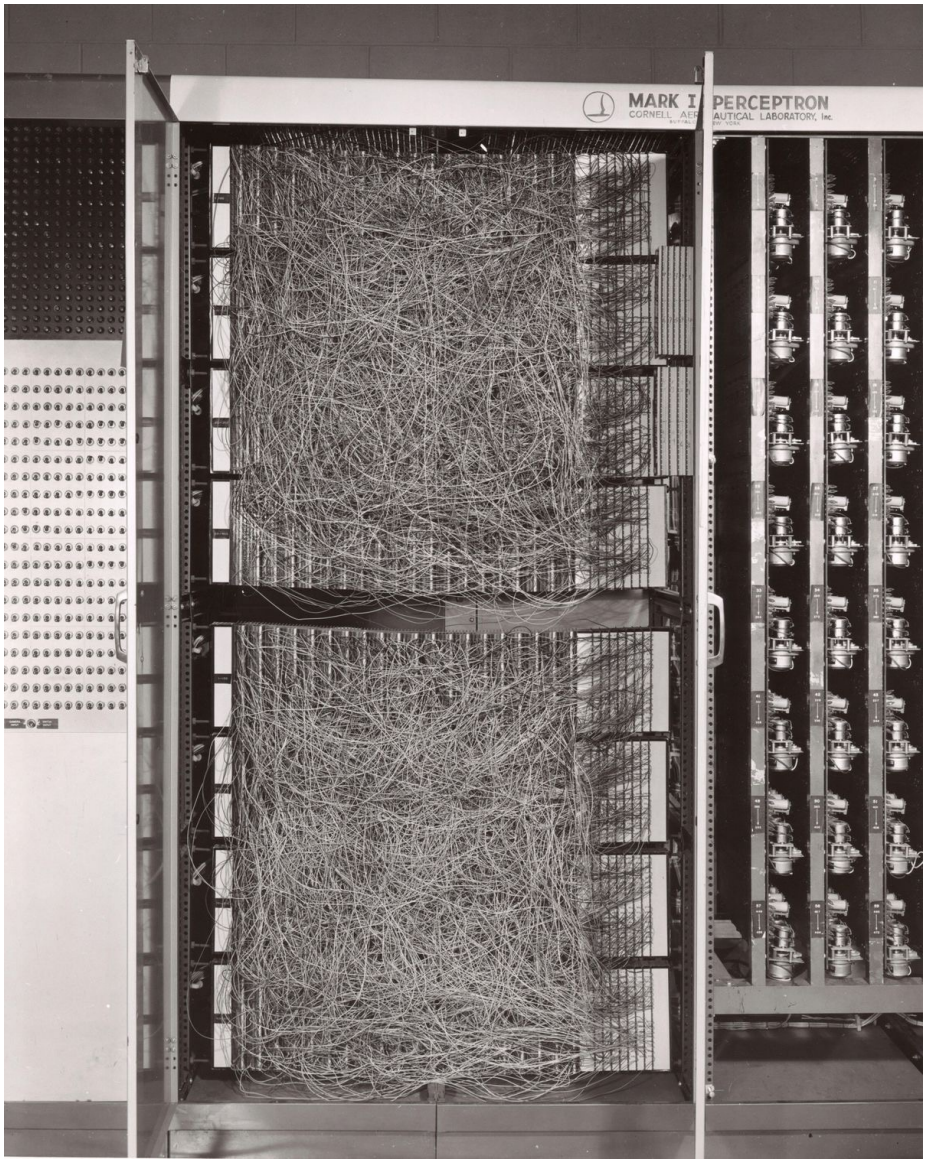
Donald O. Hebb

- Hebbian Learning:
  1949 The organization of Behavior

- “Cells that fire together, wire together”

Frank Rosenblatt

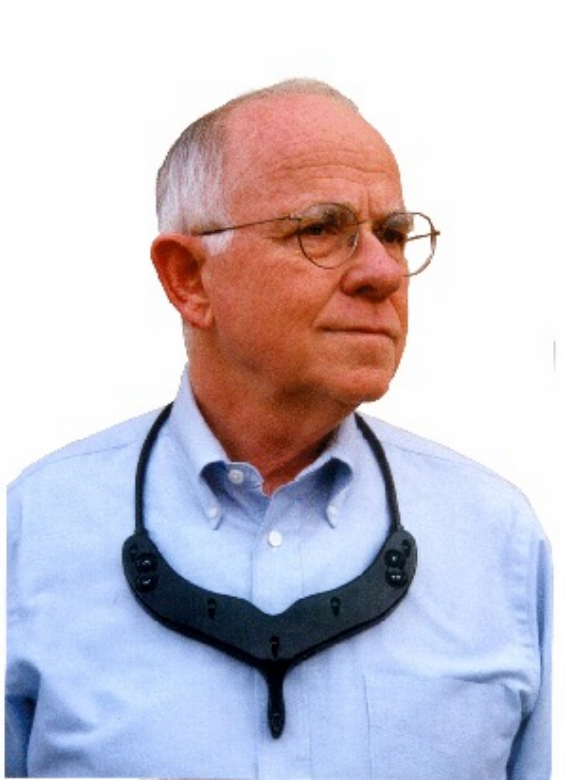
The Perceptron



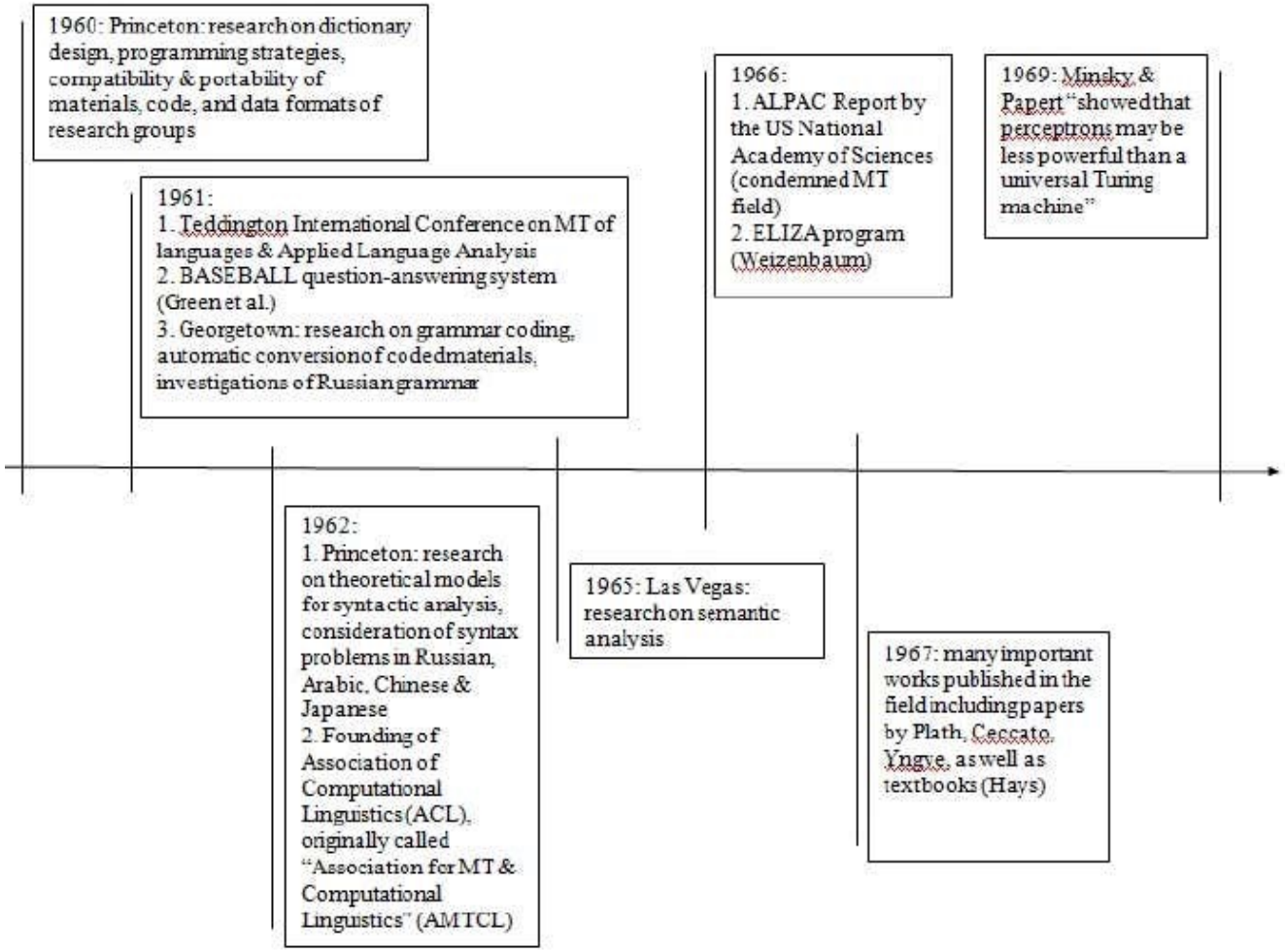
Bernard (Bernie) Widrow

- Stanford EE Faculty

- Adaline and Madaline neural network designs c. 1960
  - Precursor to backpropagation but they never got beyond one layer



NLP in the 1960s



David Glenn Hays

Raj Reddy

- Early speech work at Stanford (SAIL)
- Career in speech at CMU

Joyce Friedman

C. Ray Perrault

David Scott Warren

cs -109

A COMPUTER SYSTEM FOR WRITING AND TESTING TRANSFORMATIONAL GRAMMARS

FINAL REPORT

JOYCE FRIEDMAN
PRINCIPAL INVESTIGATOR

This research was supported in part by the United States Air Force Electronic Systems Division, under Contract F196828-C-0035.

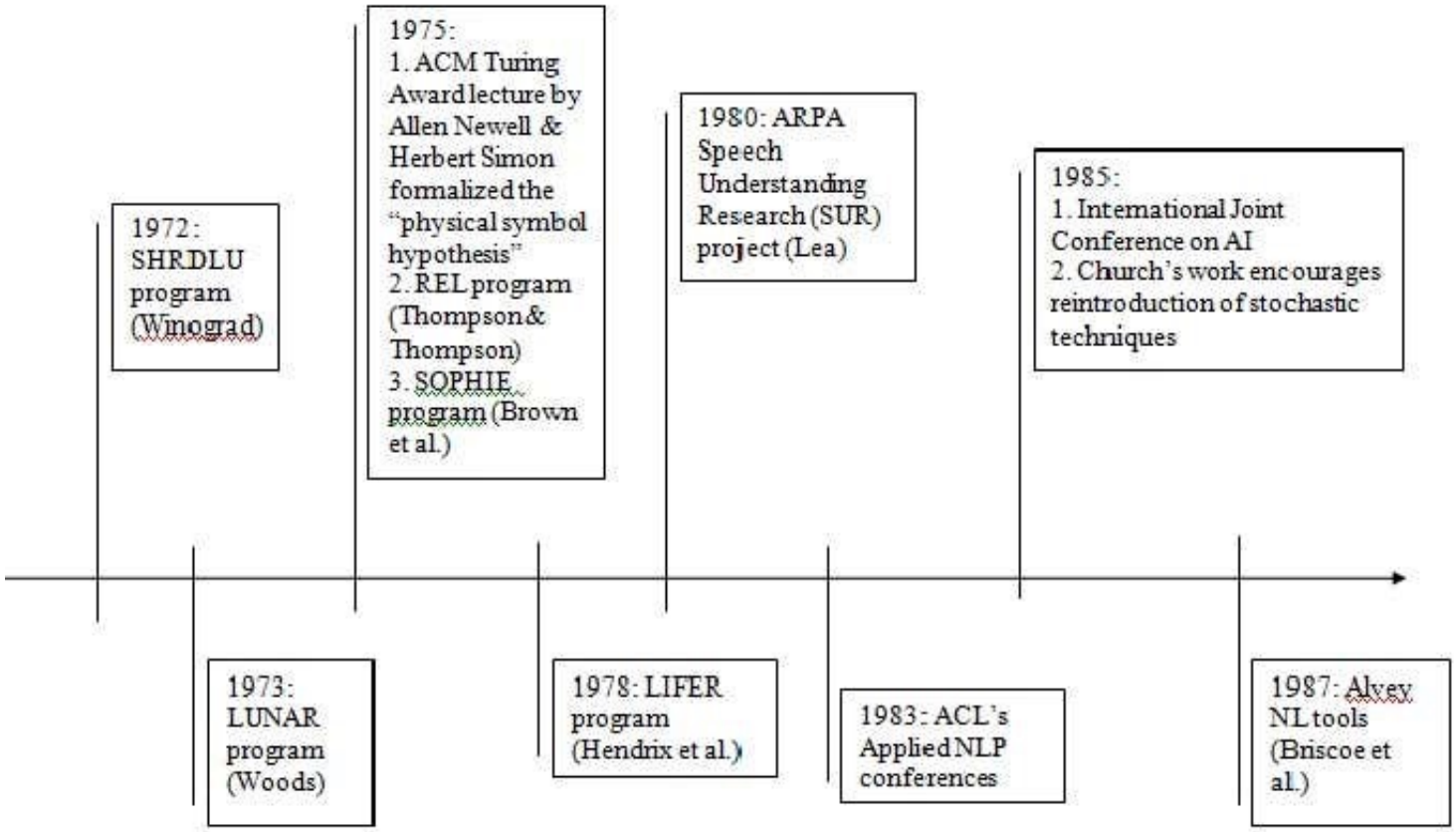
STANFORD UNIVERSITY COMPUTER SCIENCE DEPARTMENT
COMPUTATIONAL LINGUISTICS PROJECT

30 SEPTEMBER 1968

Hand-built demonstration NLP systems, of increasing formalization

1970–1992

NLP in the 1970s and 1980s



Deep Learning or Artificial Neural Networks for NLP

2013–present