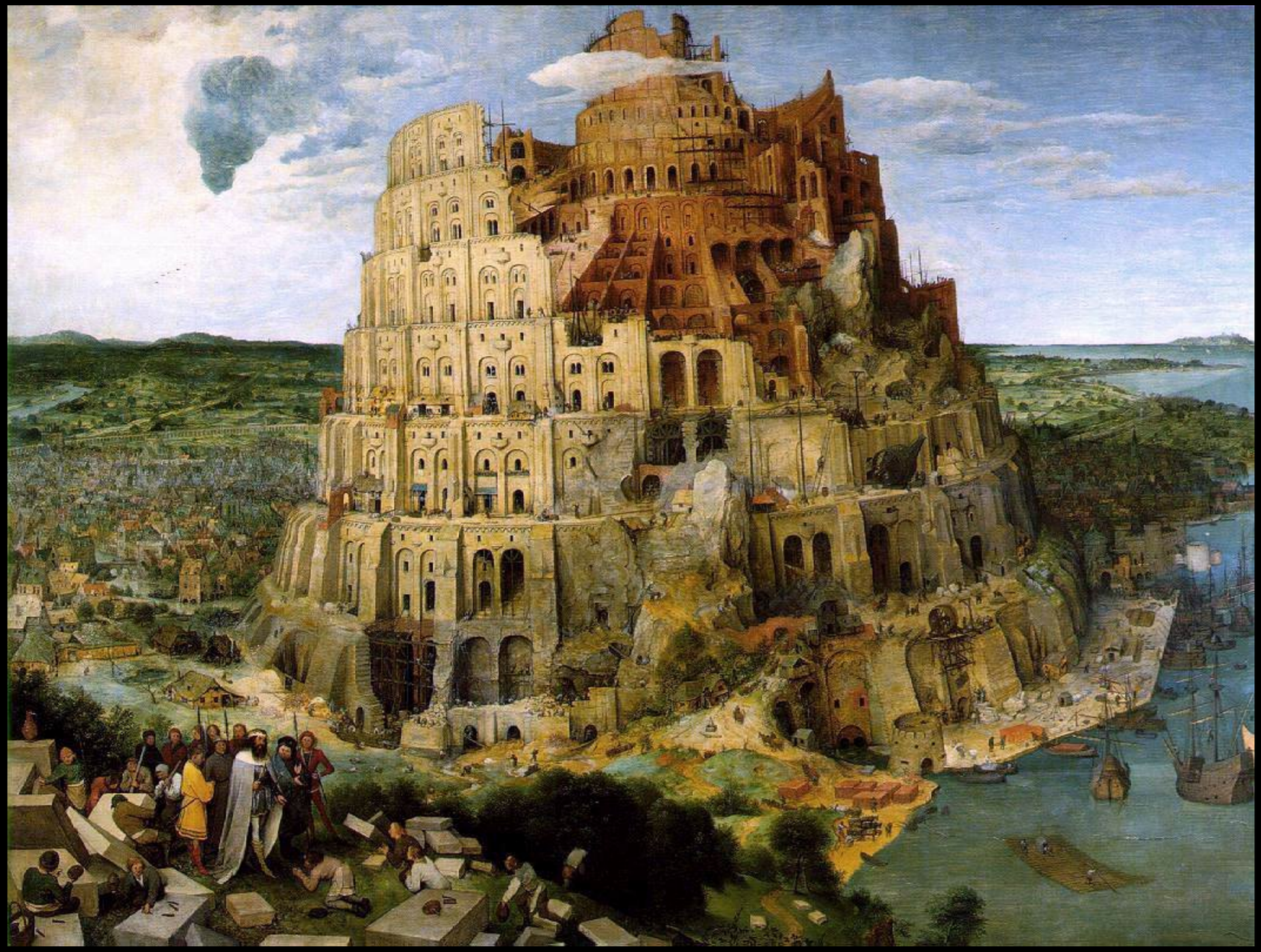
How we got here

“It takes a thousand people to invent a telegraph, or a steam engine, or a phonograph, or a telephone or any other important thing—and the last one gets the credit and we forget the others. He added his little mite—that is all he did. These object lessons should teach us that ninety-nine parts of all things that proceed from the intellect are plagiarisms, pure and simple; and the lesson ought to make us modest. But nothing can do that.”

–Mark Twain

Prologue



The Tower of Babel

Pieter Brueghel the Elder (1563)

1600s

DISSERTATIO
De
ARTE COMBI-
NATORIA,

In qua
Ex Arithmeticæ fundamentis *Complicationum* ac *Transpositionum* Doctrina novis præceptis exstruitur, & usus ambarum per universum scientiarum orbem ostenditur; nova etiam Artis Meditandi, *Seu*
Logicæ Inventionis semina sparguntur.

Præfixa est Synopsis totius Tractatus, & additamenti loco
Demonstratio
EXISTENTIÆ DEI,
ad Mathematicam certitudinem exacta.

AUTORE
GOTTFREDO GUILIELMO
LEIBNüZIO Lipsensi,
Phil. Magist. & J. U. Baccal.

LIPSIÆ,
Apud Joh. Simon. Fickium et Joh. Polycarp. Seuboldum
in Platea Nicolaea,
Literis Spörelianis.
A. M. DC. LXVI.

2° Gresh. 43

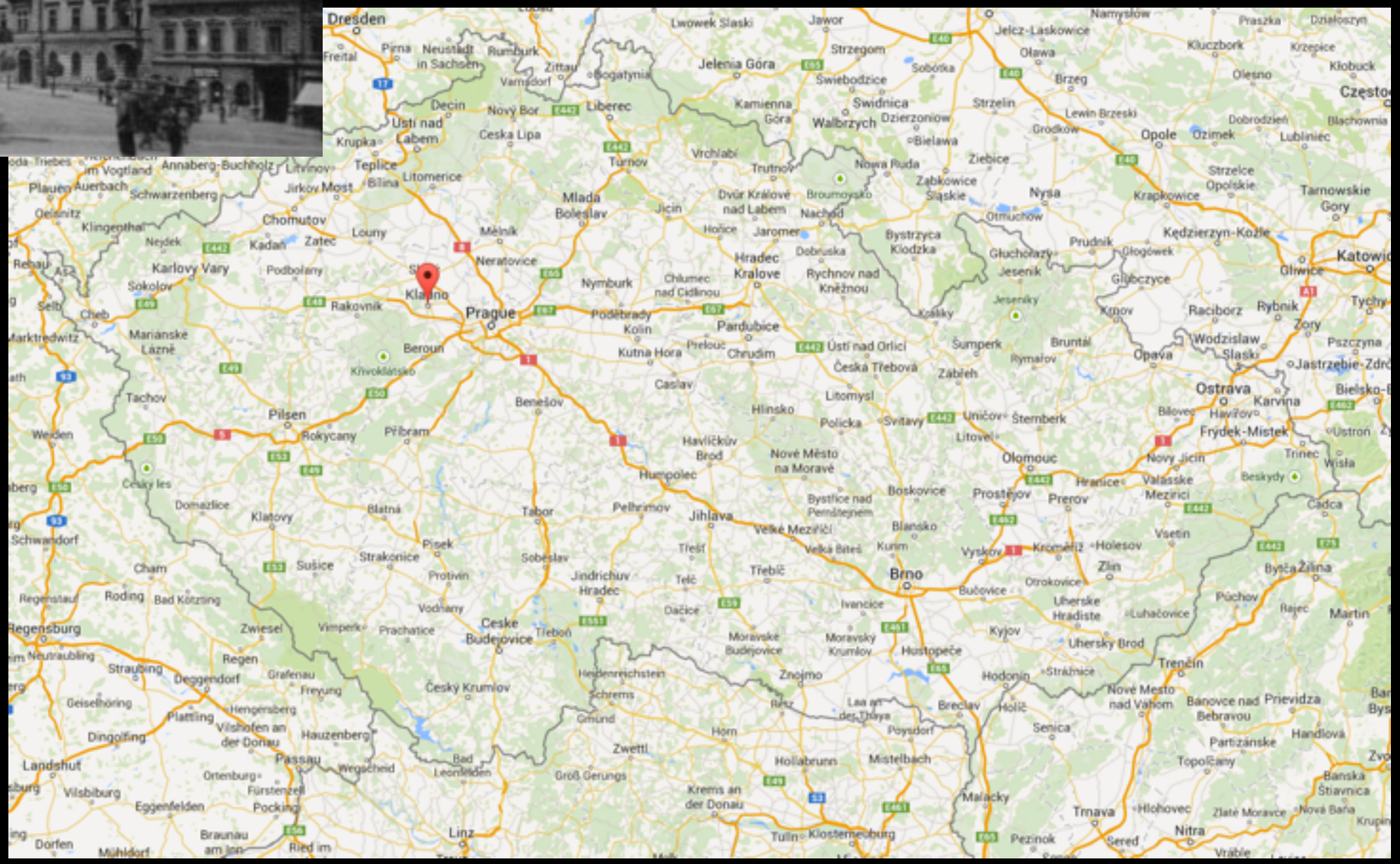
AN ESSAY
Towards a
REAL CHARACTER,
And a
PHILOSOPHICAL
LANGUAGE.

By John Wilkins D.D. *Dean of* Ripon,
And Fellow of the ROYAL SOCIETY.

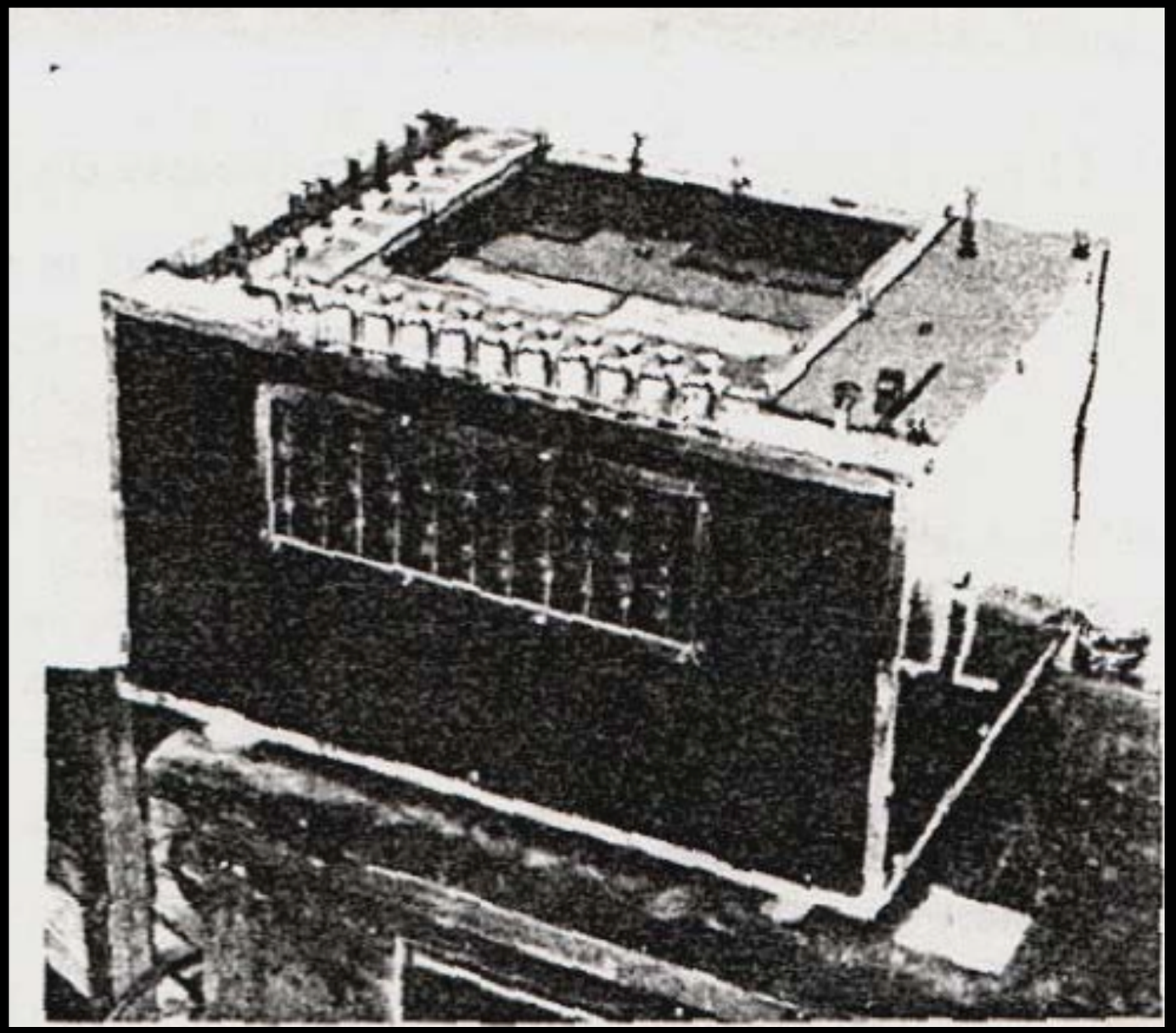
LONDON,
Printed for Sa: Gellibrand, and for
JOHN MARTYN Printer to the *ROYAL*
SOCIETY, 1668.

1932

18 November:
Bedřich Jelínek (Frederick Jelinek)
born in Kladno



1933



22 July: Artsrouni's "mechanical brain" patented in France

1933

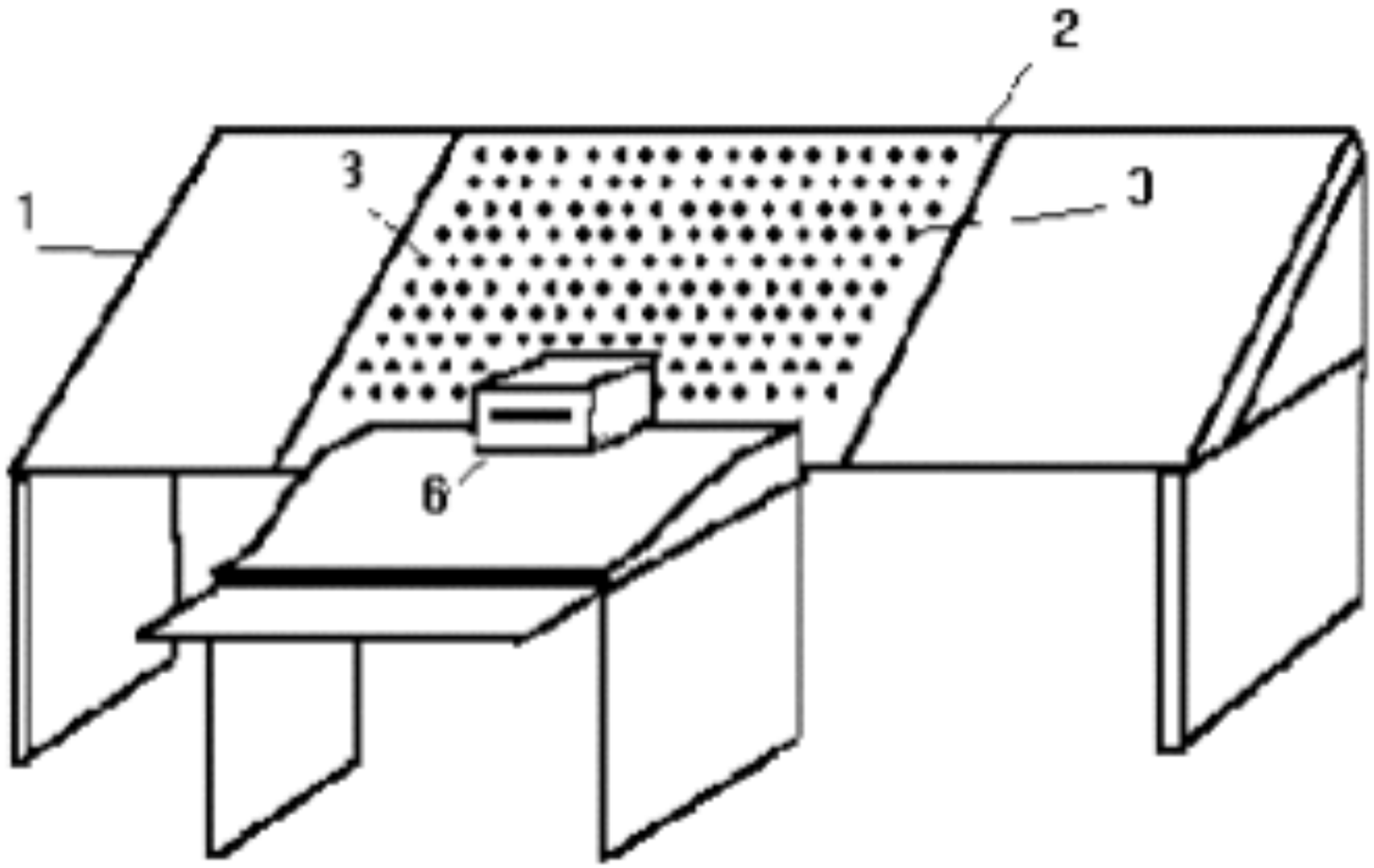
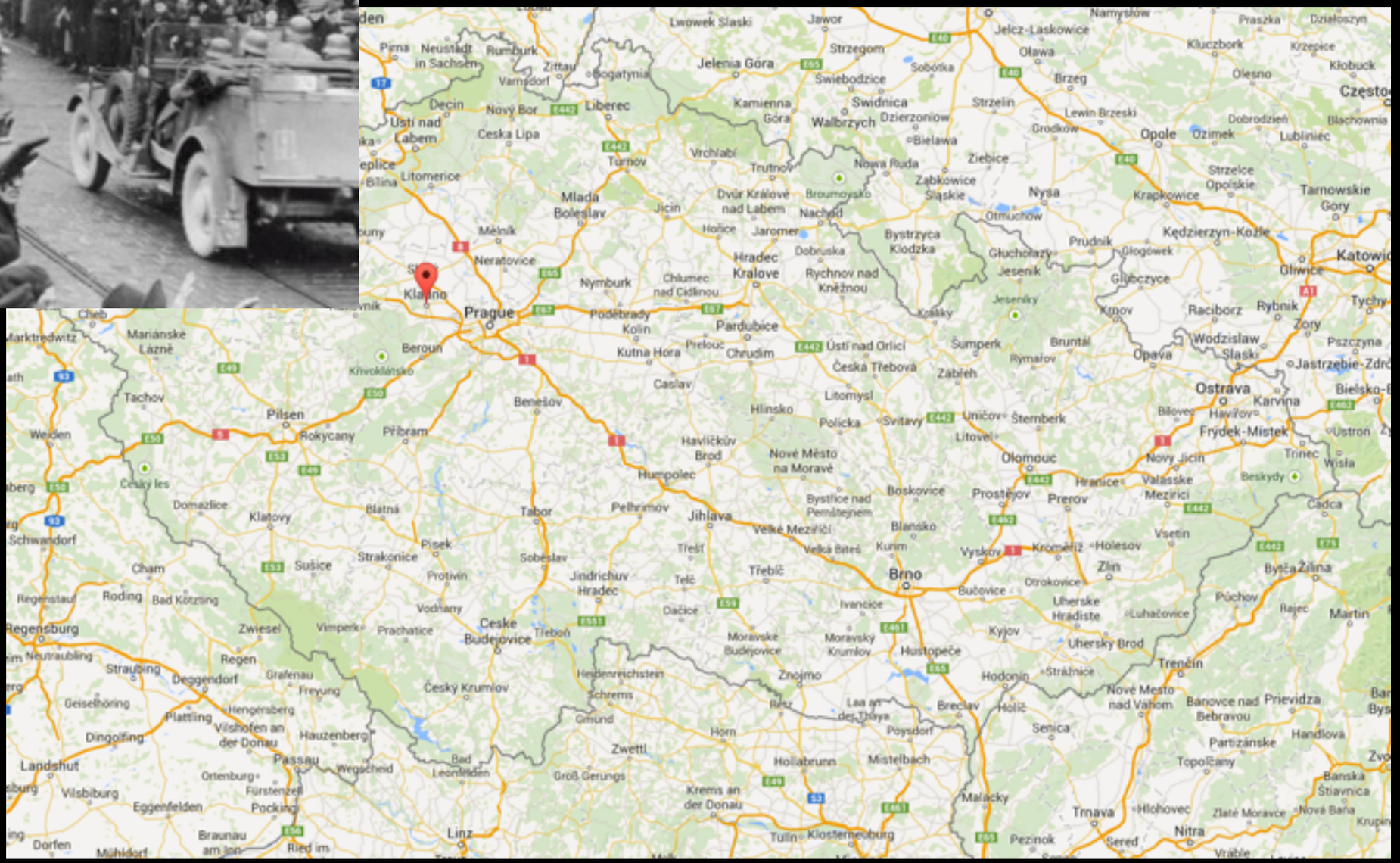


Fig. 3. Trojanskij's translating machine (*from* Bel'skaja et al. 1959)

5 September: Petr Troyanskii's device patented in Russia

1939

March:
Nazis occupy Kladno

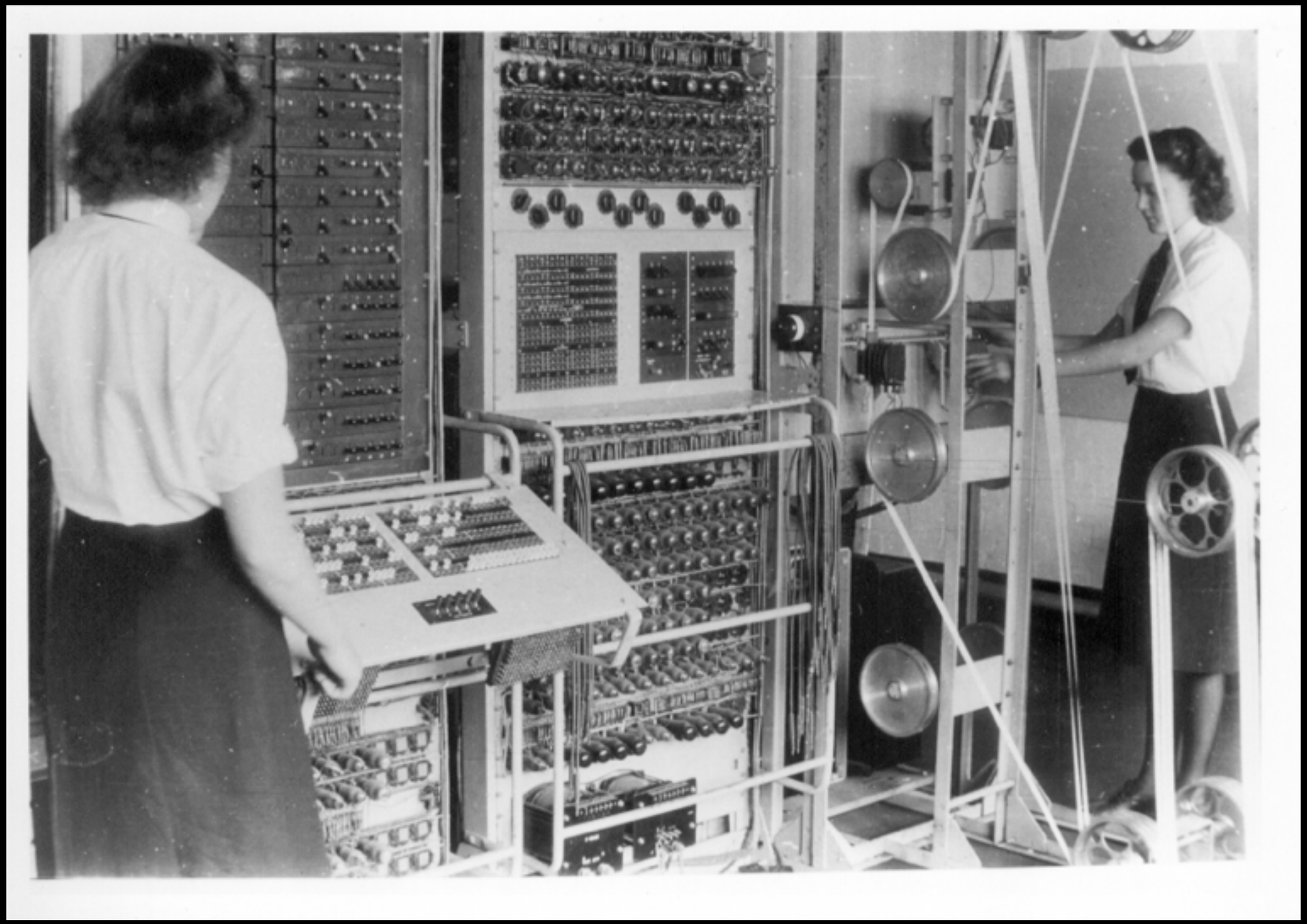


1941

1940s

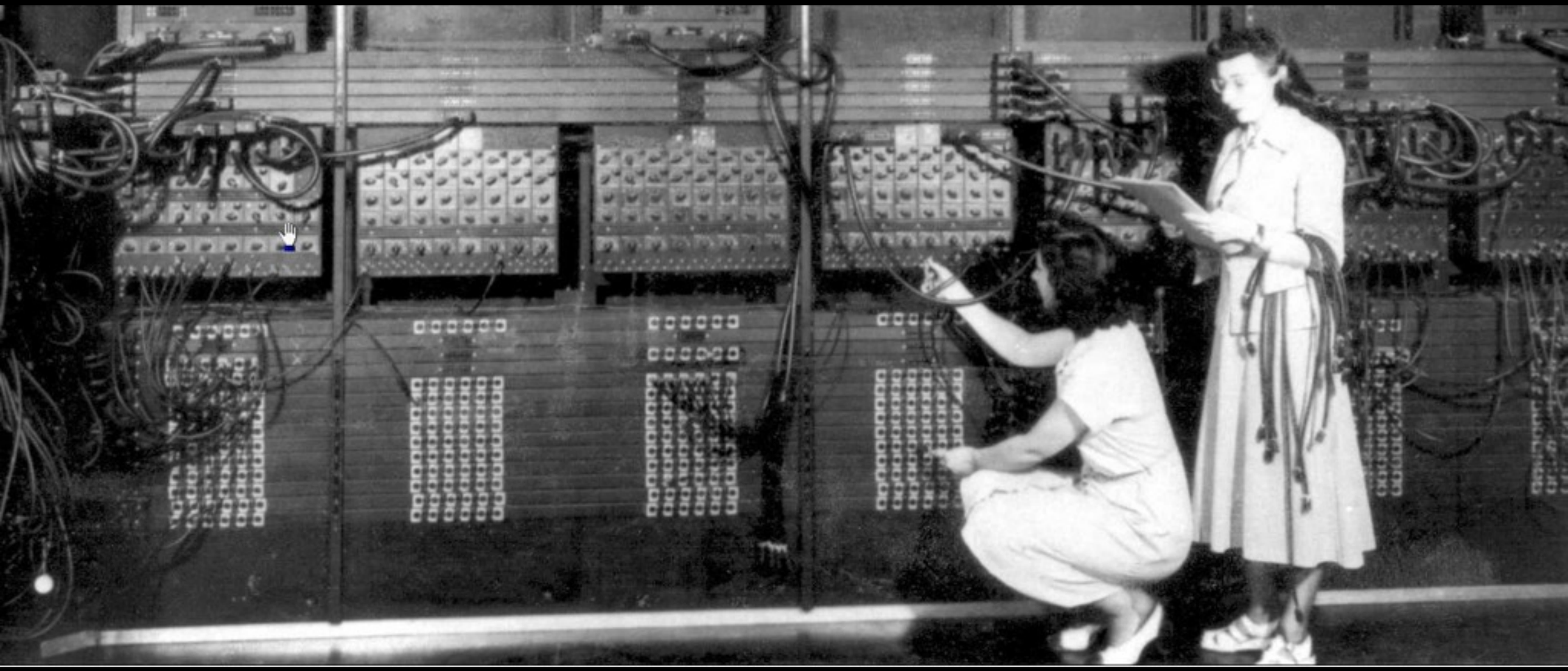
First NLP problem: the German Enigma

1944



5 February: Colossus becomes operational

1946



14 February: ENIAC announced

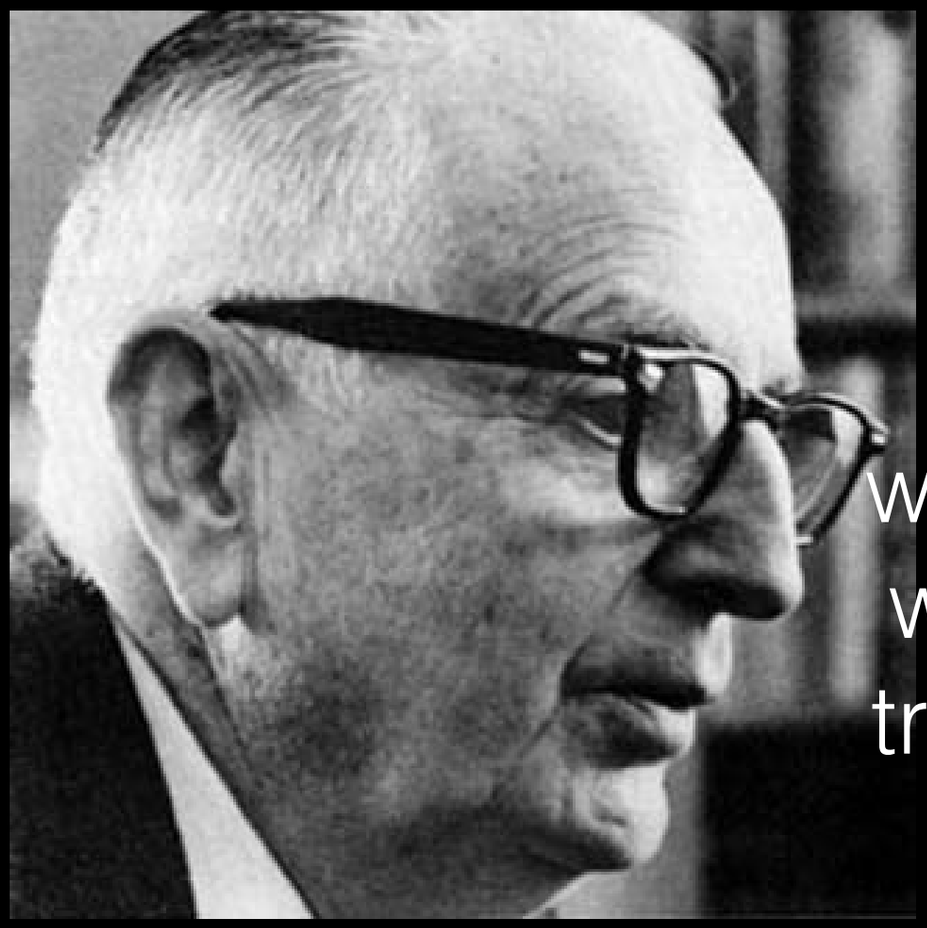
Beginnings

1947

"One thing I wanted to ask you about is this. A most serious problem, for UNESCO and for the constructive and peaceful future of the planet, is the problem of translation, as it unavoidably affects the communication between peoples. Huxley has recently told me that they are appalled by the magnitude and the importance of the translation job."

–Warren Weaver to Norbert Wiener

1947



“Recognizing fully, even though necessarily vaguely, the semantic difficulties because of multiple meanings, etc., I have wondered if it were unthinkable to design a computer which would translate. Even if it would translate only scientific material (where the semantic difficulties are very notably less), and even if it did produce an inelegant (but intelligible) result, it would seem to me worth while. ”

–Warren Weaver to Norbert Wiener

1947

“[A]s to the problem of mechanical translation, I am afraid the boundaries of words in different languages are too vague and the emotional and international connotations are too extensive to make any quasi mechanical translation scheme very hopeful. ”

–Norbert Wiener’s response

1949

"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, "Translation"

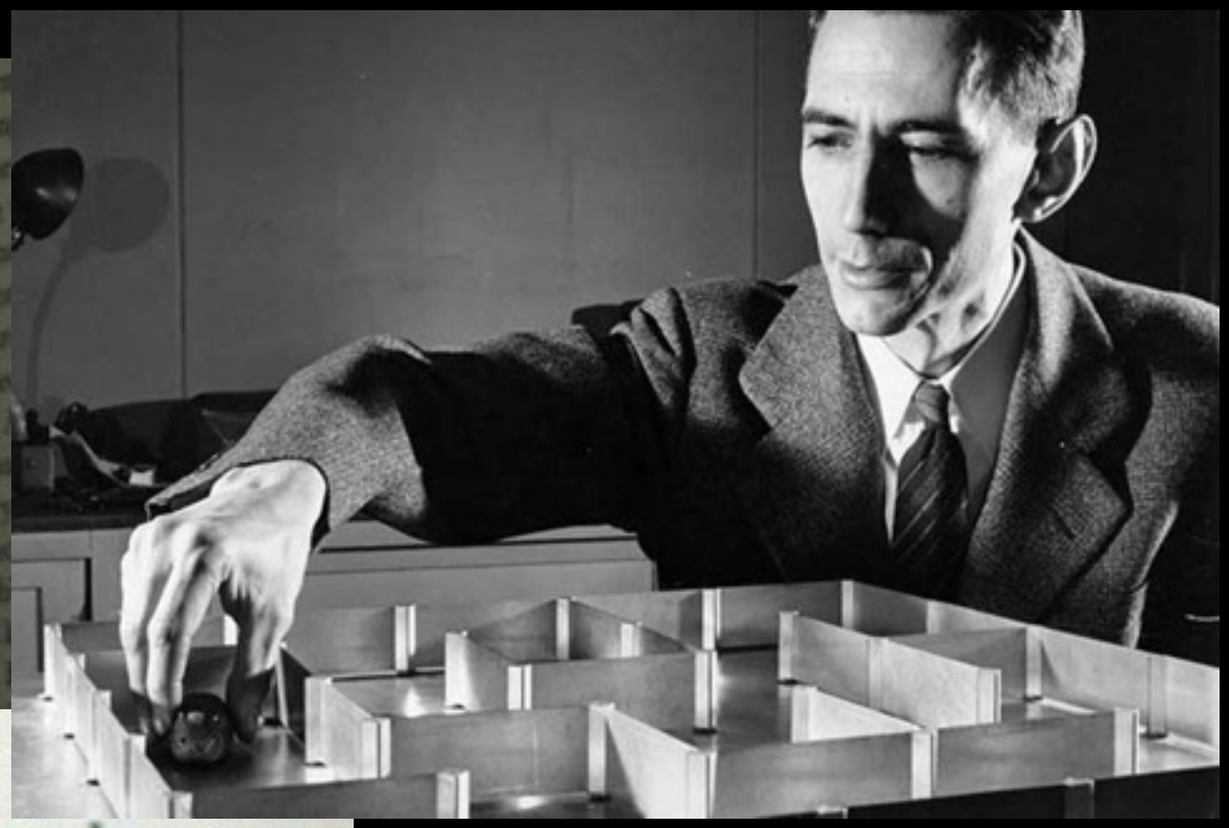
1948

Communication Theory of Secrecy Systems*

By C. E. SHANNON

1. Introduction and Summary

THE problems of cryptography and secrecy systems furnish an interesting application of communication theory.[1] In this paper a theory of secrecy systems is developed. The approach is on a theoretical level and is intended to complement the treatment found in standard works on cryptography.[2] There, a detailed study is made of the many standard types of codes and ciphers, and of the ways of breaking them. We will be more concerned with the general mathematical structure and properties of secrecy systems.



The Bell System Technical Journal

Vol. XXVII July, 1948 No. 3

A Mathematical Theory of Communication

By C. E. SHANNON

Introduction

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has in- general theory of communication. A basis for

1948

"[A]ny stochastic process which produces a discrete sequence of symbols chosen from a finite set may be considered a discrete source. This will include such cases as:

1. Natural written languages such as English, German, Chinese…"

–Claude Shannon,
A Mathematical Theory of Communication

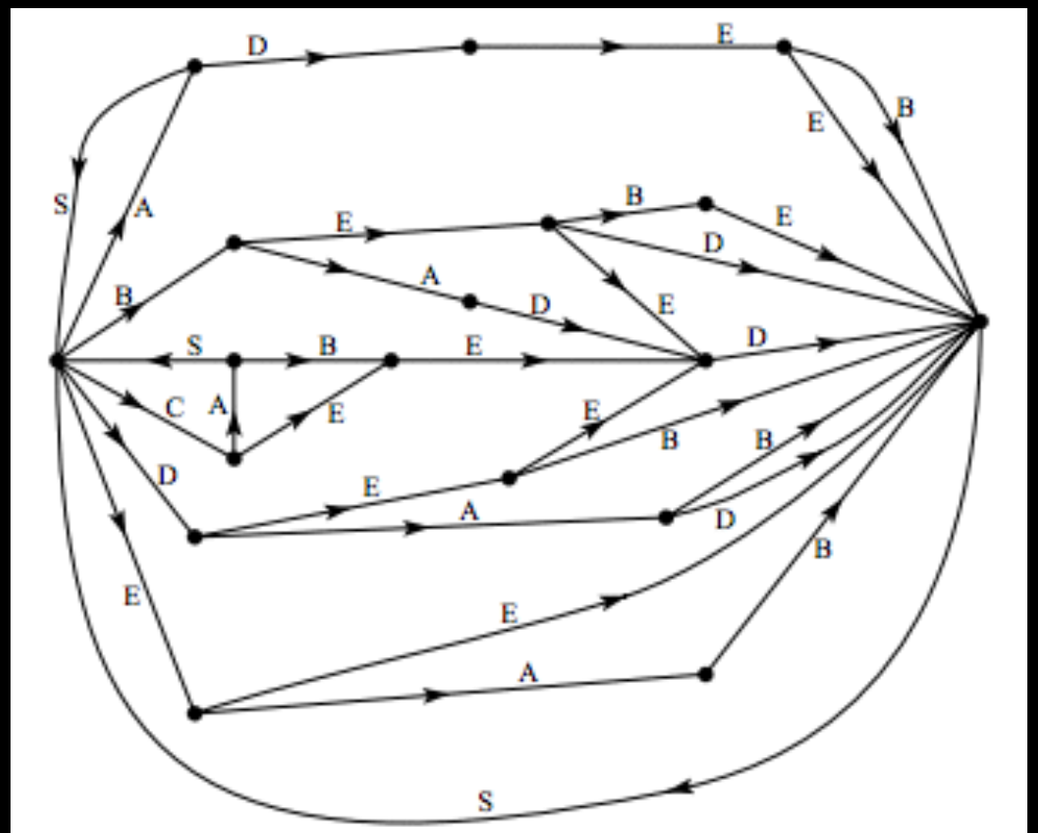


Fig. 5—A graph corresponding to the source in example D.

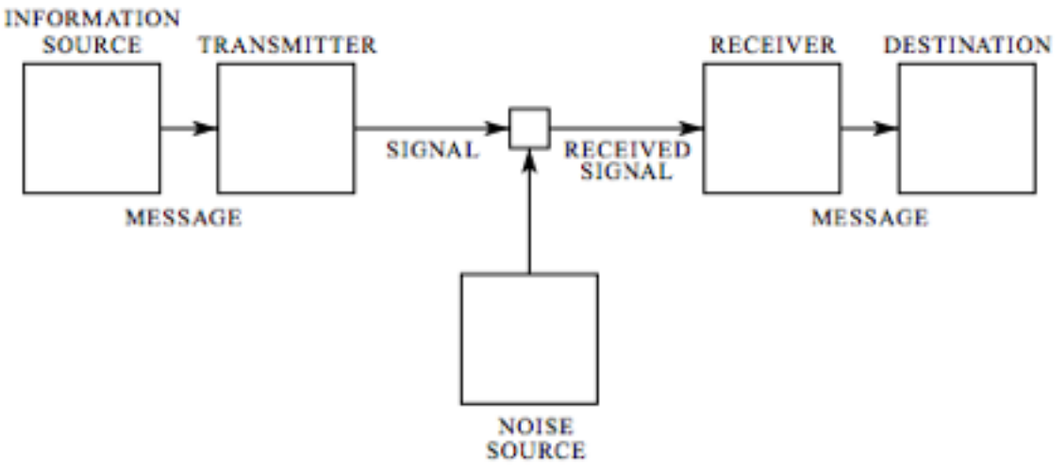
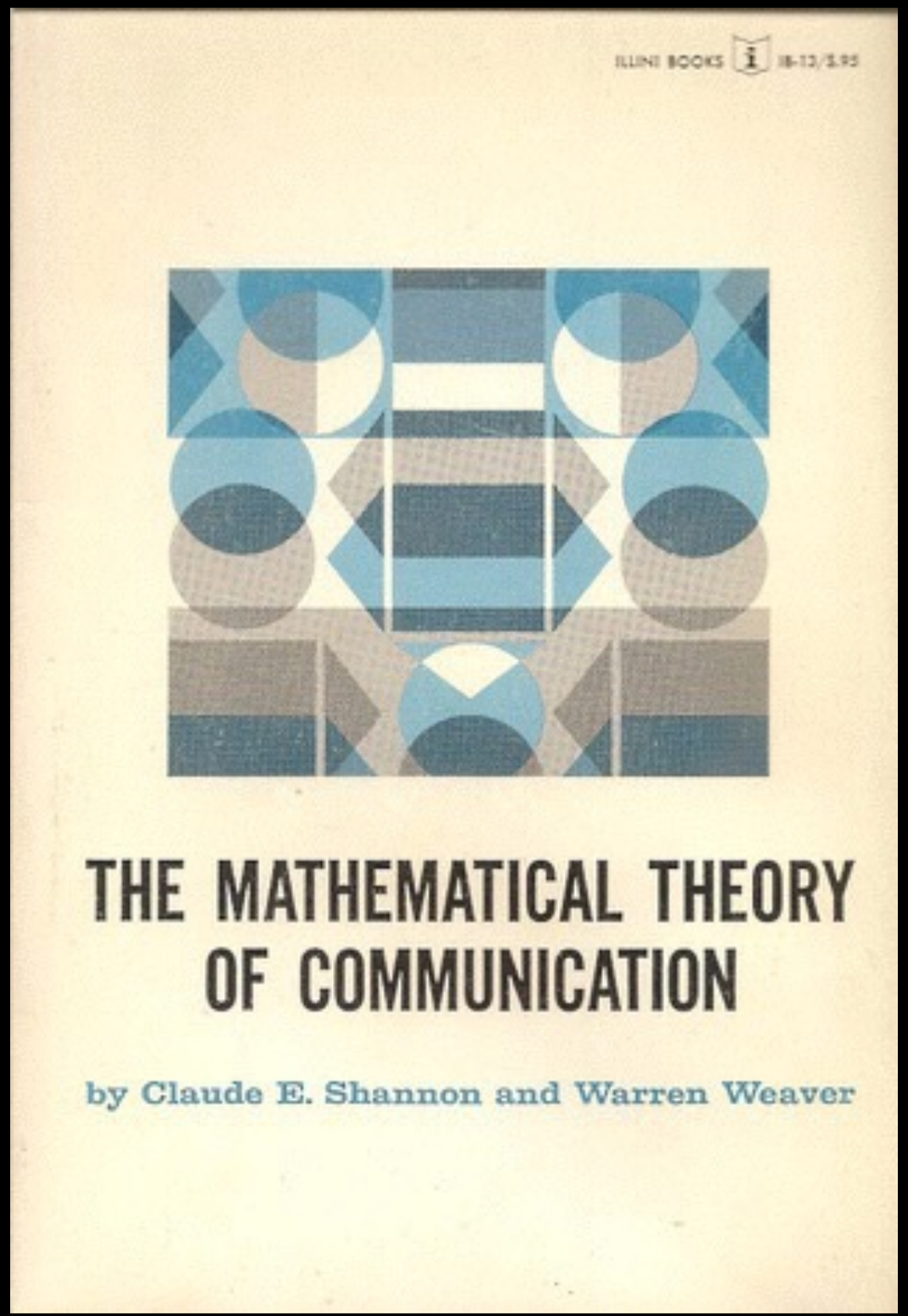
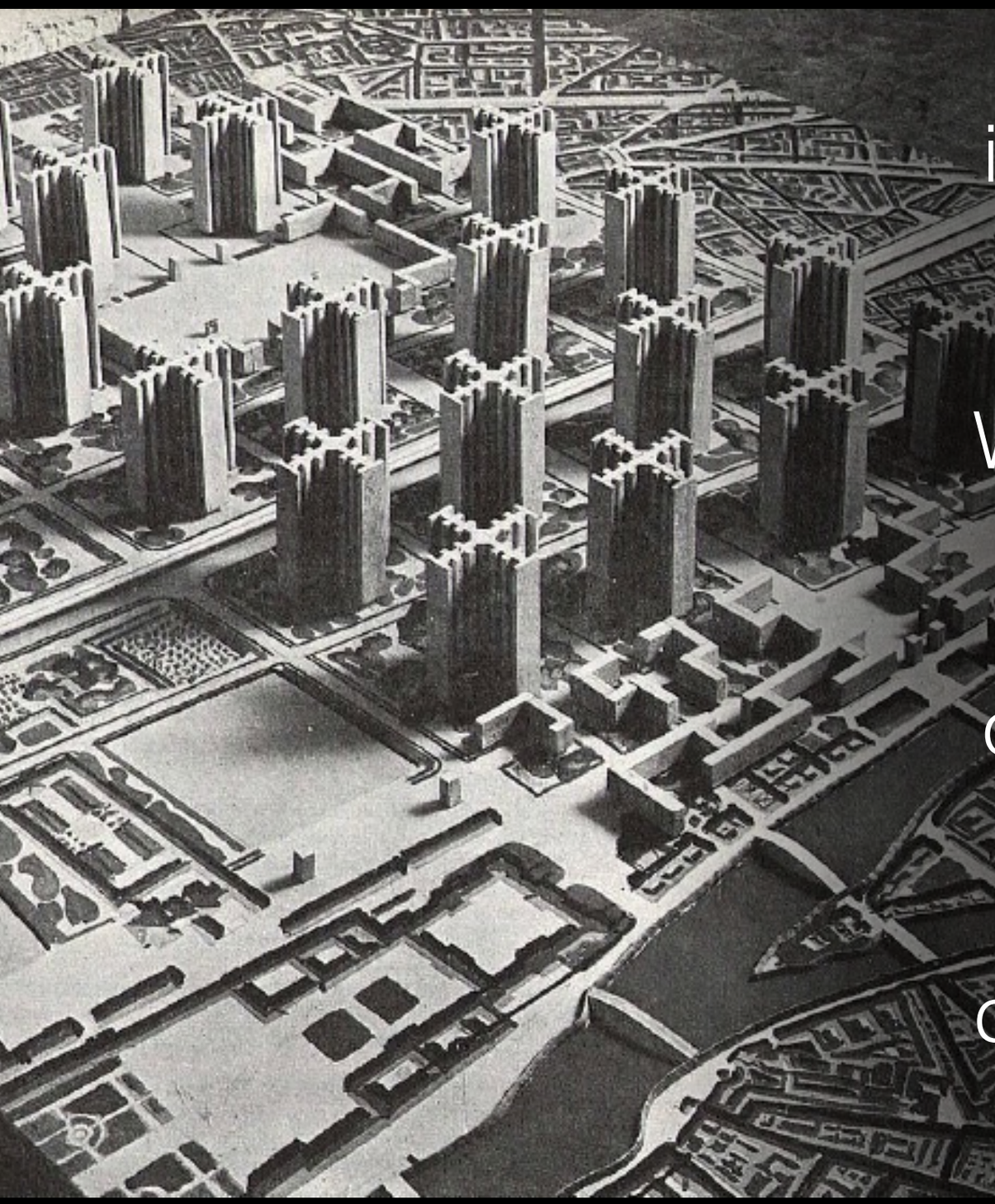


Fig. 1—Schematic diagram of a general communication system.

1949


ILLINI BOOKS
IB-13/$.95
THE MATHEMATICAL THEORY
OF COMMUNICATION
by Claude E. Shannon and Warren Weaver

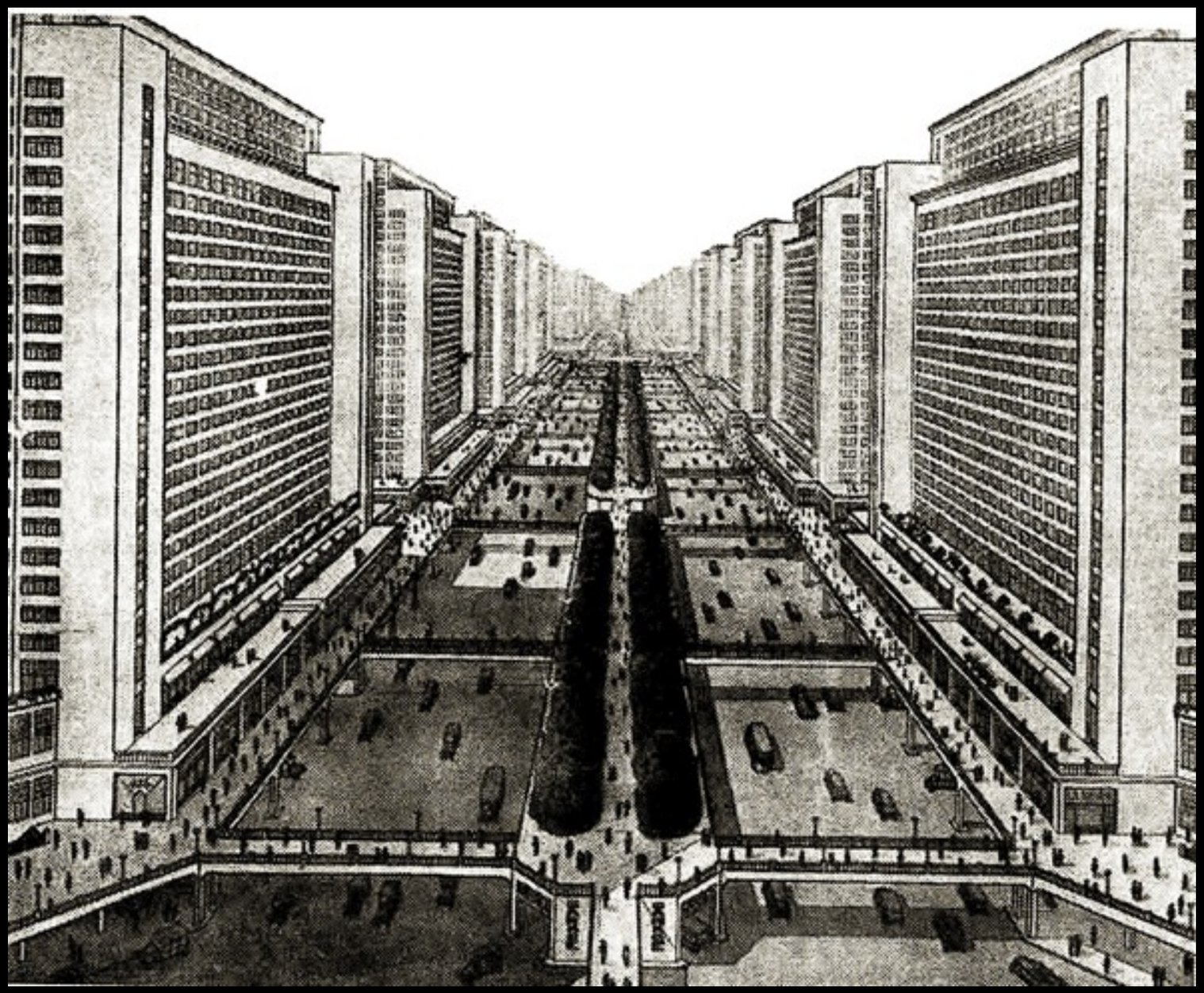
1949



"Think, by analogy, of individuals living in a series of tall closed towers, all erected over a common foundation. When they try to communicate with one another they shout back and forth, each from his own closed tower. It is difficult to make the sound penetrate even the nearest towers, and communication proceeds very poorly indeed."

–Warren Weaver, "Translation"

1949



" But when an individual goes down his tower, he finds himself in a great open basement, common to all the towers. Here he establishes easy and useful communication with the persons who have also descended from their towers. "

–Warren Weaver, "Translation"

1949

"Thus may it be true that the way to translate from Chinese to Arabic, or from Russian to Portuguese, is not to attempt the direct route, shouting from tower to tower. Perhaps the way is to descend, from each language, down to the common base of human communication - the real but as yet undiscovered universal language - and then re-emerge by whatever particular route is convenient."

–Warren Weaver, "Translation"

1949

"Thus may it be true that the way to translate from Chinese to Arabic, or from Russian to Portuguese, is not to attempt the direct route, shouting from tower to tower. Perhaps the way is to descend, from each language, down to the common base of human communication - the real but as yet undiscovered universal language - and then re-emerge by whatever particular route is convenient."

–Warren Weaver, "Translation"

1954

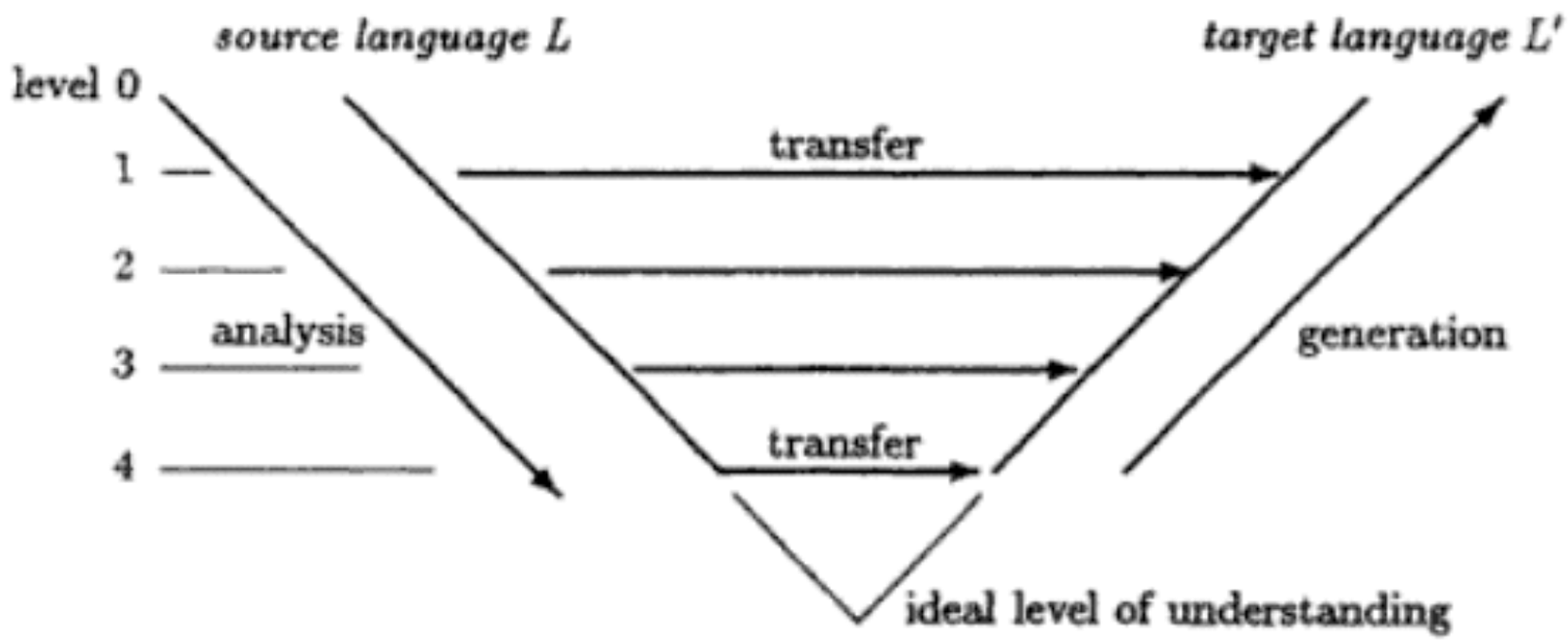


Figure 28.1

Interlingual MT (from Vauquois, 1968)

1954

Sentences in Russian are punched into standard cards for feeding into the electronic data processing machine for translation into English

Georgetown-IBM experiment

1954

Margaret Masterman founds
Cambridge Language Research Unit

Karen Sparck-Jones

Martin Kay

1962

Association for Machine Translation and Computational Linguistics
founded

1949

Trude Jelinek and family emigrate to New York

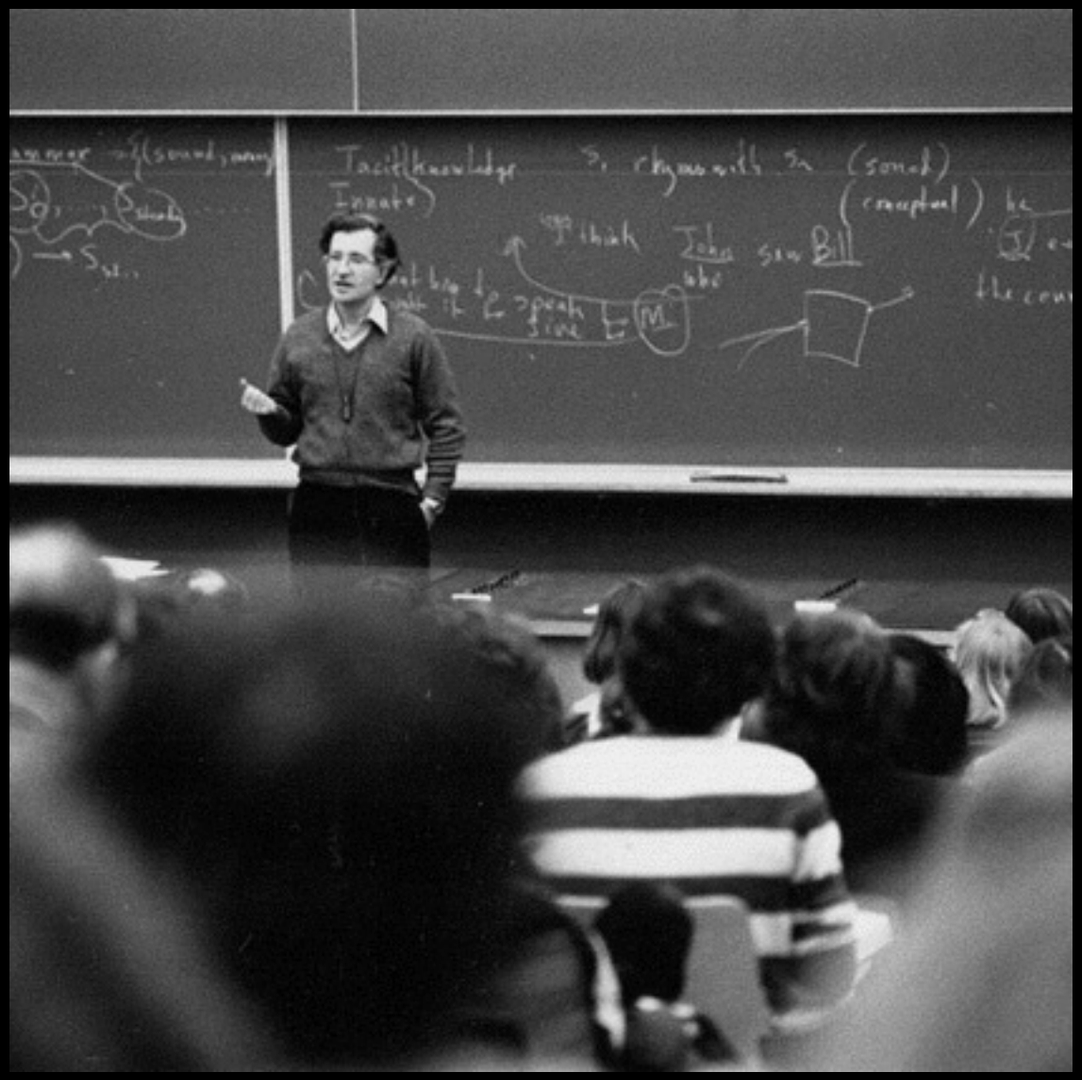
1954

Fred Jelinek moves to MIT

MIT chapel, 1955

1957

1960



Fred Jelinek considers move into linguistics

1962

"After my job talk at Cornell I was approached by the eminent linguist Charles Hockett, who said that he hoped that I would accept the Cornell offer and help develop his ideas on how to apply Information Theory to Linguistics. That decided me."

–Fred Jelinek

1962

"Surprisingly, when I took up my post in the fall of 1962, there was no sign of Hockett. After several months I summoned my courage and went to ask him when he wanted to start working with me."

–Fred Jelinek

1962

"He answered that he was no longer interested,
that he now concentrated on composing operas."

–Fred Jelinek

1962

"Discouraged a second time, I devoted the next ten years to Information Theory."

–Fred Jelinek

Winter

1956

The Bandwagon

CLAUDE E. SHANNON

INFORMATION theory has, in the last few years, become something of a scientific bandwagon. Starting as a technical tool for the communication engineer, it has received an extraordinary amount of publicity in the popular as well as the scientific press. In part, this has been due to connections with such fashionable fields as computing machines, cybernetics, and automation; and in part, to the novelty of its subject matter. As a consequence, it has perhaps been ballooned to an importance beyond its actual accomplishments. Our fellow scientists in many different fields, attracted by the fanfare and by the new avenues opened to scientific analysis, are using these ideas in their own problems. Applications are being made to biology, psychology, linguistics, fundamental physics, economics, the theory of organization, and many others. In short, information theory is currently partaking of a somewhat heady draught of general popularity.

Although this wave of popularity is certainly pleasant and exciting for those of us working in the field, it carries at the same time an element of danger. subject are aimed in a very specific direction, a direction that is not necessarily relevant to such fields as psychology, economics, and other social sciences. Indeed, the hard core of information theory is, essentially, a branch of mathematics, a strictly deductive system. A thorough understanding of the mathematical foundation and its communication application is surely a prerequisite to other applications. I personally believe that many of the concepts of information theory will prove useful in these other fields—and, indeed, some results are already quite promising—but the establishing of such applications is not a trivial matter of translating words to a new domain, but rather the slow tedious process of hypothesis and experimental verification. If, for example, the human being acts in some situations like an ideal decoder, this is an experimental and not a mathematical fact, and as such must be tested under a wide variety of experimental situations.

Secondly, we must keep our own house in first class order. The subject of information theory has cer-

1956

"Seldom do more than a few of nature's secrets give way at one time. It will be all too easy for our somewhat artificial prosperity to collapse overnight when it is realized that the use of a few exciting words like *information*, *entropy*, *redundancy*, do not solve all our problems."

–Claude Shannon, "The Bandwagon"

1957

"I think we are forced to conclude that ... probabilistic models give no particular insight into some of the basic problems of syntactic structure."

–Noam Chomsky, "Syntactic structures"

"[I]t must be recognized that the notion of 'probability of a sentence' is an entirely useless one, under any known interpretation of this term."

–in *"Challenges to empiricism"* (1969)

1960

"Let me finish… by warning in general against overestimating the impact of statistical information on the problem of MT and related questions. I believe that this overestimation is a remnant of the time, seven or eight years ago, when many people thought that the statistical theory of communication would solve many, if not all, of the problems of communication… it is my impression that much valuable time of MT workers has been spent on trying to obtain statistical information whose impact on MT is by no means evident."

–Yehoshua Bar-Hillel, "The present status of automatic translation of languages"

1966

"'Machine Translation' presumably means going by algorithm from machine-readable source text to useful target text, without recourse to human translation or editing. In this context, there has been no machine translation of general scientific text, and none is in immediate prospect."

The ALPAC report

1966

"The computer has opened up to linguists a host of challenges, partial insights, and potentialities. We believe these can be aptly compared with the challenges, problems, and insights of particle physics. Certainly, language is second to no phenomenon in importance. And the tools of computational linguistics are considerably less costly than the multibillion-volt accelerators of particle physics. The new linguistics presents an attractive as well as an extremely important challenge"

–John R. Pierce, in preface to the ALPAC report

1968

Association for Machine Translation and Computational Linguistics
changes its name to
Association for Computational Linguistics

1973

"The most notorious disappointments, however, have appeared in the area of machine translation, where enormous sums have been spent with very little useful result, as a careful review by the US National Academy of Sciences concluded in 1966; a conclusion not shaken by any subsequent developments."

–James Lighthill

1978

The Cambridge Language Research Unit
ceases operation.

The cybernetic underground

1972

Fred Jelinek goes to IBM Research for the summer

1972

Fred Jelinek goes to IBM Research for the summer
and stays there for over 20 years

1972

"IBM was worried that, with the advance of computing power, there might soon come a time when all the need for further improvements would disappear, and IBM business would dry up."

–Fred Jelinek

1972

1975

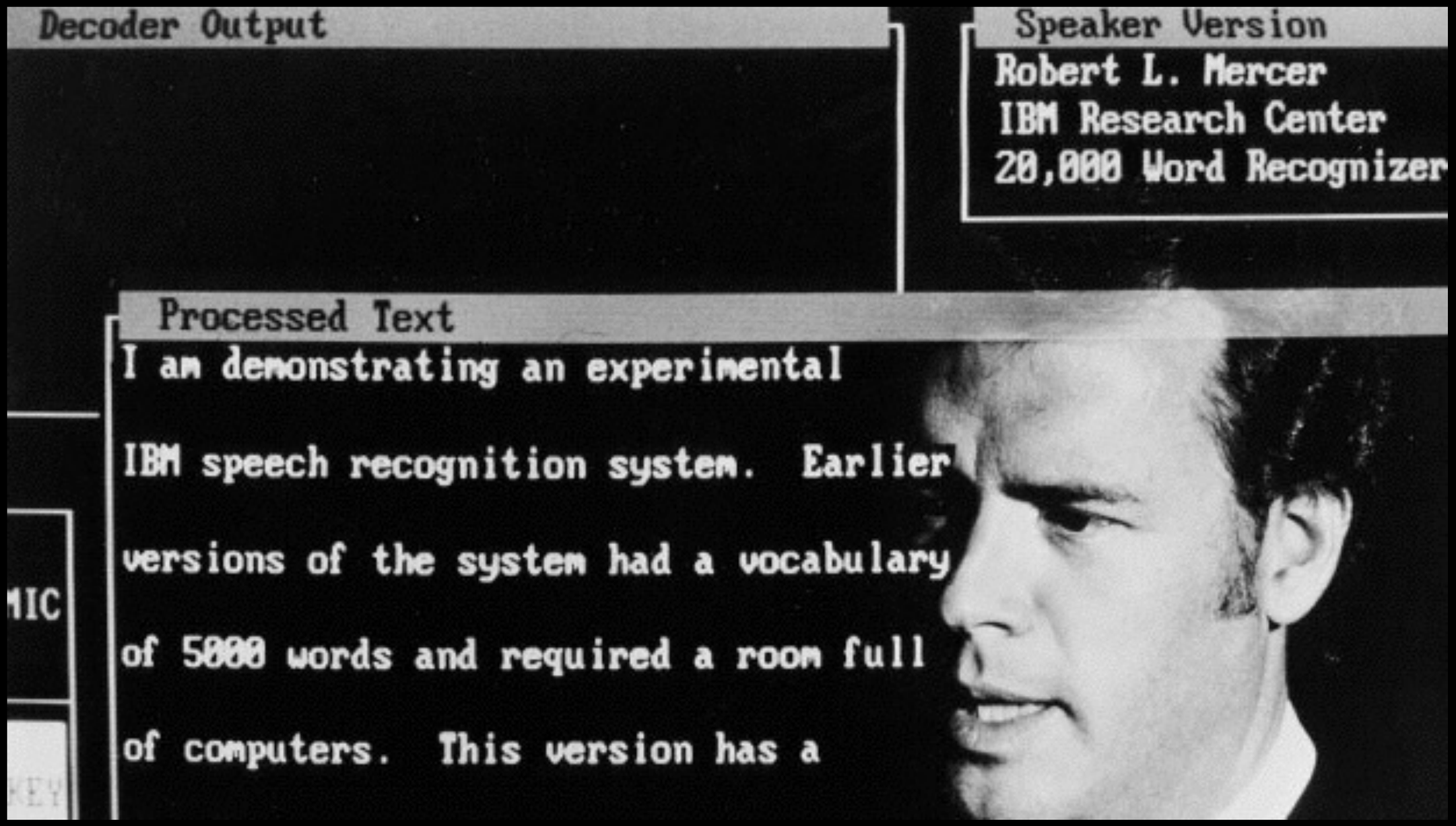
$$\hat{W} = \arg\max_{W} P(W|A) = \arg\max_{W} P(A|W)P(W)$$

"Here's something we jocularly called the fundamental equation of speech recognition at ICASSP in 1981. As I'm sure all of you very well know, this is just an application of Bayes' Rule."

–Bob Mercer

1981

1984


Decoder Output
Speaker Version
Robert L. Mercer
IBM Research Center
20,000 Word Recognizer
Processed Text
I am demonstrating an experimental
IBM speech recognition system. Earlier
versions of the system had a vocabulary
of 5000 words and required a room full
of computers. This version has a

1987

LDC formed, begins work on Penn Treebank

mid-1980s

John Cocke, inventor of RISC architecture,
finds an interesting new dataset
(John Cocke is the “C” in CKY)

1987

1987

1988

"The validity of a statistical (information theoretic) approach to MT has indeed been recognized, as the authors mention, by Weaver as early as 1949. And was universally recognized as mistaken by 1950 (cf. Hutchins, MT – Past, Present, Future, Ellis Horwood, 1986, p. 30 and references therein). The crude force of computers is not science. This paper is simply beyond the scope of COLING."

–Anonymous COLING review

1988

"The validity of a statistical (information theoretic) approach to MT has indeed been recognized, as the authors mention, by Weaver as early as 1949. And was universally recognized as mistaken by 1950 (cf. Hutchins, MT – Past, Present, Future, Ellis Horwood, 1986, p. 30 and references therein). The crude force of computers is not science. This paper is simply beyond the scope of COLING."

–Anonymous COLING review

"A statistical approach to language translation" 1988. P. Brown, J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, R. Mercer, P. Rosin. In *Proc. of COLING*.

1988

"The validity of a statistical (information theoretic) approach to MT has indeed been recognized, as the authors mention, by Weaver as early as 1949. And was universally recognized as mistaken by 1950 (cf. Hutchins, MT – Past, Present, Future, Ellis Horwood, 1986, p. 30 and references therein). The crude force of computers is not science. This paper is simply beyond the scope of COLING."

–Anonymous COLING review

"A statistical approach to language translation" 1988. P. Brown, J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, R. Mercer, P. Rosin. In *Proc. of COLING*.

"I wonder what COLING reviews look like for papers that get rejected?" *–Peter Brown*

1988

"The validity of a statistical (information theoretic) approach to MT has indeed been recognized, as the authors mention, by Weaver as early as 1949. And was universally recognized as mistaken by 1950 (cf. Hutchins, MT – Past, Present, Future, Ellis Horwood, 1986, p. 30 and references therein). The **crude force of computers** is not science. This paper is simply beyond the scope of COLING."

–Anonymous COLING review

"A statistical approach to language translation" 1988. P. Brown, J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, R. Mercer, P. Rosin. In *Proc. of COLING*.

"I wonder what COLING reviews look like for papers that get rejected?" *–Peter Brown*

1988

"The validity of a statistical (information theoretic) approach to MT has indeed been recognized, as the authors mention, by Weaver as early as 1949. And was universally recognized as mistaken by 1950 (cf. Hutchins, MT – Past, Present, Future, Ellis Horwood, 1986, p. 30 and references therein). The **force of crude computers** is not science. This paper is simply beyond the scope of COLING."

–Anonymous COLING review

"A statistical approach to language translation" 1988. P. Brown, J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, R. Mercer, P. Rosin. In *Proc. of COLING*.

"I wonder what COLING reviews look like for papers that get rejected?" *–Peter Brown*

1988

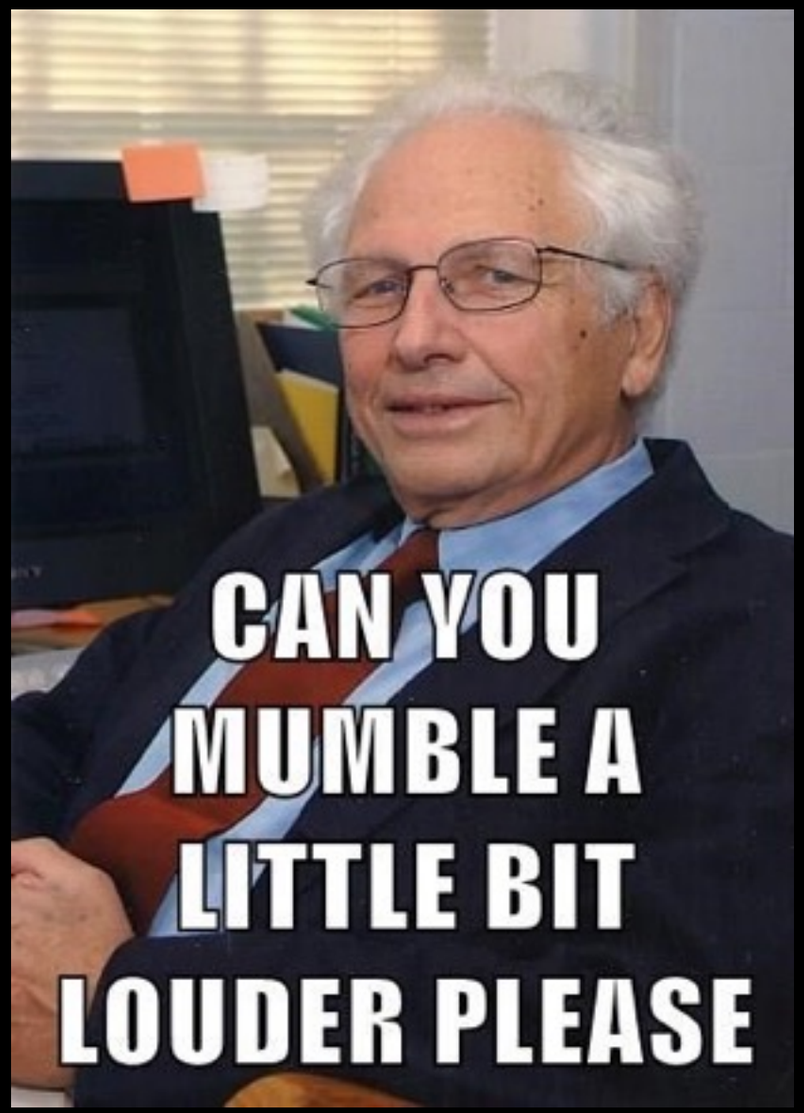
"Every time I fire a linguist the performance of the recognizer goes up."

–Fred Jelinek

1988

"Every time I fire a linguist the performance of the recognizer goes up."

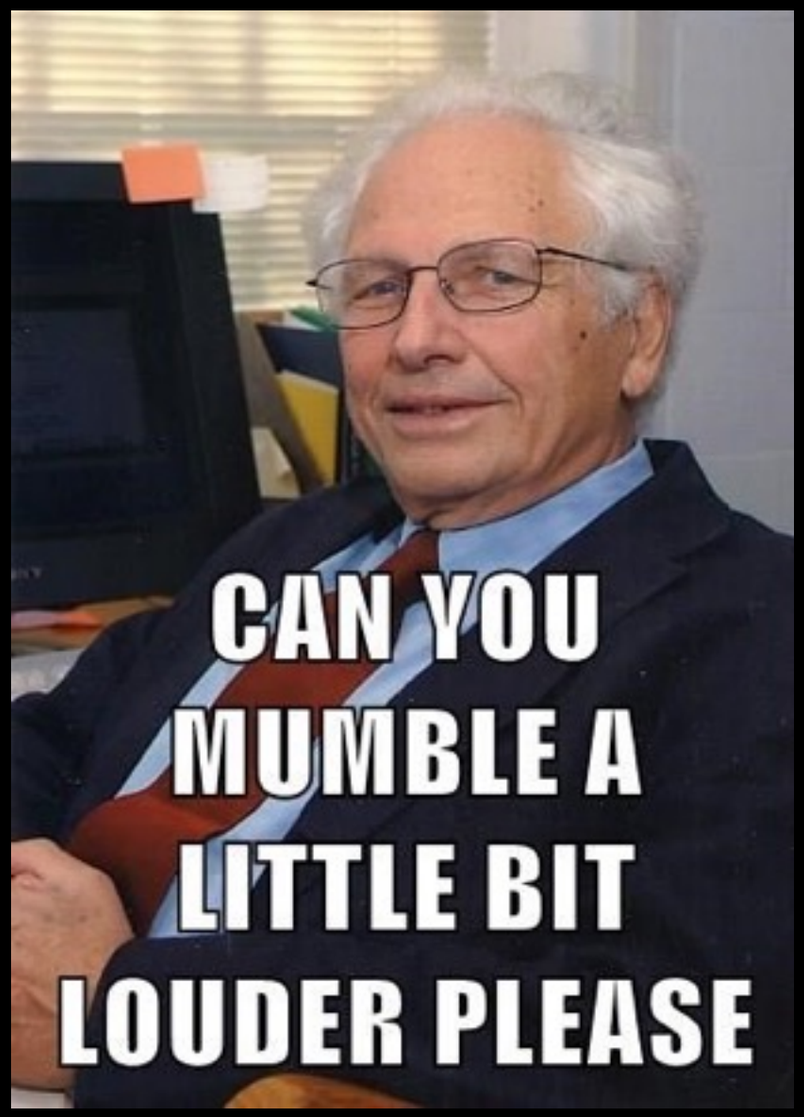
–Fred Jelinek



1988

"Every time I fire a linguist the performance of the recognizer goes up."

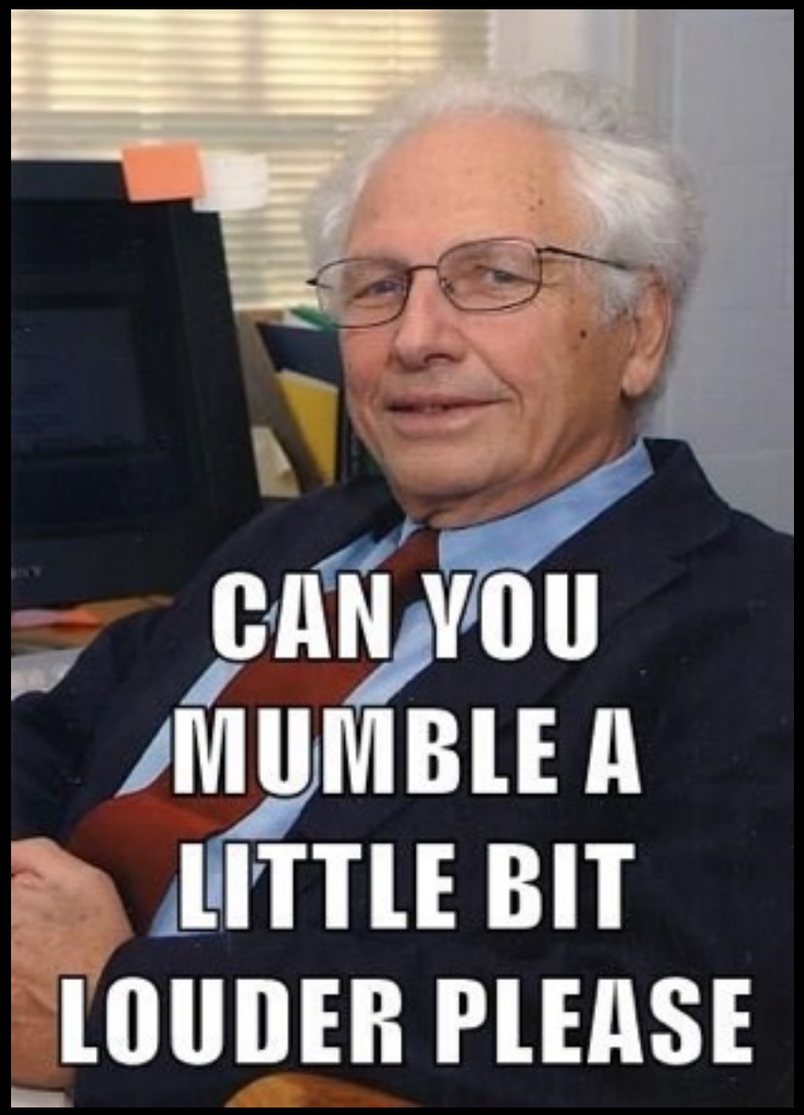
–Fred Jelinek



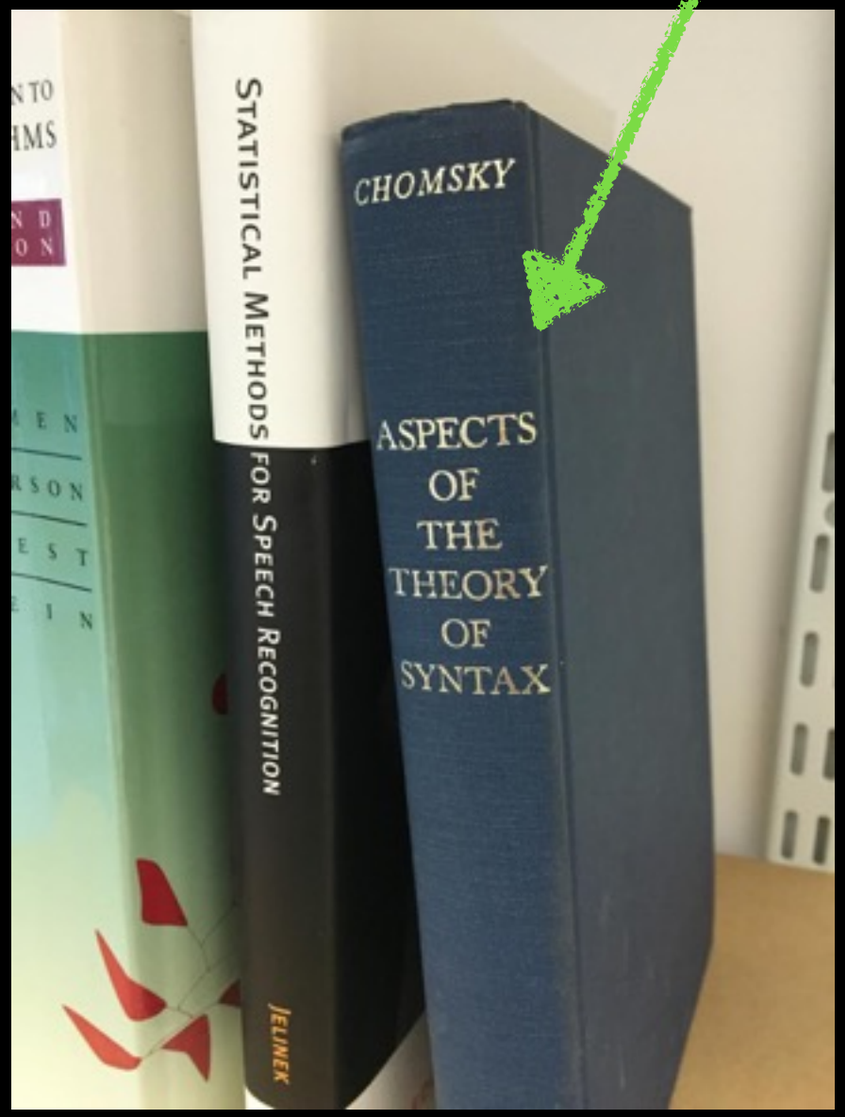
1988

"Every time I fire a linguist the performance of the recognizer goes up."

–Fred Jelinek



From Fred's library



1990

A Statistical Approach to Machine Translation

Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin

IBM
Thomas J. Watson Research Center
Yorktown Heights, NY

In this paper, we present a statistical approach to machine translation. We describe the application of our approach to translation from French to English and give preliminary results.

1 Introduction

The field of machine translation is almost as old as the modern digital computer. In 1949 Warren Weaver suggested that the problem be attacked with statistical methods and ideas from information theory, an area which he, Claude Shannon, and others were developing at the time (Weaver 1949). Although researchers quickly abandoned this approach, advancing numerous theoretical objections, we believe that the true obstacles lay in the relative impotence of the available computers and the dearth of machine-readable text from which to gather the statistics vital to such an attack. Today, computers are five orders of magnitude faster than they were in 1950 and have hundreds of millions of bytes of storage. Large, machine-readable corpora are readily available. Statistical methods have proven their value in automatic speech recognition (Bahl et al. 1983) and have recently been applied to lexicography (Sinclair 1985) and to natural language processing (Baker 1979; Ferguson 1980; Garside et al. 1987; Sampson 1986;

sentence in one language is a possible translation of any sentence in the other. We assign to every pair of sentences (S, T) a probability, $\Pr(T|S)$, to be interpreted as the probability that a translator will produce T in the target language when presented with S in the source language. We expect $\Pr(T|S)$ to be very small for pairs like (*Le matin je me brosse les dents*|*President Lincoln was a good lawyer*) and relatively large for pairs like (*Le president Lincoln était un bon avocat*|*President Lincoln was a good lawyer*). We view the problem of machine translation then as follows. Given a sentence T in the target language, we seek the sentence S from which the translator produced T. We know that our chance of error is minimized by choosing that sentence S that is most probable given T. Thus, we wish to choose S so as to maximize $\Pr(S|T)$. Using Bayes' theorem, we can write

$$\Pr(S|T) = \frac{\Pr(S)\Pr(T|S)}{\Pr(T)}$$

1993

The Mathematics of Statistical Machine Translation: Parameter Estimation

Peter F. Brown*
IBM T.J. Watson Research Center

Stephen A. Della Pietra*
IBM T.J. Watson Research Center

Vincent J. Della Pietra*
IBM T.J. Watson Research Center

Robert L. Mercer*
IBM T.J. Watson Research Center

We describe a series of five statistical models of the translation process and give algorithms for estimating the parameters of these models given a set of pairs of sentences that are translations of one another. We define a concept of word-by-word alignment between such pairs of sentences. For any given pair of such sentences each of our models assigns a probability to each of the possible word-by-word alignments. We give an algorithm for seeking the most probable of these alignments. Although the algorithm is suboptimal, the alignment thus obtained accounts well for the word-by-word relationships in the pair of sentences. We have a great deal of data in French and English from the proceedings of the Canadian Parliament. Accordingly, we have restricted our work to these two languages; but we feel that because our algorithms have minimal linguistic content they would work well on other pairs of languages. We also feel, again because of the minimal linguistic content of our algorithms, that it is reasonable to argue that word-by-word alignments are inherent in any sufficiently large bilingual corpus.

1. Introduction

The growing availability of bilingual, machine-readable texts has stimulated interest in methods for extracting linguistically valuable information from such texts. For example, a number of recent papers deal with the problem of automatically obtaining pairs of aligned sentences from parallel corpora (Warwick and Russell 1990; Brown, Lai, and Mercer 1991; Gale and Church 1991b; Kay 1991). Brown et al. (1990) assert, and Brown, Lai, and Mercer (1991) and Gale and Church (1991b) both show, that it is

1993

Techniques pioneered by the IBM group

- Statistical models of speech, translation, syntax, and part-of-speech tagging
- Triphone models (speech)
- n-gram language models and smoothing
- maximum entropy models
- Information-theoretic clustering algorithms (Brown clustering, which Peter Brown calls Mercer clustering)

1993

Mercer, Brown, and the Della Pietras leave IBM

Jelinek moves to Johns Hopkins University

The Center For Language and Speech Processing
at the Johns Hopkins University

1993

Mercer, Brown, and the Della Pietras leave IBM

Jelinek moves to Johns Hopkins University

The Center For Language and Speech Processing
at the Johns Hopkins University

Rediscovery

1999

"When we looked at this paper, written in math, we thought: 'This is really about natural language, but it has been coded in some strange symbols. We will now proceed to decode.'"

–Kevin Knight, USC/ISI

1999

GIZA toolkit (a reimplementation of the IBM models)
developed at Johns Hopkins CLSP workshop

“Late in the workshop we built an MT system for a new language pair (Chinese/ English) in a single day.”

Statistical Machine Translation

Final Report, JHU Workshop 1999

Yaser Al-Onaizan, Jan Curin, Michael Jahr, Kevin Knight,
John Lafferty, Dan Melamed, Franz-Josef Och,
David Purdy, Noah A. Smith, David Yarowsky

2002

First company to commercialize statistical machine translation

2003

Statistical Phrase-Based Translation

Philipp Koehn, Franz Josef Och, Daniel Marcu
Information Sciences Institute
Department of Computer Science
University of Southern California
koehn@isi.edu, och@isi.edu, marcu@isi.edu

Abstract

We propose a new phrase-based translation model and decoding algorithm that enables us to evaluate and compare several, previously proposed phrase-based translation models. Within our framework, we carry out a large number of experiments to understand better and explain why phrase-based models outperform word-based models. Our empirical results, which hold for all examined language pairs, suggest that the highest levels of performance can be obtained through relatively simple means: heuristic learning of phrase translations from word-based alignments and lexical weighting of phrase translations. Surprisingly, learning phrases longer than three words and learning phrases from high-accuracy word-level alignment models does not have a strong impact on performance. Learning only syntac-

method to extract phrase translation pairs? In order to investigate this question, we created a uniform evaluation framework that enables the comparison of different ways to build a phrase translation table.

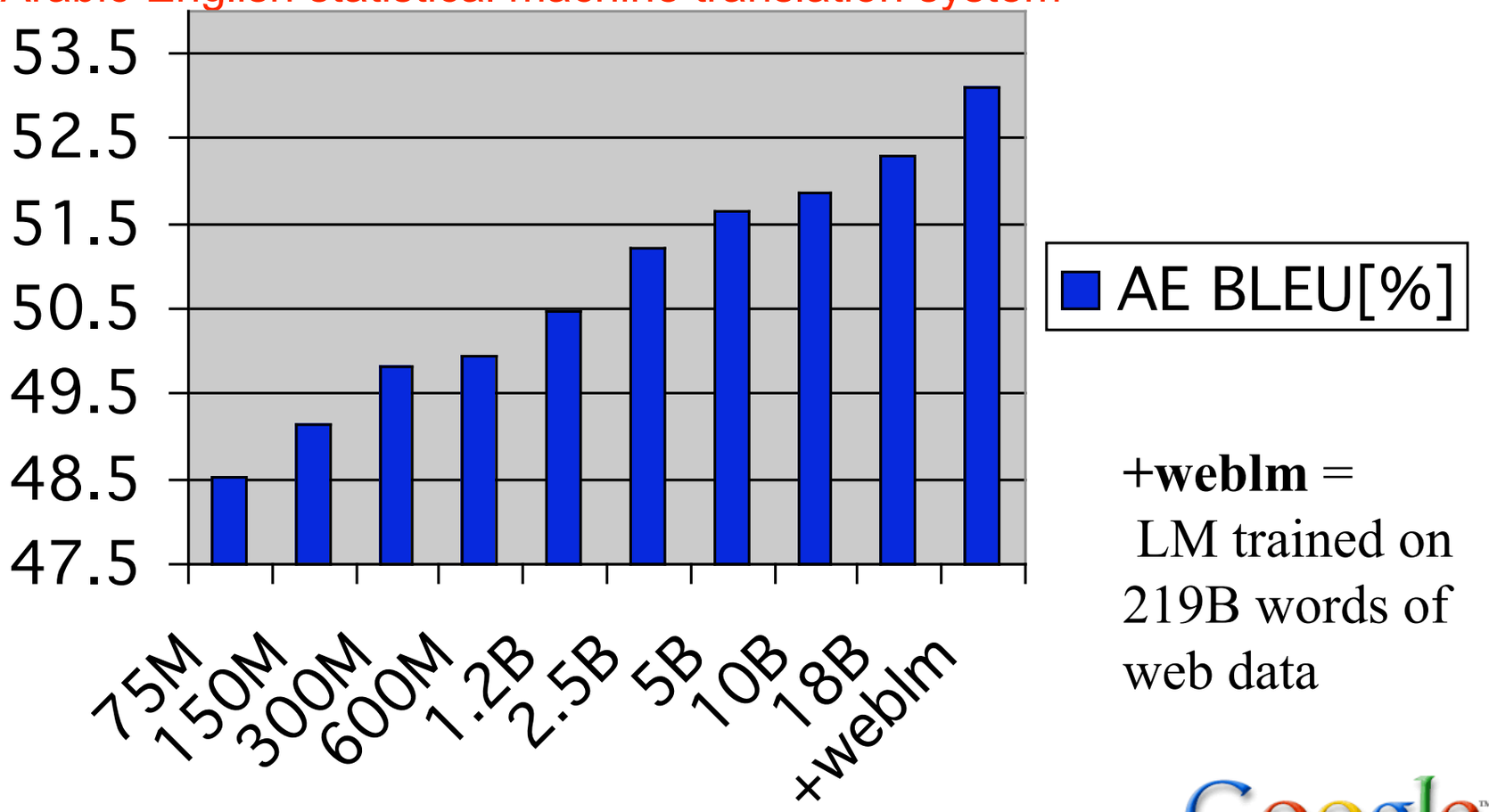
Our experiments show that high levels of performance can be achieved with fairly simple means. In fact, for most of the steps necessary to build a phrase-based system, tools and resources are freely available for researchers in the field. More sophisticated approaches that make use of syntax do not lead to better performance. In fact, imposing syntactic restrictions on phrases, as used in recently proposed syntax-based translation models [Yamada and Knight, 2001], proves to be harmful. Our experiments also show, that small phrases of up to three words are sufficient for obtaining high levels of accuracy.

Performance differs widely depending on the methods used to build the phrase translation table. We found extraction heuristics based on word alignments to be better than a more principled phrase-based alignment method. However, what constitutes the best heuristic differs from

2005

Franz Och reports first research results reported from Google's statistical machine translation system

Impact on size of language model training data (in words) on quality of Arabic-English statistical machine translation system

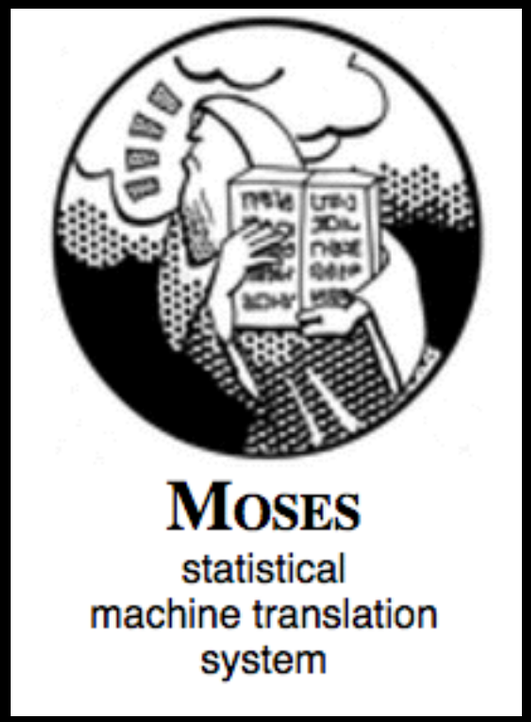


+weblm = LM trained on 219B words of web data

Google

2006

Philipp Koehn leads development of Moses machine translation toolkit at Johns Hopkins CLSP workshop



Final Report
of the
2006 Language Engineering Workshop

Open Source Toolkit for Statistical Machine Translation: Factored Translation Models and Confusion Network Decoding

http://www.clsp.jhu.edu/ws2006/groups/ossmt/

http://www.statmt.org/moses/

Johns Hopkins University
Center for Speech and Language Processing

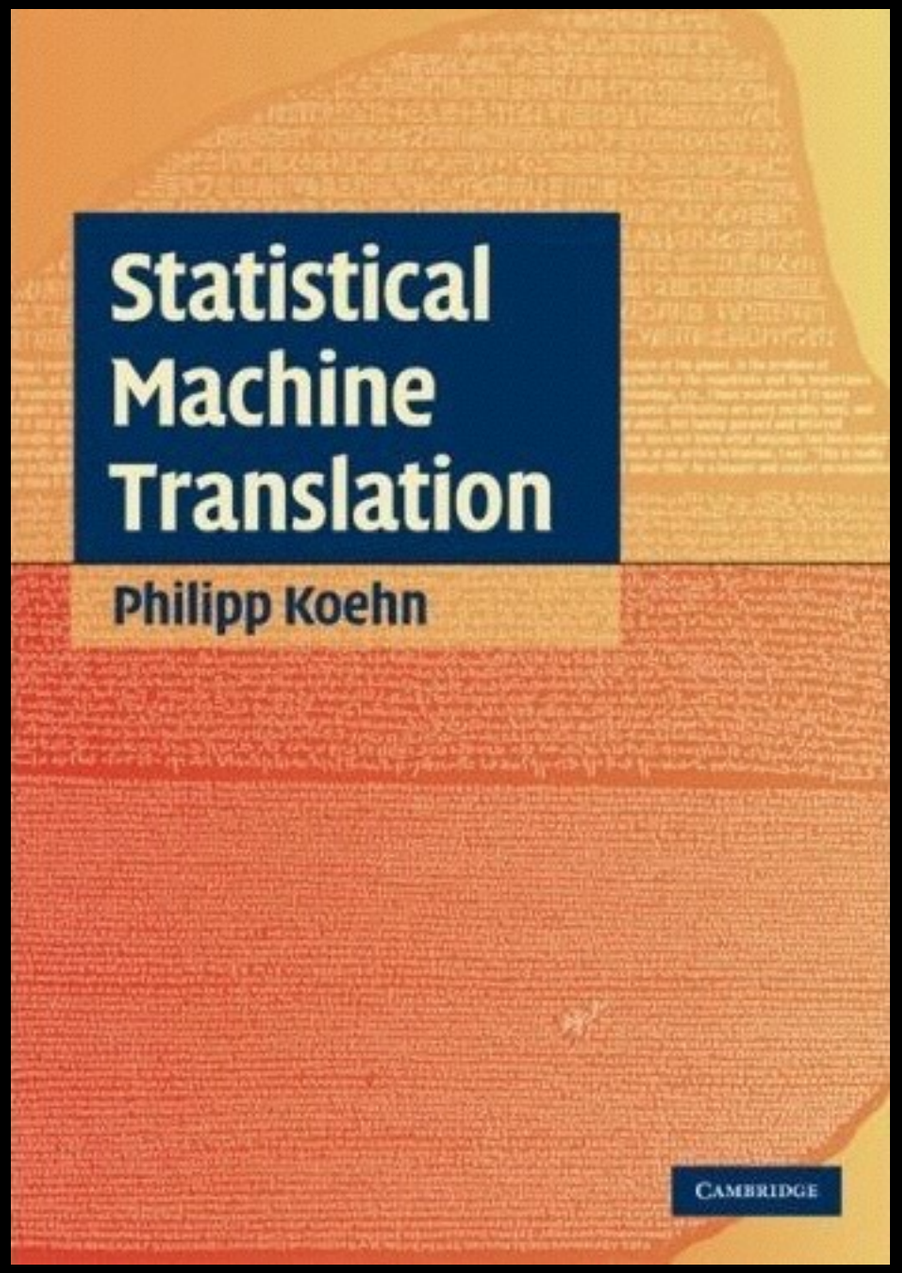
Philipp Koehn, Marcello Federico, Wade Shen, Nicola Bertoldi,
Ondřej Bojar, Chris Callison-Burch, Brooke Cowan, Chris Dyer, Hieu Hoang, Richard Zens,
Alexandra Constantin, Christine Corbett Moran, Evan Herbst

2006

28 April: Google translate goes live

2009

Philipp Koehn writes the book on statistical machine translation



2009

Fred Jelinek wins ACL lifetime achievement award in Singapore

2010

14 September
Fred Jelinek dies in Baltimore

"He was not a pioneer of speech recognition, he was **the** pioneer of speech recognition."

–Steve Young

"the BCJR algorithm (the J is for Jelinek) is a critical element in Turbo decoding. There is thus a little bit of Fred in every 3G cell phone on the planet."

–Steve Wicker

2013

18 October: “Twenty years of bitext” in Seattle

2013

18 October: “Twenty years of bitext” in Seattle

2013

18 October: “Twenty years of bitext” in Seattle

2013

20 October: first *modern* neural MT paper presented at EMNLP

Recurrent Continuous Translation Models

Nal Kalchbrenner **Phil Blunsom**
Department of Computer Science
University of Oxford
{nal.kalchbrenner,phil.blunsom}@cs.ox.ac.uk

Abstract

We introduce a class of probabilistic continuous translation models called Recurrent Continuous Translation Models that are purely based on continuous representations for words, phrases and sentences and do not rely on alignments or phrasal translation units. The models have a generation and a conditioning aspect. The generation of the translation is modelled with a target Recurrent La-

ties, linguistic or otherwise, they do not share statistical weight in the models' estimation of their translation probabilities. Besides ignoring the similarity of phrase pairs, this leads to general sparsity issues. The estimation is sparse or skewed for the large number of rare or unseen phrase pairs, which grows exponentially in the length of the phrases, and the generalisation to other domains is often limited.

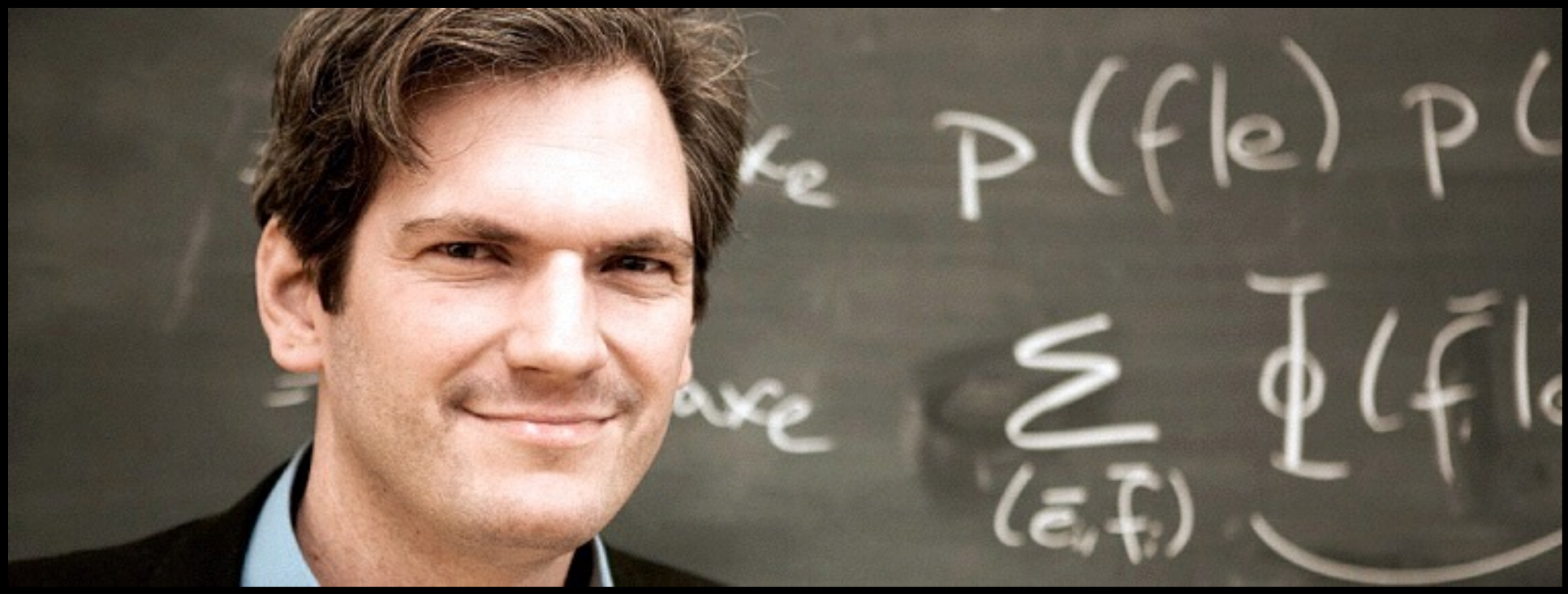
Continuous representations have shown promise

1947

"A more general basis for hoping that a computer could be designed which would cope with a useful part of the problem of translation is to be found in a theorem which was proved in 1943 by McCulloch and Pitts. This theorem states that a robot (or a computer) constructed with regenerative loops of a certain formal character is capable of deducing any legitimate conclusion from a finite set of premises."

–Warren Weaver to Norbert Wiener

2013



Philipp Koehn nominated for European inventor award for phrase-based statistical machine translation

… and accepts professorship at Johns Hopkins University

2014

Franz Och leaves Google for Human Longevity Inc.

(he now works for a cancer screening company)

2014

Bob Mercer wins ACL lifetime achievement award in Baltimore

1949

"[I]t is one of the chief purposes of this memorandum to emphasize that statistical semantic studies should be undertaken, as a necessary preliminary step. "

–Warren Weaver, "Translation"

2014

July: 1st Frederick Jelinek memorial workshop held in Prague on “Cross-lingual abstract meaning representation for MT”

2016

February: Google Translate decides to switch to neural machine translation

2016

February: Google Translate decides to switch to neural machine translation

November: Google's first neural MT systems launched.

TRANSLATE NOV 15, 2016

Found in translation: More accurate, fluent sentences in Google Translate

Barak Turovsky
PRODUCT LEAD, GOOGLE TRANSLATE

2016

February: Google Translate decides to switch to neural machine translation

November: Google's first neural MT systems launched.

The Great A.I. Awakening

How Google used artificial intelligence to transform Google Translate, one of its more popular services — and how machine learning is poised to reinvent computing itself.

BY GIDEON LEWIS-KRAUS DEC. 14, 2016

2016

The Top Early Donors in the Presidential Race

Wilks Family $15.0 million

The Wilks family of Texas — brothers Farris and Dan and their spouses, Jo Ann and Staci — earned billions in the fracking boom.

Mercer Family 11.3 million

Robert Mercer, a Wall Street hedge fund magnate, and his daughter, Rebekah Mercer.

Toby Neugebauer 10.0 million

Co-founder of a private equity firm and son of Representative Randy Neugebauer, Republican of Texas.

Kelcy Warren 6.0 million

Dallas billionaire and chief executive of an energy company.

Ricketts Family 5.1 million

Joe Ricketts founded the online brokerage TD Ameritrade and is the patriarch of the family that owns the Chicago Cubs.

Deason Family 5.0 million

Darwin Deason is a Dallas-based billionaire who made his fortune by selling a data processing company to Xerox.

Bob Mercer gets a lot of publicity

2016

Back Story

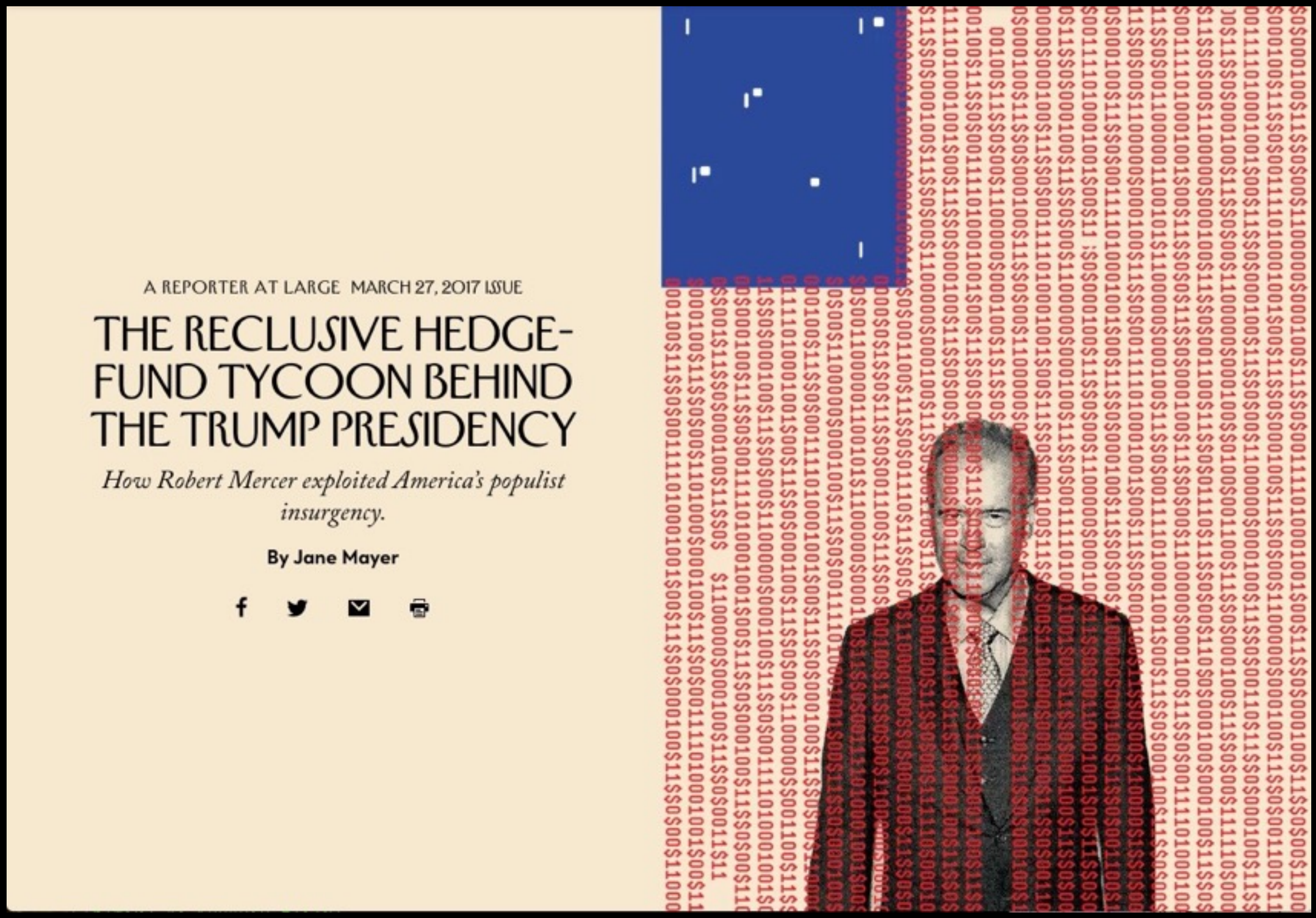
What Kind of Man Spends Millions to Elect Ted Cruz?

Jan 20, 2016 10:45 AM WET

Robert Mercer is one of the wealthiest, most secretive, influential, and reactionary Republicans in the country.

Bob Mercer gets a lot of publicity

2017



A REPORTER AT LARGE MARCH 27, 2017 ISSUE

THE RECLUSIVE HEDGE-FUND TYCOON BEHIND THE TRUMP PRESIDENCY

How Robert Mercer exploited America's populist insurgency.

By Jane Mayer

Bob Mercer continues gets a lot of publicity

2017

Currently supports 104 languages,
or 10,712 language pairs.

… a rate of nearly three per day,
since launch on 28 April, 2006.

2017

Google Translate

Currently supports 104 languages,
or 10,712 language pairs.

… a rate of nearly three per day,
since launch on 28 April, 2006.

Google plans to convert all of them to neural
MT by the end of the year.

“Research in both ASR and MT continues. The statistical approach is clearly dominant. The knowledge of linguists is added wherever it fits. And although we have made significant progress, we are very far from solving the problems. That is a good thing: We can continue accepting new students into our field without any worry that they will have to search, in the middle of their careers, for new fields of action.”

–Fred Jelinek