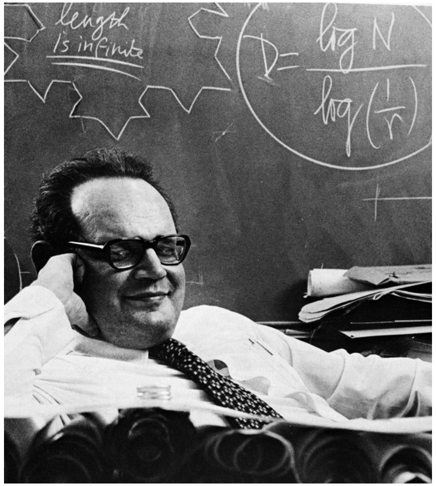
Maarten Bullynck
Dépt. de mathématiques et d'histoire des sciences
UMR 8533 IDHE.s
Université Paris 8

The crossing paths of Mandelbrot and Schützenberger. An episode of cross-overs between mathematics and computing (1953-1963).



Part I: Mandelbrot and Schützenberger in Paris

Benoît B. Mandelbrot circa 1952

Marcel-Paul Schützenberger circa 1945

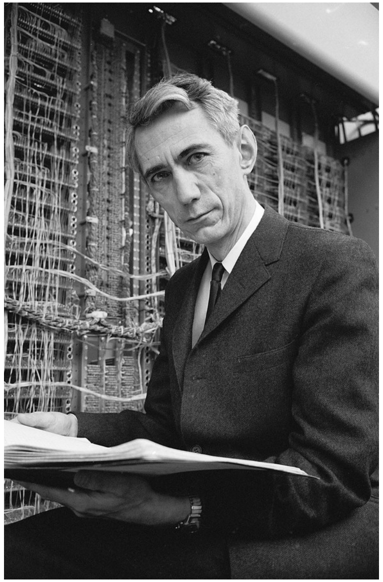
Institut Poincaré as a meeting place for probability theory in the 1950s

Since World War I, the Institut Poincaré was a key institution in research and dissemination of probability theory in France, with people like E. Borel, M. Fréchet, G. Darmois etc. and with a regular seminar with foreign guests.

After World War II, as Bourbaki gained ground in mathematics, the Institut Poincaré remained one of the (few) institutions for research and education in statistics and probability.

More info: M. Cléry, La théorie des probabilités et l'Institut Henri Poincaré (1918-1939) : construction d'un champ probabiliste parisien et pratique d'un transfert culturel. Thèse 2021.

Information theory as a new field of mathematical investigation



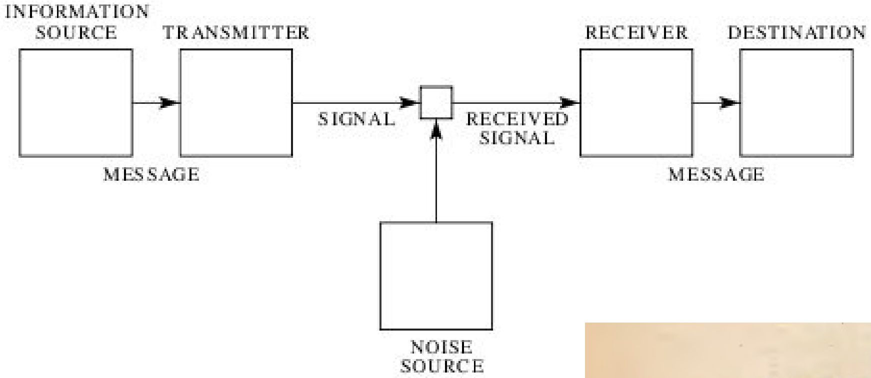
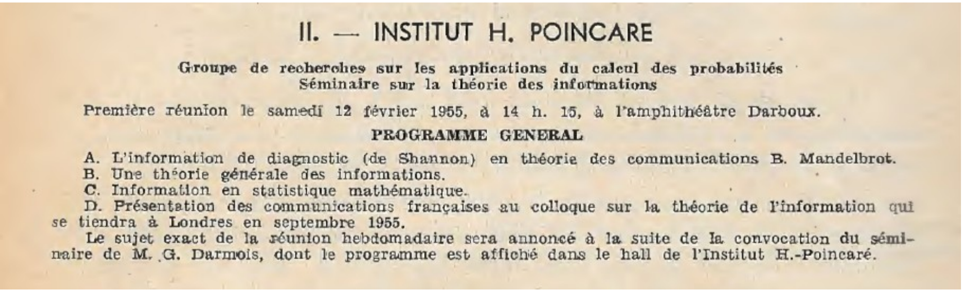


Fig. 1—Schematic diagram of a general commu



II. — INSTITUT H. POINCARE

Groupe de recherches sur les applications du calcul des probabilités
Séminaire sur la théorie des informations

Première réunion le samedi 12 février 1955, à 14 h. 15, à l'amphithéâtre Darboux.

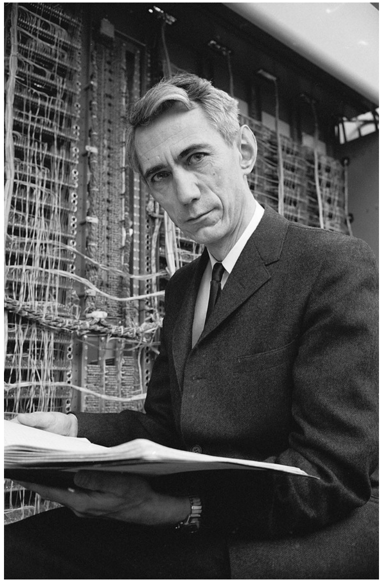
PROGRAMME GENERAL

A. L'information de diagnostic (de Shannon) en théorie des communications B. Mandelbrot.
B. Une théorie générale des informations.
C. Information en statistique mathématique.
D. Présentation des communications françaises au colloque sur la théorie de l'information qui se tiendra à Londres en septembre 1955.

Le sujet exact de la réunion hebdomadaire sera annoncé à la suite de la convocation du séminaire de M. G. Darmois, dont le programme est affiché dans le hall de l'Institut H.-Poincaré.

More info: Jérôme Ségal, Le zéro et le un; Histoire de la notion d'information au XXe siècle (2011). Axel Roch, Claude E. Shannon: Spielzeug, Leben und die geheime Geschichte seiner Theorie der Information (2010)

Classic information theoretic Model for Language



We may consider a discrete source, therefore, to be represented by a stochastic process. Conversely, any stochastic process which produces a discrete sequence of symbols chosen from a finite set may be considered a discrete Source. […] To make this Markoff process into an information source we need only assume that a letter is produced for each transition from one state to another. The states will correspond to the "residue of influence" from preceding letter"

(C.E. Shannon, A theory of Communication, 1948)

Classic Information theoretic Model for Language

3. The Series of Approximations to English

To give a visual idea of how this series of processes approaches a language, typical sequences in the approximations to English have been constructed and are given below. In all cases we have assumed a 27-symbol "alphabet," the 26 letters and a space.

1. Zero-order approximation (symbols independent and equiprobable).

   XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZL-HJQD.

2. First-order approximation (symbols independent but with frequencies of English text).

   OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL.

3. Second-order approximation (digram structure as in English).

   ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TU-COOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE.

4. Third-order approximation (trigram structure as in English).

   IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONS-TURES OF THE REPTAGIN IS REGOACTIONA OF CRE.

5. First-order word approximation. Rather than continue with tetragram, . . . , *n*-gram structure it is easier and better to jump at this point to word units. Here words are chosen independently but with their appropriate frequencies.

   REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NAT-URAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE.

6. Second-order word approximation. The word transition probabilities are correct but no further structure is included.

   THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHAR-ACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.

Mandelbrot's and Schützenberger's PhD on information theory

PUBLICATIONS DE L'INSTITUT DE STATISTIQUE DE L'UNIVERSITÉ DE PARIS

Volume 2 FASCICULES 1 et 2 1953

CONTRIBUTION A LA THÉORIE MATHÉMATIQUE DES

JEUX DE COMMUNICATION

par

BENOIT MANDELBROT

Ancien Elève de l'Ecole Polytechnique, Docteur es-Sciences
Ingénieur aux Laboratoires d'Electronique et de Physique Appliquées

THÈSE SOUTENUE LE 16 DÉCEMBRE 1952

Defended 1952

PUBLICATIONS DE L'INSTITUT DE STATISTIQUE DE L'UNIVERSITÉ DE PARIS

Volume 3 FASCICULES 1 et 2 1954

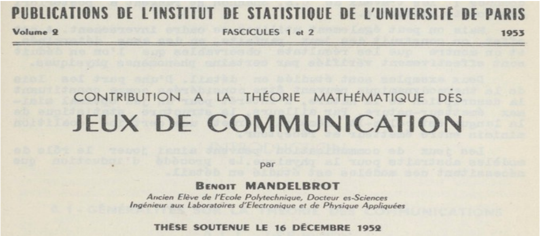
CONTRIBUTION
AUX
APPLICATIONS STATISTIQUES
DE LA
THÉORIE DE L'INFORMATION

par

M. P. SCHÜTZENBERGER

Defended 1953

Mandelbrot’s contributions to linguistic theory


PUBLICATIONS DE L’INSTITUT DE STATISTIQUE DE L’UNIVERSITÉ DE PARIS

Volume 2 FASCICULES 1 et 2 1953

CONTRIBUTION A LA THÉORIE MATHÉMATIQUE DES

JEUX DE COMMUNICATION

par

BENOIT MANDELBROT

Ancien Elève de l’Ecole Polytechnique, Docteur es-Sciences

Ingénieur aux Laboratoires d’Electronique et de Physique Appliquées

THÈSE SOUTENUE LE 16 DÉCEMBRE 1952

“I saw a golden opportunity to become the Kepler of mathematical linguistics” (Mandelbrot)

Critique of Shannon’s Model: there is exists no ideal channel without noise, there is always at least ***infinitesimal noise*** present

Transformation of the Communication Schema by introducing a ***triangular relationship between sender, channel (or nature) and receiver***, that can be studied using the theory of games
→ One particular game: sender and receiver cooperate against the channel to optimize the transmission of their messages

For the stochastic sources, Mandelbrot proposes, next to static and Markovian sources, a ***rank-frequency distribution*** for modeling language (Zipf’s law): *“les mots constituent donc les quanta naturels d'information ” (1951)*
Based on this, a noise-limiting recurrent coding is proposed that uses a space symbol

More info: Jacqueline Léon, Histoire de l’automatisation des sciences du langage (2015). English translation: Automating Linguistics (2021).

Mandelbrot’s contribution to coding theory

ON RECURRENT NOISE LIMITING CODING*

Benoit Mandelbrot**
Laboratoires d'Electronique et de Physique Appliquées,
Paris, France

Summary.
A method of word by word coding can be described by a coding tree. The study of the coding methods is equivalent to the study of their trees, considered as graphs. The number of letter strings used as codes (spellings), considered as a function of the length C of these strings, is by definition the "structure function" S(C) of the tree and of the coding method. Two coding methods having identical structure functions lead to the same cost of coding for any message, and are called equivalent. – A coding method is said to be recurrent if the decision as to whether a letter is a last letter of a word requires the knowledge of the preceding letters of this word only. The structure function of a recurrent method of coding satisfies Szilard's inequality: $\Sigma S(C) M^{-C} \leq 1$, where M is the number of letters and to any function satisfying this inequality corresponds at least one recurrent coding method. Sardinas and Patterson have shown that there are cases in which the message may be recovered from the coded string of letters, even though the identification of the ends of words requires the knowledge of the future of the message. However, the structure functions of these coding methods must still satisfy Szilard's inequality, and they can always be replaced by an equivalent recurrent coding procedure. No advantage is to be gained from non-recurrent codes using the future.

“One chooses one element to play the role of space, and defines the word as being all the sequences of the initial elements between two space symbols. The theory based upon this generation of words is conceptually parallel to Shannon's theory, but it is more useful for the description of the most important single class of statistical languages: the natural languages. […] Consider a discrete finite irreducible Markoff chain. Instead of cutting it into stretches from the outside, let it cut itself, by specializing one of the states to be spaces. […]
Let us finally remark that the crucial role which appears to be played by the symbol space, and therefore by protection against error, may be considered as completing the role which protection against noise plays in restricting language to be digital, discrete.” (Mandelbrot 1955)

Schützenberger's entry into coding theory

Séminaire Albert CHÂTELET et Paul DUBREIL
7e année : 1953/54
ALGÈBRE et THÉORIE DES NOMBRES. 2e tirage multigraphié.

MÉCANIQUE STATISTIQUE ET THÉORIE DE L'INFORMATION. — *Adaptation d message à la ligne de transmission* : I. *Quanta d'information*. Note d M. Benoit Mandelbrot, présentée par M. Jacques Hadamard.

La définition précédente fait apparaitre la théorie de codage comme une application de la théorie des demi-groupes. Il est particulièrement remarquable que les concepts fondamentaux de cette dernière, introduits par P. Dubreil [3], [4], [5] en 1941, et étudiés depuis par lui-même et son école du point de vue abstrait, aient des interprétations immédiates et importantes sur le plan de la réalisation concrète des machines codeuses ou transcodeuses.

Sous un autre angle, la théorie des événements récurrents de W. Feller [6], dans laquelle on étudie des processus stochastiques sur les suites de lettres à partir d'autres processus définis sur les suites de messages élémentaires , et réciproquement se rattache étroitement à la théorie du codage, comme l'a montré Mandelbrot qui a utilisé cette analogie dès 1951 [7], [8] : les processus récurrents sont des codes unitaires au sens que nous donnerons plus loin à ce terme. En retour, l'extension par Feller lui-même de sa théorie à des

Schützenberger’s entry into coding theory

Faculté des Sciences de Paris
-:-:-:-
Séminaire P. DUBREIL et C. PISOT
(ALGÈBRE et THEORIE DES NOMBRES)
Année 1955/1956
-:-:-:-

27 février 1956

Exposé n° 15

UNE THÉORIE ALGÉBRIQUE DU CODAGE,

par M.P. SCHÜTZENBERGER.

1956 SYMPOSIUM ON INFORMATION THEORY

held at

Massachusetts Institute of Technology
Cambridge, Massachusetts

September 10-12, 1956

INSTITUT HENRI POINCARE

SEMINAIRE DE CALCUL DES PROBABILITES

THEORIE DU CODAGE
ET DES EVENEMENTS RECURRENTS

Exposé de M.P. SCHUTZENBERGER
du 16 mars 1956

* Une théorie algébrique du codage. In Séminaire Dubreil-Pisot, année 1955-56, Exposé No. 15, 27 février 1956, 24 pages. Inst. H. Poincaré, Paris, 1956
* Théorie du codage et des événements récurrents. In Séminaire de calcul des probabilités, 16 mars 1956, 11 pages. Publ. Inst. Statist. Univ. Paris, Inst. H. Poincaré, Paris, 1956.
* Une théorie algébrique du codage. C. R. Acad. Sci. Paris, 242:862–864, 1956.
* On the application of semigroup methods to some problems in coding. IRE Trans. Inf. Theory, IT-2:47–60, 1956.

Schützenberger's entry into coding theory

Faculté des Sciences de Paris
-:-:-:-
Séminaire P. DUBREIL et C. PISOT
(ALGÈBRE et THÉORIE DES NOMBRES)
Année 1955/1956
-:-:-:-

27 février 1956

Exposé n° 15

UNE THÉORIE ALGÉBRIQUE DU CODAGE,

par M.P. SCHÜTZENBERGER.

"L et A étant respectivement les demi-groupes libres engrendrés par L_0 et A_0, une application C de L_0 sur une partie P_0 de A sera un code si et seulement si l'extension de C à L détermine un isomorphisme de L sur les sous-demi-groupe P de A engendré par les suites de lettres constituant P_0. La définition précédente fait apparaitre la théorie de codage comme une application de la théorie des demi-groupes."

Equivalence syntactique: pour tout x,y en A:
$xby \in K \Rightarrow xay \in K$ et vice versa,
alors a est équivalent à b dans A par rapport à K

Schützenberger's entry into coding theory
1956 IRE Convention

ON an APPLICATION of SEMI GROUPS METHODS
TO SOME PROBLEMS in CODING

By M.P. Schützenberger
(C.N.R.S. Paris)

It is proper at this place to acknowledge the contributions of three authors who influenced deeply the building of the theory :

Sardinas and Patterson[1] who discussed first on a logical basis the general coding process.

B. Mandelbrot[2] who recognised and studied extensively the role of "word units" in communication theory and related the problem to Feller's recurrent events.

P. Dubreuil[3] and his school whose pionnering work on discrete semi groups has provided many basic concepts and arguments as it will be seen below.

3 families of codes:

* uniform codes: all words have the same length
* unitary codes: codes in which no word is left divisor of another word (= Mandelbrot's recurrent codes or today's instantaneous or prefix-free codes)
* "natural" codes: with a special letter at the end of a word

"Mandelbrot has shown that unitary code is, at least asymptotically, as good from the point of view of economy of length as any other one." (Schützenberger 1956)

Schützenberger first visits to MIT (1956-7 + 1959)

Still much working on information theory and coding and contributing to R.S. Marcus PhD on “Discrete noiseless coding” (1957)

* On the quantization of finite dimensional messages. Information and Control, 1:153–158, 1958.

* A characteristic property of certain polynomials of E. F. Moore and C. Shannon. Quarterly Progress Report of the Research Lab. of Electronics, MIT, 66:117–118, 1959.

* Sur certains sous-demi-groupes qui interviennent dans un problème de mathématiques appliquées. Publ. Sci. Univ. Alger Sér. A, 6:85–90, 1959.

* with R. S. Marcus. Full decodable code-word sets. IRE Trans. Inf. Theory, IT-5:12–15, 1959.

* Un problème de la théorie des automates. In Séminaire Dubreil-Pisot, année 1959-60, Exposé No. 3, 6 pages. Inst. H. Poincaré, Paris, 1960.

From 1960 onwards, however, his main attention will go to obtaining theoretical results linking the different formal models for languages.

Part II “Les Passeurs” – Going to the U.S.

Benoît B. Mandelbrot circa 1955-1958

Marcel-Paul Schützenberger circa 1962

Mandelbrot moves to the U.S
Schützenberger visits the U.S.

Mandelbrot

1953-54 MIT

1954-55 IAS, Princeton

[1955-57 Lille+Paris+Geneva]

1958-87 IBM Fellow

1961-62 invited professor Harvard

Schützenberger

1956-7 MIT

1959 MIT

1960-61 MIT + Chapel Hill, NC

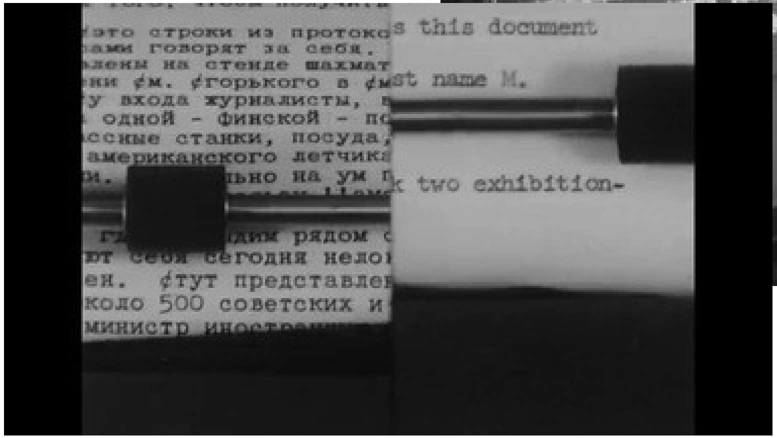
1961-62 Harvard

1962-3 IBM

The 1950s search for automatic translation

Warren Weaver’s 1949-report

- **Financial investments in automatic translation**
- **Use of digital computers**
- **→ Probabilistic approaches to language (information theory)**
- **→ Grammatical approaches to language (structural linguistics)**



More info: Jacqueline Léon, Histoire de l’automatisation des sciences du langage (2015). English translation: Automating Linguistics (2021).

MIT’S RLE Lab mechanical translation group

Experience shifts the focus

QPR 33: “*Workers in the field of mechanical translation have expressed surprise that a word-for-word translation is as good as it is. [...] Such experiments have focused attention on the error correcting nature of language.*” (Yngve, 1954)
QPR 34: “*The structure of language can be considered as a deviation from randomness; thus it is that the comparison of gap distributions actually observed with the distributions expected on a random basis can lead to information, obtained entirely by the use of statistical techniques, on the structure of language.*” (Yngve, 1954)
QPR 36: “*Our most recent effort at devising such objective methods consists of comparing a sample of the language with certain features of a simple statistical model of language in such a way that the constraints of grammar and syntax show up as deviations from randomness.*” (Yngve, 1955)
QPR 40: “*In addition, work is being done on the theory of language and on general considerations of machine capabilities and how they impinge on the syntactic problems.*“

MIT’S RLE Lab mechanical translation group: Chomsky arrives in the group

Experience shifts the focus

QPR 41: “*It may be, in fact, that every natural language can be regarded as a finite state language [...] However, when we actually attempt to construct grammars of the specified kinds for natural languages, we find that this description, though perhaps possible, is so complex that it is practically useless. Investigating the situation more closely, we find that some of the complexity is due to the presence of a large but finite number of conditions on utterances (e.g., parallelism of constructions) […] This suggests that essentially new conceptions of linguistic structure are necessary, along with more extensive methods for generating sentences from given sentences and for taking into account the history of derivation (constituent structure) of the given sentences.*” (Chomsky, 1956)

MIT'S RLE Lab mechanical translation group
the problem of nesting or embedding

(QPR 42) "If S contains an m-termed dependency set, then at least 2m states are necessary in the finite-state grammar that generates the language L that contains S. Hence, a necessary condition on finite-state languages is that there must be a finite upper limit. With this condition in mind, we can easily construct many nonfinite-state languages. […] Turning now to the English language, we find that there are infinite sets of sentences with just the mirror-image properties of L1 . For example, let S1, S2, S3 , ... , be declarative sentences. Then the following are all English sentences:

(1) (i) If S1, then S2

(ii) Either S3 , or S4

(iii) The man who said that S5 , is arriving today.

These sentences have dependencies between "if" and "then," "either" and "or," "man" and "is." But we can choose S 1, S 3 , and S 5 in (1) as (i), (ii), or (iii) themselves. Proceeding to construct sentences in this way, we arrive at sentences with dependency sets of more than any fixed number of terms, just as in the case of L1 . English is therefore not a finite-state language. […] The question of the literal possibility or impossibility of a phrase-structure description of English therefore remains open." (Chomsky 1956)

Chomsky at the 1956 IRE Convention

THREE MODELS FOR THE DESCRIPTION OF LANGUAGE*

Noam Chomsky
Department of Modern Languages and Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts

”Although we have found that no finite-state Markov process that produces sentences from left to right can serve as an English grammar, we might inquire into the possiblity of constructing a sequence of such devices that, in some nontrivial way, come closer and closer to matching the output of a satisfactory English grammar. […] there is no significant correlation between order of approximation and grammaticalness. ”

Chomsky at the 1956 IRE Convention

Three Models

1. Finite-state grammars (FSG)
2. Phrase-structure grammars (PSG)
3. Transformational grammars

Classes of counter examples:

(i) L1, ab, aabb, aaabbb ….

(ii) L2, aa, bb, abba, baab, aabbaa, bbaabb, …

(iii) L3, aa, bb, abab, baba, aabaab, baabaa, …

(i) and (ii) are terminal languages not generated by FSG, (iii) is not terminal and not generated by a PSG.

Also: Passives cannot be generated in phrase-structure grammars, therefore new transformation rules have to added to obtain transformational grammars. Only the last model also explains ambiguity of interpretation e.g. the shooting of the hunters

Noam Chomsky’s attack on probabilistic models for language

“*In short, the approach to the analysis of grammaticalness suggested here in terms of a finite state Markov process that produces sentences from the left to the right, appears to lead to a dead end […] If a grammar of this type produces all English sentences, it will produce many non-sentences as well. If it produces only English sentences, we can be sure that there will be an infinite number of true sentences etc. which it will simply not produce*.” (Chomsky 1957)

“ *In fact, there is little reason to believe that there is any interesting relation between the statistics of language use [] and the grammatical pattern. […] It would, incidentally, not be particularly surprising if statistical models turns out to be of little relevance to grammar*.” (Chomsky 1957)

“*Now, assuming that some probability model, such as we have described, did account properly for text statistics [...] either we must suppose that linguistic texts ARE actually generated by a finite-state automaton, i.e. that the micro-description underlying our macro-statistics is just this combinatorial, information-maximizing, or space-symbol randomizing process, and that macro-behavior is calculable from the detailed behavior of that process, or else we must say that the statistical properties of a text are consistent with the true but nonprobabilistic or nonstochastic micro-process, but are not to be calculated from it in the manner of the explanations under review. If the first alternative is correct, we need seek no further for the grammatical structure of texts-it is simply some matrix of conditional probabilities. It is not necessary to emphasize again in a linguistic journal why such a conception of language is absurd. We are left with the second choice*.” (Lees 1959)

Mandelbrot’s reaction

“Nous voulons revenir sur la simplification, dont nous avons dit qu’elle intervient, lorsqu’on accepte de remplacer une description microlinguistique, de type grammatical, devenue trop compliquée, par une description macrolinguistique et probabiliste. Nous avons en effet trop rapidement passé alors sur une profonde difficulté méthodologique, que comporte ce passage: ”tout simplement“, rigoureusement parlant, logique et probabilité sont, dans ce contexte, incompatibles. […] [i]l y a donc conflit entre la caractère probabiliste de la structure du signal et celui de sa transmission”

Footnote: “N. Chomsky (1956) a particulièrement insisté sur ce point important“ (Mandelbrot 1957)

“the “finite-state “ model appears as rather shocking because of the well-known existence of some long-range influences such as those studied by grammar. [...] [We propose] a [synchronic] markovian model, in which the “memory” reaches back to some origin of time [diachronic evolution]. [...] We have discussed [...] statistical and grammatical models. Strictly speaking, they are contradictory, but this does not mean that either is wrong.” (Mandelbrot 1961)

“In 1953, I gained durable praise from linguists for having shown that a straight rank-size plot for word frequencies is devoid of meaning for linguistics; there is nothing in it for syntax or semantics. However, Zipf's law proved interesting in probabilistic terms and somehow started me on a path that led, first, to finance and economics, and eventually to fractals.” (Mandelbrot 1997)

1960-1964 Automatic translation disappoints

Bar-Hillel's report (1960) and the ALPAC-report (1964) review the mechanical translation efforts very negatively

Consequences:

→ Funding disappears from automatic translation projects (though not from documentation projects)

→ Probabilistic models of language fall out of favour (until bigger memories and HMM in the 1970s)

→ Chomsky and his followers start to dominate linguistics for two generations

More info: Jacqueline Léon, Histoire de l'automatisation des sciences du langage (2015). English translation: Automating Linguistics (2021).

Building up the Chomsky hierarchy (1959-1963)

Schützenberg

* Un problème de la théorie des automates. In Séminaire Dubreil-Pisot, année 1959-60, Exposé No. 3, 6 pages. Inst. H. Poincaré, Paris, 1960.
* Some remarks on Chomsky's context-free languages. Quarterly Progress Report of the Research Lab. of Electronics, MIT, 68:155–170, 1961.
* A remark on finite transducers. Information and Control, 4:185–196, 1961.
* On the definition of a family of automata. Information and Control, 4:245–270, 1961.
* Finite counting automata. Information and Control, 5:91–107, 1962.
* On probabilistic push-down storages. In Self-Organizing Systems, Proceedings, pages 205–213. Spartan Books, Washington, 1962.
* On an abstract machine property preserved under the satisfaction relation. Technical Report NC-167, IBM Thomas Watson Research Center, 1962.
* On the minimum number of elements in a cutting set of words. Technical Report NC-173, IBM Thomas Watson Research Center, 1962.
* On a family of formal power series. 11 pages, manuscrit, mars 1962.
* Certain elementary families of automata. In Proc. Sympos. Math. Theory of Automata (New York, 1962), pages 139–153. Polytechnic Press of Polytechnic Inst. of Brooklyn, Brooklyn, New York, 1963.
* On context-free languages and push-down automata. Information and Control, 6:246–264, 1963.

Chomsky

With Miller, G. A. Finite state languages. Information and Control, 1958, 1,91-112.

On certain formal properties of grammars. Information and Control, 1959,2, 137-167. (a)

A note on phrase structure grammars. Information and Control, 1959, 2, 393-395. (b)

On the notion "Rule of grammar." In R. Jakobson (Ed.), Structure of language and its mathematical aspects, Proc. 12th Sympos. in Appl. Math. Providence, R.I.: American Mathematical Society, 1961. Pp. 6-24.

Context-free grammars and pushdown storage. RLE Quart. Prog. Rept. No. 65. Cambridge, Mass.: M.I.T. March 1962. (a)

The logical basis for linguistic theory. Proc. IXth Int. Cong, of Linguists, 1962.

with Miller, G.A. Introduction to the formal analysis of natural languages, pp. 269-322.

Formal Properties of Grammars, pp.323.418.

With Miller, G.A. Finitary models of Language Users, pp. 419-528.In: Bush, Galanter, Luce (eds.) Handbook of Mathematical Psychology, Volume II. 1963.

Noam Chomsky and Marcel-Paul Schützenberger. The algebraic theory of context-free languages. In P. Braffort and D. Hirschberg, editors, Computer Programming and Formal Systems, pages 118–161. North-Holland, Amsterdam, 1963.

Building up the Chomsky hierarchy (1959-1963)

The importance of ***pushdown storage*** (PSD) is due to the search for formal description of the syntax of ***programming languages*** around 1960

→ pushdown storage registers in the list-processing language IPL (Newell, Shaw, Simon, 1957-1960)

→ the use of Post production systems to define the syntax of ALGOL (1959)

Recognition by Rice & Ginsburg that the ALGOL syntax is equivalent to Chomsky's ***context-free grammars*** (and the context-sensitive grammars to Myhill's linearly bounded automata)

Indeed, the "algebraic" generating functions are associated with homomorphisms of F into a free group, i. e., with a special case of a so-called "push down storage." [37, 31]

In Section IV we list several problems concerning the supports which have been proved to be unsolvable.

Another presentation of this material but with a definitely different emphasis is given elsewhere by N. Chomsky and myself. [13] In fact, most of the remarks developed here (and especially the ones dealing with push down storage) are results of this collaboration over a period of many years.

Schützenberger 1962

[3] The results in this section and Sec. 4.2 are the product of work done jointly with M. P. Schützenberger. For a concise summary, see Chomsky (1962a). See Schützenberger (1962a,b,d) for generalizations and related results.

Chomsky 1963

Building up the Chomsky hierarchy (1959-1963)

"It is the special merit of the structures discovered by Kleeene and of those by Chomsky that, having been found at so many cross-roads, they are the object of so many theorems. If the most serious authors only see the utensil virtues of finite automata and of cell memories, I must remind you that their definition, as we now know it, could be the same one as for finite monoids and free groups.

[…]

Like all applications of mathematics, the theory being considered has tasks which may be regrouped as follows: to orient research by classifying the problems, by extracting the proper concepts and by unifying the arguments; to put to use the essential results accumulated by the relevant branches of mathematics; and to allow the latter to profit from a restated problematics and from intuition born of experience and of the thorough study of special cases it requires." (Schützenberger 1964)

New horizons for Mandelbrot: Error clustering

Based upon the empirical observation that errors in transmission tend to cluster, Berger and Mandelbrot at IBM developed a new model that is neither a model without memory, nor a markovian model with a finite number of transition states.

"This [markovian] approach does indeed give rise to certain qualitative feature ascribed to the data; but quantitatively the fit is poor as was shown especially clearly in Mertz. The most striking difficulty is to be found m the fact that "first-order" bursts are clustered in "second-order' bursts, and so on. One readily imagines that any finite set of empirical data can be accounted tor with arbitrary accuracy, if one agrees to introduce a enough number of hierarchical levels and hence of independent parameters. However, such models are analytically unmanageable, and include explicitly in the input all the features that they hope to obtain in their output"

"Generally speaking, the mathematical theory of coding, that is, the theory of information as understood in the strictest sense, consists in evaluating the various "pre-correcting" codes suggest by inventors, and in comparing them with an ideal of performance associated with certain probability limit theorems. The theory of information was divided by Shannon into two parts, according to the presence or absence of nois. Actually, this division is somewht of an oversimplification, because the theory of noiseless transmission is not the limit of the theory of transmisson in the presence of vanishingly small noise. […] this type of limit behavior is frequently observed in engineering." (Berger & Mandelbrot 1963)

Jay M. BERGER & M 1963. A new model for the clustering of errors on telephone circuits. IBM Journal of Research And Development: 7, 224-236.

1965c. Self-similar error clusters in communications systems and the concept of conditional stationarity. IEEE Transactions on Communications Technology: COM-13, 71-90.

1965. Ensembles de multiplicité aléatoires (Jean-Pierre Kahane & M). Comptes Rendus. Paris, FR: 262, 3931-3933.

New horizons for Mandelbrot: Coin tossing and the rediscovery of the Cantor set

"both pure and applied mathematics have so far concentrated upon chance phenomena such that something is happening much of the time, in the sense that $Pr\{|X(t) > 0\}$ *is non-zero. Moreover, different portions cut from the same function are somewhat similar in appearance to each other. It now appears that adherence to these properties excessively restricts the practical scope of the theory of random functions. It excludes many well known mathematical examples which can in no way be considered pathological, and it also appears to exclude the ill-understood processes that underlie communication errors"*

"It is a simple matter to program a computer to generate sequences of errors according to the self-similar model; its most practical consequence is therefore that it becomes a simple matter to Monte Carlo the comparison between codes." (Mandelbrot 1965)