

Markovketten - Variationen im Computer

Maarten Bullynck

Statistik & Sprache 1900 Vor 1900 streift die statistische Technik nur selten die Sprache. Nur zweimal um 1900 gab letztere Anlass für ein statistisches Beispiel, oder aber, um genauer zu sein: Zweimal wird Dichtung Material eines theoretischen Gedankenspiels in Sachen statistischer Spitzentechnologie. Einmal 1885 in *Methods of statistics*, dem zweiten Artikel einer vierteiligen Reihe, in der F. Y. Edgeworth die Fehlerbehandlung bei Wahrnehmungsdaten auch auf andere Daten anzuwenden versucht.¹ Ein ander Mal in 1913, als A. A. Markov *Démonstration du second théorème-limite du calcul des probabilités par la méthode des moments* veröffentlicht, in dem als einfachstes Beispiel abhängiger Variablen Markovketten und im besonderen Vokal-Konsonant-Verkettungen auftauchen.²

Im Gegensatz zur nachrichtentechnischen Verschmelzung von Sprache und Statistik unter dem Stichwort *Information* in der Nachfolge Shannons, gibt es bei Edgeworth und Markov eher die Zufallsbegegnung. Beide ihren Arbeiten sind Momente in der Erfindung stochastischer Variablen, d.h., abstrakter Daten die mit der Apparatur zur Fehlerminimierung kommensurabel sind.³ Beide stärken auch das zentrale Grenzwerttheorem, das alle Variablen unter dem Joch der gaussischen Normalkurve einzuebnen weiß, und nicht zuletzt zentral heißt, weil es Grund und Anfang aller wissenschaftlichen Zufallsabschätzung ist. Von Seiten des Materials, der Sprache, scheint gerade Dichtung sich zur Stärkung mathematischer Wahrheiten besonders zu eignen: Edgeworth untersucht die Verteilung von lang und kurz bei Vergil, Markov die Abhängigkeit von Vokalen und Konsonanten bei Puschkin.

Warum aber kommen gerade jene "zufälligen Fehler, die dem Calcül nicht unterworfen werden können",⁴ aber mit Kombinationslehre minimiert werden sollten, gegen 1900 in die Buchstaben an? Zum einen weil Buchstaben seit Ebbinghaus und Fechner Zufallswert, aber "in gewissen Grenzen eingeschlossen"⁵, errungen hatten.⁶ Zum anderen galt es die "bekannte Art", wie unbekannte Größen zusammenhängen - nämlich die "hypothetische" Funktion e^{-hx} - für andersartige Größen (nämlich abhängige Variablen) auch zur Geltung und Rech-

¹Vgl. Stigler 1986, S. 309.

²Markov 1913.

³Gegen 1900 umfasste diese Apparatur die kleinste Quadratenmethode und die Momentenmethode.

⁴Gauss 1821, S. 95-96.

⁵Gauss 1821, S. 96.

⁶Vgl. Kittler 1985/87, II. Teil; Stigler 1986, S. 239-261.

nung zu bringen. Das Gleichgewicht von Silben und Betonungen ergab angeblich ein schönes Beispiel für solch halbgeordnete Mengen.

Deswegen, dass gerade zu der Zeit, als de Saussure lateinische Gedichte anagrammatisch in Wiederholungen von Buchstaben zerlegte, Statistiker dieselbe Poesie, mit ihrer symmetrischer Verteilung von Klängen nach traditionsreichen Reimschemata, als gelegentliches Material heranzogen. In den Händen von Statistikern werden autorspezifische Abweichungen, die in der Rhetorik bislang als Alliteration, Konsonanz, Enjambement oder Inversion bekannt waren, alle zu "zufällige[n] Fehler[n]"⁷, und das heißt: Formalisierbar als Gelungen/Misslungen nach dem Beispiel des Münzwerfens bei Bernouilli in 1713. Deswegen die binäre Aufteilung der Sprache in kurz/lang oder Vokal/Konsonant, denn auch die Sprachreihen dienen nur dem zentralen Grenzwerttheorem, das jene Technik um das Mittelmaß begründen sollte. In diesem Rahmen - und d.h. unter der schönen gaussischen Glocke - sind die Markovketten 1913 anzusiedeln, Sprache als Randbeispiel von *hardcore* statistischer Theorie, rhetorische Figuren als Abweichungen der Wahrnehmung - seien sie auch ästhetische Fehler⁷ - im Text.

Shannon 1948 Wenn aber 1947-48 vor dem Substrat der Bell Telephonlinien Shannons *Mathematical Theory of Communication* Gestalt bekommt, kollidieren auf einmal Markovketten mit saussureschen Selektions- und Kombinationsreihen. Denn nicht nur gilt, "a discrete source [can] be represented by a stochastic process" - jene Entdeckung Edgeworths und Markovs - sondern mehr noch: "Conversely, any stochastic process which produces a discrete sequence of symbols chosen from a finite set may be considered a discrete source."⁸ Eine fast zu glatte Umkehrung, die Sprache als Grenzwert "of a series of of simple artificial languages"⁹ in Markovketten wirft. Die allmähliche Verkettung der Gedanken - nach Kleist - wird ein Markovgenerator, die Lautverschiebung unter Einfluss der Umgebung - nach J. Grimm - Übergänge im Markovprozess:

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 letters.¹⁰

Produktion von Lettern unter "branching"-Bedingungen macht die Gleichung "finite state Markoff process" mit "natural language" komplett, schaltet die poetischen Randbeispielen Markovs und Edgeworths in die neue, prosaische Wirklichkeit namens Computer ein, der somit praktischer Fokus dieser neuen und noch unstillen Gleichung wird.

Als Shannon einmal gezeigt hatte, wie Information über die logarithmische Verkopplung von Code-Elementen mit Zahlen quantifiziert werden konnte, wurden die alphabetischen Codes schnell Opfer der Arithmetisierung. Die 1953-

⁷Der Diskurs um Konvention und Abweichung bei den russischen Formalisten ist nicht zuletzt auch Erbe dieser literaturwissenschaftlichen Tradition.

⁸Shannon/Weaver 1947-48, S. 11.

⁹Ebenda, S. 13.

¹⁰Ebenda, S. 15.

Ausgabe der Zeitschrift *Language* zeugt von der vielschichtigen Rezeption Shannons in der Sprachwissenschaft. In ihr ist nicht nur der Review von Shannons Werk enthalten, sondern auch die Aufbereitung der Sprache in Übergängen von Markov-kontrollierten Phonemen durch C. Cherry, M. Halle und R. Jakobson¹¹, sowie auch eine computerorientierte arithmetische Beschreibung der Syntaxis durch Y. Bar-Hillel.¹² Ersterer Artikel ist empirisch auf die statistische Verteilung von Phonemen zugeschnitten, letzterer versucht ein "operational syntax whose instructions could be carried out satisfactorily by a digital computer"¹³ zu konstruieren.

Logik-Statistik-Computer Diese Dualität, die sich hier nach Transposition in die Ebene der natürlichen/künstlichen Sprachen wiederholt, zieht sich durch die ganze Informationstheorie und deren Derivate hindurch. Shannons Umkehrung des mathematisch-stochastischen Gedankengangs kann die Perspektive des Decodeurs, der die Botschaft mittels statistischer Analyse entziffert, in die Perspektive des Encodeurs umsetzen, der seine Symbolen mit dem stochastischen Markovprozess generiert.¹⁴ Aber wie Jakobson in seiner "Linguistik und Kommunikationstheorie" betont, dreht mit diesem Schritt die Hierarchie zwischen Encodierung "from the lexico-grammatical to the phonological level" und Decodierung "from sound to meaning" um.¹⁵

Es sind diese beiden Gangarten der Information, die Kolmogoroff in einer Übersicht aus 1965 im Begriff Information diagnostiziert. Es gibt die **kombinatorische** Seite, die "grammatische" Verbindung von Zeichen, es gibt die **probabilistische**, das Handelbarmachen großer Datenmengen.¹⁶

Historisch das erste Beispiel dieser praktischen Divergenz ist Axel Thues Arbeit über Veränderungen von Zeichenreihen, die als zeitgenössisches Pendant von Markovs Untersuchungen zur statistischen Zeichenverteilung gelten kann. Thues Frage, "if Q can be obtained from P by [repeated] replacing [of] a substring [of P]"¹⁷ verfährt zwar auch über Übergangsregeln zwischen *substrings* - aber die Frage von Thue ist "wie kommt man von P nach Q ?", während Markovs Frage als "Was liegt so alles in der Umgebung von P nach n Übergängen?" umschrieben werden kann. Um 1900 gehören beide Fragen noch unterschiedlichen Disziplinen an, nach 1945 aber stehen sie immer mehr von Angesicht zu Angesicht.

Denn drittens gibt es die **algorithmische** Gangart der Information: "Basically, the most meaningful approach has to do with the amount of information in "something" (x) about "something else" (y)."¹⁸ Dieser der Definition der Metapher so ähnlich klingende Formel hat auch ihr Mittelterm: "the program

¹¹Cherry et al. 1953: "Toward the logical description of languages in their phonemic aspect."

¹²Bar-Hillel 1953: "A quasi-arithmetical notation for syntactic description."

¹³Bar-Hillel 1953, S. 47.

¹⁴Oder auch: Einmal Statistik im Sinne Edgeworths und Markovs, einmal Stochastik im Sinne Shannons.

¹⁵Jakobson 1961, S. 249.

¹⁶Kolmogorov 1965, S. 188.

¹⁷Post 1947, S. 1.

¹⁸Kolmogorov 1965, S. 188.

p [that transforms] the object x into the object y ¹⁹ und dieser Programm p steht für nichts anderes als die Komplexität oder das Maß der *randomness* bzw. *pattern-ness*. In den Interferenzen von Markovketten und Texttheorie, wie sie sich unter dem Zeichen der Kybernetik zwischen 1953 und 1963 entwickeln würden, steht dann auch letztendlich nichts anderes als der Programm zentral, der die mathematische Idee der Markovketten computertechnisch variiert. Und um diese Variationen wiederholen sich die Fragen: Statistik? Logik?

Im Weiterdenken der markov'schen Gedanken während der 50'er Jahre haben sich auch Benoît Mandelbrot und Noam Chomsky verdient gemacht. Obwohl Mandelbrots fraktale Geometrie und Chomskys generative Grammatik heutzutage angeblich keine Verwandtschaft mehr aufweisen, waren ihre Erfinder 1955-1957 in Auseinandersetzungen verwickelt, in denen die Keime ihrer Grundideen wiederzufinden sind. Auf den ersten Blick wiederholt sich hier mit ihren Diskussionen die Verschiedenheit Markov-Thue, auf den zweiten Blick aber ist die Lage viel praktischer. Wer lässt die Programme laufen und sieht sich deren Ergebnisse erst dann an? Wer will die Programme so schreiben, dass sie ein bestimmtes Ergebnis zeitigen?

Mandelbrot 1953 In seinem 1952 angefangenen *thèse doctoral ès sciences* verband Mandelbrot die damals modische Kybernetik Wieners sowie auch die Spieltheorie von Neumanns mit der Zufälligkeit, dass er während des 2. Weltkriegs in der Mülltüte seines Onkels, des Mathematikers Szolem Mandelbrojt, eine Rezension von Zipfs statistischen Untersuchungen über Wortverteilungen gefunden hatte. Es führte zu dem merkwürdigen Werk, *Contribution à la théorie mathématique des jeux de communication*. In seiner Einführung schon ordnet Mandelbrot seine Theorie unter die neue aber nie verwirklichte Wissenschaft de Saussures ein, unter die Semiologie:

”Sémiologie“ serait donc synonyme de ”théorie des communications“. Le déplacement de l'accent de l'information au signe est d'ailleurs conforme à leurs importances relatives: l'information n'est qu'un certain nombre rattaché au signe, et l'œuvre de Shannon un instrument d'étude.²⁰

Kronstück in dieser neuartigen Semiologie war jenes Gesetz, das Mandelbrot als das erste synchrone Gesetz in der Linguistik bezeichnete und heute als das Mandelbrot-Zipf-Gesetz bekannt ist. Es formuliert die exponentielle Verteilung der Wörter in einer Sprache.²¹ Wie Mandelbrot in einem Interview selber besagte, hoffte er mit diesem Gesetz eine heliozentrische Wende zustande zu bringen.

¹⁹Ebenda, S. 192.

²⁰Mandelbrot 1953, S. 10.

²¹Das Gesetz lautet folgendermaßen::

$$p_r = P(r + \rho)^{-B} \quad \text{mit} \quad B = \log M / \log M' > 1$$

Der zentrale Exponent B ist das Quotient der Logarithmen von M und M' , wobei M' die Anzahl der ”Innen“-Buchstaben und $M-M'$ der ”Außen“-Buchstaben (d.h. Leerzeichen oder ”spaces“) ist.

Indem er die Sprache unter makrolinguistische statt mikrolinguistische Perspektive analysierte, gedachte er sogar "the Kepler of mathematical linguistics"²² zu werden.

Die Neuheit des mandelbrotschen Ansatzes bestand darin, dass er die Diskontinuitäten des Kommunikationsspiels als Eichpunkte seiner Arbeit sah und nicht sosehr die Kontinuität des Textes. Wesentlich in dieser Organisation der Zeichen unter makrolinguistische Perspektive des Kommunikationsingenieurs ist die Einführung des Leerzeichens ("space"), ein Zeichen das nicht innerhalb eines Wortes vorkommen kann. Unabhängig von den umgebenden Worten/Texten hat dieses Zeichen eine absolute Probabilität, oder einfacher gesagt: Das Leerzeichen deutet die Grenzen zwischen Worten an und ist deshalb nicht von den Nachbarbuchstaben abhängig.

Bei Shannon werden alle N Elemente oder Bits in die Markovkombinatorik geschickt, um 2^{NH} Wörter²³ zu generieren, Mandelbrot aber schlägt vor, um "instead of cutting [the Markov chain] into stretches from the outside [i.e. in telephone-line based bits], to let it cut itself, by specializing one of the states to be spaces."²⁴ Als Ergebnis wird das Wort das zentrale stochastische Intervall, das einfach als "all possible sequences of letters between two successive spaces"²⁵ definiert werden kann. Solche Stochastik entspricht der Aussage, das Wort - und nicht sosehr das Phonem - sei "un morceau naturel d'information, adapté au codage arithmétique", die nachrichtentechnische Übersetzung der saussureschen Behauptung, das Wort sei die konkrete Einheit der Sprache.²⁶

Mandelbrots Kritik an der Theorie Shannons war, dass diese eine Idealisierung der physischen Tatsachen darstellt. Weil ein Kanal ohne Störung realistischere nicht bestehe, wählt Mandelbrot einen Kanal mit Infinitesimalstörung als Ansatzpunkt seiner Ableitungen. Das Leerzeichen ist nicht mehr als die zwar *stricto sensu* redundante aber praktisch äußerst effiziente Antwort auf diese Störung: Es wahrt die Lesbarkeit der Botschaft, den kleinen Unterbrechungen jedwelcher Botschaft mit einbegriffen. In vollem Bewusstsein injiziert Mandelbrot dann auch " " in die Gleichungen der Kanäle:

Let us finally remark that the crucial role which appears to be played by the space-symbol, and therefore by protection against error, may be considered as completing the role which protection against noise plays in restricting languages to be digital, discrete. This discreteness could not, (no more than the statistics of the words) be considered as being influenced by some structure of the universe of ideas to be represented by the language. The only "explanation" for it seems to be that languages need to be relayed a very great number of times,

²²Barcellos 1984, S. 3.

²³ $H = -\sum p_i p_j \log p_{ij}$ ist der Markov-Operator.

²⁴Mandelbrot 1955, S. 213. Mandelbrot 1955 ist eine code-orientierte Fassung von Mandelbrot 1953. Es wurde während seines Aufenthalts am Princeton Centre for Advanced Studies unter Betreuung von John von Neumann geschrieben und begründete seinen Ruf im angelsächsischen Kulturkreis.

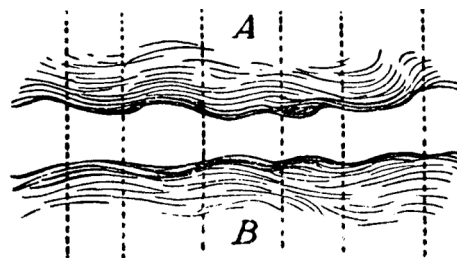
²⁵Mandelbrot 1955, S. 213.

²⁶Mandelbrot 1953, S. 17.

and that in non-discrete languages the accumulation of noise at the relays could not be resisted, and that these languages could not fulfill any aim.²⁷

Es macht auch der historischer und linguistischer Restwert der mandelbrot'schen Theorie aus, dass sie konsequent eine informationstheoretische Übersetzung von de Saussures Semiologie zu formulieren versucht. Als Beispiel kann folgender Ausschnitt aus de Saussures *Cours de Linguistique générale* gelten, in dem Mandelbrot bloß Zeichen für Ideen, Signalen für Klänge und Code für Sprache einsetzen muss - und seine Fragestellung aus 1953 wiederfindet:

Nous pouvons donc représenter le fait linguistique dans son ensemble, c'est-à-dire la langue, comme une série de subdivisions contiguës dessinées à la fois sur le plan indéfini des idées confuses (A) et sur celui non moins indéterminé des sons (B); c'est ce qu'on peut figurer très approximativement par le schéma:



Le rôle caractéristique de la langue vis-à-vis de la pensée n'est pas de créer un moyen phonique matériel pour l'expression des idées, mais de servir d'intermédiaire entre la pensée et le son, dans des conditions telles que leur union aboutit nécessairement à des délimitations réciproques d'unités.

Was de Saussure noch "ce fait en quelque sorte mystérieux, que la "pensée-son" implique des divisions"²⁸ nennen musste, wird für Mandelbrot eine Frage, die er ruhig der Statistik zuweisen kann.

Chomsky 1956 Nur ein Jahr später war das M.I.T. Symposium on Information Theory 1956 dem Thema formaler Sprachen aufgrund Transformationsregeln gewidmet. *Chairman* der Konferenz war John McCarthy, dessen Programmiersprache LISP die Implementierung von A. Churchs Kalkültheorie bot und Grundlage der *Artificial Intelligence*-Forschung wurde. Mit diesem Symposium über künstliche Sprachen brachten einerseits Noam Chomsky und andererseits Mandelbrots Landsmann und Freund Schützenberger entscheidende Beiträge zu jenen Theorien, die heute als generative Grammatik bzw. Compilertheorie bekannt sind.

²⁷Mandelbrot 1955, S. 218.

²⁸de Saussure 1916, p. 156.

Zuerst gab es das grundlegende *exposé* von Noam Chomsky, "Three models for the description of language", das die Grundideen zur generativen Grammatik in lapidarer Form enthält. In diesem Artikel beweist Chomsky, dass "no finite-state Markov process that produces symbols with transition from state to state can serve as an English grammar."²⁹ Der Beweis dazu ist erhellend, trotz seiner hoch formalen, aber mathematisch eher naiven Schlussfolgerungsweise:

2.3 Turning now to English, we find that there are infinite sets of sentences that have dependency sets with more than any fixed number of terms. For example, let S_1, S_2, \dots be declarative sentences. Then the following are all English sentences:
 (13) (i) If S_1 , then S_2 .
 (ii) Either S_3 , or S_4 .
 (iii) The man who said S_5 , is arriving today.
 These sentences have dependencies between "if-" then", "either-" or", "man-" "is". But we can choose S_1, S_3, S_5 which appear between the interdependent words, as (13i), (13ii), or (13iii) themselves. Proceeding to construct sentences in this way we arrive at subparts of English with just the mirror image properties of the languages L_1 and L_2 of (12)³⁰. English is not a finite-state language.³¹

Wie trocken und formalisiert auch, der springende Punkt ist überdeutlich vorgeführt: Markovketten können weder verschachtelte noch eingebettete Strukturen, so wie Sätze natürlicher Sprachen, generieren - Kurz: Markovketten können die hierarchischen Ebenen der Grammatik nicht erzeugen. Chomsky selbst macht die noch schärfere Aussage, "there is apparently no significant approach to the problems of grammar in this direction [i.e. statistical approximation]."³²

Mit der unendlichen Unterordnung von strukturähnlichen Nebensätzen konstruiert Chomsky nichts weniger als das grammatische Pendant zum Mengenparadox der Logik und dessen Einführung in die strukturelle Selbstreferenz. So wie z.B. Alfred Tarski "es schneit" logisch analysiert als:

ein Ausdruck, der aus zwei Worten gebildet ist, von denen das erste aus den zwei aufeinander folgenden Buchstaben: E, Es, das zweite aus den sieben aufeinander folgenden Buchstaben: Es, Ce, Ha, En, E, I, Te besteht, ist eine wahre Aussage dann und nur dann, wenn es schneit.³³

so generiert Chomsky grammatische Ebenen. Und obwohl die Logik bzw. Grammatik solche Strukturen zulässt, stehen die auswuchernden Sätze ohne Grund da - bei Tarski ohne Wahrheitsbegriff, bei Chomsky ohne Zuhörer, denn die Verweisstrukturen lassen sich nicht mehr auflösen.³⁴ Was Chomsky aber beweist,

²⁹Chomsky 1956, S. 113.

³⁰Das heißt, der Form ABCCBA.

³¹Ebenda, S. 115.

³²Ebenda, S. 116.

³³Tarski 1931, S. 454.

³⁴In Tarski 1931 wird z.B. die unendliche Klammer-Analyse von "A" ist A, in "A" ist A" ist "A" ist A, usw. als Beispiel aufgeführt, dass nicht-formale Sprachen von Widersprüchen nicht heilbar seien bzw. keinen Wahrheitsbegriff gewähren. Als Lösung konstruierte Tarski zusammen mit Lukasewicz die linkslaufende Anschreibeweise für Logik, auch bekannt als die polnische Anschreibeweise. Eine andere Parallele zu den Beispielen in Chomsky 1956 (iii) und zu den Klammermethoden in der Logik sind die Hollerith-strings in FORTRAN, die aus der

ist, dass Markovketten eine wirkende, logische Verweisstruktur nicht einmal generieren können.

Lösung für die vorgeführte Defizienz der Markovketten sind die "phrase-structure grammars", Vorläufer seiner generativen Grammatik:

In summary, then, we picture a language as having a small possibly finite kernel of basic sentences with phrase structure in the sense of §1, along with a set of transformations which can be applied to kernel sentences or to earlier transforms to produce new and more complicated sentences from elementary components.³⁵

Scheitern die markovkettenbasierten Informationsquellen in 1956, so scheitern in den nächsten Jahren auch andere Formalisierungen grammatischer Kompetenz, die Chomsky für seine regelbasierte Neubelebung von Port Royal auf die Probe nahm. Denn als Cartesianische Linguist³⁶ versteht sich Chomsky, was zur Folge hat, dass die logische Satzebene als Substrat der Linguistik dienen soll, und nicht die kommunikative Oberflächenstatistik Mandelbrots.

Schützenberger und der Compiler Retrospektiv gehören sowohl Chomskys als auch M.P. Schützenbergers Arbeiten zur entscheidenden Phase in der Entwicklung der Theorie von Programmiersprachen, und im besonderen von Compilern, die die kleinsten maschinensprachlichen Instruktionen zu "high-level" programmiersprachlichen Makrobefehlen zusammenbinden. Kritischer Auftakt der Compiler-Standardisierung war 1959-1960, als die Rapporte von J. Backus und P. Naur über eine *International Algebraic Language* erschienen und die verschiedensten Ansätze in die kurzlebige Standardsprache ALGOL60 kanalisierten.³⁷ Verdeckt von den sehr kurzgehaltenen Bibliographien dieser beiden Rapporte, ist aber das eclecticische und vielschichtige Vorspiel zu ALGOL, die vielen Verzweigungen aus den Jahren vor 1960, die nur sehr langsam aber allmählich zusammenzuwachsen begannen.

Einer der Konzepte, die in die theoretische Informatik aufgehen, ist Chomskys "context-free language"³⁸, andere sind die Automaten, die Semigruppen³⁹ usw. M.P. Schützenbergers Arbeiten waren grundlegend für die Verbindung der vielen Theorien:

Pursuing his mathematical agenda led Schützenberger to generalize his original problem and thereby to establish an intersection point

ALGOL-Sprachbeschreibung weggelassen wurden, als die Backus-Naur-Form für Programmiersprachen Standard wurde.

³⁵Chomsky 1956, p. 124.

³⁶Vgl. dazu, Chomskys *Cartesianische Linguistik* (1969), in welchem er seine geistige Verfahren erörtert, und die umfassen neben Descartes vor allem die Logik und Grammatik von Port Royal.

³⁷Siehe Backus u.A. 1959 und Naur u.A. 1960.

³⁸Aber erst nach der Einführung von ALGOL als Standard, nämlich in Ginsburg/Rice 1962.

³⁹Post 1944 und 1947 formalisierte die Schemata der Turingmaschinen als Semigruppen. Wie oben angegeben, greift Post in diesen Arbeiten Thues Problematik wieder auf. Auch A.A. Markov des Jüngeren Algorithmentheorie stützt sich weitgehend auf diesen Formalismus und Chomsky/Miller 1957 berufen sich ebenfalls auf Posts Theorie.

not only with Chomsky's linguistic agenda, but also with the agenda of machine translation and with that of algebraic programming languages. The result was the equivalence of "algebraic" formal power series,⁴⁰ context free languages, and the pushdown (or stack) automaton.⁴¹

Auch Schützenbergers 1956-Beitrag zu den *Proceedings*, "On an application of semi-group methods to some problems in coding", leistet solche Zusammenführung.

Im Gegenteil zu dem idiosyncratischen Mandelbrot hantierte Schützenberger die algebraische Methoden der Gruppe Bourbaki, Semigruppen, um Codes zu konstruieren. Seine glashellen Analysen verschiedener Codes zeigen - zwar als Nebeneffekt - mathematisch explizite die von Chomsky diagnostizierten Defizienzen der Mandelbrot-Codes auf. Als Freund Mandelbrots mit dessen Werken gut vertraut, übernimmt Schützenberger das Leerzeichen als Trennzeichen zwischen Symbolketten. Aber während es bei Mandelbrot bloß als besonderes Zeichen innerhalb normaler Zeichen fungierte, reißt Schützenberger das Leerzeichen aus der Reihe von Buchstaben und transformiert es in einen (metasprachlichen) algebraischen Operator, "the scansion symbol /" oder Querstrich.⁴² Wie beim klassischen Skandieren der Verse kennzeichnen gepaarte Querstriche eine Verszeile, ein vollständige Botschaft, ungepaart aber die Unvollständigkeit des Ungereimten.

Anhand dieses Skandierungsprozesses wird vorgeführt, wie natürliche Codes ergodische Verse produzieren, d.h. wenn es einen Fehler oder eine Störung in der Botschaft gibt, werden nur lokal die Wortgrenzen verstört, die globale Botschaft aber bleibt lesbar. Diese "noise absorption" hängt aber zusammen mit "the problem of reading "backward" messages with an inverted code."⁴³ Die Tatsache, dass eine gestörte Botschaft dank des redundanten Leerzeichens lesbar bleibt, bedeutet, dass Mandelbrots natürliche Codes einen statistischen Faktor in den Prozess des Encodierens/Decodierens bringen. Und diese statistische Unterbrechung ist einerseits dem präzisen, diskreten algebraischen Codieren fremd, und gefährdet andererseits die kryptoanalytischen Eigenschaften eines Codes, denn die Unterbrechung beim Encodieren bleibt auch dem Decodeur lesbar - was alle ergodische Codes relativ einfach decodierbar macht. Natürliche Sprachen sind also nicht invertierbar und deswegen codemäßig doppeldeutig. Weil diese Eigenschaft die gewünschte Eins-zu-Eins-Korrespondenz der algebraischen Übersetzung hindert, verbannt Schützenberger die ergodischen Mandelbrotcodes und beschränkt sich auf supersymmetrische unitäre und uniforme Codes.

Gerade die Eigenschaft, die Mandelbrot als den Vorteil der natürlichen Codes empfand, nämlich, dass "first, the correspondence between words and ideas, which each person builds for himself, is to a great extent arbitrary, and [that] there are always alternative ways of expressing ideas; second, that the choice of

⁴⁰Diese Potenzreihen zielen auf A. Churchs mit Turingmaschinen äquivalenten λ -Kalkül.

⁴¹Mahoney 1999, S. 23.

⁴²Schützenberger 1956, S. 49.

⁴³Ebenda, S. 53.

the words actually is not only governed by the need for this particular word, but also by an unconscious matching of the frequencies of words to their codes somewhere in the circuit of communication“⁴⁴, gerade diese bewegliche Adaptabilität der Wortcodes an ihre Kanäle führt zu Kurzschlüssen, wenn die Codes in eine Logik, Grammatik oder Geheimbotschaft integriert werden müssen. Demzufolge, nach Chomskys preskriptiver Perspektive, generieren Mandelbrotcodes nur Wortsalat. Nicht von ungefähr fügten sich gerade Mischsprachen, *pidgins* und Aussagen von Schizophrenen - alle mit minimalem logisch-grammatischem Apparat - am besten in Mandelbrots statistische Verteilungen ein.⁴⁵

Variationen des Computers/Variationen im Computer Nicht nur die erfolgreiche ALGOL-Fügung der unterschiedlichen Ansätze im Rahmen der *symbol manipulation* wird gegen 1960 erreicht, sondern auch die halbbetretenen Holzwege innerhalb dieses Gebiets müssen sich notgedrungen des sich standardisierenden Hauptwegs trennen und fangen ihre manchmal langwierige Emanzipation an. Neue Disziplinen wie Chaostheorie oder Komplexitätstheorie ordnen sich an die Mathematik- oder Physik Institute ein, andere werden zu historischen Curiosa.

Zu den Erfolgreichen gehört auch Chomsky. In Ginsburg/Rice 1962 werden seine kontextfreien Sprachen als der Backus-Naur-Form äquivalent und der Programmierpraxis nützlich erwiesen, überdies überlagert Chomsky nach 1956 seine Phrasenstrukturmodelle mit neuen Regeln, bis sie letztlich in die generative Grammatik einmünden.

Im gegensatz zur Markov-Grammatik [von links nach rechts] erzeugt die Phrasenstrukturgrammatik ihre Sätze ”von oben nach unten“ im Zuge einer Ableitung nach Maßgabe einer eindlichen Zahl von Regeln⁴⁶

Die Phrasenstrukturgrammatik ist eine Umformungsmaschine für Ketten, die Transformationsgrammatik ist eine Umformungsmaschine für Bäumchen - so könnte man demnach den Hauptunterschied formulieren.⁴⁷

Unter dem Schlagwort, ”we must know not only the final shape of these sentences, but also their ’history of derivation“⁴⁸, wird ”a mapping of P-markers into P-markers“⁴⁹ unternommen, oder jene oben angeführte Überführung von Bäumchen in Bäumchen.

So bleiben die Phrasenstrukturen ”als ”Basis“ der generativen Transformationsgrammatik beibehalten“, und wird ”mit einem Durchlaufen dieser Basis“⁵⁰ eine Ähnlichkeit mit der empirischen Performanz der Sprache sichergestellt.

⁴⁴Mandelbrot 1955, S. 217.

⁴⁵Vgl. Mandelbrot 1953, S. 17.

⁴⁶S. 190.

⁴⁷Hermanns 1977, S. 192.

⁴⁸Chomsky 1957, S. 37.

⁴⁹Chomsky 1961, S. XX

⁵⁰Hermanns 1977, S. 155-156.

Zweidimensional bleiben die Tiefenstrukturen, von denen "die "Oberflächenstrukturen" nur ein Derivat sind, das Informationen gegenüber den Tiefenstrukturen nichts Neues enthält."⁵¹ In dieser Art und Weise bloß noch über seine philosophischen Theorien der Kompetenz/Performanz mit der sich damals im Schatten des Computers konstituierenden Kognitionswissenschaften verbunden, erlebt seine generative Grammatik vor allem an den linguistischen Instituten riesigen Erfolg. Genauso wie Port Royal setzte Chomsky mit seinen Strukturen einen Standard für eine dem 20. Jahrhundert gemäße Grammatikbeschreibung.

Neue Horizonte muss sich aber Mandelbrot suchen. Schon in 1957 widmet er implizite dem Chomsky-Problem einen Aufsatz:

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*.⁵²

Unter Hinzufügung einer Fussnote, "N. Chomsky (1956) a particulièrement insisté sur ce point important"⁵³, wird die ganze Debatte explizite. Der Kommentar, "[i]l y a donc conflit entre la caractère probabiliste de la structure du signal et celui de sa transmission",⁵⁴ ist in der Tat auch der Konflikt inkommensurabler Sichtweisen, wie z.B. dieser der Grammatiker und Bourbaki-Algebraiker gegenüber den Ingenieuren.

Auf der Suche nach Ideen weicht Mandelbrot 1957 auf Metaphern aus, die einerseits das Problem der grammatikalischen Hierarchie in der Sprache verbildlichen und andererseits Lösungsansätze zu bieten scheinen. Zum einen die alte Metapher des Kommunikationsspiels, "mais imaginons un autre jeu, en supposant que les règles de la conversation [qui sont celles de la grammaire] ne sont elles-mêmes qu'une solution à un jeu d'ordre supérieur"⁵⁵, zum anderen eine Idee, die von Schützenberger zu stammen scheint. Es ist die Metapher von Routinen und Subroutinen im Computer: "on établit un compromis entre ces deux limites, en dressant un catalogue de sous-routines, morceaux de calcul codés à l'avance [...], à utiliser comme "briques" dans la construction de tout code."⁵⁶ Genau am diesen Punkt kommt die Biographie ins Spiel, denn ab 1957 wurde Mandelbrot als Kommunikationsingenieur am IBM Thomas J. Watson Research Centre in New York angenommen, wo er verschiedene Computer, Programme

⁵¹Hermanns 1977, S. 155-156.

⁵²Mandelbrot 1957, S. 13-14.

⁵³Mandelbrot 1957, S. 15.

⁵⁴Ebenda, S.15.

⁵⁵Ebenda, S. 46. Diese Fragestellung wurde 1965-1975 von Harvey Sacks neu aufgegriffen und teilweise gelöst, vgl. Sacks u.a. 1978. Obwohl auch dieser Aufsatz eine Reaktion gegen die chomskyanische Linguistik war, hatten Sacks' Arbeiten keine direkte Verbindung zu Mandelbrot.

⁵⁶Ebenda, S. 44.

so wie auch Programmeure zu seiner Verfügung hatte. Es würde eine Wiederholung des gleichen Ingenieursproblems veranlassen, aber jetzt in einem neuen Medium des Experiments - jenseits der Metapher.

Es sind andere ephemere Phenomene, denen sich Mandelbrot nach 1957 zuwendet. Nicht mehr Sprache, sondern Turbulenzen, Chaotik und Störungen - kleine Störungen, die er schon in seiner Dissertation in die shannon'schen Kanäle hineinpostuliert hatte. In "A new model for the clustering of errors on telephone circuits" (1963) entwirft er zusammen mit J.M. Berger ein stochastisches Langzeitgedächtnismodell, das - im Gegensatz zu den kurzgedächtnis-emulierenden Markovketten - die unvermeidlichen Störungshäufungen und -verteilungen von realen Telephonlinien befriedigend vorhersagen lässt. Wie Mandelbrot in 1965 über dieses Problem ausführt:

The most striking difficulty is to be found in 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 for with arbitrary accuracy, if one agrees to introduce a large enough number of hierarchical levels and hence of independent parameters. However, such models are analytically unmanageable, and include explicitly in their input all the features that they hope to obtain in their output.⁵⁷

Obzwar das Problem ingenieurstechnisch angeschrieben wird, kann man un schwer diese Worte genauso gut auf das Chomsky-Problem anwenden.

Nicht mit sich überlagernden Markovprozessen aber, sondern mit einer iterativen Skalierung der Intervalle von Nicht-Störung löst Mandelbrot das stochastische Modellierungsproblem. Mathematisch füsst die Skalierung auf eine Konstruktion von Georg Cantor. Die Ideen in diesem Artikel veranlassten nach Mandelbrots eigenen Worten seinen "breakthrough in the fractal world"⁵⁸, so dass sich Cantors Konstruktion jetzt als *Cantor Set* in die fraktale Geometrie allgemeiner Bekanntheit erfreuen darf. Praktischen Ursprung dieser Methode erwähnt Mandelbrot hingegen nicht, aber die Anmerkungen, "[i]t is a simple matter to program a computer to generate sequences of errors according to our [approximation models]", sowie auch "these models make it a simple matter to compare codes by the Monte Carlo method"⁵⁹, sind Klartext. Die Heuristik läuft über den Computer, das Modell ist *a posteriori* aus diesem abstrahiert worden - und mathematisch-theoretische Probleme sind für später.

Identität Symptom des theoretischen Denkens im Medium des Computers ist die Neuprägung des Gleichheitszeichens =.

Mathematicians would traditionally present the control mechanisms of algorithms informally, and the computations involved would be

⁵⁷Mandelbrot 1965, p. 170.

⁵⁸Mandelbrot 1999, S. 69.

⁵⁹Mandelbrot 1963, S. 147. Vgl. mit Fig. 2.

expressed by means of equations. There was no concept of assignment (i.e. of replacing the value of some variable by a new value); instead of writing $s \rightarrow -s$ one would write $s_{n+1} = -s_n$, giving a new name to each quantity that would arise during a sequence of calculations.⁶⁰

Just an dem Punkt, wo im Computer "aus einer "assertion" eine "instruction" wird"⁶¹, finden sich die Variationen in Hard- und Software. Wie winzig auch, dreht sich alles um die Interpretation - d.h. Manipulation - des Identitätszeichens "=". Ob Transformation oder Iteration, nach jeder Wiederholung begegnet man diesem Zeichen mit der erneuten Wissenschaft, dass gleich nur gleich ist⁶², weil es unter Befehl steht. Über Weierstrass' epsilon-delta-Kalkül verwandelt Mandelbrot ein Assignment wieder in mathematische Gleichheit um⁶³; "in der Ersetzung des Gleichstrichs [...] durch den Pfeil"⁶⁴ liegt die wesentliche Revolution, die Harris' strukturelle Grammatik in Chomskys Kalkül umsetzt. Instanzen eines "Zweiwegeverkehr[s]" zwischen Computer und Wissenschaft, die in buchstäblichstem Sinne nur eine "bloß fraktale Wiederkehr des Gleichen auf Stufen niederer Skalierung"⁶⁵ ist, aber manchmal Effekte erzeugt, und Wissen schafft.

Literatur

- [1] Backus, J.W. (Hg.) (1959): "The syntax and semantics of the proposed International Algorithmic Language of the Zürich ACM-GAMM conference." *Proceedings of the International Conference on Information Processing*, UNESCO, Paris, S. 125-131.
- [2] Barcellos, Anthony (1984): "Interview of B.B. Mandelbrot." In: *Mathematical People*. Boston.
- [3] Bar-Hillel, Y. (1953): 1953: "A quasi-arithmetical notation for syntactic description." *Language* Vol .29 Nr. 1, S. 47-63.
- [4] Cherry, C, Halle, M. und Jakobson, R. (1953) "Toward the logical description of languages in their phonemic aspect." *Language* Vol .29 Nr. 1, S. 34-46.

⁶⁰Knuth/Pardo 1977, p. 423.

⁶¹Hermanns 1977, S. 154.

⁶²Gleich ist gleich heißt in der generativen Grammatik: Oberflächenstruktur wiederholt Tiefenstruktur. In der fraktalen Geometrie gesteht man der Binsenwahrheit *omne comparatio claudicat* einige Rechte ein und nennt "gleich" dann abgeschwächt "self-similar".

⁶³Aus:

$$\Pr\{U/h_1 \geq u \mid U/h_1 > 1\} = \Pr\{U/h_2 \geq u \mid U/h_2 > 1\}$$

macht Mandelbrot $\Pr\{U \geq u\} = (u/\epsilon)^{-D}$ (Mandelbrot 1965, p. 173). Und umgekehrt kann man aus der Konditionalbedingung, die $h_{1,2}$ in ϵ umsetzt, und aus der "Absorption" der Indices wieder einfach auf ein *loop* mit *conditional jump* schließen.

⁶⁴Hermanns 1977, S. 154.

⁶⁵Kittler 1998, S. 129.

- [5] Chomsky, Noam (1956): "Three Models for the Description of Language." *IRE Transactions on Information Theory* Vol. 2 Nr. 3, S. 113-124.
- [6] Chomsky, Noam und Miller, George A. (1958): "Finite State Languages." *Information and Control*, Nr. 1, S. 91-112.
- [7] Gauss, Carl Friedrich (1821): "Anzeige der *Theorie Combinationis observationum erroribus minimis obnoxiae, pars prior.*" *Göttingische gelehrte Anzeigen*, 26. Februar 1821.
- [8] Ginsburg, S. und Rice, H.G. (1962): "Two families of languages related to ALGOL." *Journal of the ACM*, Vol. 9, S. 350-371.
- [9] Hermanns, Fritz (1977): *Die Kalkülisierung der Grammatik*. Heidelberg.
- [10] Jakobson, Roman (1961): "Linguistics und Communication Theory." In: *Structure of Language and its Mathematical Aspects*, hg. v. R. Jakobson, S. 245-252. Providence, Rhode Island.
- [11] Kittler, F.A. (1985/1987): *Aufschreibesysteme 1800/1900*. München.
- [12] Kittler, F.A. (1998): "Hardware, das unbekannte Wesen." In: *Medien Computer Realität*. Hg. v. S. Krämer. Frankfurt-am-Main.
- [13] Knuth, D. und Pardo, L. (1977): "Early Development of Programming Languages." *Encyclopedia of Computer Science and Technology*, Vol. 7. Hg. v. Belzer, Holzmann and Kent. New York, Basel.
- [14] Kolmogoroff, A. A. (1958): "To the definition of algorithms." In: *Selected Works III*, S. 62-85.
- [15] Kolmogoroff, A. A. (1965): "Three approaches to the notion of amount of information." In: *Selected Works III*, S. 184-193.
- [16] Mandelbrot, Benoît B. (1953): *Contribution à la théorie mathématique des jeux de communication*. (Thèse doctoral). Paris.
- [17] Mandelbrot, Benoît B. (1955): "On recurrent noise limiting coding." *Proceedings of the Symposium on Information Networks*, Vol. III, S. 205-222. Hg. von E. Weber. Brooklyn, N.Y.
- [18] Mandelbrot, Benoît B. (1957): "Linguistique statistique macroscopique." In: *Logique, Langage et Théorie de l'Information*, S. 1-78. Hg. v. Apostel, Mandelbrot und Morf. Paris.
- [19] Mandelbrot, B and Berger, J.M. (1963): "A new model for the clustering of errors on telephone circuits." *IBM Journal of Research and Development* 7, pp. 224-236. Reprinted in: Mandelbrot 1999, S. 132-165.
- [20] Mandelbrot, Benoît B. (1965): "Self-similarity and conditional stationarity." *IEEE Transactions on Communication Technology: COM-13*, pp. 71-90. Reprinted in: Mandelbrot 1999, S. 166-205.

- [21] Mandelbrot, Benoît B. (1999): *Multifractals and 1/f Noise*. New York, Berlin ...
- [22] Markov, A. A. (1913): *Démonstration du second théorème-limite du calcul des probabilités par la méthode des moments*. Sankt-Petersburg.
- [23] Naur, Peter (Hg. von) (1960): "Report on the Algorithmic Language ALGOL 60." *Communications of the ACM*, X, S. 299-318.
- [24] Post, Emil (1944): "Recursively enumerable sets of positive integers and their decision problems." *Bull. Amer. Math. Soc.* Vol. 40, pp. 284-316.
- [25] Post, Emil (1947): "Recursive unsolvability of a problem of Thue." *Journal of Symbolic Logic*, Vol. 12, Nr. 1, S. 1-11.
- [26] Sacks, Schegloff and Jefferson (1978): "A simplest systematics for the organization of turn taking in conversation." In: *Studies in the organization of conversational interaction*, hg. v. Schenkein, S. 7-56. New York.
- [27] de Saussure, Fernand (1916/1967): *Cours de linguistique générale*. Hg. v. T. de Mauro. Paris.
- [28] Schützenberger, M.P. (1956): "On an application of semi-group methods to some problems in coding." *IRE Transactions on Information Theory*, September 1956.
- [29] Shannon, Claude E. and Weaver, Warren (1949): *A mathematical theory of communication*. In: *Collected Papers*, hg. v. Sloane and Wyner, S. 5-83. New York: 1990.
- [30] Stigler, Stephen M. (1986): *The history of statistics. The measurement of uncertainty before 1900*. Cambridge (MA) and London (UK).
- [31] Thue, Axel (1914): "Probleme über Veränderungen von Zeichenreihen nach gegebenen Regeln." *Skrifter utgiv av Videnskapsselskapet i Kristiania*, I. Matematisk-naturvidenskabelig klasse, Nr. 10.
- [32] Tarski, A. (1930-1931): "Der Wahrheitsbegriff in den formalisierten Sprachen." In: *Logik-Texte*, Hg. v. Berka, Kreiser. Berlin.