Andrej A. Markov

Berechenbare Künste
Mathematik, Poesie, Moderne

Herausgegeben von
Philipp von Hilgers und Wladimir Velminski

diaphanes

Gedruckt mit freundlicher Hilfe der
Geschwister Boehringer Ingelheim Stiftung für Geisteswissenschaften
in Ingelheim am Rhein

1. Auflage
ISBN 978-3-935300-69-8

www.diaphanes.net

Satz und Layout: 2edit, Zürich
Druck: Stückle, Ettenheim

Inhalt

Vladimir A. Uspenskij

Die revolutionäre Bedeutung von A. A. Markovs Untersuchungen zur Buchstabenalternierung in literarischen Texten

Die Tatsache, dass die Erforschung von statistischen und wahrscheinlichkeitsbezogenen Gesetzmäßigkeiten innerhalb von Texten für die philologische Analyse von zentraler Bedeutung ist, ist eine Banalität.[1]

Man kann von den Lexemen über die grammatischen Fälle bis hin zu ganzen syntaktischen Redewendungen und rhythmischen Konstruktionen jedes beliebige Detail eines Textes einer statistischen Wahrscheinlichkeitsanalyse unterziehen. Der einfachste Fall ist wohl derjenige, wenn die Statistik auf einzelne Buchstaben und ihre Zusammenstellung in einem Text angewendet wird.[2]

Das erste und bekannteste Beispiel für eine solche Untersuchung ist wohl der 1913 in den Schriften der Akademie der Wissenschaften zu St. Petersburg erschienene Aufsatz des Akademiemitglieds Andrej Andreevič Markov d. Ä. (1856-1922) mit dem Titel *Beispiel einer statistischen Untersuchung am Text »Evgenij Onegin« zur Veranschaulichung der Zusammenhänge von Proben in Ketten.* Der Text wurde auf der Sitzung der physikalisch-mathematischen Abteilung der Akademie vom 23. Januar 1913 vorgetragen.[3]

In seiner Veröffentlichung nahm sich Markov Aleksandr Puškins (1799-1837) Poem *Evgenij Onegin* (1823-1831) vor und schrieb zunächst die ersten zwanzigtausend Buchstaben des Textes ab.[4] Dann fand er auf experimentellem Wege die Wahrscheinlichkeit heraus, mit der ein zufällig ausgewählter Buchstabe dieses Textes ein Vokal bzw. ein Konsonant ist.[5] Die Aufeinanderfolge von Vokalen und Konsonanten in Puškins Text bestimmte Markov wie folgt: Die Wahrscheinlichkeit, dass ein Buchstabe ein Vokal ist, hängt davon ab, ob vor ihm ein

1. Statistische Charakterisierungen von Texten können z. B. als Basis für Autorschaftsbestimmungen dienen. Vgl. den Beitrag von Wladimir Velminski in diesem Band 101ff.
2. Einer wichtigen Idee Kolmogorovs zufolge muss die Untersuchung der gesetzmäßigen Wahrscheinlichkeiten von Texten unbedingt der Untersuchung ihrer künstlerischen Verfahren (Kunstgriffe) vorangehen, weil man ansonsten versucht sein könnte, eine solche unveränderliche statistische Gesetzmäßigkeit für einen Kunstgriff zu halten. Kolmogorov/Prochorov 1964, 75-94; Kolmogorov 1968. Es ist klar, dass sich das Erscheinen einer bestimmten rhythmischen Konstruktion in einem in Jamben geschriebenen russischen Text dann nicht als Kunstgriff bezeichnen lässt, wenn diese Konstruktion allgemein den russischen Jambus auszeichnet, wenn also diese Konstruktion in jedem in Jamben geschriebenen russischen Text auftaucht. Wenn, wie Michail L. Gasparov gezeigt hat, der Gebrauch des Anapäst für eine bestimmte Schaffensperiode des Dichters Osip Mandelštam allgemein typisch ist, dann kann man sein Auftauchen in dem Gedicht »Verse über einen unbekannten Soldaten«, das aus derselben Periode stammt, wohl kaum für einen literarischen Kunstgriff halten. Vgl. Gasparov 2001.
3. Man könnte also das Markov gewidmete Symposium, das im April 2003 in Berlin stattfand und dessen Ergebnisse in dem hier vorliegenden Band veröffentlicht werden, als eine Art Widmung an den 90. Jahrestag jener Akademie-Sitzung verstehen.
4. Siehe Markovs »Berechnungen zu ›Evgenij Onegin‹« in diesem Band 41ff.
5. Zusammen ergeben diese beiden Wahrscheinlichkeiten eine Einheit, so dass es ausreicht, eine von ihnen zu finden.

Vokal oder ein Konsonant steht. Man kann natürlich auch die Abhängigkeit von noch früheren Buchstaben verfolgen, wie dies Markov in seinem Aufsatz auch tut. Zu Beginn des Textes beschreibt der Autor sein Verfahren folgendermaßen:

> Unsere Untersuchung basiert auf einer Folge von 20 000 russischen Buchstaben aus A. S. Puškins Versroman »Evgenij Onegin« – mit Ausnahme (des Härte und Weichheitszeichens) ъ und ь – die das gesamte erste Roman-Kapitel und sechzehn Strophen des zweiten umfassen *(обнимают)*. Die Folge ergibt 20 000 zusammenhängende Proben – jede weist entweder einen Vokal oder einen Konsonanten auf.
>
> Demzufolge nehmen wir die Existenz einer unbekannten, beständigen Wahrscheinlichkeit p eines Buchstabens an, der ein Vokal ist. Diese annähernde Größe der Zahl p versuchen wir zu bestimmen, indem wir die auftauchenden Vokale und Konsonanten addieren. Abgesehen von der Zahl p finden wir zudem, ebenfalls durch genaues Beobachten, eine Annäherung von Größen der beiden Zahlen p_1 und p_0 sowie die vier Zahlen $p_{1,1}$, $p_{1,0}$, $p_{0,1}$, $p_{0,0}$, die folgende Wahrscheinlichkeiten darstellen:
>
> p_1 – Vokal folgt Vokal,
> p_0 – Vokal folgt Konsonanten,
> $p_{1,1}$ – Vokal folgt zwei Vokalen,
> $p_{1,0}$ – Vokal folgt einem Konsonanten, dem ein Vokal vorangeht,
> $p_{0,1}$ – Vokal folgt einem Vokal, dem ein Konsonant vorangeht und
> $p_{0,0}$ – Vokal folgt zwei Konsonanten.
>
> Diese Bezeichnungen stimmen mit denen überein, die ich bereits in dem Aufsatz *Über einen Fall von Untersuchungen, die in einer Kette zusammengefasst sind* verwendet habe; bei Vergleichen mit dem Artikel *Untersuchung eines bemerkenswerten Falles abhängiger Proben* muss man p_0 und p_2 gleichsetzten. Die Wahrscheinlichkeit, dass Buchstaben Konsonanten sind, bezeichnen wir durch den Buchstaben q mit den gleichen Indizes.[6]

Markovs statistische Ermittlung der hier aufgeführten Wahrscheinlichkeiten geschieht in drei Etappen. Zuerst wird eine Hypothese aufgestellt. Es wird angenommen, dass die beobachtete Realität den Gesetzen der Wahrscheinlichkeit unterliegt und dass es demzufolge möglich ist, über diese uns unbekannte Wahrscheinlichkeit zu sprechen. Dies kann zum Beispiel die Wahrscheinlichkeit eines Vokals nach einem Konsonanten sein. Danach wird die Häufigkeit ausgerechnet, mit der der entsprechende Fall eintrifft.[7] Zuletzt wird die ausgerechnete Häufigkeit als Annäherungswert für die gesuchte Wahrscheinlichkeit übernommen. Aufgrund dieses Schemas findet Markov folgende Wahrscheinlichkeiten für die Buchstaben in *Evgenij Onegin*:

> 1) Für die Wahrscheinlichkeit, dass ein beliebig dem Text entnommener Buchstabe ein Vokal ist: $p \approx 0{,}432$; für einen Konsonanten: $q \approx 0{,}568$
> 2) Für die Wahrscheinlichkeit, dass der entnommene Buchstabe einem Vokal folgt und selbst ein Vokal ist: $p_1 \approx 0{,}128$; für einen Konsonanten nach einem Vokal: $q_1 \approx 0{,}872$

6. Markov, in diesem Band 75.
7. Mit dem Ziel, die Häufigkeit der gesuchten Wahrscheinlichkeit anzunähern, kann die Häufigkeit gewissen gemittelten Verfahren unterzogen werden, auf die an dieser Stelle nicht näher eingegangen werden kann.

3) Für die Wahrscheinlichkeit eines Vokals nach einem Konsonanten: $p_0 \approx 0{,}663$; für einen Konsonanten nach einem Konsonanten: $q_0 \approx 0{,}337$

4) Für die Wahrscheinlichkeit eines Vokals nach zwei aufeinander folgenden Vokalen: $p_{1,1} \approx 0{,}104$

5) Für die Wahrscheinlichkeit eines Konsonanten nach zwei aufeinander folgenden Konsonanten: $q_{0,0} \approx 0{,}132$

6) Für die Wahrscheinlichkeit eines Vokals nach zwei Konsonanten: $p_{0,0} \approx 0{,}868$.[8]

Markovs Untersuchungen stellen in Russland und vermutlich auf der ganzen Welt einen der ersten Versuche dar, die Mathematik auf die Analyse von literarischen Texten anzuwenden. Markovs Aufsatz ist außerdem das wohl früheste Beispiel für den Versuch, einen (russischen) literarischen Text zum Objekt der Statistik und Wahrscheinlichkeitstheorie zu machen. In der Geschichte der Wissenschaft war es bis dahin noch nicht vorgekommen, dass ein mathematischer Begriff seine erste Anwendung in der Philologie fand. Man kann daher annehmen, dass Markov durch die Beobachtung der Aufeinanderfolge von Buchstaben in real geschriebenen Texten in einer natürlichen Sprache zu seiner Auffassung kam. Dies ist umso plausibler, als Markov noch im selben Jahr (1913) eine analoge Untersuchung am Beispiel des Romans *Obryv* von Ivan Aleksandrovič Gončarov (1814-1873) und *Die Kindheit des Enkels Barkov* von Sergej Timofeevič Aksakov (1791-1859) durchführte, und zwar diesmal anhand einer Buchstabenmenge von nicht 20 000, sondern 100 000.

Wahrscheinlichkeitsbezogene, gesetzmäßige Prognosen werden auf der Basis von statistischen Erhebungen durchgeführt, die sich auf die Vergangenheit beziehen. Wenn diese Prognosen richtig hergeleitet sind, treffen sie mit einem hohen Grad an Präzision zu, wenn die Zukunft zur Vergangenheit wird, und verwandeln sich dann in statistische Daten.[9] Der Begriff der Statistik und derjenige der Wahrscheinlichkeit sind eng miteinander verbunden, unterscheiden sich aber bezüglich der Zeit, auf die sie sich beziehen. Die Statistik richtet sich auf die Vergangenheit; sie konstatiert die zahlreichen Besonderheiten von bereits beendeten Ereignissen, und zwar vor allem deren Häufigkeit. Als Beispiel könnte man die Häufigkeit nennen, mit der ein bestimmtes Wort in einem von uns gelesenen Text vorkommt. Die Wahrscheinlichkeitstheorie dagegen ist auf die Zukunft gerichtet: Sie berechnet unsere Erwartungen in Bezug auf noch kommende Ereignisse, und zwar vor allem bezogen auf deren Häufigkeit. Ein Beispiel hierfür wäre die Häufigkeit eines bestimmten Wortes in einem Text, den wir erst zu lesen beabsichtigen.

Der Begriff der Häufigkeit gehört der Statistik an, während derjenige der Wahrscheinlichkeit der Wahrscheinlichkeitstheorie zuzuordnen ist. Beide werden durch eine Zahl ausgedrückt, die zwischen 0 und 1 liegt, wobei auch die Zahlen 0 und 1 selber auftreten können. Zwischen Häufigkeit und Wahrscheinlichkeit existiert dennoch ein wesentlicher Unterschied. Häufigkeit ist eine

8. Markov gibt die Wahrscheinlichkeiten $p_{1,0}$ und $p_{0,1}$ nicht an, obwohl er dies an früherer Stelle ankündigt.

9. Vgl. Kolmogorov 1987.

objektive Besonderheit des untersuchten Materials; sie zeigt den Anteil derjenigen Fälle an, in denen eine für uns interessante Erscheinung aufgetreten ist. Wenn z. B. in einem Hain mit 500 Bäumen 400 Birken wachsen, dann ist die Häufigkeit, mit der in diesem Hain eine Birke auftritt 400:500, d. h. 0,8. Wenn in einer Reihe von untersuchten Texten mit einem durchschnittlichen Umfang von 400 000 Wörtern das Wort »fast« 204-mal auftritt, dann ist die Häufigkeit dieses Wortes in diesem Textkonvolut 204:400 000, d. h. ungefähr 0,0005.[10]

Anders liegt der Fall bei der Wahrscheinlichkeit. Sie ist ein abstrakter Indikator der Intensität, mit der ein Ereignis erwartet wird. Es macht natürlich keinen Sinn, bei *jeder* Art von Ereignis über diese Erwartung zu sprechen, geschweige denn, sie in Zahlen auszudrücken. So wäre es zum Beispiel unsinnig, darüber nachzudenken, wie hoch die Wahrscheinlichkeit ist, dass der Leser dieser Zeilen, wenn er bis zu dem Wort *niest* vordringt, tatsächlich selber niest. Die Schwierigkeit hier ist weniger die Frage, wie man diese Wahrscheinlichkeit zu bewerten hat, sondern vielmehr die Frage, was man hier überhaupt unter Wahrscheinlichkeit zu verstehen hat.

Wenn wir von der Wahrscheinlichkeit eines Ereignisses sprechen, dann übernehmen wir automatisch die Hypothese, dass eine solche Wahrscheinlichkeit auch wirklich existiert. Die praktische Bedeutung der Wahrscheinlichkeit besteht in erster Linie darin, dass man sie als a-priori-Bewertung einer erwarteten Häufigkeit betrachten kann. Wenn wir also annehmen, dass in dem schon erwähnten Wald die Wahrscheinlichkeit des Auftretens einer Birke 0,8 beträgt, dann nehmen wir diese Größe als die erwartete Häufigkeit an und hoffen, dass auf einem Grundstück mit 200 Bäumen im Durchschnitt 160 Birken vorkommen. Wenn uns dagegen die Häufigkeit schon bekannt ist, die Wahrscheinlichkeit aber noch nicht, dann können wir die Häufigkeit als einen Richtwert für die gesuchte Wahrscheinlichkeit verwenden.

Auf eben diese Weise, d. h. mit Hilfe eines statistischen Experiments, ermittelt auch Markov die Wahrscheinlichkeiten in seinen Untersuchungen zu Puškins *Evgenij Onegin*. Markov hat diese Untersuchungen keineswegs für nebensächlich gehalten, sondern er maß ihnen ganz im Gegenteil große Bedeutung bei. Dies wird von zwei Umständen bekräftigt. Erstens übernahm Markov sein Experiment in die dritte und vierte Ausgabe seiner berühmten Monographie über die Wahrscheinlichkeitsrechnung.[11] Zweitens wandte er seine Untersuchungen auch auf andere Texte an. So lesen wir in seinem Buch *Zur Wahrscheinlichkeitsrechnung* (1913) Folgendes: »Während dieses Buch gedruckt wird, stellte ich eine

10. Ein weiteres Beispiel: Folgt man Tab. 8, S. 95, dann tritt in Puškins Gedicht »Die Fontäne von Bachtschisoraj« der vierfüßige Jambus, bei dem die Betonungen auf allen Versfüßen außer dem ersten liegen, mit einer Häufigkeit von 0,038 auf. In »Evgenij Onegin« beträgt diese Häufigkeit 0,067. Wenn wir also bei dem entsprechenden Gedicht die Anzahl der Verse, in denen der vierfüßige Jambus auftritt, durch die Gesamtzahl der Verse des Gedichts teilen, dann erhalten wir mit Hilfe einer Näherung auf der dritten Stelle nach dem Komma die genannten Werte 0,038 und 0,067.

11. So erscheinen auf den Seiten 363, 365 und 366 in der dritten und auf den Seiten 570, 572 und 573 in der vierten Auflage seines Buches zur Wahrscheinlichkeitsrechnungen die uns bereits bekannten Werte für die 0,432, 0128 und 0,663 für die Wahrscheinlichkeiten p, p_1 und p_2. Mit p_2 bezeichnet Markov in seiner Monographie die Wahrscheinlichkeit eines Vokals nach einem Konsonanten, d. h. jene Wahrscheinlichkeit, die er 1913 als p_0 bezeichnet hatte.

ähnliche Untersuchung anhand von Sergej Aksakovs *Die Kindheit des Enkels Barkov* an. Die Ergebnisse dieser Untersuchung, die sich auf insgesamt 100.000 Buchstaben bezog, sind auf den folgenden Tafeln dargestellt.«[12] Daraufhin gibt Markov die von ihm ermittelten Näherungswerte für die Wahrscheinlichkeiten eines Vokals in dieser Textprobe mit $p = 0{,}44898$; $p_1 = 0{,}147$; $p_2 = 0{,}695$ an.[13]

Man kann die Bedeutung von Markovs Aufsatz aus methodologischer, mathematischer, informationswissenschaftlicher und philosophischer Perspektive betrachten. Methodologisch demonstriert Markovs Arbeit die Möglichkeiten der Mathematik für die Analyse von Textstrukturen in natürlichen Sprachen, und zwar vor allem in Bezug auf literarische Texte. Darüber hinaus gibt es Grund zu der Annahme, dass wir es hier zum ersten Mal mit dem seltenen Fall zu tun haben, dass ein linguistisches Objekt zur Grundlage für das Erscheinen eines neuen mathematischen Begriffes wird. Dieser neue mathematische Begriff ist die *Markovkette.* Als *Markovkette* wird eine Folge von aufeinander folgenden Ereignissen bezeichnet, in der die Wahrscheinlichkeit jedes Ereignisses durch das ihm unmittelbar vorangehende Ereignis bestimmt wird. Die mathematische Idee einer Markovkette erscheint in Markovs Publikationen schon im Jahre 1907, obwohl Markov hier noch nicht von Ketten spricht, sondern lediglich von einem »bemerkenswerten Fall abhängiger Versuche«. Allerdings fehlen in den Texten aus dem Jahre 1907 noch die konkreten Beispiele, so dass der Begriff hier eine rein theoretisch-abstrakte Konstruktion bleibt. Erst in seinem Aufsatz zu *Evgenij Onegin* demonstriert Markov das Konzept einer Markovkette zum ersten Mal an einem konkreten Beispiel aus der Literatur. Heute gilt der Begriff *Markovkette* als ein Zentralbegriff der Wahrscheinlichkeitstheorie.[14]

Aus der Perspektive der Informatik stellt Markovs Forschung den ersten erfolgreichen Versuch dar, strikte Gesetzmäßigkeiten in einem als Serie diskreter Signale verstandenen Text zu entdecken.

Der philosophische Aspekt von Markovs Revolution schließlich liegt darin, dass er schriftliche Texte in einer natürlichen Sprache nicht nur als Träger eines semantischen Inhalts betrachtet, sondern erstmalig als Objekte der physikali-

12. Zum Wort »Buchstaben« fügt der berühmte Autor eine Fußnote ein, die es verdient hier aufgeführt zu werden. Sie lautet: »Die Untersuchung wurde an einem von mir abgeschriebenen Text durchgeführt, der sich etwas vom Original unterscheidet aufgrund der eingeschlichenen Fehler, die aber aufgrund ihrer geringen Anzahl und ihrer Unvorsätzlichkeit auf die Schlussfolgerung keine Auswirkung haben dürften. In der ersten Untersuchung habe ich sehr viel Zeit und Arbeit darauf verwendet, solche Fehler zu beseitigen. Die Rechnungen sind in beiden Fällen mit großer Sorgfalt durchgeführt.« Markov 1924, 331.

13. p, p_1 und p_2 geben die Wahrscheinlichkeiten des Erscheinens eines Vokals an, und zwar a) an einem beliebigen Ort, b) nach einem Vokal und c) nach einem Konsonanten. Der Leser kann die Werte dieser Wahrscheinlichkeiten für Puškins und Aksakovs Roman miteinander vergleichen und dann daraus die entsprechenden Schlüsse ziehen.

14. Markovketten und vor allem ihre Verallgemeinerung, die Markovprozesse, deren allgemeine Theorie und Klassifikation Kolmogorov in den 30er Jahren aufstellte, kommen auf breiter Front sowohl in der Naturwissenschaft als auch in der Technik zum Einsatz. Vgl. Kolmogorov 1968, 145–167. Doch die erste ausführliche Anwendung der Markovketten war mit der schönen Literatur verknüpft. Es ist nicht ausgeschlossen, dass selbst der Begriff Markovkette aus Markovs Beobachtung von Buchstabenfolgen in literarischen Texten stammt. Wenn diese Hypothese richtig ist, haben wir es hier mit einem bemerkenswerten Beispiel dafür zu tun, wie die Analyse eines Textes zur Entstehung eines wichtigen mathematischen Begriffes führen kann.

schen Welt, mit ihrer eigenen Struktur, eigenen Charaktereigenschaften und Gesetzmäßigkeiten.

Kann man Markovs Forschungen wirklich als erste Anwendung mathematischer Methoden auf die Sprachforschung betrachten? Eben haben wir Markovs Arbeiten vorsichtig als »eines der ersten Beispiele« einer solchen Anwendung bezeichnet. Diese Zurückhaltung erklärt sich daraus, dass Markov gewisse Vorgänger hatte, wie zum Beispiel die russischen Versspezialisten Lev Polivanov und Andrej Belyj.[15] Der Literaturwissenschaftler Michail Gasparov schreibt Folgendes über die Arbeiten beider zur Evolution des russischen Versrhythmus: »Im Jahre 1892 zählte Polivanov die Häufigkeit bestimmter Arten von Zäsur im sechsfüßigen Jambus bei 27 Dichtern. Andrej Belyj zählte die *Betonungsauslassungen* (пропущенные ударения) im vierfüßigen Jambus bei 20 Dichtern.«[16] Gasparov deutet an, dass ähnliche Zählungen auch schon früher durchgeführt wurden, allerdings nicht beim russischen Vers, sondern anhand antiker Verse.[17]

Ich neige dazu, Untersuchungen dieser Art der Vorgeschichte mathematischer Methoden in der Sprachwissenschaft zuzuordnen, während die wahre Geschichte dieser Methoden mit Markov beginnt. Denn die Verslehre beschäftigt sich mit mündlichen Texten und behandelt den geschriebenen Text lediglich als schriftliche Fixierung eines mündlichen. Bei Markov treffen wir dagegen zum ersten Mal auf den Fall, dass zur Erforschung der Sprache ein rein mathematisches Modell benutzt wird. Die Rolle dieses Modells übernimmt dabei die Markovkette.

Oft wird mit Recht auf die prinzipielle Begrenztheit mathematischer Modelle verwiesen. Unter Begrenztheit wird dabei die Unfähigkeit dieser Modelle verstanden, das von ihnen beschriebene Phänomen in seiner Gesamtheit zu umfassen, und zwar vor allem dann, wenn es sich um ein linguistisches Phänomen handelt. Es wäre aber ungerecht, in dieser Begrenztheit eine Schwäche zu sehen. Vielmehr muss sie als eine Stärke gelten. Ein mathematisches Modell muss einfach sein und ist daher grob. Ich möchte hierfür ein Beispiel anführen. Jedermann weiß, dass die Erde eine Kugel ist. Menschen mit einem höheren Bildungsstand wissen darüber hinaus, dass die Erde ein Ellipsoid ist, der sich um seine eigene Achse dreht und an seinen Polen eingedrückt ist. Ein Geodät weiß ferner, dass die Erde ein Geoid ist, d. h. eine geometrische Figur, deren Oberfläche – wenn man von kleinen Details wie z. B. Bergen absieht – mit der Oberfläche der Erde genau übereinstimmt.[18] Wir haben es hier mit drei mathematischen Modellen zu tun, die mit ansteigender Genauigkeit ein von ihnen modelliertes Objekt beschreiben, nämlich die Form des Planeten Erde. Das wichtigste dieser drei Modelle ist das erste, obwohl es zugleich das ungenaueste ist.

15. Vgl. den Beitrag von Julia Kursell in diesem Band 137ff.
16. Gasparov 2001, 34.
17. Gasparov 2001, 34.
18. Genauer gesagt stimmt die Oberfläche des Geoiden mit jener Oberfläche überein, die ein weltbedeckender Ozean gebildet hätte, wenn alle Kontinente und Inseln unter Wasser versunken wären, oder, genauer gesagt, wenn alle Kontinente und Inseln bis auf das Niveau dieses Ozeans zugeschnitten wären.

Die Rolle eines mathematischen Modells für den Linguisten kann man mit einem Skelett vergleichen, das einem Künstler hilft, einen Menschen zu zeichnen. Der Künstler stellt nicht das Skelett dar, denn dieses bleibt ihm wie auch dem Betrachter des Bildes verborgen.[19] Um die Figur des Menschen richtig darzustellen, ist es dennoch nützlich, sie sich als ein skelettartiges Gestell vorzustellen, das mit Fleisch umgeben ist. Wenn wir *Begrenztheit* so verstehen wie hier beschrieben, dann sehen wir, dass auch der Begriff der Markovkette, definiert als mathematisches Modell realer Texte, einigermaßen begrenzt ist. Diese Begrenztheit erweist sich zumindest in den folgenden beiden Fällen.

Erstens erschöpft sich die Struktur eines Textes nicht einfach in der Aufeinanderfolge von Buchstaben oder gar in deren Eigenschaften, obwohl letztendlich ein Modell ja immer nur eine Seite eines erforschten Phänomens beschreibt. Zweitens hängt in realen Texten die Wahrscheinlichkeit des Eintretens eines bestimmten Buchstabens nicht nur vom *unmittelbar* vorangehenden Buchstaben ab, sondern auch von dem Buchstaben, der dem unmittelbar vorangehenden unmittelbar vorangeht, und sogar von noch früheren Buchstaben – im Extremfall vom gesamten vorangehenden Textabschnitt. Dies hat Markov sehr wohl verstanden und führt in seinem Artikel die Wahrscheinlichkeit für einen auf zwei Vokale folgenden Vokal und einen auf zwei Konsonanten folgenden Konsonanten an. Aus der Sicht der Informatik ist die Einbeziehung dieser, wie auch der noch ferneren Vergangenheit von entscheidender Bedeutung.

Das Modell der Markovkette ist von großer Schlichtheit. Es versteht sich von selbst, dass man durch die Komplizierung des Modells, d. h. durch den Übergang von der Abhängigkeit von nur einem vorangehenden Buchstaben zur Abhängigkeit von einer größeren Anzahl von Buchstaben, genauere Modelle erhält. Diese genaueren Modelle stellen eine Weiterentwicklung der Markovkette dar. Markov selber hat auf eine solche Entwicklung hingewiesen.[20]

Warum hat Markov in seinem Experiment mit *Evgenij Onegin* nicht die Aufeinanderfolge der Buchstaben selber untersucht, sondern nur ihre Binarität? Dafür gibt es einen technischen, einen mathematischen und einen ideologischen Grund. Der technische Grund liegt darin, dass es zu Beginn des 20. Jahrhunderts, als es noch keinen Computer gab, sehr mühselig und vielleicht unmöglich gewesen wäre, die Aufeinanderfolge *aller* Buchstaben des Alphabets zu errechnen, ohne sich auf die Eigenschaft »Vokal oder Konsonant« zu beschränken. Der mathematische Grund ist die Tatsache, dass bei einem Textumfang von 20.000 Buchstaben die experimentell ermittelten Wahrscheinlichkeiten erheblich genauer sind, wenn nur zwei aufeinander folgende Elemente in Betracht gezogen werden und nicht alle Buchstaben des Alphabets. Der ideologische Grund schließlich liegt in der Tatsache, dass Markov die Anwendbarkeit seiner mathematischen Konstruktion an einem einfachen und anschaulichen Beispiel zeigen wollte; ein Beispiel, das mit nur zwei aufeinander folgenden Elementen operiert, ist von allen denkbaren Beispielen das einfachste.

19. Kolmogorov 1961.
20. Vgl. den Beitrag von Wladimir Velminski in diesem Band 101ff.

Mathematisch gesprochen haben wir es bei Markovs Beispiel mit einer Markovkette mit zwei Zuständen zu tun. Ein wirklich großes mathematisches Modell kann immer anhand eines sehr einfachen Beispiels illustriert werden. Wir können uns die Sache so vorstellen, dass Markov das gesamte russische Alphabet auf die beiden (Quasi-)Buchstaben Vokal und Konsonant reduziert. Indem er nun jeden Buchstaben des Textes durch den ihm entsprechenden Quasi-Buchstaben (Vokal/Konsonant) ersetzt, untersucht er die Aufeinanderfolge dieser beiden Quasi-Buchstaben.

Ich möchte nun zum mathematischen Aspekt der Markovschen Entdeckung übergehen. Das Konzept der Markovkette ist eine Verallgemeinerung der Vorstellung von der Aufeinanderfolge unabhängiger Ereignisse. Das einfachste Beispiel einer solchen Aufeinanderfolge ist eine Serie zufälliger Münzwürfe, bei der das Ergebnis jedes einzelnen Wurfes nicht von den ihm vorangehenden Würfen abhängt. Wenn eine Münze zehnmal hintereinander auf Zahl fällt, nimmt die Wahrscheinlichkeit, dass sie das nächste Mal auf Kopf fällt, nicht aus diesem Grunde zu, während sie für Zahl entsprechend abnimmt. Statt dessen bleiben die Wahrscheinlichkeiten immer dieselben. Wenn die Münze echt ist, dann sind beide Wahrscheinlichkeiten – Zahl oder Kopf – gleich groß und haben jeweils denselben Wert, nämlich einhalb.

Die Wahrscheinlichkeit, dass bei einer echten Münze zehnmal Zahl geworfen wird, ist freilich kleiner als ein Tausendstel. Wenn wir andererseits annehmen, dass von 9000 Menschen ein jeder zehn Würfe macht, dann wird die Wahrscheinlichkeit, dass kein einziger von ihnen zehnmal Zahl wirft, verglichen damit noch geringer. Eine analoge Situation entsteht, wenn man einen Affen vor eine Schreibmaschine setzt und ihn damit beauftragt, einen aus kyrillischen Buchstaben bestehenden russischen Text zu schreiben, den er weder verstehen noch sprechen kann. Wenn wir annehmen, dass der Affe jeden Anschlag unabhängig vom vorausgehenden Anschlag macht, dann erscheint jeder Buchstabe des geschriebenen Textes mit eben derjenigen Wahrscheinlichkeit, die für den entsprechenden Buchstaben charakteristisch ist, und zwar unabhängig davon, welcher Buchstabe vorher getippt wurde.

Bei real existierenden russischen Texten sieht die Sache dagegen ganz anders aus. In solchen Texten sind bestimmte Verbindungen von Buchstaben schlicht unmöglich. So ist z. B. die Wahrscheinlichkeit, dass in einem russischen geschriebenen Wort nach einem Vokal das so genannte »harte Zeichen« auftaucht, gleich null. In realen Texten beeinflusst jeder einzelne Buchstabe die Wahrscheinlichkeit, mit der ein bestimmter anderer Buchstabe auf ihn folgt.

Wenn Informatiker eine Aufeinanderfolge von Signalen untersuchen, taucht häufig die Frage auf, ob es möglich sei, ein erwartetes Signal auf der Grundlage ihm vorangehender Signale vorherzusagen. In der Regel geht es dabei nicht um eine absolut richtige Vorhersage, sondern um eine Vorhersage von lediglich relativer Präzision. Wenn das erwartete Signal nicht von den vorher eingegangenen Signalen abhängt, führt die Kenntnis der vorangegangenen Signale nicht zur Verbesserung der Qualität der Vorhersage.[21] Im Falle eines Textes in einer natürlichen Sprache, in dem als Signale Buchstaben bzw. deren Eigenschaften auftreten, hängt die Wahrscheinlichkeit des Auftretens eines erwarteten Signals dagegen von den Signalen ab, die ihm vorangehen. In diesem Falle erlaubt es die Kenntnis der Vorgeschichte des Prozesses, dessen weiteren Verlauf mit einer bestimmten Wahrscheinlichkeit vorauszusagen. Bei der Markovkette ist die Vorgeschichte auf ein unmittelbar vorangehendes Signal begrenzt, wiewohl die Vorhersage desto aussagekräftiger und genauer wird, je länger der von uns berücksichtigte Zeitraum der Vorgeschichte ist.

Eine weitere wichtige Aufgabe für den Informatiker ist die Komprimierung einer Signalfolge, indem man sie durch eine andere, kürzere Folge so ersetzt, dass dabei keine Information verloren geht. Schon die Kenntnis der »absoluten«, von dem vorherigen Buchstaben unabhängigen Wahrscheinlichkeiten einzelner Buchstaben erlaubt eine solche Komprimierung. Man muss sich dabei vor Augen halten, dass jeder Buchstabe nach dem Muster des Morsealphabets von einer Kette doppelter Zeichen codiert wird. Der Informatiker codiert nun die Buchstaben, die oft vorkommen, mit kürzeren Ketten, und die Buchstaben, die relativ selten vorkommen, mit längeren. Dabei kann es in seltenen Fällen vorkommen, dass manche Texte (vor allem solche, die aus seltenen Buchstaben bestehen) nicht kürzer, sondern länger werden. Solche Fälle sind jedoch sehr selten; im Allgemeinen wird eine Komprimierung erzielt. Eine größere Komprimierung wird freilich erreicht, wenn man die Wahrscheinlichkeiten für alle möglichen Paare benachbarter Buchstaben kennt oder, was auf dasselbe hinausläuft, die bedingte Wahrscheinlichkeit des Auftretens eines Buchstabens, von dem man weiß, welcher Buchstabe ihm unmittelbar vorangeht. Wenn auch nur ein Buchstabe es erlaubt, mit einer Wahrscheinlichkeit, die nicht gleich Null ist, den auf ihn folgenden Buchstaben vorauszusagen, dann bedeutet das, dass der erste Buchstabe eine Information über den zweiten enthält. Man kann daher die Information, die in dem Paar aufeinander folgender Buchstaben steckt, kürzer anschreiben, als mit zwei Buchstaben.[22] Wenn die benachbarten Buchstaben voneinander abhängig sind, lässt der Text als ganzes eine zusätzliche Komprimierung zu. Hierbei gilt ebenso wie beim Problem der Vorhersage, dass der Text umso stärker (im Idealfall maximal) komprimiert werden kann, je länger der in Betracht gezogene Zeitraum ist.

Auf dem bisher Gesagten gründen sich einige zentrale Konzepte Claude Shannons und Andrej Kolmogorovs. Shannon hat den Begriff der Sprachentropie als quantitativen Messwert für die Flexibilität einer Sprache eingeführt.[23] Dieser Wert erlaubt es, die Entropie einer Sprache auf der Grundlage von Experimenten zur Vorhersage der Fortsetzung eines Textes zu bewerten. Kolmogorov hat sich mit ähnlichen Experimenten befasst und angeregt, die in einem Objekt enthaltene Menge an Informationen als Länge der kürzest möglichen Beschreibung dieses Objekts zu verstehen, und zwar mit Hilfe von adäquaten Mitteln der

21. Kolmogorov 1987.
22. Natürlich handelt es sich hier um eine Vereinfachung. »Kürzer anschreiben« bedeutet in diesem Falle nicht »eineinhalb Buchstaben«. Es geht vielmehr um die Komprimierung.
23. Vgl. Shannon 1963.

Beschreibung, die Kolmogorov »optimal« nennt.[24] Für den Fall, dass das informationstragende Objekt ein Text ist, fällt die Suche nach der kürzest möglichen Beschreibung mit seiner maximalen Komprimierung zusammen.

Abschließend bleibt noch der philosophische Aspekt der Markovschen Entdeckung zu beschreiben. In der Geschichte des menschlichen Wissens von der Welt tauchen immer wieder neue Erkenntnisobjekte auf. Als Beispiele könnte man die Entdeckung neuer Planeten, Tier- und Pflanzenarten anführen. So wurde 1846 Neptun entdeckt, und 1930 Pluto. Beide Objekte gehören zu einer bereits bekannten, älteren Kategorie, derjenigen der Planeten. Aus diesem Grund hatten diese Entdeckungen keine philosophische Bedeutung. Philosophische Bedeutung könnte man höchstens dem Umstand zuschreiben, dass die Existenz beider Planeten schon früher vorausgesagt wurde, im Falle Plutos sogar ganze 15 Jahre vor seiner eigentlichen Entdeckung. Anders verhält es sich mit neuen, zuvor nicht bekannten Arten von Objekten, wie den Elementarteilchen, den Viren oder den schwarzen Löchern. Die Entdeckung der ersten Elementarteilchen (Elektron 1897, Proton 1919, Neutron 1932) hat die schablonenhafte Vorstellung von der Unteilbarkeit des Atoms widerlegt, die in der griechischen Etymologie des Wortes »Atom« enthalten war. Die Entdeckung der Viren in den neunziger Jahren des 19. Jahrhunderts führte zur Revision der traditionellen Ansicht, dass die Zelle der kleinste Träger des Lebens sei. Im 20. Jahrhundert verbreitete sich zunehmend die Einsicht in die Existenz schwarzer Löcher, kosmischer Objekte mit der paradoxalen Eigenschaft, dass jede Information über die Vorgänge, die sich innerhalb der Grenzen dieser Objekte abspielen, diese Grenzen nicht überschreiten kann. Die gnoseologischen Konsequenzen dieser Eigenschaft schwarzer Löcher sind offensichtlich: Sie widerlegt alle naiv-materialistischen Dogmen bezüglich dessen, was wir über die Welt wissen können.

Wirkliche Umwälzungen im philosophischen Verständnis der Wirklichkeit produzieren diejenigen Entdeckungen, die uns dazu zwingen, Objekte und Begriffe, die uns bekannt erscheinen, aus einer neuen Perspektive zu betrachten. In einem solchen Fall kommt es zu einem Wechsel der mentalen Koordinaten, bei dem sich der Sinn der verwendeten Wörter plötzlich ändert. Besonders prägnante Beispiele hierfür bieten die Relativitäts- und die Quantentheorie. Beide Theorien machen Begriffe und Vorstellungen sinnlos, die lange als selbstverständlich und unanfechtbar galten. So entzieht die Relativitätstheorie der etablierten Auffassung von der Gleichzeitigkeit den Boden. Es zeigt sich, dass die Behauptung, dass zwei Ereignisse sich an verschiedenen Orten gleichzeitig bzw. nicht gleichzeitig abspielen nur dann Sinn hat, wenn man festlegt, von welchem Beobachterstandpunkt aus diese Behauptung aufgestellt wird. Beide Ereignisse können vom Standpunkt des einen Beobachters aus zeitgleich sein, während sie es von dem des anderen aus nicht sind. Die Quantentheorie führt paradoxerweise dazu, dass es bei einigen physikalischen Phänomenen keinen Sinn macht, die Werte gleich mehrerer ihrer Parameter – d. h. der Parameter in ihrer Gesamtheit – zu bestimmen. Dabei gilt, dass die Bestimmung des Wertes jedes

24. Kolmogorov 1987, 124-167; Vgl. auch Uspenskij 2002, 734.

einzelnen Parameters durchdacht ist. Hierin besteht das berühmte Prinzip der Unschärferelation.

In den zitierten Beispielen aus der Relativitäts- und Quantentheorie wird der uns so vertraute Begriff des Betrachters einer Umcodierung unterzogen. Das Vorhandensein eines Betrachters ist nunmehr eine ebenso unabdingbare Eigenschaft eines physikalischen Phänomens wie dessen raum-zeitliche Determinanten. In der Physik des 20. Jahrhunderts tritt der Beobachter nicht als subjektiver Erzähler eines Ereignisses auf, sondern als dessen nicht hintergehbarer Teilnehmer. Im Fall des Ereignisses »Gleichzeitigkeit« leuchtet die Beobachterrolle unmittelbar ein. Bei der Unschärferelation wird diese Rolle hingegen erst dann deutlich, wenn wir uns vor Augen halten, dass gemäß der Quantenphysik der Prozess der Beobachtung eines Ereignisses Einfluss auf dieses Ereignis selber hat. Dieser Umstand hat beträchtliche Folgen für jeden Versuch, ein Ereignis »objektiv« zu beschreiben, und erzeugt solche ungewohnten Gesetze wie das Prinzip der Unschärferelation.

Diese Beispiele aus der Physik sollten assoziativ den mentalen Koordinatenwechsel beleuchten, zu dem Markovs Untersuchungen geführt haben. Deren Innovation besteht wie gesagt darin, dass hier zum ersten Mal der schriftliche Text mit seinen charakteristischen Eigenschaften als diskretes, materielles Objekt zum Gegenstand einer unabhängigen Untersuchung wurde. Ähnlich wie die physikalischen Theorien des 20. Jahrhunderts die Rolle des Beobachters verändert haben, so verändert sich nach Markov die Rolle des schriftlichen Textes. Vor Markov hielt man einen geschriebenen Text für nicht viel mehr als ein Mittel zur Fixierung eines mündlichen Textes, und den mündlichen Text als ein Mittel zur Fixierung von Sinn. Nach Markov erscheint hingegen der schriftliche Text als ein selbstständiges Objekt mit seinen eigenen räumlich-geometrischen Charaktereigenschaften.

Aus dem Russischen von Sven Spieker und Wladimir Velminski

Literatur

Gasparov (2001): Гаспаров, М.Л.: *Русский стих начала XX века в комментариях*, Москва.

Kolmogorov (1961): Колмогоров, А.Н.: Автоматы и жизнь: Тезисы доклада // *Машинный перевод и прикладная лингвистика.* Вып. 6, 3–8.

Kolmogorov (1968): Колмогоров, А.Н.: Пример изучения метра и его метрических вариантов // *Теория стиха*, Ленинград, 145–167.

Kolmogorov (1987): Колмогоров, А.Н.: *Теория информация и теория алгоритмов*, Москва.

Kolmogorov/Prochorov (1964): Колмогоров, А.Н. / Прохоров, А.В.: О дольнике современной русской поэзии (Статистическая характеристика дольника Маяковсково, Багрицкого, Ахматовой) // *Вопросы языкознания.* № 1, 75–94.

Markov (1913): Марков, А.А.: *Исчисление вероятностей.* СПб.
Markov (1924): Марков, А.А.: *Исчисление вероятностей.* Москва.
Markov (1951): Марков, А.А.: *Избранные труды. Теория чисел. Теория вероятностей.* Ленинград.
Shannon (1963): Шеннон, К.: *Работы по теории информации и кибернетике*, Москва.
Uspenskij (2002): Успенский, В.: *Труды по НЕматематике*, Москва.