*Science in Context* **19**(4), 561–589 (2006). 
doi:10.1017/S0269889706001062 Printed in the United Kingdom

# Chains to the West: Markov's Theory of Connected Events and Its Transmission to Western Europe*

*David Link*

*Academy of Arts and the Media, Cologne*

## Argument

At the beginning of the twentieth century, the Russian mathematician Andrey A. Markov extended the laws of the calculus of probability to trials that were dependent on each other, purely theoretically. In two articles from 1913, which are presented here in English translation, he applied his theory for the first time to empirical material, namely text. After a presentation of Markov's methods, results, and possible inspirations, the introduction investigates the dissemination of his ideas to Western Europe and North America in detail. The experimental application of his method to various types of text finally determines its scope.

Since World War II, the concept of Markov chains has played an important role in the theoretical and the applied branches of modern science. These include, but are not confined to, the calculation of the dynamics of gases, liquids, and radioactive material in statistical physics, the measurement of the amount of information contained in text in communication theory, pattern-recognition, and more specifically speech- and optical-character-recognition. In recent years, they have also proven to be valuable in genetic sequencing. Among the more popular applications are the page-ranking algorithm of the internet search-engine Google and the T9 text-input system for mobile phones. Some of the applications will be discussed in more detail in section IV.

Just one example may be briefly described here to demonstrate the fundamental change Markov introduced into probability theory. Francis Galton invented the "quincunx" in 1877 as a pedagogic illustration of the fundamental laws that govern the statistics of independent events. After the concept of dependent trials had been transmitted to Western Europe, Günther Schulz from the Institute for Applied Mathematics in Berlin found out in 1934 that he could approximate the empirical distributions on the board much better if he assumed the events were internally

* *Science in Context* transliterates Russian names and words according to the BGN/PCGN system of 1947. This system generally corresponds to the latest version of rules prescribed by the Library of Congress, Washington, D.C.

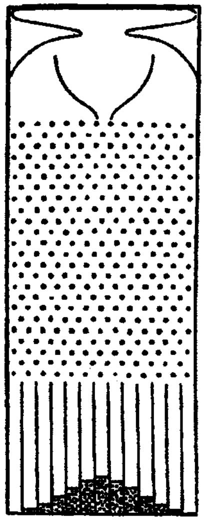


**Fig. 1a.** Galton's drawing of his board.

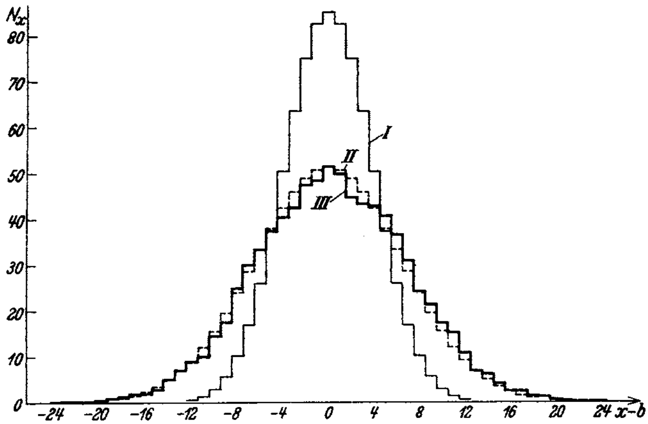



**Fig. 1b.** Schulz's results of the comparison of the empirical distribution on the quincunx (curve III) with the theoretical expectations for dependent (II) and independent trials (I).

connected. A tool that had regularly illustrated probability theory for over fifty years was shown to follow laws falling outside of classical stochastics! (see fig. 1).

The basic idea of the probabilities of connected events was developed by the Russian mathematician Andrey Andreyevich Markov, Sr. (1856–1922) in a series of articles written between 1906 and 1912 (Markoff 1910; Markov 1906, 1907a–b, 1908, 1910, 1911a–b, 1912). Purely theoretically, he showed that the weak law of large numbers and other important results of the calculus of probability were valid not only for independent events, as assumed by classical stochastics, but also for samples that were connected in simple or multiple chains. Here the outcome of the current trial was dependent on that of one or more preceding ones.[1] In those six years, Markov could not think of a practical illustration and empirical verification of his deductions, or was not interested in finding one.[2] But then, in 1913, the bi-centenary of Jakob Bernoulli's first formulation of this law in his *Ars Conjectandi* (Bernoulli 1713), he finally applied his results, namely to the 20,000 letters at the beginning of the most important work by Russia's national writer Aleksandr S. Pushkin (1799–1837), *Eugene Onegin* (Pushkin [1833] 1975; Markov 1913a). Soon afterwards, he conducted a second, even more extensive count of the first 100,000 letters contained in the work, *Childhood Years of Bagrov's Grandson*, by the lesser-known Russian author Sergey T. Aksakoff (1791–1859) (Aksakoff [1858] 1960).[3] Unfortunately, those important investigations have never been published in a satisfactory way in any language other than Russian and it is not astonishing that the numerous references in scientific literature always point to the original article on Pushkin, which very few people can read. I have therefore translated it into German and included it as an appendix in my doctoral dissertation on early text generation algorithms (Markov 1913b).[4]

English translations of these two articles are presented here in order to enable Western researchers to study the primary sources of the application of Markov chains. This commentary on the two articles can be divided roughly into five sections. In the first, I will present the investigation, explaining the methods used and the results obtained. Since the two articles are quite similar, I shall concentrate on the first. After reviewing the early propagation of Markov's ideas, I will try to reconstruct what it was that inspired the Russian mathematician to think that text could be a suitable material to empirically verify his assumptions. In the fourth section, I will follow the obscure way by which the theory of connected events was imported to Western Europe and North America. Finally, I will present some experiments to determine the scope of this type of mathematical analysis.

## I. Russian Word Count

Markov arranged Pushkin's text in tables of 10 rows of 10 letters each, then determined the number of vowels in every column of letters and added the first and the sixth, the

[1] The law of large numbers predicts that the relative frequency of occurrence of random events, when observed in great quantity, converges to their common expectation and is fundamental to probability theory.

[2] Cf. Markov to Chuprov, 7 December 1910: "I shall not go a step out of that region where my competence is beyond any doubt" (Ondar 1981, 52).

[3] Markov 1924, 577–581, but already included in the appendix of the third edition (Markov 1913c).

[4] A rather clumsy German translation was published in the German journal *Exakte Ästhetik* and there is a French translation of the abridged version of the lecture in the Appendix of the third edition of *Calculus of Probability* (Markov 1913d, Markow [1913] 1969, Markoff 1915). More recently, Micheline Petruszewycz has paraphrased in French both the article on Pushkin and on Aksakoff, together with Markov 1907a and 1911b, in her detailed study *Les chaînes de Markov dans le domaine linguistique* (Petruszewycz 1981).

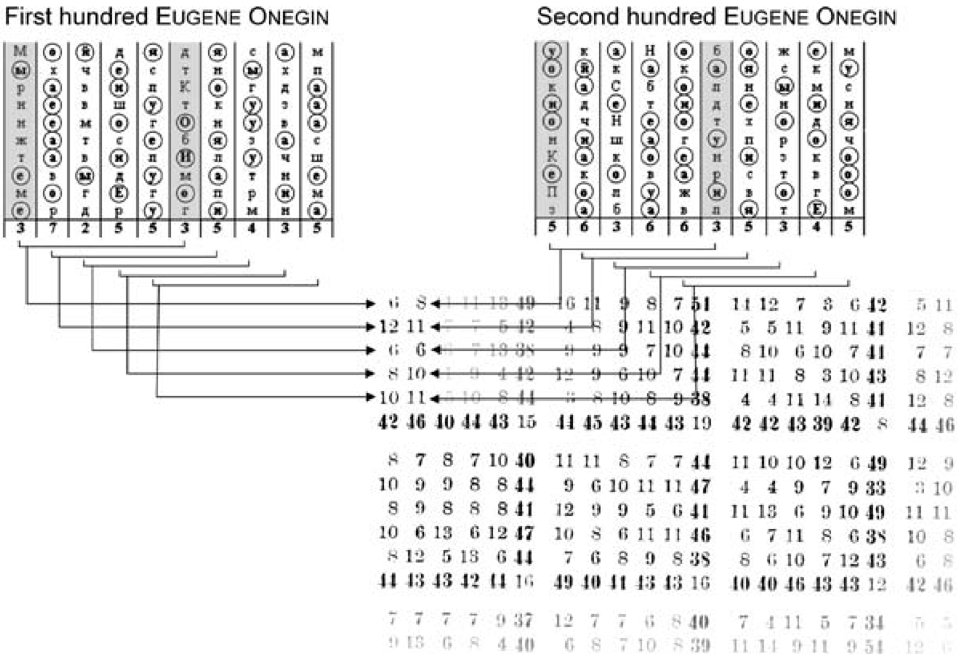



**Fig. 2.** Markov's counting method.

second and the seventh, the third and the eighth, etc. (see fig. 2).[5] In this way, he observed the frequency of vowels in letters that were separated by four others in the text. He entered those five numbers into the columns of 40 little tables of five rows and five columns, thus representing 500 letters of the novel in each table. He added the quantities in the rows and inserted the sums in bold into a sixth column, then added the numbers in the columns and wrote the results in bold into a sixth row. Finally, to check the correctness of the two counts, he added the last rows and last columns, and since it was the same 500 letters that were examined, the two values had to agree. He entered it into the bottom right corner, to save space, minus 200. As a result of the rather complex operation, the 200 quantities in the last rows (the "horizontal count") showed the number of vowels in continuous sequences of 100 letters of the novel, whereas the 200 in the last columns (the "vertical count") represented the same value for 100 letters that were separated by four others in the novel (see pages 591–595). Markov compared the two arrangements of the text material using the techniques of probability theory of his time. In the first 200 samples of 100 signs, letters

[5] The reader not familiar with the general concepts of probability theory might have a look at Keynes 1921, a rather philosophical, but detailed and accessible introduction to most of the mathematical concepts Markov employed.

followed each other in their natural sequence, while in the second, the connection was destroyed.

He first concentrated on the horizontal count and made a table of how often the numbers 37 to 49 appeared; he calculated the common average as 43.2 and consequently the mean probability of a single sign to be a vowel as 0.432. Then, he analyzed the dispersion of the empirical values for one hundred letters from their common theoretical expectation using the method of least squares.[6] The sum of the squares of their deviations from 43.2 turned out to be 1,022.8, divided further by the number of samples, 200, the result was 5.114. This number represented the mathematical expectation of the square of deviation (the so-called variance) of each of the 200 trials from the mean. Since in the groups of 100 consecutive letters that were analyzed only the last few had an influence on the beginning of the next sequence, their connection was fairly weak and the dispersion thus followed the normal distribution that could be expected for independent random events.[7] Under this curve, usually half of the values fell in the 0.67-fold of the mean deviation, and Markov demonstrated that it also held true here.[8] The independence of the quantities for the groups of 100 could be seen as well from the fact that the mean dispersion did not change considerably if they were combined in two, four, or five. If there had been dependence, the operation would have weakened it and would have altered the mean deviation (see pages 595f.).

Inside the groups of 100 letters though, the signs were connected in their natural order. Markov computed the mean dispersion from the average on the level of the single letter by dividing the value found, 5.114, by 100, and compared it to its theoretical expectation. To calculate the latter, he used Newton's formula for the variance of the binomial distribution, "n p (1 − p)", and obtained, after dividing by n, 0.245376, which differed strongly from the experimental value 0.05114.[9] The coefficient of dispersion that related the empirical to the theoretical quantity equaled approximately

[6] The method of least squares serves to determine the deviation of single samples from a common mean, the "dispersion." It was first published in 1805 by Adrien-Marie Legendre in *Nouvelles méthodes pour la détermination des orbites des comètes* but is, however, linked with Carl Friedrich Gauss' name, who elucidated it in 1809 in his *Theoria motus corporum coelestium*. The deviation of each value from the arithmetic mean is squared, and thus positive; next, the average of these differences is calculated by multiplying them by their frequency of occurrence, adding them, and finally, dividing them by their total number. In the discussion of methods from the history of probability, I follow closely the thorough work of Hald 1990 and 1998.

[7] For a detailed discussion of the irritating fact that language can effectively be described as a random process, see my other article on the subject, Link 2006.

[8] In 1733, Abraham de Moivre developed the normal curve in connection with problems relating to games of chance. It approximated the expected binomial distribution for independent random elementary errors, which was too complicated to calculate for large quantities of samples (Moivre [1733] 1738). As the properties of this curve were known, it was possible to calculate the number of values that fell in certain regions to the left and the right of the mean. The interval of 0.67 was used for the first time in 1815 by Friedrich Wilhelm Bessel, director of the Royal Observatory in Königsberg (now Kaliningrad), in his studies on the position of the pole-star to estimate inaccuracies in astronomical measurements and was termed "probable error" (Bessel 1815).

[9] The formula is discussed in the German edition of Markov's *Calculus of Probability* (Markoff 1912, 27).

1/5.[10] The mean deviation on the level of single letters was thus five times smaller than could be expected for independent random events. Markov explained the fact with the connectedness of the single letters in the groups and went on to show that by applying a formula from his research on simple chains, the dispersion coefficient could be approximated theoretically. He counted the frequency of vowels that were preceded by another vowel, calculated from this the number of the sequences consonant–consonant and consonant vowel and derived the following simple, but momentous, empirical result: "As we can see, the probability of a letter to be a vowel changes considerably depending upon which letter – vowel or consonant – precedes it" (see page 597). He entered the values found into a formula from his paper "Research on a Remarkable Case of Dependent Samples" and calculated a theoretical dispersion coefficient of 0.3, which was relatively close to the empirical one, 0.208 (Markov 1907a, in the French version, Markoff 1910, 100).[11] Thus, the experimental deviation of the quantities could be predicted much more accurately if the text was regarded as a chain in which the probability of the current element was linked to the preceding one, rather than as completely independent samples (see page 597f.).

In 1911, Markov had also investigated the case of multiple chains, where the probability of every trial was connected with the preceding two, again purely theoretically. With the aim of applying the results obtained in "On a Case of Samples Connected in Complex Chain" and approximating the observed deviation even closer, he counted the frequency of three vowels and three consonants in sequence, entered them in his formula for the dispersion coefficient for complex chains and found it to be 0.195 which was even closer to the empirical one of 0.208 (Markov 1911b, 179; (see page 598f.)).

To contrast the results, Markov turned his attention to the vertical count in the last columns of the tables, which contained the number of vowels in the hundred letters that were separated by four others in the text, and for this arrangement calculated the mean dispersion from their mathematical expectation, which equaled 43.2, as before. The sum of the squares of their deviations from this value turned out to be much higher, 5,788.8, because every letter in one sequence of 100 preceded one letter in the next one and the groups were thus strongly dependent on each other. Quoting from his *Calculus of Probability*, Markov showed that the mathematical expectation of the variance of each quantity could be obtained by simply dividing by the number

[10] Dissatisfied with contemporary common practice in statistics, whereby material was assumed uncritically to be normally distributed, Wilhelm Lexis (1837–1914), a mathematician and social scientist, was the first to compare the empirically found standard deviation with the theoretically predicted one and to form their coefficient Q. If this was bigger than one, he said the dispersion was "supernormal" and explained it by the fact that the base probabilities of the series under consideration were fluctuating. A value equal to one was called "normal" dispersion. If Q was smaller than one, Lexis regarded the series to be internally connected in some way and, different from Markov, banished it from the field of probability theory (cf. Lexis [1879] 1903, 183). Thus, Lexis would have repudiated the Galton board, had he ever investigated it empirically.

[11] It is interesting to note that Markov always refers to "cases" when investigating purely theoretical constructions.

of samples as before, 200, even if the hundreds were now connected (in the German edition, Markoff 1912, 209). The chaining of the sequences could also be seen from the fact that if now the numbers were combined in two, four, or fives, the mean dispersion decreased considerably. Because of this addition, progressively fewer letters in the sample (relative to its size) preceded those in the next, the dependence became weaker and consequently, the deviation normalized. On the level of the single letter the dispersion coefficient was now close to unity, because the signs in the groups were no longer connected, as they were separated by four others in the text (see pages 599ff.).

## II. Early Propagation

Since Markov wrote a letter to Charles Hermite (1822–1901) in Paris in 1886 (Markoff 1886), which was published in the *Annales Scientifiques de l'École Normale Supérieure*, pointing out some theorems from his dissertation (Markov 1885), he had been in contact with the French mathematician, the teacher of Jacques Hadamard, who had met his teacher, Pafnutiy L. Chebyshëv (1821–1894), when Chebyshëv visited Paris in 1852 (cf. Delone [1947] 2005, 4). In the years that followed, Hermite regularly published Markov's results, which were communicated to him mostly by letters, in the *Comptes rendus de l'Académie des Sciences* (Markoff 1889, 1891a–c, 1892, 1894). His 1892 paper was reviewed in the *Jahrbuch über die Fortschritte der Mathematik* (JFM) by a pupil of Felix Klein, Adolf Hurwitz, who appointed George Pólya in 1914 as private lecturer at Technical University Zurich and wrote around 400 abstracts for JFM between 1879 and 1899 (cf. Preussische Akademie der Wissenschaften et al. 1868–1942).[12]

In 1896, Felix Klein in Göttingen encouraged a former pupil of Markov, Theophil Friesendorff, to translate Markov's *Calculus of Finite Differences* into German. They felt the "lack of appropriate textbooks" in this area and had read about the Russian mathematician's work in the JFM (Markoff 1896, iv). The result was the publication of Markov's first German monograph, *Differenzenrechnung*, an assembly of Markov 1891a and b. Heinrich Liebmann, professor of mathematics at Munich University, who had previously translated two works of Nikolay I. Lobachevskiy ([Lobatschefskij] [1835] 1904, [1856] 1902) and written a book on non-euclidean geometry in 1905 (Liebmann 1905), subsequently published a German edition of Markov's *Calculus of Probability* in 1912, probably inspired by Felix Klein's former assistant, Arnold Sommerfeld, a theoretical physicist who taught in Munich from 1906.[13]

[12] Since all *Jahrbuch* reviews are easily accessible on the internet, I will not provide detailed references for them in the following. When querying the database, attention has to be paid to the different possible transliterations of Russian names, like "Markov," "Markoff," and even "Markow."

[13] Encouragement might have come also from Friedrich Engel, a pupil of Felix Klein in Leipzig, who had also translated two of Lobachevskiy's works in 1898 (Lobatschefskij 1898) and reviewed Liebmann's 1912 edition in JFM.

While the main text of *Wahrscheinlichkeitsrechnung* was not concerned with connected samples, it did include two articles on the topic in the appendix: "Extension of the Limit Theorems of Probability Theory to a Sum of Values Connected in a Chain" from 1908 (Markov 1908) and "On Connected Values that Do not Form a Real Chain" from 1911 (Markov 1911a). In his review in JFM, Heinrich Emil Timerding completely ignored the relevance of the investigations: "Three papers of a mathematical character form the conclusion. The task of these works consists mainly in giving 'strict proofs for the fundamental limit theorems of the calculus of probability and to generalize them as far as possible'." At this point, Markov, aged 56, had already been reviewed in JFM around 90 times and had published 21 articles in Western European journals such as *Mathematische Annalen* (9), *Acta Mathematica* (4), *Comptes Rendus de l'Académie des Sciences* (6), and even the *Bulletin of the American Mathematical Society* (1), but only one about connected trials, "Research on a Remarkable Case of Dependent Samples" (Markoff 1910).[14] So three works on the subject were available in Western European languages in 1912, mostly ignored by mathematicians outside Russia. His research on *Eugene Onegin* was published in 1913, in the *Bulletin* of St. Petersburg Academy (Markov 1913a) as well as in a separate print of the appendix of the third Russian edition of *Calculus of Probability* commemorating the bi-centenary of Bernoulli's law of large numbers (Markov 1913d). The latter contained "On a Remarkable Case of Samples Connected in a Chain," an assembly of condensed versions of Markov's articles from 1907 and 1913 (Markov 1907a, 1913a).

Two years later, the poet, astronomer and revolutionary Nikolay A. Morozov published an investigation in the St. Petersburg *Bulletin* trying to use the frequency of auxiliary words to tell different literary writers apart and to determine the authorship of Plato's works (Morozov 1915). He combined his counts in "spectra," "like the astronomer easily and reliably determines the chemical composition of celestial bodies inaccessible to our flight vehicles from them." After presenting the frequency of single letters and also bigrams in German and French, he "signals that diplomats and serious conspirators . . . use 'keys' without 'spaces' and that vocatives and phrases of courtesy can permit one to find the key." Markov criticized the article in 1916, and recommended to the author that his investigation be extended from 5,000 to 100,000 words to yield more significant results (Markov 1916; Petruszewycz 1981, 140ff.). This clearly shows his interest in literary analysis and explains why he felt the need to extend his own research to the same quantity of samples with his count of the text from Aksakoff in the third edition of *Calculus of Probability*.

## III. Word Crosses from the Underground

The inspirations for Markov's experiment are difficult to reconstruct, especially because he does not hint at them at all. There are, however, a number of reasonable sources. As

[14] For a very extensive bibliography of Markov, cf. my website, http://alpha60.de/research/markov/biblio.html.

early as 1847 Victor Y. Bunyakovskiy (1804–1889), author of the first Russian book on probability theory and the teacher of Markov's teacher, Chebyshëv, recommended that language be viewed as a statistical phenomenon and that stochastics be applied to questions of linguistics, such as etymology (cf. Petruszewycz 1981, 133ff. for details).

Another inspiration might have been Jan Baudouin de Courtenay (1845–1929), the founder of the Kazan school of language research, who became professor for general linguistics at St. Petersburg University in 1900 and taught there until 1918. In his 1895 "Essay of a Theory of Phonetic Alternations" he described the diachronic changes of language through transformation of its elements, the single phonemes, and made extensive use of mathematical notation to present his findings:

> The formula for anthropophonic changes ... looks like the following: $x + n\Psi$, where x designates some primal phoneme, $\Psi$ some arbitrary historical-phonetic change in a certain direction, n the coefficient of this change. (Baudouin de Courtenay [1895] 1984, [31])

The transformations were regarded as linked to each other: "The features of articulation that are alternating together are not individual or independent features of anthropophonetic varieties ... of the corresponding phoneme ..., but they are solely determined combinatorically, i.e. they depend on the connection with other phonemes." He advanced the formula "$x = f(y)$"; x being the changing element and a function of y, the "conditions of connection in general" (ibid., [48]).

The development of language as a whole he described as "not so much a gradation, but rather a continuous fluctuation, continuous oscillations" (ibid., [34]).[15] The founder of the Kazan physico-mathematical society and its first chair, Aleksandr V. Vasil'yev (1853–1929), who reviewed numerous articles of Markov in JFM and corresponded with him at the latest from 1898 on, might have mediated between the two scientists (cf. Markov 1898).[16]

It can be seen indirectly, from the techniques employed by Markov and his unusual interpretation of language as a connected chain of events, that he was mainly inspired by cryptography and cryptology. Radiotelegraphy, invented by Nikola Tesla in 1893, but successfully marketed by Guglielmo Marconi, changed civil and military communication considerably around 1900. Because of the completely public transmission, all messages that were to remain private had to be encrypted. To this end, the French military, then allies of the Russians, introduced "SD modèle 1912" (SD meaning "sans dictionnaire," without dictionary) around 1912. Unlike earlier systems

[15] Baudouin de Courtenay met the Swiss linguist Ferdinand de Saussure in Paris in December 1881 and corresponded with him from 1889 on. Saussure had the German translation of the aforementioned book in his personal library (cf. Sljusareva 1972, 12, 15). This connection sheds light on the similarity of the results of Markov and Saussure noted in my earlier article on the subject (cf. Link 2006).

[16] Vasil'yev also lectured at the bi-centenary celebration at St. Petersburg in 1913, along with Markov and Aleksandr Chuprov (cf. Markov to Chuprov, 31 January 1913, Ondar 1981, 69).

that relied on codebooks or methods of substitution, it employed a transposition cipher using diagonals, with a simplified version, a pure transposition. Also the Germans obscured their transmissions by two cryptographic systems, a single transposition for messages of secondary relevance and a double transposition for important ones, code-named "Übchi" by the French. An exclusive radiotelegraphic connection was established between Paris (using the Eiffel tower as antenna) and Bobruisk in 1911. The Russian office for this line was located in St. Petersburg (cf. Ollier 2002, 81ff.).

The techniques were used not only by the military in some European countries, but also by the Russian revolutionaries attempting to kill the Tsar, in the form of the so-called "Nihilist transposition cipher." The plaintext was arranged in a square by rows (or diagonals); a keyword switched the rows; a same or different mnemonic exchanged the columns, and the resulting ciphertext was removed by columns or by one of 40 or more routes out of the square (cf. Gaines 1989, 17ff.). The cipher emanated from Russian prisons where a checkerboard was used to convert letters into numerals that could be transmitted in a variety of ways (cf. Kahn 1967, 619ff.).

The similarities between simple columnar transposition and Markov's method are striking. In both, the text was written in the rows of a square, then read out in columns. By this mixing operation, what was obscured was not the letters themselves, as in systems of substitution, but rather the connection between them. Consequently, to break this cipher, the cryptanalyst first had to determine the number of columns of the square and then use common letter combinations to guess at their original order. One of the first to present bi- and trigram frequency tables for English and Spanish (and the solution of many substitution and transposition systems) was the American Colonel Parker Hitt in 1916 (Hitt [1916] 1976). Transposition methods like the "rail fence" and "route cipher" were also very common in the American Civil War (cf. Kahn 1967, 214ff.).

The author of the above-mentioned scientific article "Linguistic Spectra," Nikolay Morozov, obviously was a very impressive figure in 1876. Ol'ga Lyubatovich, whom he married later, recalled:

> Kravchinskii [Sergey Kravchinskiy] came by with his close friend Nikolai Morozov, whom he introduced as "our young poet." Morozov blushed like a girl. Apart from his literary work, he was one of the party's most ardent advocates of partisan revolutionary warfare – terrorist struggle – and he was always heavily laden with guns, nearly bent over from their weight. (Engel 1975, 152)

After the splitting of the revolutionary movement "Land and Liberty" in 1879, he became a prominent member of "People's Will," a group that favored a policy of armed resistance. In 1880 he left Russia for Geneva with his wife and wrote one of the first terrorist pamphlets, "The Terrorist Struggle" (Morozov [1880] 1987). When he returned to his home country in order to distribute his work, he was arrested and sentenced to life imprisonment in 1881. He spent about 25 years in the notorious

"Schlüsselburg" and other political jails. Here, probably by code-talking with his fellow convicts, he acquired an impressive knowledge of diverse fields of science and wrote his ideas down in a number of notebooks. Upon his release after the revolution of 1905, apart from his literary work, he started to publish scientific articles and books on a wide range of topics, including astronomy, physics, chemistry, biology, mathematics, etc. (cf. Fomenko 1994, 89). In 1907, he published *Revelation in Tempest and Storm*, explaining the Book of Revelation through meteorological phenomena. It sold 16,000 copies in the first year and was widely discussed in the newspapers. He gave public lectures in several major Russian cities. In the same year, he was appointed to teach chemistry at St. Petersburg University (cf. Morozow 1912, xixf.). Even if Markov perhaps did not know his colleague before 1913, the biography of Morozov gives a clear picture of the milieu at St. Petersburg University in those years. It must have been very easy here for the "militant academician" to acquire knowledge of the most basic operations of revolutionary encryption techniques. Already when Lenin's close friend and collaborator Maksim Gorkiy became an honorary member of the Academy and the government declared the election invalid in 1902, "Andrew the Furious" refused to accept tsarist awards in protest.[17]

It is a curious coincidence that in the same year that Markov's article on *Eugene Onegin* was published, an American immigrant from Liverpool, Arthur Wynne, invented the crossword puzzle under the name "word-cross." It was published 21 December 1913 in the newspaper *New York World* and led to a popular craze from 1924 on (cf. Costello 1988, 21).

## IV. Transmission to the West

Thirty-five years after the publication of the article on Pushkin, Markov's ideas reappeared prominently, but transformed, in Claude E. Shannon's (1916–2001) "Mathematical Theory of Communication," another work developed in close proximity to questions of cryptography.[18] It was used here to simulate the properties of textual information transmitted over a channel with noise. To show the effectiveness of his approach, Shannon employed Markov's method for the first time not to analyze, but to generate text, with the following, well-known result:

> THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD

[17] Cf. Seneta 1996, 258. These were Markov's nicknames in the press, cf. Sheynin 1988, 340.

[18] Shannon developed his seminal paper in parallel with his "Communication Theory of Secrecy Systems," that was first published one year later (cf. Kahn 1967, 744, and Shannon [1948] 1993, 1949). The reader not acquainted with the main figures in probability theory in the twentieth century might want to read the biography of George Pólya to get an overview (Alexanderson 2000).

FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED. (Shannon [1948] 1993, 14)

Shannon did not reference one of Markov's articles, but Fréchet's book *Méthode des fonctions arbitraires* from 1938 (Fréchet 1938). The exact way in which the idea of chains permeated into Western Europe and North America was unknown until recently when the French science historian Bernard Bru published a very detailed account of the early diffusion of the Russian researcher's work (Bru 2003). According to Bru, the main transfer happened on the International Congress of Mathematics in Bologna, 1928: "It is hardly questionable, in any case, that the interest of the schools of Paris and Moscow for Markov's theory dates from Bologna and that Fréchet, Khinchin and Kolmogorov have been initiated . . . by Hostinský" (ibid., 166). The transmission of the theory of chains before 1928, when it reached Bohuslav Hostinský (1884–1951), the Czech mathematician corresponding with Fréchet from 1919 on, is reconstructed rather vaguely along the lines of Bernshteyn – Pólya – Hadamard – Hostinský – Fréchet. I will try to complement Bru's account in the following, since at least part of the transfer happened before 1928. Figure 3 presents an overview.

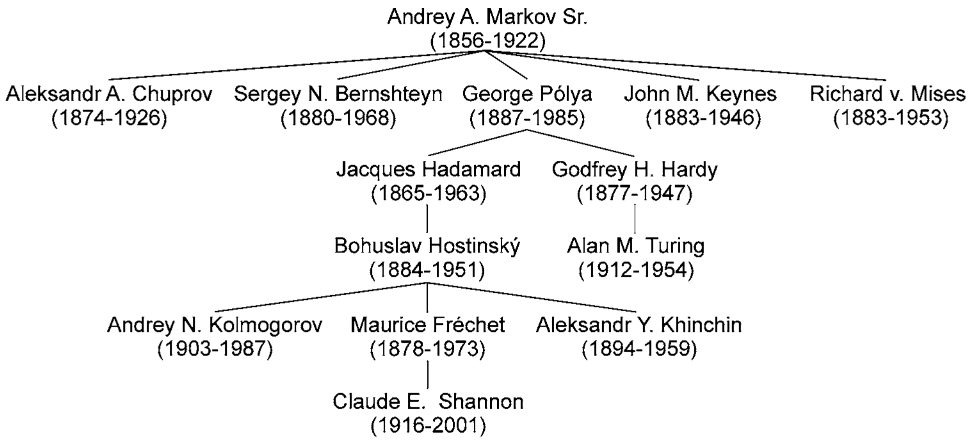



**Fig. 3.** The probable transmission of Markov's ideas to the West.

The young Kharkov mathematician Sergey N. Bernshteyn (1880–1968) was in fact one of the first to understand the relevancy of Markov's results and to extend them. According to Bru, he had already included the theory of chains in his lectures on probability calculus in 1917 and published numerous articles on the subject between 1922 and 1941.[19] But his influence in propagating the idea was limited, perhaps because

[19] Cf. Bru 2003, 141, and personal communication, 20 April 2006: "In the archives of Fréchet at the Academy of Sciences is the program of the course of Bernshteyn for the year 1917, where one finds explicitly mentioned that Bernshteyn will treat the generalization of the law of the large numbers to the case of connected events by

of his remote location. He taught from 1908 to 1933 at Kharkov University, and later in Leningrad and Moscow (cf. Seneta 1982, 475).

The statistician Aleksandr A. Chuprov (1874–1926), with whom Markov conducted a lively correspondence from 1910 to 1917, definitely read and understood his colleague's results, but due to his early death he did not have much chance to spread them. From 1917 to 1920, he stayed mostly in Stockholm, moved to Dresden in 1920, and in 1925 to Prague. In September of that year he attended the International Congress of Statistics in Rome, where Maurice Fréchet held an (albeit not very long) lecture about "A New Analytical Representation of Revenues." He died shortly after, on 19 April 1926 (cf. Fréchet 1926 and Sheynin 2004, 84ff.). He was also primarily interested in applied probability theory. This might be the reason that Markov started his last letter to Chuprov with the following words: "Your work frightens me with its abundance of complicated calculations which are difficult for me to look through in all their plentitude, all the more since I have glaucoma in one eye" (Markov to Chuprov, 27 February 1917, Ondar 1981, 134f.).

The above-mentioned special edition of the appendix of *Calculus of Probability* was translated into French in 1915 (Markoff 1915) and this time found a more sensible reviewer in Adolf Hurwitz's pupil, George Pólya (1887–1985), private lecturer for Higher Mathematics at the Technical University (ETH) in Zurich. He followed the extensive reading habits of his teacher and wrote around 100 reports in JFM between 1914 and 1920, but this was only the second time he contributed to the year-book. Having obtained his Ph.D. at the University of Budapest in 1912, under the advisor Leopold Fejér, with a work on probability theory (Pólya 1913), Pólya read the publication attentively and commented on the third part:

> A sequence of non-independent samples is connected in such a manner that the probability for the arrival of an event left to coincidence is $p_1$, if in the preceding sample the event occurred, however $p_2$, if it did not occur. Such an "event" is e.g. the occurrence of a consonant in the successive letters of a book; since a consonant follows more easily a vowel than a consonant, $p_1$ is $< p_2$. The author treats this simplest special case of his samples "connected in a chain" in detail and *ab ovo*. The results of the calculation are illustrated and proven by a statistic that extends to the first 20,000 letters of Pushkin's novel *Eugene Onegin*.

In 1919 and more relevantly, 1921, he developed his theory of "Irrfahrten" (random walks) and kept mentioning Markov's work (Pólya 1919, 1921). Gerald Alexanderson provides an anecdotal account on how the mathematician discovered the topic:

> One day while out on his walk he [Pólya] encountered one of his students strolling with his fiancée. Somewhat later their paths crossed again and even later he encountered them once again. He was embarrassed and worried that the couple would conclude that he

Markov." The first articles are Bernstein [Bernshteyn] 1922, 1927, 1928; for a more complete bibliography, cf. Bru 2003, 209.

> was somehow arranging these encounters. This caused him to wonder how likely it was that walking randomly through paths in the woods, one would encounter others similarly engaged.[20]

Random walks are today considered special cases of Markov chains and an approximation to Brownian motion (cf. Spitzer 1976, vii).

In Germany, Richard von Mises (1883–1953), director of the Institute of Applied Mathematics in Berlin from 1920 and founder of the journal *Zeitschrift für angewandte Mathematik und Mechanik* one year later, was one of the first to notice the new theory. On 3 July 1918, he wrote in his diary: "Read a bit of Markov. Especially tried to follow the problems of probabilities connected in a chain." He had discovered the works in the appendix as early as August 1913 and included the findings of the Russian mathematician in his lectures in the late 1920s.[21]

Another early author to mention "the very beautiful work which we owe to these three Russians" (Chebyshëv, Markov, and Chuprov) was the British economist John M. Keynes (1883–1946) in his fellowship dissertation at King's College, Cambridge, accepted in 1909, but published only in 1921, *A Treatise on Probability* (Keynes 1921, 358). By absorbing more publications on stochastics "than it would be good for any one to read again" (ibid., 432), he acquired a relatively broad overview of the field in his time. Keynes' very extensive bibliography contained eleven works of Markov, of which he could read five.[22] As can be seen from the unusual translations, most of them are copied from the references in the German edition of the Russian mathematician's *Calculus of Probability*. It is surprising that Keynes covered some of the themes Markov was concerned with in the later parts of his book. He mentions a precursor in the numerical treatment of text, Francis Y. Edgeworth, who tried to criticize Lexis by applying his theory "to the frequency of dactyls in successive extracts of the *Aeneid*" in 1885 (ibid., 400). The response to this criticism was simply that the series was not sufficiently long. Keynes concluded that

> these authorities are at fault in the principles, if not of Probability, of Poetry. The dactyls of the Virgilian hexameter are, in fact, a very good example of what has been termed *connexité*, leading to sub-normal dispersion. The quantities of the successive feet are not

[20] Alexanderson 2000, 51. While most of this publication is anecdotal, it provides a good bibliography and "Pólya's work in Probability" by Kai Lai Chung in appendix I which presents an overview of Pólya's work in this field.

[21] Reinhard Siegmund-Schultze, personal communication, 21 May 2006. He also found the following entry in Mises' scientific diary for the year 1913: "17.VIII. Quantities connected in a chain. Markov, Appendix II" (personal communication, 8 June 2006; cf. Siegmund-Schultze 2004).

[22] The listing that was obviously extended prior to the publication in 1921 included Markov 1900b (Russian), 1906 (R), 1907a (R, Keynes did not note the French translation in *Acta Math.*, Markoff 1910), 1907b (R), 1908 (R, German), 1910 (R), 1911a (R,G), 1911b (R), Markoff 1912 (G), 1915 (F). Both the 1907 and the 1908 version of the work are located in the Personal Papers Archive of King's College, Cambridge, UK, catalogue ref. JMK/13/TP.

> independent, and the appearance of a dactyl in one foot *diminishes* the probability of another dactyl in that line. ... But Lexis is wrong if he supposes that a *super-normal* dispersion cannot also arise out of connexité or organic connection between successive terms. It might have been the case that the appearance of a dactyl in one foot *increased* the probability of another dactyl in that line. (ibid., 401; emphasis in original)[23]

These considerations are very similar to those in the correspondence between Markov and Chuprov.[24] It is possible that Keynes was inspired by the appendix of *Calculus of Probability*, but did not dare to reference the work in this context because he could not read so many of the other articles.[25] Instead, by using the term *connexité* he referred implicitly to the works of the ostracized Louis Bachelier.[26] Keynes' book was translated into German in 1926 by Friedrich M. Urban, like Hostinský and Mises a mathematician from Brno in Czechoslovakia (Keynes 1926).

As an example of the different dispersions caused by connected quantities, we might consider two sorts of extreme Markov processes. If the probabilities for a vowel after a vowel and a consonant after a consonant are 1, generation of chains of six letters yields only two results, "CCCCCC" or "VVVVVV." The overall mean probability of a sign to be a vowel is 0.5, but the deviation is very high, since none of the samples contains half vowels and half consonants. If, on the other hand, the probability of a consonant after a vowel is 1, and vice versa, then the process only produces "CVCVCV" or "VCVCVC" with the same average, and zero dispersion. The deviation considered "normal" is located between those two extremes. In his analysis of Aksakov, Markov included one count of groups in which letters separated by one other letter in the text were put together. Here, the probability of a vowel was higher after a preceding vowel than after a consonant. Consequently, the dispersion coefficient turned out to be highly supernormal, 1.45 (see pages 603f.).

[23] Chuprov, who was asked by Markov on 15 January 1913, concerning the count of Pushkin, to "please let me know whether or not you have come across a similar example elsewhere," mentioned Edgeworth's article in a letter on 10 March 1916 giving wrong references, but only wrote about his results on the measurement of heights, not on poetry (Ondar 1981, 67f., 85).

[24] Cf. Markov to Chuprov, 24 November 1910, concerning "the true value of Q," the dispersion coefficient: "No one has proved that its mathematical expectation is equal to one." 3 December 1910: "Lexis's case of supernormal dispersion is inseparably linked to the opposite case of supernormal stability" (Ondar 1981, 41, 50).

[25] In his review of *Calculus of Probability* for the *Journal of the Royal Statistical Society* he wrote in contrast that it was "the work of a pure mathematician, who avoids philosophical difficulties on the one hand and practical applications on the other" and that "Professor Markoff's book does not contain much that is notably new" (Keynes 1913, 116).

[26] Already on 10 November 1910, Chuprov wrote to Markov that Bachelier had previously covered the topic of dependent samples (Ondar 1981, 6). Markov reacted resolutely on Bachelier's work on 15 November 1910: "I do not attempt to judge its significance for statistics but with respect to mathematics, it has no importance in my opinion."

Together with one of his students, Florian Eggenberger, George Pólya published "On the Statistics of Events in a Chain" in 1923, in the journal on applied mathematics mentioned above, von Mises' *Zeitschrift für angewandte Mathematik und Mechanik* (Eggenberger and Pólya 1923). Unlike Markov, who obviously found it difficult to think of an application for his theory, the authors explicitly referenced the two articles on connected events in the appendix of the German edition of *Calculus of Probability* and showed that the mathematics of "simple chains" could be successfully used in diverse practical cases (ibid., 280). The first they mentioned was the statistics of railway accidents and industrial explosions: "For the lives of occupants of the same train show 'solidarity' to a great extent: The death of one person because of railway accident must be regarded as an extreme degradation of the chances of all fellow passengers" (ibid., 279). A number of major railway accidents occurred in the 1910s, among them the Great Train Wreck of 1918, in Nashville, Tennessee, that killed 101 people and injured 171 (cf. Aldrich 1993). A catastrophic industrial explosion happened in 1921 at a BASF plant in Oppau, now part of Ludwigshafen, Germany, killing at least 500 persons and injuring about 2000 more (cf. Flynn and Theodore 2001, 5). Eggenberger and Pólya treated in detail the statistics of epidemics. The "Spanish" flu killed over 30 million people worldwide in 1918 and 1919 and the sixth pandemic of cholera (1899–1923) was almost over in 1923, striking mainly Russia (cf. Phillips and Killingray 2003, i; Ranger and Slack 1992, 151). It was naturally interesting for insurance mathematicians to be able to grasp such phenomena more precisely. The two authors analyzed the tables of lethal smallpox in Switzerland between 1877 and 1900 and showed that "the picture of probability theory agrees much better with reality if augmentation of chance by success is assumed rather than independence of trials" (Eggenberger and Pólya 1923, 284; cf. fig. 4).

The second problem treated by Eggenberger and Pólya emanated from biology, the spreading of plants on a two-dimensional surface, or, more generally, "the distribution of a set of points in space." As the authors noted, it was also interesting for physics as a means to understand phenomena like radioactive decay and Brownian motion (ibid., 279). Again, Markov's theory proved to be more appropriate than the assumption of independence, since "the neighboring segments are in a manner of speaking 'infected'" (ibid., 289). The title of Eggenberger's dissertation one year later was "The Contagion of Probability" (Eggenberger 1924).[27]

The theory of connected events allowed the more exact treatment of problems from different fields of science, being an abstract meta-theory like probability calculus itself. Lexis' critique of the assumption of normal distribution was transformed by Markov into a generalization of stochastics, in which the independence of the trials was just a special case of connected events. Formally speaking, in addition to the free

[27] Here, Markov's theory is successfully applied to the statistics of deaths due to explosions of steam boilers, smallpox, and scarlet fever.

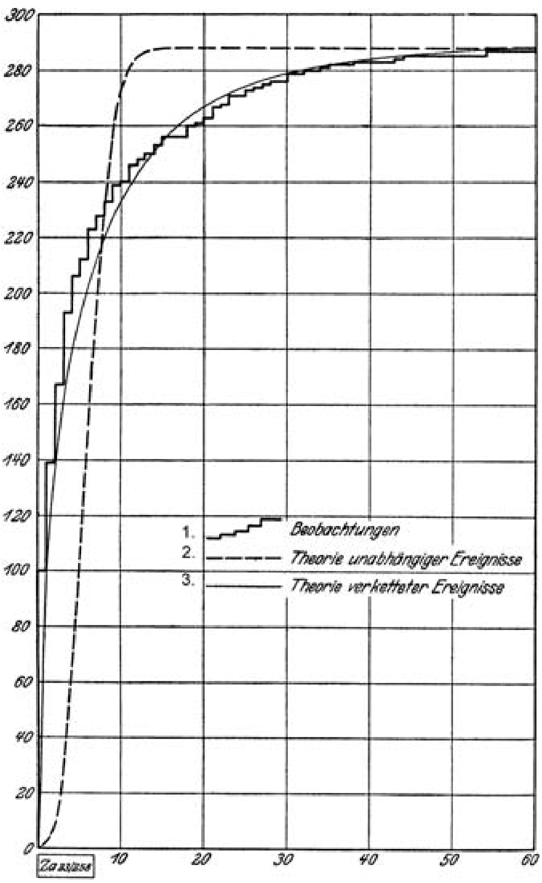



**Fig. 4.** Eggenberger/Pólya's comparison of the empirical values for lethal smallpox (1) with the predictions from the theories of independent (2) and connected events (3).

variable that indicated the probability of the outcome, the equations of stochastics now contained a second free variable that determined the connectedness between the samples. The authors noted that "the $P_r^*$ [classical probability calculus] contains only one adjustable parameter, whereas the $P_r$ [Markov's theory] contains two such parameters" (Eggenberger and Pólya 1923, 284). The case of independence was included in the new formulas when the second variable equaled zero.

In his dissertation the next year, reviewed by Pólya and Hermann Weyl, Alfred Aeppli treated "Markov Chains of Higher Order" where the current trial depended on several preceding ones (Aeppli 1924). The Russian mathematician's results were derived using Pólya's concept of "random walks" in a square network. The decision on which way to walk depends on one or more previous decisions. Aeppli quoted all of Markov's articles on chains published up to that point in Western European languages.

Andrey N. Kolmogorov (1903–1987), successor to Markov's son as chair of the Department of Mathematical Logic at Moscow State University, reflected in 1947: "Perhaps the slow adaptation of the ideas of the Petersburg school by the West can indeed be partly explained by the fact that it was very remote from statistical physics, so that Markov illustrated the application of his theory . . . of Markov chains only by considering the distribution of vowels and consonants in the text of *Eugene Onegin*" (Kolmogorov 1947, 147).

Many phenomena were discovered in the beginning of the twentieth century that required a treatment different from that of classical physics: Brownian motion, the dynamics of gases, of fluids, and of radioactivity (cf. Brown 1828, Boltzmann [1877] 1909; Gibbs 1902; Einstein 1905). With their explanation by interacting molecules, the scientific point of view changed from the observation of relatively few objects in the macroscopic world to the calculation of the behavior of very numerous, albeit finitely many, particles on the microscopic level.

The Austrian theoretical physicist Ludwig Boltzmann (1844–1906) was one of the first to see that a thorough understanding of the thermodynamics of gas could only be achieved by means of probability calculus. In his model from 1877, space is occupied and completely determined by a finite number of particles that are each in one of a finite number of "states." The molecules interact in a succession of collisions and change their state accordingly. Due to the complications of a thermodynamic theory which based its explanations on the behavior of large (if finite) numbers of molecules, Boltzmann felt the need to introduce and even partly develop probabilistic methods which were able to say something on statistical averages. He conducted a purely combinatorial estimation of the possible number of state distributions and founded the second law of thermodynamics on stochastics: "the distribution of states, which is the most probable of all, corresponds . . . to the state of thermal equilibrium" (Boltzmann [1877] 1909, 193). He also introduced the logarithmic connection between entropy and probability that later served Shannon to quantify the amount of information in discrete communication.

Because these were important theoretical problems, many mathematicians sought for examples of the repeated application of elementary operations that would transform the distribution of a given set in a succession of steps. The paradoxical demand from the second law of thermodynamics was that a number of reversible procedures had to lead to an irreversible result, the overall increase of entropy. Jacques Hadamard (1865–1963) proposed card-shuffling as a model in 1906 when discussing the problem of mixing liquids by Josiah W. Gibbs (cf. Gibbs 1902; Hadamard 1906, 205) that was subsequently treated by Henri Poincaré and Émile Borel in 1912. Hadamard and Hostinský took up the problem in 1927 and discussed it in a number of notes, culminating in Hadamard's lecture at the International Congress of Mathematicians in Bologna, "On the Shuffling of Cards and its Relations with Statistical Mechanics," in 1928 (cf. Bru 2003, 141ff.).

In a footnote, Hadamard credited Polyá for having introduced him to Markov's work that he regarded as similar to his own: "As has been pointed out to me by Mr. Pólya, the question of 'enchained quantities', treated by Markoff ... and, more recently by Mr. Serge Bernstein ... is proximate to the one of this text" (Hadamard 1928, 133).

Shortly after his occupation with connected trials, in 1924, Pólya traveled to Cambridge, England, to work with Godfrey H. Hardy (1877–1947). In his office, the English mathematician "kept papers methodically in piles on a large table in front of a window. I remember him saying 'a pile for Landau and a pile for Pólya'" (Cartwright 1981, quoted in Alexanderson 2000, 75). Hardy read and corrected most of Pólya's papers from 1923 on.[28] Their collaboration culminated in the book "Inequalities" in 1934. It treated the inequality of Chebyshëv, extended by Markov, in detail (Hardy, Littlewood, and Pólya 1934, 43ff.). One of the students of Hardy at that time was Alan M. Turing (1912–1954), who attended King's College from 1931 to 1934 and obtained his fellowship with a dissertation on probability theory, "On the Gaussian Error Function," which discusses Aleksandr M. Lyapunov's works of 1900 and 1902 in Appendix B.[29] In these articles, Lyapunov proved the central limit theorem under lesser conditions than Chebyshëv and thus extended its mathematical range considerably. He noted that the "question ... was taken up again by M. Markov who gave all the development to Chebyshëv's method that it required" (Liapounoff 1900b, 126). Two years later, Turing wrote his fundamental paper "On Computable Numbers" (Turing 1936), conceptualizing universal calculating machinery theoretically.

To estimate the similarity of Markov's and Turing's constructions, we will attempt a detailed formal description of both in the following. A Markov process starts by reading the current sign, say, a consonant ("C"). The indexing of the probabilities in the Russian mathematician's articles shows that we are dealing with a matrix. For the simple chain in Pushkin, it could be written as follows.

| | | Next letter | |
|---|---|---|---|
| | | Vowel ("p") | Consonant ("q") |
| Preceding letter | Consonant (index "0") | 0.663 | 0.337 |
| | Vowel (index "1") | 0.128 | 0.872 |

[28] Cf. Alexanderson 2000, 75: "Hardy went over Pólya's manuscripts carefully, particularly those submitted for publication in English journals ... in one [of the collections of papers] is included a manuscript for a paper of 1923 ... extensively marked in Hardy's distinctive hand."

[29] Lyapunov 1900, 1902; Turing 1935, 41ff. Like Markov, Lyapunov was one of the famous pupils of Chebyshëv in St. Petersburg. He became professor at Kharkov University in 1893. For a detailed study of Turing's dissertation, cf. Zabell 1995.

By looking up the corresponding row in the matrix, we find that the probability for a vowel as the next letter is 0.663, and that for another consonant 0.337. A suitable random generator (a book for example) is used to pick one of the two choices with the right frequency.[30] The character, say, a vowel ("V") is appended to the last sign. The process starts from the beginning with the letter just obtained. In modern conception, Markov chains are probabilistic "Finite State Machines."[31] A simple question might help to illustrate the difference between this algorithm and Turing's universal computer: Why is a Markov process not capable of performing the simplest arithmetical operation, addition? It would need to transform "…CCCVCC…" (3 + 2) to "…CCCCCV…" (5). More specifically, the separating "V" has to be converted to "C" and the last "C" to "V."[32] As can be seen from the example, the machine needs to write the field it is currently reading, instead of the following one. Only then can chains of symbols be transformed in the desired way. The second capability missing in Markov's construction is that of not only stepping forward, to the right, but also backwards, to review or change the signs already written. To transform the last "C" to "V" involves moving to the right until a "V" is encountered, then one to the left, and to write another "V." The third difference is that in Markov's process the signs on the paper are identical with the "state." With the application of the theory of connected samples to statistical physics, the Russian mathematician's "events" were renamed "states" (*Zustand*) (cf. Kolmogoroff [Kolmogorov] 1931, 415). In Turing's conception, the state becomes independent of the read character and is instead transferred into the machine as an internal property, replacing the probabilities that determine the next sign to be written in Markov's algorithm.[33] The cell of the matrix to be executed is now selected dependent on the symbol on the tape and the "state" of the machine. In it, the algorithm finds the properties we required above: which sign to write to the current field ("0" or "1"), to be set arbitrarily, the next "state" to enter ("A" or "B"), and whether to advance to the left or to the right ("L" or "R").[34]

[30] In this way, Shannon generated his examples of Markov chains: "One opens a book at random and selects a letter at random on the page. This letter is recorded. The book is then opened to another page and one reads until this letter is encountered. The succeeding letter is then recorded. Turning to another page this second letter is searched for and the succeeding letter recorded, etc." (Shannon [1948] 1993, 15). The dependence on the previous letter can be regarded as a primitive sort of conditional branching.

[31] Cf. Wolfram 2002, 958: "Simple finite automata have implicitly been used in electromechanical machines for over a century. A formal version of them appeared in 1943 in McCulloch-Pitts neural network models. (An earlier analog had appeared in Markov chains.)"

[32] For a more detailed treatment on the fundamentals of Turing Machines, cf. Link 2005.

[33] The origin of the machine's "state" in statistical physics might have been the reason why Tibor Rado replaced the states with cards in his 1962 paper: "It appeared that terms like state, internal configuration, and the like had a mysterious connotation for beginners" (Rado 1962, 878).

[34] The following table is a two-state-machine that prints four "1s" on the tape and then stops, a "busy beaver" (cf. Rado 1962). The notion of the "stop"-sign is not part of Turing's original idea, but due to Post 1936.

| | | "state" A | B |
|---|---|---|---|
| sign read | 0 | 1-B-R | 1-A-L |
| | 1 | 1-B-L | 1-Stop |

One of the first truly empirical applications of Markov chains also arose in statistical mechanics. At the end of World War II, Stanislaw Ulam and Janós (John) v. Neumann (having received his degree from Fejér and taught by Pólya at the ETH in Zurich), then working in Los Alamos to build nuclear weapons before the Germans, resuscitated the old method of statistical sampling to solve problems related to stochastic neutron diffusion and the chain reactions to be expected: "The essential feature of the process is that we avoid dealing with multiple integration or multiplication of the probability matrices, but instead sample single chains of events" (Metropolis and Ulam 1949, 339).[35] The physical reality was modeled by a Markov process, and then a random generator was used to produce concrete possible outcomes in a succession of steps by transition probability matrices. It was an inversion of the usual method of simulation, in that it was not used to test a previously understood deterministic problem, but treated deterministic problems by first finding a probabilistic analog, anticipating the availability of automatic calculating machinery. After a huge number of runs, the overall properties and behavior of the system in question could be estimated statistically, from simulative "experience" rather than from a theoretical, analytical effort. Nicholas Metropolis, one of the pioneers of "Markov Chain Monte Carlo," wrote later that it was interesting to "note the emergence ... of experimental mathematics, a natural consequence of the electronic computer. ... At long last, mathematics achieved a certain parity – the twofold aspect of experiment and theory – that all other sciences enjoy" (Metropolis 1987, 129f.). The practice of executing matrices as simulations might have been the reason why it seemed natural to Shannon in 1948 to "run" the transition probabilities of English text.

A further extension of the method was the concept of "Hidden Markov Models," developed by Leonard Baum and Ted Petrie in 1966 (Baum and Petrie 1966). The system being simulated was assumed to be a Markov process with unknown properties, and the technique consisted of determining the probable hidden parameters from the observable ones. It allowed the very successful application of the theory of connected events to problems of speech-, optical character-, and in general pattern-recognition starting in the middle of the 1970s (Raviv 1967; Baker 1975; Jelinek, Bahl, and

[35] Cf. p. 337: "Mathematically, this complicated process [of interacting particles] is an illustration of a so-called Markov chain." Enrico Fermi, physicist and constructor of the first nuclear reactor, apparently used similar methods already in the 1930s, but never published them.

Mercer 1975),[36] and to genetic sequencing in the second half of the 1980s, when the field of bioinformatics emerged (cf. Blaisdell 1984; Borodovsky [Borodovskiy] et al. 1986).[37]

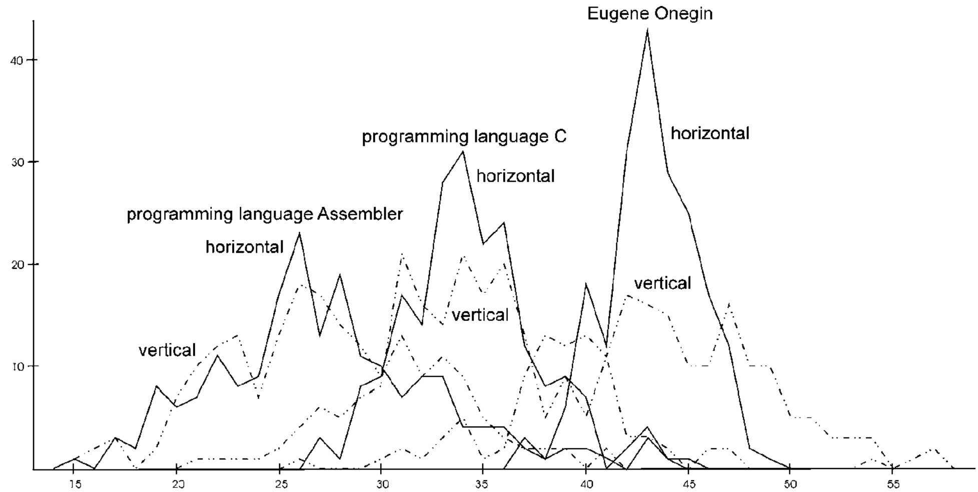
## V. The Limits of Chains

Markov's analysis finds in writing what it supersedes, in the twofold Hegelian sense of annihilation and conservation: orality, or more precisely its organ, the mouth.[38] The statistical phenomenon that the probability of a vowel is higher after a consonant and vice versa is due to the necessities of articulation. Therefore, written languages that are rarely spoken tend to use the recombinatoric potential of the alphabet to a higher extent. C, the programming language developed by Dennis Ritchie in 1972, uses tongue-twisters like "strstr" ("obtain first substring of a string") or "strrstr" ("obtain last substring of a string") in its utmost rationality and is similar to the letter sequences produced by Markov in the vertical count. The need of programmers to name a multitude of variables and functions as short as possible, because they will have to write those symbols over and over, results in an arbitrary language that comes very close to Shannon's ideal of the perfectly informative use of letters.[39] As in Hebrew or in American street signs ("Bklyn-Qns Expwy"), the first victims of the rationalization are the vowels. Text that is no longer spoken by a mouth to be heard by an ear marks one of the limits of the theory of chains. The following table shows the results of a Markovian analysis of the kernel of the Linux operating system, programmed in C ("C", from <http://www.kernel.org>), and of game software for the Atari in Assembler ("A", from <http://www.atariarchives.org/agagd/index.php>), along with the counts of *Eugene Onegin* (EO).[40] In all cases a sample of 20,000 letters was used (fig. 5).

In the analysis of the C source code, the dispersions are still different for the horizontal and the vertical count, but already less than in the literary text. Since the length of expressions in C is not limited, programmers use a lot of mnemonics in natural language:

"unsignedlongunusedacctneedcheckstaticintcheckfreespacestructfilefilestructstatfssbuf-intresintactlockkernelresacctactiveiffileacctneedcheckgotooutunlockkernelifvfsstatfsfile-fdentrydinodeisbsbufreturnresifsbuffbavailsuspendsbuffblocks." In Assembler, "words"

[36] Cf. Rabiner 1989, for a good introduction.
[37] Again, as can be seen from the names of the authors and the journal, the impulse originated from Russia.
[38] Cf. Link 2006 for a detailed development of this idea.
[39] As Shannon already pointed out, the redundancy of language is related to the existence of crossword puzzles. Cf. Shannon 1948, 25: "If the redundancy is zero any sequence of letters is a reasonable text in the language and any two-dimensional array of letters forms a crossword puzzle."
[40] Since a modern version of the novel was used, the results differ slightly from those of Markov.



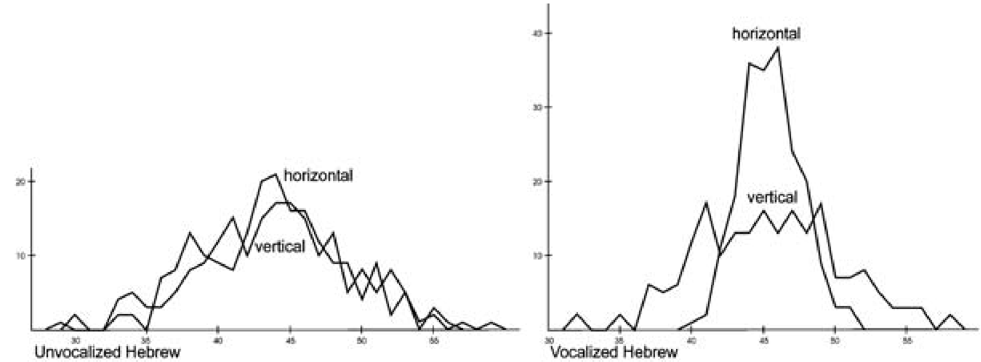

| | arithm. mean | sum of the squares of deviations horizontal | dispersion coefficient horizontal | sum of the squares of deviations vertical | dispersion coefficient vertical |
|---|---|---|---|---|---|
| EO | 43.19 | 1,022.8 | 0.21 | 5,788.8 | 1.18 |
| C | 34.39 | 2,117.3 | 0.47 | 4,721.35 | 1.04 |
| A | 27.29 | 7,016.7 | 1.77 | 5,660.75 | 1.42 |

**Fig. 5.** Comparison of the frequency of vowels in "Eugene Onegin" and in the programming languages C and Assembler.

are not allowed to be longer than a very few letters, so the text is much less readable and does not produce significant results in the numerical analysis:

"mphbccaaldastascloaaclcldaschiadctemphstaschistandlistclcldascloadcxsstasclostandlist-ldyaaclcldascloadcstasclostandlistyldaschiadcstaschistandlistyinyinyinycpyblaartsorgshpl-equshphequshpmlequshpmhequshpmolequshpmohequpmadrequmaplequ."

Markov's method also reaches its limits in languages where writing and speaking differ substantially. In non-vocalized Hebrew only the consonants are set down. In the following analysis of the Thora, the letters י, ו, ה, א and ע that usually accompany vowels are counted against all other signs in an edition without diacritics and compared to the same reckoning in the vocalized text, with the diacritics as vowels. The coefficients of dispersion only vary insignificantly in non-vocalized Hebrew. Therefore, Markov's method also determines the degree to which writing represents orality (fig. 6).




| | arithm. mean | sum of the squares of deviations horizontal | dispersion coefficient horizontal | sum of the squares of deviations vertical | dispersion coefficient vertical |
|---|---|---|---|---|---|
| A | 43.78 | 4,684.8 | 0.95 | 6,110.87 | 1.24 |
| B | 45.49 | 857.9 | 0.17 | 5,235.95 | 1.05 |

**Fig. 6.** Comparison of the frequency of vowels in non-vocalized (A) and vocalized Hebrew (B).

## References

Aeppli, Alfred. 1924. *Zur Theorie verketteter Wahrscheinlichkeiten. Markoffsche Ketten höherer Ordnung*. Zürich: Leemann.

Aksakoff, Serge [Aksakov, Sergey Timofeyevich]. [1858] 1960. *Years of Childhood* [Childhood Years of Bagrov's Grandson]. Translated by Alec Brown. New York: Vintage Books.

Aldrich, Mark. 1993. "Combating the Collision Horror: The Interstate Commerce Commission and Automatic Train Control, 1900–1939." *Technology and Culture* 34(1):49–77.

Alexanderson, Gerald L. 2000. *The Random Walks of George Pólya*. Washington: Mathematical Association of America.

Baker, James K. 1975. "The Dragon System – an Overview." *IEEE Transactions Acoustic Speech Signal Processing* (ASSP) 23(1):24–29.

Baudouin de Courtenay, Jan. [1895] 1984. "Versuch einer Theorie phonetischer Alternationen." In *Jan Baudouin de Courtenay, Ausgewählte Werke*, edited by Joachim Mugdan, 51-180. München: Fink.

Baum, Leonard E., and Ted Petrie. 1966. "Statistical Inference for Probabilistic Functions of Finite State Markov Chains." *Annals of Mathematical Statistics* 37:1554–1563.

Bernoulli, Jakob. 1713. *Ars Conjectandi, Opus Posthumum, Accedit Tractatus de Seriebus infinitis, et Epistola Gallice scripta de ludo Pilae recticularis*. Basileae: Impensis Thurnisiorum, Fratrum.

Bernstein, Serge [Bernshteyn, Sergey Natanovich]. 1922. "Sur le théorème limite du calcul des probabilités." *Mathematische Annalen* 85:237–241.

Bernstein, Serge. 1927. "Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes." *Mathematische Annalen* 97:2–59.

Bernstein, Serge. 1928. "Sur les sommes de quantités dépendantes." *Comptes rendus de l'Académie des Sciences de l'Union des Républiques Soviétiques Socialistes* (A):55–60.

Bessel, Friedrich Wilhelm. 1815. "Ueber den Ort des Polarsterns." In *Astronomische Jahrbücher für das Jahr 1818*, 233–241. Berlin: Königliche Akademie der Wissenschaften.

Blaisdell, B. Edwin. 1984. "Markov Chain Analysis Finds a Significant Influence of Neighboring Bases on the Occurrence of a Base in Eucaryotic Nuclear DNA Sequences both Protein-Coding and Noncoding." *Journal of Molecular Evolution* 21(3):278–288.

Boltzmann, Ludwig. [1877] 1909. "Über die Beziehung zwischen dem zweiten Hauptsatz der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht." *Sitzungsberichte der Kgl. Akademie der Wissenschaften in Wien, Mathematisch-Naturwissenschaftliche Klasse* 76:373–435. Reprinted in *Wissenschaftliche Abhandlungen von Ludwig Boltzmann*. Vol. 2: 1875–1881, edited by Fritz Hasenöhrl, 164–223. Leipzig: Barth.

Borodovsky [Borodovskiy], Mark Y., Y. A. Sprizhitskii [Sprizhitskiy], E. I. Golovanov, and A. A. Aleksandrov. 1986. "Statistical Patterns in Primary Structures of the Functional Regions of the Genome in Escherichia Coli. II. Nonuniform Markov Models." *Molecular Biology* (Molekulyarnaya Biologiya) 20:1024–1033.

Brown, Robert. 1828. "A Brief Account of Microscopical Observations Made in the Months of June, July and August, 1827, on the Particles Contained in the Pollen of Plants; and on the General Existence of Active Molecules in Organic and Inorganic Bodies." *Philosophical Magazine* 4:161–173.

Bru, Bernard. 2003. "Souvenirs de Bologne." *Journal de la Société française de statistique* 144(1–2):135–226.

Costello, Matthew J. 1988. *The Greatest Puzzles of All Time*. New York: Prentice Hall.

Davis, Martin, ed. 1965. *The Undecidable. Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*. Hewlett, NY: Raven.

Delone, Boris N. [1947] 2005. *The St. Petersburg School of Number Theory*. History of Mathematics, vol. 26. Providence, RI: American Mathematical Society.

Eggenberger, Florian. 1924. *Die Wahrscheinlichkeitsansteckung. Ph. D. diss., Technische Hochschule, Zürich. Mitteilungen der Vereinigung Schweizerischer Versicherungs-Mathematiker*. Bern: Stämpfli.

Eggenberger, Florian and George Pólya. 1923. "Über die Statistik verketteter Vorgänge." *Zeitschrift für Angewandte Mathematik und Mechanik* 3(4):279–289.

Einstein, Albert. 1905. "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen." *Annalen der Physik* 17:549–560.

Engel, Barbara Alpern. 1975. *Five Sisters: Women against the Tsar. The Memoirs of Five Young Anarchist Women of the 1870s*. New York: Knopf.

Flynn, Ann Marie and Louis Theodore. 2001. *Health, Safety and Accident Management in the Chemical Process Industries*. New York: Decker.

Fomenko, Anatoli T. 1994. *Empirico-Statistical Analysis of Narrative Material and Its Applications to Historical Dating*. Vol. 1: *The Development of Statistical Tools*. Dordrecht: Kluwer.

Fréchet, Maurice. 1926. "Une nouvelle représentation analytique de la répartition des revenus." *Bulletin de l'Institut International de Statistique* 22(3):547–549.

Fréchet, Maurice. 1938. *Méthode des fonctions arbitraires. Théorie des événements en chaîne dans le cas d'un nombre fini d'états possibles*. Paris: Gauthier-Villars.

Gaines, Helen Fouché. 1989. *Cryptanalysis. A Study of Ciphers and Their Solutions*. Mineola, NY: Courier Dover.

Galton, Francis. 1877. "Typical Laws of Heredity." *Nature* 15:492–495,512–514,532–533.

Gauss, Carl Friedrich. 1809. *Theoria motus corporum coelestium in sectionibus conicis solem ambientum*. Hamburg: Perthes and Besser.

Gibbs, Josiah W. 1902. *Elementary Principles in Statistical Mechanics, Developed with Especial Reference to the Rational Foundation of Thermodynamics*. New York: Scribner.

Hadamard, Jacques. 1906. "La mécanique statistique." *Bulletin of the American Mathematical Society* 12:194–210.

Hadamard, Jacques. 1928. "Sur le battage des cartes et ses relations avec la mécanique statistique." In *Atti del Congresso internazionale dei Matematici, Bologna* 5:133–139.

Hald, Anders. 1990. *A History of Probability and Statistics and Their Applications before 1750*. New York: Wiley.

Hald, Anders. 1998. *A History of Mathematical Statistics from 1750 to 1930*. New York: Wiley.

Hardy, Godfrey H., John E. Littlewood, and George Pólya. 1934. *Inequalities*. Cambridge: Cambridge University Press.

Hitt, Parker. [1916] 1976. *Manual for the Solution of Military Ciphers*. Laguna Hills, CA: Aegean Park Press.

Jelinek, Frederick, L. R. Bahl, and R. L. Mercer. 1975. "Design of a Linguistic Statistical Decoder for the Recognition of Continuous Speech." *IEEE Transactions on Information Theory* IT-21:250–256.

Kahn, David. 1967. *The Codebreakers. The Story of Secret Writing*. New York: Macmillan.

Keynes, John M. 1913. "Review of 'Wahrscheinlichkeitsrechnung' (Markoff 1912)." *Journal of the Royal Statistical Society* 76:113–116.

Keynes, John M. 1921. *A Treatise on Probability*. London: MacMillan.

Keynes, John M. 1926. *Über Wahrscheinlichkeit*. Translated by F. M. Urban. Leipzig: Barth.

Kolmogoroff, A. [Kolmogorov, Andrey Nikolayevitch]. 1931. "Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung." *Mathematische Annalen* 104:415–458.

Kolmogorov, A. N. 1947. "The Role of the Russian Science in the Development of the Theory of Probability." In *Sheynin* 1998, 139–156.

Legendre, Adrien-Marie. 1805. *Nouvelles méthodes pour la détermination des orbites des comètes*. Paris: Courcier.

Lexis, Wilhelm. [1879] 1903. "Über die Theorie der Stabilität statistischer Reihen." In *Abhandlungen zur Theorie der Bevölkerungs- und Moralstatistik*, edited by Wilhelm Lexis, 170–212. Jena: Fischer.

Liebmann, Heinrich. 1905. *Nichteuklidische Geometrie*. Leipzig: Göschen.

Link, David. 2004. "Poesiemaschinen / Maschinenpoesie." Ph.D. diss., Humboldt-University, Berlin and Academy for Arts and the Media, Cologne. <http://edoc.hu-berlin.de/docviews/abstract.php?lang=ger&id=25251>.

Link, David. 2005. "while(true). On the Fluidity of Signs in Hegel, Gödel, and Turing." In *Variantology 1. On Deep Time Relations of Arts, Sciences and Technologies*, edited by Siegfried Zielinski and Silvia Wagnermaier, 261–278. Cologne: König.

Link, David. 2006. "Traces of the Mouth. Andrei Andreyevich Markov's Mathematization of Writing." *History of Science* 44(145):321–348.

Lobatschefskij, N. J. [Lobachevskiy, Nikolay Ivanovich]. [1835] 1904. *Imaginäre Geometrie und Anwendung der Imaginären Geometrie auf einige Integrale*. Translated by H. Liebmann. Leipzig: Teubner.

Lobatschefskij, N. J. [1856] 1902. *Pangeometrie. Ostwalds Klassiker der exakten Wissenschaften*, Nr. 130. Translated by H. Liebmann. Leipzig: Engelmann.

Lobatschefskij, Nikolaj Iwanowitsch. 1898. *Zwei geometrische Abhandlungen*. Translated by Friedrich Engel. Leipzig: Teubner.

Lyapunov, Aleksandr. 1900. "Sur une proposition de la théorie des probabilités." *Bulletin de l'Académie Impériale des Sciences de St.-Pétersbourg* (5) 13(4):359–386.

Liapounoff, A. [Lyapunov, Aleksandr Mickhailovich]. 1900. "Sur un théorème du calcul des probabilités." *Comptes rendus de l'Académie des Sciences de St.-Pétersbourg* 132:126–127.

Lyapunov, Aleksandr M. 1902. *Neue Form des Theorems über die Grenze der Wahrscheinlichkeit. Mémoires de l'Académie Impériale des Sciences de St.-Pétersbourg* (8) 12(5). In Russian.

Markoff, A. [Markov, Andrey Andreyevich]. 1886. "Extrait d'une lettre adressée à M. Hermite." *Annales Scientifiques de l'École Normale Supérieure* (3) 3:81–88.

Markoff, André. 1889. "Sur les séries . . . " *Comptes rendus de l'Académie des Sciences, Paris* 59:934–935.

Markoff, André. 1891a. "Sur la théorie des équations différentielles linéaires." *Comptes rendus de l'Académie des Sciences, Paris* 63:790–791.

Markoff, André. 1891b. "Sur la théorie des équations différentielles linéaires." *Comptes rendus de l'Académie des Sciences, Paris* 63:1024–1025.

Markoff, André. 1891c. "Sur les équations différentielles linéaires." *Comptes rendus de l'Académie des Sciences, Paris* 63:685–688.

Markoff, André. 1892. "Sur la série hypergéometrique." *Comptes rendus de l'Académie des Sciences, Paris* 64:54–55.

Markoff, André. 1894. "Démonstration d'un théorème de Tchébycheff." *Comptes rendus de l'Académie des Sciences, Paris* 70:1032–1034.

Markoff, A. A. 1896. *Differenzenrechnung*. Translated by Th. Friesendorff and E. Prümm. Leipzig: Teubner.

Markoff, André. 1910. "Recherches sur un cas remarquable d'épreuves dépendantes." *Acta Mathematica* 33:87–104.

Markoff, A. A. 1912. *Wahrscheinlichkeitsrechnung*. Translated by H. Liebmann. Leipzig: Teubner.

Markoff (Markov), A. 1915. *Bicentenaire de la loi des grands nombres. 1713–1913. Démonstration du second théorème-limite du calcul des probabilités par la méthode des moments. (Supplément à la 3ième édition russe du "Calcul des probabilities"*. St.-Petersbourg: Impr. de l'Académie imp. des sciences. In French.

Markov, Andrey A. 1885. "On Some Applications of Algebraic Continued Fractions." Ph.D. diss., Académie Impériale des Sciences de St.-Pétersbourg. In Russian.

Markov, Andrey A. 1891a. *Calcul des différences finies. Section I. L'interpolation*. St. Petersburg. In Russian.

Markov, Andrey A. 1891b. *Calcul des différences finies. Section II. Equations aux différences finies et leurs sommations*. St. Petersburg. In Russian.

Markov, Andrey A. 1898. "The Law of Large Numbers and the Method of Least Squares. Excerpt from a Correspondence with Prof. A. Wassilief [Vasil'yev]." *Bulletin de la société physico-mathématique de Kasan* (2) 8:110–128. In Russian.

Markov, Andrey A. 1900a. *Calculus of Probability*. St. Petersburg. In Russian.

Markov, Andrey A. 1900b. "Sur une probabilité a posteriori." *Communications de la société mathématique de Kharkov* (2) 7(1):23–26. In Russian.

Markov, Andrey A. 1906. "Extension de la loi de grands nombres aux événements dependants les uns de autres." *Bulletin de la société physico-mathématique de Kasan* (2) 15(4):135–156. In Russian.

Markov, Andrey A. 1907a. "Recherches sur un cas remarquable d'épreuves dépendantes." *Bulletin de l'Académie Impériale des Sciences de St.-Pétersbourg* (6) 1(16):61–80. In Russian. Slightly different version in French Markoff 1910.

Markov, Andrey A. 1907b. "Sur quelques cas des théorèmes sur les limites de probabilité et des espérances mathématiques." *Bulletin de l'Académie Impériale des Sciences de St.-Pétersbourg* 1(16):707–714. In Russian.

Markov, Andrey A. 1908. *L'extension des théorèmes limites du calcul des probabilités à une somme des valeurs en chaîne. Mémoires de l'Académie Impériale des Sciences de St.-Pétersbourg* (8) 22(9). In Russian. German translation: "Ausdehnung der Sätze über die Grenzwerte in der Wahrscheinlichkeitsrechnung auf eine Summe verketteter Größen." In Markoff 1912, Appendix 2, p. 272–298.

Markov, Andrey A. 1910. *Recherches sur le cas général d'épreuves liées en chaîne. Mémoires de l'Académie Impériale des Sciences de St.-Pétersbourg* (8) 25(3). In Russian. English translation: "An Investigation of the General Case of Trials Connected into a Chain." In Sheynin 1998, 169–193.

Markov, Andrey A. 1911a. "Sur les valeurs liées qui ne forment pas une chaîne véritable." *Bulletin de l'Académie Impériale des Sciences de St.-Pétersbourg* (6) 5(2):113–126. In Russian. German translation: "Über verbundene Größen, die keine eigentlichen Ketten bilden." In Markoff 1912, Appendix 3, 299–311.

Markov, Andrey A. 1911b. "Sur un cas d'épreuves liées en chaîne multiple." *Bulletin de l'Académie Impériale des Sciences de St.-Pétersbourg* (6) 5(3):171–186. In Russian.

Markov, Andrey A. 1912. "Sur les épreuves liées en chaîne par les événements laissés sans observation." *Bulletin de l'Académie Impériale des Sciences de St.-Pétersbourg* (6) 6(8):551–572. In Russian.

Markov, Andrey A. 1913a. "Essai d'une recherche statistique sur le texte du roman 'Eugène Onĕgin', illustrant la liaison des épreuves en chaîne." *Bulletin de l'Académie Impériale des Sciences de St.-Pétersbourg* (6) 7(3):153–162. In Russian.

Markov, Andrey A. 1913b. "Ein Beispiel statistischer Forschung am Text 'Eugen Onegin', zur Verbindung von Proben in Ketten." In Link 2004, 187–203.

Markov, Andrey A. 1913c. *Calculus of Probability*, 3rd edn. St. Petersburg. In Russian.

Markov, Andrey A. 1913d. *Bicentenaire de la loi des grands nombres. 1713–1913. Démonstration du second théorème-limite du calcul des probabilités par la méthode des moments. (Supplément à la 3ième édition russe du "Calcul des probabilités."* St.-Petersbourg: Impr. de l'Académie imp. des sciences. In Russian.

Markov, Andrey A. 1916. Sur une application de la méthode statistique. *Bulletin de l'Académie Impériale des Sciences de St.-Pétersbourg* (6) 10(4):239–243. In Russian.

Markov, Andrey A. 1924. *Calculus of Probability*, 4th edn. Moscow: Editions de l'Etat. In Russian.

Markow, Andrej A. [1913] 1969. "Beispiel einer statistischen Untersuchung über den Text 'Eugen Onegin', das den Zusammenhang von Merkmalen in der Kette veranschaulicht." *Exakte Ästhetik* 6:52–59.

Metropolis, Nicholas. 1987. "The Beginning of the Monte Carlo Method." *Los Alamos Science* Special Issue 15:125–130.

Metropolis, Nicholas, and Stanislaw Ulam. 1949. "The Monte Carlo Method." *Journal of the American Statistical Association* 44(247):335–341.

Moivre, Abraham de. [1733] 1738. "Approximatio ad summam terminorum binomii $(a + b)^n$ in seriem expansi." In *The Doctrine of Chances*, 2nd ed., edited by Abraham de Moivre, 243–59. London: Woodfall.

Morozov, Nikolai A. [Nikolay Aleksandrovich] [1880] 1987. "The Terrorist Struggle." In *The Terrorism Reader*, edited by Walter Laqueur and Yonah Alexander, 72–84. New York: Penguin.

Morozov, Nikolai A. 1915. "Spectres linguistiques." *Bulletin de l'Académie Impériale des Sciences de St.-Pétersbourg, Section of Russian Language* 20(3):93–134. In Russian.

Morozow, Nikolaus [Morozov, Nikolay Aleksandrovich]. 1912. *Die Offenbarung Johannis. Eine astronomisch-historische Untersuchung*. Stuttgart: Spemann.

Newton, Isaac. [1669] 1711. "De analysi per aequationes numero terminorum infinitas." In *Analysis per quantitatum series, fluxiones, ac differentias*, edited by Isaac Newton, 1–22. London: Pearson.

Ollier, Alexandre. 2002. *La cryptographie militaire avant la guerre de 1914*. Paris: Lavauzelle.

Ondar, Kh. O., ed. 1981. *The Correspondence between A. A. Markov and A. A. Chuprov on the Theory of Probability and Mathematical Statistics*. New York: Springer.

Petruszewycz, Micheline. 1981. *Les chaînes de Markov dans le domaine linguistique*. Geneva: Slatkine.

Phillips, Howard and David Killingray, eds. 2003. *The Spanish Influenza Pandemic of 1918–19*. Routledge, New York: New Perspectives.

Pólya, George. 1913. "Über einige Fragen der Wahrscheinlichkeitsrechnung und über bestimmte Integrale." *Math. és. phys. lapok* 22:53–73, 163–219. In Hungarian.

Pólya, George. 1919. "Wahrscheinlichkeitstheoretisches über die 'Irrfahrt'." *Mitteilungen der Physikalischen Gesellschaft Zürich* 19:75–86.

Pólya, George. 1921. "Über eine Aufgabe betreffend die Irrfahrt im Strassennetz." *Mathematische Annalen* 84:149–160.

Post, Emil. 1936. "Finite Combinatory Processes. Formulation I." In Davis 1965, 288–291.

Preussische Akademie der Wissenschaften and Deutsche Mathematiker-Vereinigung, eds. 1.1868(1871) – 68.1942(1944). Jahrbuch über die Fortschritte der Mathematik. Berlin: de Gruyter. Online version: Electronic Research Archive for Mathematics (ERAM): *Jahrbuch Database*. 12 April 2006<http://www.emis.de/MATH/JFM/full.html>.

Pushkin, Aleksandr S[ergeyevich]. [1833] 1975. *Eugene Onegin: A Novel in Verse*. Translated by Vladimir Nabokov, 2 vols. Princeton: Princeton University Press.

Rabiner, Lawrence R. 1989. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition." *Proceedings of the IEEE* 77:257–285.

Rado, Tibor. 1962. "On Non-Computable Functions." *Bell System Technical Journal* 41:877–884.

Ranger, Terence O., and Paul Slack, eds. 1992. *Epidemics and Ideas: Essays on the Historical Perception of Pestilence*. Cambridge: Cambridge University Press.

Raviv, J. 1967. "Decision Making in Markov Chains Applied to the Problem of Pattern Recognition." *IEEE Transactions on Information Theory* 3:536–551.

Schulz, Günther. 1934. "Zur Theorie des Galtonschen Brettes." *Zeitschrift für Physik* 92:747–754.

Seneta, Eugene. 1982. "Bernstein, Sergei Natanovich." In *Encyclopedia of Statistical Sciences*. Vol. 1, edited by Samuel Kotz and Norman Johnson, 475–476. New York: Wiley.

Seneta, Eugene. 1996. "Markov and the Birth of Chain Dependence Theory." *International Statistical Review* 64:255–263.

Shannon, Claude. [1948] 1993. "A Mathematical Theory of Communication." In *Claude Elwood Shannon. Collected Papers*, edited by Neil A. Sloane and Aaron D. Wyner, 5–83. Piscataway, NJ: IEEE Press.

Shannon, Claude Elwood. 1949. "Communication Theory of Secrecy Systems." *Bell System Technical Journal* 28(4):656–715.

Sheynin, Oscar. 1988. "A. A. Markov's Work on Probability." *Archive for History of Exact Sciences* 39:337–377.

Sheynin, Oscar, ed. 1998. *From Markov to Kolmogorov. Russian Papers on Probability and Statistics*. Engelsbach: Hänsel-Hohenhausen. Mikrofiche.

Sheynin, Oscar, ed. 2004. *Alexandr Chuprov. Statistical Papers and Memorial Publications*. Berlin: NG Verlag. 10 April 2006<http://www.sheynin.de/download/C3_Chuprov.pdf>.

Siegmund-Schultze, Reinhard. 2004. "A Non-Conformist Longing for Unity in the Fractures of Modernity. Towards a Scientific Biography of Richard von Mises (1883–1953)." *Science in Context* 17:333–370.

Sljusareva, Natalja A., ed. 1972. "Lettres de Ferdinand de Saussure à J. Baudouin de Courtenay." *Cahiers Ferdinand de Saussure* 27:7–17.

Spitzer, Frank. 1976. *Principles of Random Walk*, 2nd ed. New York: Springer.

Turing, Alan M. 1935. "On the Gaussian Error Function." Fellowship dissertation, King's College. Unpublished. 10 April 2006<http://www.turingarchive.org/browse.php/C/28>.

Turing, Alan M. 1936. "On Computable Numbers, with an Application to the Entscheidungsproblem." *Proceedings of the London Mathematical Society* (2) 42:230–265.

Wolfram, Stephen. 2002. *A New Kind of Science*. Champaign, Ill.: Wolfram Media.

Zabell, Sandy L. 1995. "Alan Turing and the Central Limit Theorem." *American Mathematical Monthly* 102:483–494.