Hist. Sci., xliv (2006)

TRACES OF THE MOUTH: ANDREI ANDREYEVICH MARKOV'S MATHEMATIZATION OF WRITING

David Link
Centre for Art and Media Technology, Karlsruhe

INTRODUCTION

This article discusses in detail two works by the Russian mathematician Andrei Andreyevich Markov Sr (1856–1922). They represent an early and momentous attempt to understand the phenomenon of language in mathematical terms.[1] Outside of the strictly mathematical field, Markov's achievements are only very rarely discussed.[2] In the works, he counted the frequency of vowels and consonants in Pushkin's *Eugene Onegin* and another text, and analysed the results with the mathematical tools of the probability theory of his time. In what follows, I give a brief account of the role that letters played in probability theory up to this point. The understanding of language in these concepts was so weak that it did not even allow very simple problems to be solved. I then describe Markov's analysis in detail.

Since 1906, Markov's work had extended certain concepts of probability theory, which were considered to apply only to independent trials, to the field of dependent variables on a purely theoretical basis.[3] In Pushkin's text Markov found for the first time material to verify his assumptions empirically. Since this was his primary interest, he made no further comment on the meaning of his findings. The first astonishing result was that the distribution of vowels and consonants followed a 'normal' distribution. Although Markov did not say as much, this means that at the source of language lies a random process. I attempt to find an explanation for this in the lectures of the Swiss linguist Ferdinand de Saussure, which he gave at approximately the same time and which offer a helpful theory on the collective genesis of language. Even though it is probable that Markov and Saussure were unacquainted with each other's work, they shared a strong interest in formalization and an approach that is differential rather than substantial. By applying Markov's analysis to randomly selected words, I demonstrate first that it is not the individual style of an author that produced the observed randomness. It is the fact, as stated by Saussure, that language is formed in an unconscious, collective process. Certain individuals begin to speak differently and their changes to the language may or may not be accepted by others.

Markov's second result was that the dispersion of this random distribution is much smaller than would be expected. Again, Markov applied the theoretical formulae of his earlier work only to verify their validity, and did not enquire as to the reason for this phenomenon. Saussure's theory provides an explanation: the few individuals who start to speak differently are subject to the physical constraints imposed by the mouth and thus cannot recombine letters completely at random. Therefore,

0073-2753/06/4403-0321/$10.00

Markov's method also determines the degree to which written text represents orality and this allows a much firmer grip on language than probability theory had achieved before Markov.

To generate a completely random text as a comparison and to destroy any dependence between the letters, Markov wrote the text row by row into a table and read out the columns vertically. I show that this technique was inspired by the cryptography of his time and that even Pushkin, the author studied by Markov, used this technique to encrypt the politically dangerous tenth chapter of *Eugene Onegin*.

A SHORT HISTORY OF THE MATHEMATIZATION OF WRITING

"There are very few things, which we know, which are not capable of being reduced to a mathematical reasoning; and when they cannot, it's a sign our knowledge of them is small and confused." If this statement by the Scottish mathematician and physician John Arbuthnot, written in 1692, is correct, then our knowledge of the human use of language remains limited and confused.[4] Despite the fact that numerous different cultures use the same symbols for letters and numbers, there is a deep divide between the two domains. Numerals can be expressed in words but not vice versa. For mathematics, language is a system of a non-describable order, although through probability calculations it has the ability to discover regularities even in highly chaotic data.[5] The belief in the universality and power of probability theory, for example, comes out very strongly in the following statement by Francis Galton in 1889:

> I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the 'Law of Frequency of Error'. The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along.[6]

However, it has not the power to explain manifestations of subjectivity that are characterized by a radically different order. Since François Viète, letters had been used to denote numbers,[7] but the other way round was unknown. Before 1913, the theory of probability had little cognisance of the letters of the alphabet. In the early years, it was concerned with calculating the odds of winning at games of chance.[8] A further application was astronomy. As Anders Hald writes, "Observational and mathematical astronomy give the first examples of parametric model building and the fitting of models to data. In this sense, astronomers are the first mathematical statisticians."[9] In a parallel development, John Graunt initiated the application to mass phenomena with his examination of death statistics in London 1604–62, with particular reference to the effects of the Black Death. In later years, this gave rise to attempts at putting life insurances on a reliable footing by calculating the probable life expectancy for different age groups.[10] Through the work of Adolphe Quetelet, Francis Galton, and Wilhelm Lexis, the application of statistics was expanded to include a large proportion of the phenomena of human life; not, however, to the social phenomenon of language.[11]

Before Markov, the symbols of the alphabet were considered in only two ways. First, in the period when their main subject was games of chance, they were the material for combinatorial calculations. In 1718, de Moivre interrogated their natural order in "Problem 35" of his *Doctrine of chances*:

> Any number of letters a, b, c, d, e, f, & c, all of them different, being taken promiscuously as it happens: to find the probability that some of them shall be found in their places according to the rank they obtain in the alphabet; and that others of them shall at the same time be displaced.[12]

This derived from the convention of assigning capital letters to the various possible results of a sample. When their probability was calculated *a priori*, that is, leaving aside empirical material, words appeared merely as exceptional cases of randomness. Each letter represented one selection among 26 equally possible choices and it was immaterial in which order they appeared. For this reason, in 1770 d'Alembert raised doubts about probability calculation as a whole:

> In order to expand my idea with yet more clarity and precision, I suppose that we find on a table some printed characters arranged in this way: C o n s t a n t i n o p o l i t a n e n s i b u s, *or* a a b c e i i i l n n n n n o o o p s s s t t t u, *or* n b s a e p t o l n o i a u o s t n i s n i c t n, these three arrangements contain absolutely the same letters: in the first arrangement they form a known word; in the second they form no word at all, but the letters are disposed according to their alphabetical order, and the same letters are found as many times in sequence as they are found in turn in the twenty-five characters which form the word; finally, Constantinopolitanensibus in the third arrangement, the characters are pell-mell, without order, and at random. Now it is first certain that, *mathematically speaking*, these three arrangements are equally possible. It is not less that all sane men who would cast a glance on the table where these three arrangements are supposed to be found, will not doubt, … that the first is not the effect of chance.[13]

That stochastically a meaningful word was just as probable as a meaningless one, ran counter to common sense. Six years later, in 1776, Laplace provided an answer to this problem. In a direct reply to the "very fine objections" of "Mr d'Alembert" he wrote:

> Suppose that on a table we find letter types arranged in the order INFINITÉSIMAL; the reason which leads us to believe that this arrangement is not the effect of chance can come only from this that, physically speaking, it is less possible than the others, because, if the word *infinitésimal* were not used in any language, this arrangement would be neither greater, nor less possible, and yet we would suspect then no particular cause. But, since this word is in use among us, it is

> incomparably more probable that some person has thus arranged these letters than that this arrangement is due to chance.[14]

This argument of Laplace's eloquently glossed over the circumstance that for statistics, linguistic combinations were actually little more than very improbable combinations, and a person's choice of one word rather than another was merely a statistical phenomenon or an inexplicable fact. This branch of mathematical science could only state that a word is "in use among us"; it could not predict or understand it. In d'Alembert's reservations, however, there is a sense of what probability calculation could not grasp, or only very imperfectly: the order of language, including the individual that uses it.

The second application of numerals to letters developed in cryptology. Around A.D. 850, the Arab scholar Al-Kindi described such a method:

> One way to solve an encrypted message, if we know its language, is to find a different plaintext of the same language long enough to fill one sheet or so, and then we count the occurrences of each letter. We call the most frequently occurring letter the 'first', the next most occurring letter the 'second', the following most occurring the 'third', and so on, until we account for all the different letters in the plaintext sample.
>
> Then we look at the cipher text we want to solve and we also classify its symbols. We find the most occurring symbol and change it to the form of the 'first' letter of the plaintext sample, the next most common symbol is changed to the form of the 'second' letter, and so on, until we account for all symbols of the cryptogram we want to solve.[15]

Known today as frequency analysis, this procedure makes use of the insight, which can only be gained *a posteriori*, that in empirical documents single letters are not equally probable, but appear with differing frequencies. This made early substitution codes, like the Caesar cipher, easy to crack. In this mathematization, writing is a sequence of characters, which — if one takes sufficiently long sequences — appear in characteristic frequencies and thus remain recognizable in spite of their substitution with other characters. However, this method also did not permit differentiation between words that mean something and nonsensical words that contain the same letters. Up to the twentieth century, techniques for detecting regularities in cryptograms continued to be refined and eventually also considered the frequency of letter combinations.[16] However, Markov was the first to develop a complete theory that takes into account the connections between letters.

MARKOV'S MATHEMATICAL CROSSWORDS

On 23 January 1913, Markov gave a lecture to the physical-mathematical faculty of the Imperial Academy of Sciences in St Petersburg. The occasion was to celebrate the bicentenary of the book that had formulated the fundamental theorem of probability: Jakob Bernoulli's *Ars conjectandi*, virtually completed by 1690, but only published posthumously in 1713 by his nephew Niklaus because of family quarrels. However, in 1913, Russia was celebrating first and foremost a different anniversary: three centuries of the ruling house of Romanov. Obviously, this gave the physical-mathematical faculty's event a political dimension. Usually, mathematicians wrote texts about numbers, in which the former merely served to elucidate the latter. Markov, however, shifted the periphery of his field to centre stage and lectured on calculations that he had performed upon a literary text.[17] In the years before, he had made theoretical extensions to probability theory, which traditionally applied only to independent events, to the field of trials that are dependent on each other. In a literary text he found for the first time material that would allow him to verify his assumptions empirically. His object of study was the most important work by Russia's national writer A. S. Pushkin, *Eugene Onegin*. Applying the rigorous tool of statistics to this famous literary work, Markov went one step further than Pushkin, whose realistic style was directed against Romanticism, and delivered an analysis where signifiers were taken as the basis of all meaning. Several years later, Markov published a second analysis of a literary work in the Appendix to the fourth edition of his *Probability theory*: *Childhood years of Bagrov's grandson* (1858) by the lesser-known Russian author Sergei Timofeevic Aksakov.[18]

Both texts are autobiographical memoirs — Pushkin's work deals with the recollections of an 18-year-old and Aksakov's with memories of early childhood. Both writers conceal this by giving their main protagonists fictitious names: Eugene Onegin and Bagrov's Grandson. Memory binds the present moment to the past; there are no independent samples in life. Aksakov distinguishes between "fragmentary memories" (single images) and "connected memories", which were at the same time the titles of the first two chapters.[19] A further circumstance that the two writers had in common is that they both wrote their works while living in a situation of isolation. Because of his political poetry, Pushkin was banished from St Petersburg in 1820 and exiled to Kishinev in contemporary Moldavia. He had begun work on *Eugene Onegin*, when he was sent to Odessa (in today's Ukraine) and then to his mother's country seat at Mikhaylovskoe in Northern Russia. Aksakov's isolation resulted from the fact that he had been blind for ten years and dictated the memoir to his daughter.[20]

The passage from Aksakov's book containing 100,000 letters of the alphabet, which Markov analysed, describes a journey in a carriage to Orenburg, where the mother, who is gravely ill, wishes to consult a doctor. The author lauds this form of transport, because it evokes thoughts of one's own life-story being a series of events in the form of a chain:

> Travel — what a wonderful thing it is! Its powers are … snatching him out of the environment he is in …, first directing his attention to himself, then to the remembrance of the past, and finally, to dreams and hopes that lie ahead.[21]

Under the aegis of the here and now, travel enables the human subject to step outside the sequence of the syntagmatic present and into thoughts of one's own biography; in this way, the subject is liberated from time.

A recurrent theme is the Russian language — Bagrov, the son of the family and the narrator, is an avid reader. His emotions while fishing are described thus: “With what a keen eye and what curiosity I gazed at those objects, new to me, how quickly I grasped their significance, and how easily and soundly I learned all their names!”[22] However, Russia is a multinational state and language is by no means uniform. Already at the beginning of their journey, the family encounters Chuvashes, Tatars, and Bashkirs. Statistically, this is to be expected as these groups represent the largest ethnic minorities. The family’s servant attempts to adapt to the situation verbally: “Thinking to make himself more understandable, Evseich now began to distort his Russian frightfully, mixing in Tatar words. He wanted them to tell him where we could find worms for fishing.” He gets the worms as well as the following answer: “Ai-ai! very much good fish catch here!”[23] Language is not presented as a static system, but rather as a structure that permanently changes, especially through contact with other languages and people’s inability to find the right word. In *Eugene Onegin*, too, there are instances of language reshaping. In verse 35 of the first chapter, early one morning while the hero is still asleep, of course, at the German baker’s the “васисдас” (transcribed in the Latin alphabet as “Wasisdas” — Whatsit) opens. The editor of the German translation of *Eugene Onegin* comments on this word, obviously derived from German, thus: “This word form was introduced in Russia during the foreign invasion 1812 and was used to denote the little flapping or sliding windows of the German bakeries in Petersburg.”[24] In French, the loan-translation “vasistas” is documented since 1798.[25]

In these texts, Markov investigated the frequency of vowels and consonants and the possible connections between them. Since his research on Pushkin is more detailed, I shall focus on this paper and mention his analysis of Aksakov’s text only where the techniques or results differ. Markov divided the first 20,000 letters of Pushkin’s novel into 200 groups of one hundred letters each and wrote each group in a square table with ten rows and ten columns (see Figure 1). Not included were the hard signs (ъ) and soft signs (ь), which are not pronounced independently but modify the pronunciation of the preceding letter, punctuation, and spaces. He counted the vowels in each column, combined two at a time (the first and the sixth, second and seventh, third and eighth, fourth and ninth, fifth and tenth), and wrote down the five sums underneath each other in a vertical column. Thus, each sum represented the number of vowels among 20 letters, which were separated by four letters in the text. Combined, they showed the number of vowels in a sequence of 100 letters of the novel. Markov made tables each with five of these columns, that is, representing 500 letters from the novel, and represented the entire text in 40 such tables, which fill an entire page. Here, he first added the numbers vertically and entered the total at the bottom in boldface. The square matrix of the table also permitted addition of the numbers horizontally and thus a calculation of how many vowels there were per 100 letters that were separated by four letters in the text. He entered these totals, also in bold, in the final column. Finally, Markov added the vertical and horizontal sums together and, to save space, entered the total minus 200 in the bottom right-hand corner of each table. As it was

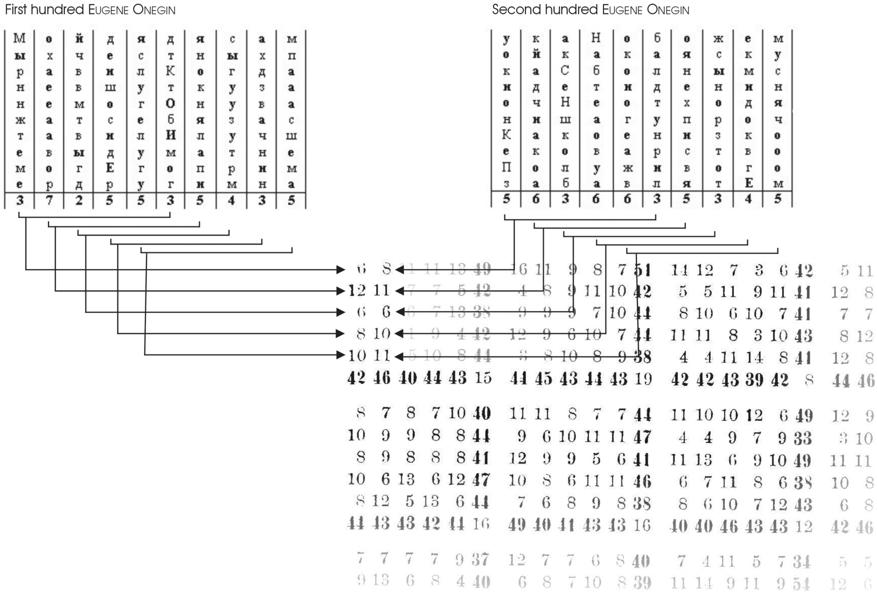


FIG. 1. Markov's counting method (analysis of diagram in Markov, "Essai" (ref. 1)).

the same 500 letters that were being counted using different methods, the addition of the last lines and last columns gave the same result.

The text was broken up according to a specific scheme — at every tenth letter. This imitated the text's form, that of a novel in verse. The existing caesuras in the language were replaced by the regular and artificial ones of mathematical counting. Originally, the Latin word *versus* meant the turn of the plough at the end of a furrow and derives from the verb *vertere*, to turn. In contrast to the style of 'prose',[26] where the only direction is forward, the flow of letters in a poem — particularly in rhyming verse — returns constantly to the beginning of the line. This facilitates a second reading of the text in addition to the syntagmatic one: one that breaks its linearity. The written form of language inherits linearity from the spoken word and, at the same time, integrates it, for on book pages, stone tablets, or papyrus rolls, the acoustic thread along which perception makes its way is turned into the two-dimensional. In a similar way to how the novel's traveller is liberated from being a slave to time on the journey and can think in free associations, the written form gives the reader the freedom to choose his or her own way through the rectangular field of letters, to go back or forward, skim through, or read every which way. Unlike speaking, writing is not a linear medium.[27]

The constraints exerted by orality are two-fold and concern the organs of the mouth and the ear. In the case of that which is mistakenly taken for the transmitter, this means that always only one word at a time — and thus one word after another — can be spoken. Jacques Derrida termed this restriction of speech 'angustia'.[28]

Prerequisite for progress in a straight line is the Hegelian fading away of the voice, or, expressed differently, the fact that the medium — air — cannot store information.[29] Only an echo realizes memory, similar to how the engineers of the first computers did: as a self-propagating impulse that is continually refracted and mirrored between two points in a liquid carrier medium and in this way survives.[30] The nightmare of a continually echoing space, which, since the 1960s, is possible to create using taped sound loops and other special effects devices, allows the opposite of a soothing, idealistic fade — the expression of just one single sound that will continue to sound for all eternity and is thus stored.[31] Because of its constraints, the field of what is audible has either no or only one storage space. During actual experience, time is zero-dimensional, because its expansion is present only in its passing. What is voiced by the mouth disappears immediately in the general medium or silences everything — *out of memory*.

This restriction applies equally to the ear. A listener cannot re-wind or fast-forward the medium of transmission. It has no storage capability nor can it be accessed by him. Thus, he is obliged to remain slavishly in the constant now of speech, which runs by him. Sorry, what did you say?

The 'return' breaks the linearity of Chronos on and in space. Language has always been set down in the form of cut strips. The make-up of text in a second dimension probably took place *c*. 1500 B.C. and this is already reflected in the Bible. There are numerous instances of acrostics, where the initial letters of the lines form the complete alphabet and were an aid, for example, in the Psalms, in committing texts to memory.[32] Besides mnemotechniques, others that use proper names served as dedications to other persons, self-praise by the author, or for concealing messages in general. With the advent of computers and word processing programmes, the function of 'return' changed from its original one of line advance and return to the line's beginning into the final confirmation of a command.[33]

A pre-condition of Markov's operation was not only the temporality of speech made up upon a book page. Columns are formed only when the characters are distributed discretely and regularly spaced on the page's area. This is a quality that oral expression lacks and, to a certain extent, also handwriting. The organic flow protects the line from being read vertically. The first typewriter with Cyrillic letters, the Modell 8, was exported as of 1903 by the German Adler Company.[34] Like cryptography, the mathematician's studies are a writing game with discrete characters.[35]

The expansion of language into the second dimension, which is effected by the transcription of the voice into writing, Markov implemented as a paper-automaton and employed it to calculate "other connections"[36] within the text material with the aim of bringing out the properties of the natural syntagmatic sequence.

Markov's interest centred first on the results of the consecutive calculation. A table shows how often the numbers from 37 to 49 occurred in the final rows of the small tables, which contained the number of vowels in the sequences of 100 letters. The modal, in statistics, the most frequent value, was 43. From this, Markov calculated the arithmetic mean. He multiplied the differences to the modal with the frequency of their occurrence, added the results, and divided this by the total of tests. The mean deviation from the most frequent number thus obtained was added to this and gave the mean sought, 43.19, as the average number of vowels in 100 letters. The unusual method of deriving this from the modal had the advantage — for someone doing calculations by hand — of working with much smaller quantities than is customary today, by multiplying the values themselves. After dividing by 100, p — the mean probability of any single letter's being a vowel — was approximately 0.432.[37]

However, the actual number of vowels in the various groups of 100 letters differed. The value that indicates how far single numbers deviate from their mean is called 'dispersion' in mathematics. To determine this dispersion, Markov used a technique, which was first published in 1805 by Legendre in *Nouvelles méthodes pour la détermination des orbites des comètes* but is, however, linked with the name of Gauss, who elucidated it in 1809 in his *Theoria motus corporum coelestium*: the method of least squares.[38] The deviation of each value from the arithmetic mean of 43.19 was squared, and thus positive; next, the average of these differences was calculated by multiplying them by their frequency of occurrence, adding them, and finally, dividing them by their total number. The sum of the squares of deviations was 1022.8; division by 200 gave 5.114. By drawing the root from the result, one usually arrives at the so-called standard deviation as a relative measure of the dispersion. The mathematician, however, simply worked with the square of this value, the so-called variance. The number of vowels per 100 letters thus differed on average approximately $\sqrt{5.114} = 2.26$ from its mean.[39]

LANGUAGE AS A RANDOM PROCESS

In order for it to be permissible to apply the method of least squares, traditionally and until the work of Markov the quantities studied had to be independent of each other. This held true in this case, because the number of vowels in the first 100 letters had virtually no influence on how many there were in the second 100. There was dependence only between the last letters of the first batch and the first letters of the second for they were considered in linear sequence. This was why the dispersion followed the normal distribution — the first surprising discovery presented in Markov's lecture (see Figure 2). In 1733, de Moivre developed the normal distribution in connection with problems relating to games of chance. It approximated the expected binomial distribution for independent random elementary errors, which was complicated to calculate for large numbers of samples.[40] As its properties were known, it was possible to calculate the probability that values fell in certain intervals. For example, half of them usually fell in the range of the 0.6745-fold of the standard deviation to the left and the right of the arithmetic mean. This discovery 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 errors in astronomical measurements and was termed 'probable error'.[41] Thus, it could be assumed that 100 of Markov's 200 values were $2.26 \times 0.67 = 1.5$ away from the average. And actually there were 103 values between $43.2 - 1.5 = 41.7$ and

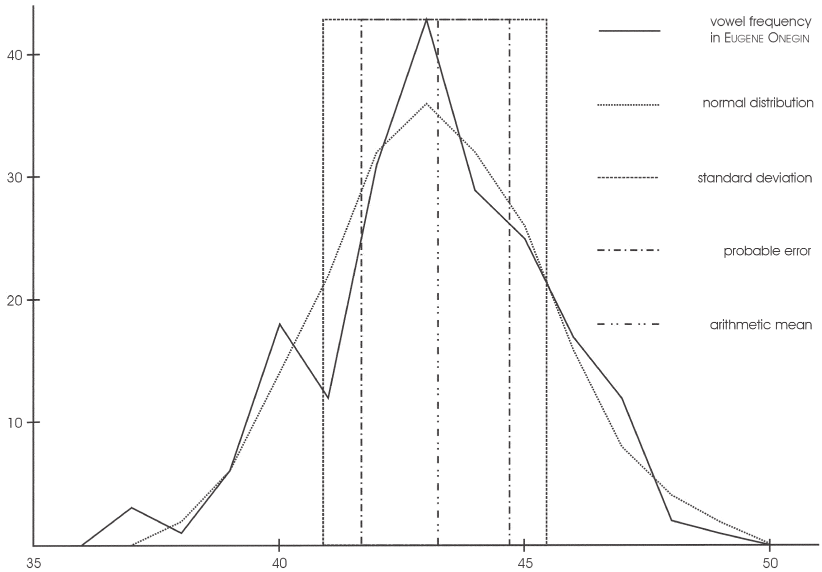


FIG. 2. The values of the horizontal calculation compared to the normal curve.

43.2 + 1.5 = 44.7, which was a good fit with the theory.

The method of least squares serves usually to correlate measurements better with theoretical models, for example, geometrical ones in the case of geodesy, and in this way it subtracts out of the data the instruments together with their inevitable inaccuracies.

> However carefully one takes observations of the magnitudes of objects in nature, the results are always subject to larger or smaller errors.… Such errors come from the imperfections of our senses and random external causes, as when shimmering air disturbs our fine vision. Many defects in instruments, even the best, fall in this category; e.g., a roughness in the inner part of a level, a lack of absolute rigidity, etc.[42]

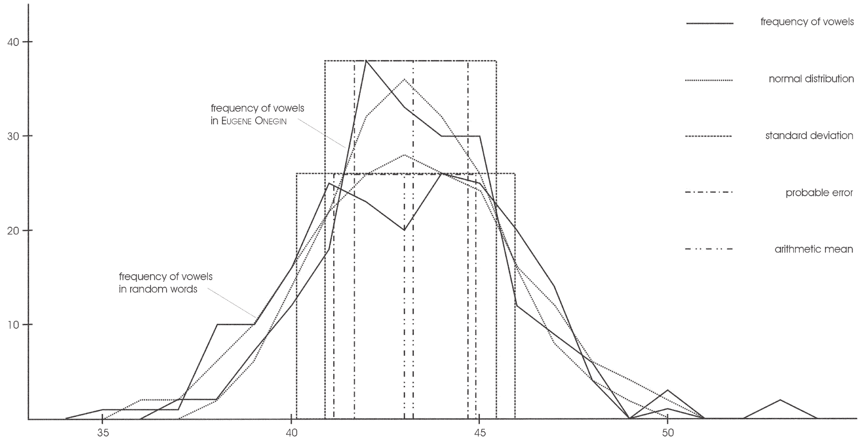
The law of error allowed empiricism to be corrected by theory, in that it converted numerical measured values into possible ranges of results and thus weakened them. The observed quantities were typically continuous and thus could be made more precise and approximated infinitely. By contrast, Markov's material was precisely defined on the level that he was dealing with. Obviously, printed letters exhibit certain irregularities, but this did not concern this enquiry. Markov did not measure, he counted various discrete entities. He might make mistakes in this, but these too were not thematic. The deviation of the number of vowels from the average was simply understood as error. Whereas in other cases it was a question of using the Gaussian method to make the real fit closer to the symbolic, here the symbolic was taken as the real and its internal irregularity re-interpreted as an error of measurement.

That the distribution of the number of vowels in the groups of 100 letters corresponded to a random dispersion is something that demands an explanation. For language is certainly not determined by the contingency of an individual wielding tools, but is the intentional expression of a human subject. The subject can act within the oral space freely and without resistance so that he even masters it without a feed-back loop. The accident of real collisions or mishaps, as evidenced by measurement errors, is precisely that which the oral space excludes in an incomparable manner.[43] Why is language statistically subject to the same contingency as the meeting of physical objects with an observer or balls falling in the Japanese game of Pachinko?[44] This might be plausible for deviations in pronunciation, but not for counts of discrete letters of the alphabet. Markov explained his astounding finding simply with the independence of the samples. The number of vowels in the first batch of 100 letters does not determine how many there will be in the second. This explanation, however, is not sufficient to explain the observed normal distribution. That the frequencies are dispersed in this manner forces one to advance the radical thesis that the source of language is a random process. This is not to be expected. In their critique of the application of the normal curve to social and psychological phenomena, Fashing and Goertzel, who think that it serves as the foundation of a theory of social inequality, write:

> the normal bell curve is 'normal' only if we are dealing with random errors. Social life, however, is not a lottery, and there is no reason to expect sociological variables to be normally distributed.[45]

At the level of the speaking or writing human individual, the random theory does appear completely absurd, because people implement their intentions in such actions unhindered and say what they want to say. However, if one looks at the social aspect — that if one wants to be understood, one has to use the language everyone else is using — then there is an explanation. Although Pushkin was free to choose what he expressed and which chains of letters he strung together, he utilized a repertoire of well-formed words that was already established at the time he was writing. Moreover, the manner in which words were put together was to a certain extent governed by syntactic and grammatical rules. Pushkin, or anyone, expressed personal thoughts in a general medium. In a text there are two forming processes at work, which differ considerably in their temporality, spatiality, and origination: the collective process of language formation over time and, conditioned by this, the individual process to articulate specific contents. At first, one is unable to decide where the randomness might come in.

The following experiment provides clarification in this matter. From an alphabetical list of Russian words, as used by hackers and system administrators to test the safety of passwords, random entries are chosen until the total number of characters reaches 20,000.[46] The result is a completely nonsensical string of valid expressions



	Arithm. Mean	Sum of the Squares of the Deviations	Standard Error	Range of Error	Number of Values in this Range
EO	43,19	1045,3	1,5	41,6 – 44,7	101
RW	43,06	1.702,3	1,9	41,1 – 45,0	94

FIG. 3. Comparison of the frequency of vowels in *Eugene Onegin* (*EO*) and random words (RW).

in Russian, devoid of all syntax or grammar. The act of an author, which possibly produces randomness through selecting certain words to formulate a specific content, is omitted. The sample thus produced is then examined using Markov's method. The results are shown in Figure 3.[47]

The results of the calculations performed on the random words hardly differ in their proximity to the normal curve from those of Markov on *Eugene Onegin*. This applies equally to all of Markov's other calculations. Thus, they do not concern a particular author's style and manner of arranging words but language in general. The conclusion, therefore, from the dispersion of values is that the process of formation, which generates the ensemble of all allowable chains of letters, is a random process.

Since Markov made no comment whatsoever on this randomness, we must search for an explanation in the linguistic theory of his time. From 1907 to 1911, the Swiss linguist Ferdinand de Saussure held a series of lectures at the University of Geneva, which were later published posthumously by his students as *Course in general linguistics* and became one of the most influential language theories of the twentieth century. As far as we know, there was never any direct contact between the Russian mathematician and the Swiss linguist, nor did they know each other's work.[48] Therefore, the close correspondences described below should be considered as a kind of 'parallel invention' that sheds considerable light on the meaning of Markov's calculations.

Saussure compared diachronic changes of language to chess and established the following agreements:

(a) One piece only is moved at a time. Similarly, linguistic changes affect isolated elements only.
(b) In spite of that, the move has a repercussion upon the whole system. It is impossible for the player to foresee exactly where its consequences will end....
(c) In a game of chess, any given state on the board is totally independent of any previous state of the board. It does not matter at all whether the state in question has been reached by one sequence of moves or another sequence.

In the following passage, he declared the process of language formation to be random:

> There is only one respect in which the comparison is defective. In chess the player *intends* to make his moves and to have some effect upon the system. In a language, on the contrary, there is no premeditation. Its pieces are moved, or rather modified, spontaneously and fortuitously.… If the game of chess were to be like the operations of a language in every respect, we would have to imagine a player who was *either unaware of what he was doing or unintelligent*.[49]

The origin of these spontaneous changes lies in changes to spoken language initiated by a few people:

> [*E*]*verything which is diachronic in language is only so through speech.* Speech contains the seed of every change, each one being pioneered in the first instance by a certain number of individuals before entering into general usage.… But not all innovations in speech meet with the same success.[50]

The question why these persons begin to speak differently "is one of the most difficult tasks in linguistics" and cannot be explained completely.[51] "[B]lind forces of change" have an effect on the organization of the system of signs.[52] As the main source of mutations, Saussure identified changes to the sound of words, to which there are no limits.

> If one attempts to evaluate the effect of these changes, one soon sees that it is unrestricted and incalculable. It is impossible, in other words, to foresee where they will end.[53]

For Saussure, the "most important factor of change" was "imagination over a gap in memory".[54] When speakers do not know how to answer the above question about the "wasisdas", they transform the language and make modifications. These are spontaneous in the sense that they do not follow a particular goal. They do not know whether their modifications will be taken over by others or what further effects they may elicit. Such actions are not purposive nor can their causality be predicted. It is, therefore, plausible that in total this results in a random distribution. A further point against the above-mentioned caveat by Fashing and Goertzel — not to apply the normal curve to social phenomena — is that even collective processes, where the actors possess greater clarity regarding the goals and effects of their actions, result in Gaussian dispersions, for example, events on the stock exchange.[55]

The randomness of this genesis leaves its marks on the chains of letters, which can be discerned in Markov's calculations but which he did not comment on. The findings of the mathematician would have pleased the linguist Saussure, who described language as a "mechanism, which involves interrelations of successive terms", "like the functioning of a machine in which the components all act upon one other even though they are arranged in one dimension only".[56]

CONSTRAINTS ON RANDOMNESS

Markov now turned his attention from the groups that were independent of each other to the dependence of the single letters in these groups. He analysed the dispersion on this level to see whether the connectedness had an influence on it. He calculated the deviation of each letter from the mean probability of being a vowel (0.432) and then divided the variance of the sample (5.114) by the number of letters it contained (100). The result was 0.05114. For independent events, however, theoretically one would expect a considerably higher dispersion. Random samples, which are independent of each other and only have two possible outcomes, approach the binomial distribution. The expected variance is calculated from the multiplication of the probabilities and, in this particular case, was $0.432 \times 0.568 = 0.245376$. When the theoretical result was compared with the empirical result derived from the text, the coefficient of dispersion, which relates the two values, was 0.208. Thus, the variance of each single letter was five times smaller than one would expect (see Figure 4).

This method of analysing the stability of statistical series was developed in 1879

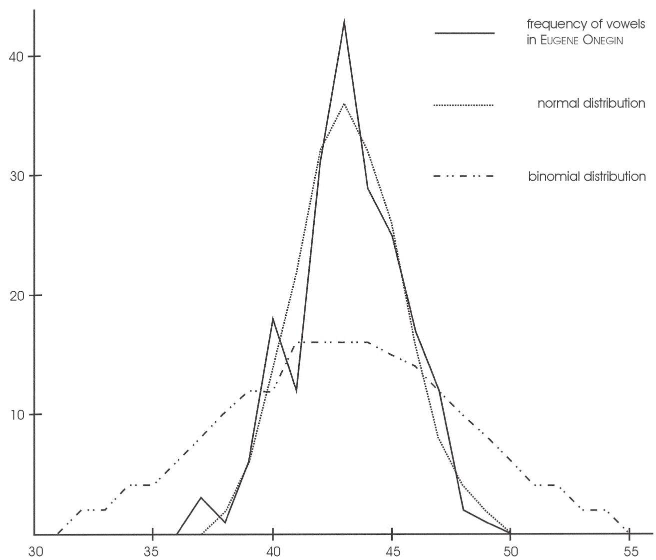


FIG. 4. The frequency of vowels in *EO* compared to the binomial distribution.

by Wilhelm Lexis, a mathematician, social scientist, and economist. Dissatisfied with contemporary common practice in statistics, whereby material was assumed uncritically to be normally distributed, Lexis was the first to compare the empirically-found standard deviation with the theoretically predicted deviation and to form their coefficient Q. When the two corresponded, that is, when $Q = 1$, Lexis spoke of a "normally random distribution": "It evolves only because a constant base probability becomes apparent with only the incertitude in the observed numbers which is admissible by analogy with a correct game of chance."[57] He judged such a series to be maximally stable as concerns the underlying probability. However, if the observed dispersion was higher than the theoretical dispersion, Q was greater than 1. Lexis termed this "supernormal" and explained it with "the fact that the 'normal random' fluctuations of the base probabilities combined with the physical ones", that is, changes of the probabilities themselves. For an "undernormal" dispersion, where Q was smaller than 1, as with Markov, Lexis gave his verdict that "it would indicate that the mass phenomenon under consideration is internally connected or subject to certain regulating interventions or norms. It would more or less belong to the field of systematic order or commanding laws",[58] that is, not that of probability calculation.

This was precisely so in the case of single letters, which we are looking at here. In the first counting method, consecutive series of 100 letters formed groups, which were dependent on one another in the sense that a certain number of preceding consonants forced the next letter to be a vowel and vice versa. This led to the circumstance that their total number fluctuated much less than lots drawn in a game of chance with constant probabilities. Unlike Lexis, who banned this area from the magisterium of stochastics, Markov attempted to grasp it mathematically.

To this end, he counted the frequencies of the sequences vowel–vowel and consonant–vowel. It turned out that the probability of a vowel varied according to which type of letter preceded it. When this was a consonant, it was 0.663; when a vowel, only 0.128. Since 1906, Markov had studied simple chains as groups of samples in which each member determined the next; however, he did so entirely theoretically because of a lack of suitable empirical material.[59] Using an equation from his *Investigation of a notable case of dependent samples*, Markov calculated from the difference δ of the two values, which he called p_0 and p_1, a theoretical coefficient of dispersion of 0.3, which was very close to the empirical one of 0.208, and in this way confirmed for the first time his mathematical deductions experimentally.[60] Thus the empiric dispersion could be explained much better if one assumed that in Pushkin's text the sequences of letters were chains rather than independent entities.[61] In the 'simple chain' that he postulated for the purpose of approximation, each element was dependent on the preceding one.[62]

However, the value could be even better approximated following the hypothesis that each letter was dependent on the pair of letters preceding it. This was investigated by Markov in his paper of 1911, "On a case of samples connected in multiple chain". He counted the frequency, $p_{1,1}$, of sequences of three vowels and of three consonants, $q_{0,0}$, in Pushkin's text. After inserting these values in an equation from the above paper, he found the theoretical coefficient of dispersion to be 0.195, which was even closer to the empirical coefficient of dispersion.[63] The first arrangement, therefore, gave 200 nearly independent tests each with 100 enchained letters.[64]

The dispersion was normally distributed; however, at the level of individual letters it was five times smaller than would be expected for a random process with this mean. To employ the main metaphor of stochastics, this resembled "an irregular game where the results of the single series of experiment are pushed closer to the mean value as would be expected in a game of chance with constant probabilities by voluntary influence. Typical quantities of this sort do not belong to the quantities of probability but are only of the same form".[65] Language appeared as a game with a marked deck of cards. The random generator was formed atypically in such a way that it produced a distribution that was too constrained. Similar methods can be utilized to detect manipulated games of chance without having to take the mechanism, for example, roulette, into account. The case found here was incomprehensible to probability theory, because the oscillation of the probabilities themselves cancelled out the Gaussian noise. Lexis underlined this by demonstrating that in his equation for calculating the total dispersion, $R = \sqrt{(r^2 + p^2)}$, when this was smaller than the random dispersion, r, the factor p, which gave the physical deviation, became imaginary.[66]

For the observed mathematical limitation of variance, reasons are also found in Saussure's lectures. He not only described the ceaseless proliferation of signifiers through modification of sounds, but also phenomena that, in turn, constrained this:

> For the entire linguistic system is founded upon the irrational principle that the sign is arbitrary. Applied without restriction, this principle would lead to utter chaos. But the mind succeeds in introducing a principle of order and regularity into certain areas of the mass of signs. That is the role of relative motivation.[67]

Saussure differentiated between a lexicological and a grammatical pole between which the process of formation moved. Different languages exhibited different trends. Chinese, for example, embodied the lexicological pole, and Indo-Germanic languages and Sanskrit the grammatical. Concerning the production of grammaticality, he stated: When a single modification of sound entered general usage, alternation and analogy produced a series of similar sounding phenomena and in this way created subsequent regularity. Similar to Freudian rationalization, these processes inscribed a spontaneous, non-purposive, yet collective mutation with an apparent regularization through first creating examples. However, one cannot see how these phenomena might give rise to a more regular distribution of vowels.

This is accomplished by a self-evident limitation, which Saussure did not mention, that constrained sound changes from the outset. The few persons, who begin spontaneously to speak in a different way, are constrained by the mouth, the organ of utterance. The semiologist Saussure mentioned this in his observations on phonetics:

> For it is not always within our power to pronounce as we had intended. Freedom to link sound types in succession is limited by the possibility of combining the right articulatory movements.[68]

Saussure demanded strict mathematization, similar to that which Markov carried out a few years later:

> To account for what happens in these combinations, we need a science which treats combinations rather like algebraic equations. A binary group will imply a certain number of articulatory and auditory features imposing conditions upon each other, in such a way that when one of them varies there will be a necessary alteration of the others which can be calculated.[69]

This "chain of speech"[70] was bracketed together with the organs of speech:

> [A] normal continuous sequence … is characterized by a succession of graduated abductions and adductions, corresponding to the opening and closing of articulators in the mouth.[71]

THE VERTICAL *LALULA* AND CRYPTOGRAPHY

To verify his findings concerning the connection between the observed dispersion and the dependence of the letters, Markov then proceeded to analyse a completely different arrangement of the text material in which the connection between the letters was destroyed (see Figure 5). He turned his attention to the sums in the last columns of the main table, which gave the number of vowels per 100 letters, and which one got when one looked only at every fifth letter of the 500 letters that each small table represented. The elements of this group were only weakly mutually dependent because they were far apart in the text. However, the five sums within one batch of five hundred were strongly mutually dependent because the individual letters stood

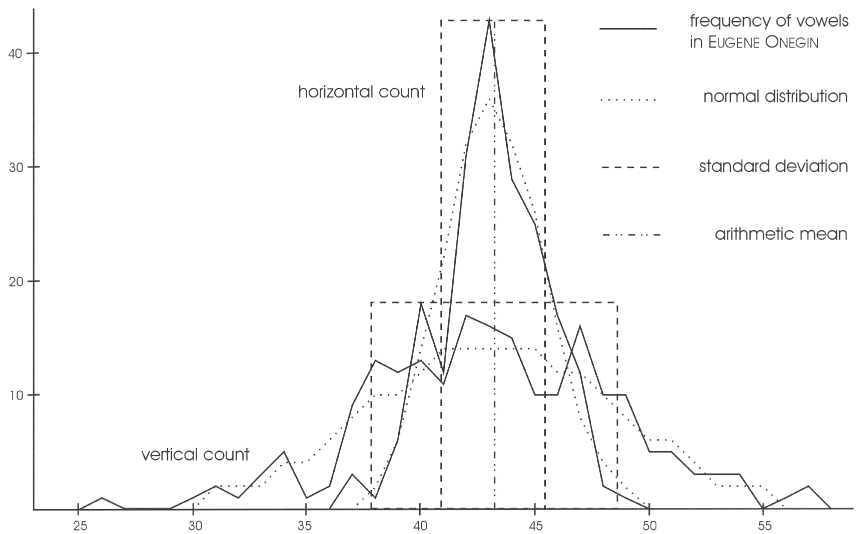


FIG. 5. The horizontal and vertical count compared.

in close proximity. The first contained the letters 1, 6, 11, 16, 21, 26, and so on; the second contained the letters that followed immediately after — 2, 7, 12, 17, 22, 27, etc. All 100 letters of the second were dependent on the first 100. Between the tables there existed weaker connections. The first 100 of the second table contained only 50 letters (in column 1), which followed the last group of the first (those in column 9). The second arrangement thus gave 200 strongly dependent samples of 100 non-enchained letters. The number of vowels per 100 letters fluctuated between 26 and 57. The values were now much more dispersed; the table had to be split up because it was now too wide to fit on the page. The arithmetic mean was still 43.2 — the 20,000 letters examined were the same ones. The sum of the squares of deviations from this number was much higher: 5788.8. By modifying an equation from his *Probability theory*, Markov demonstrated that, contrary to the traditional view, the law of large numbers as well as the Gaussian method could also be applied to dependent quantities.[72] This expansion of stochastics to include cases that went against the dogma of independent events and were usually excluded for that reason was pursued by Markov in numerous papers throughout his life. Through dividing the above sum by the number of samples, 200, the mean variance of the numbers in the last columns was calculated to be 28.944. It was more than five times greater than in the first arrangement. Divided further by 100 (each value describes 100 letters), the result was 0.28944 mean dispersion of each single letter, which differed only slightly from the theoretical expectation of the binomial formula, $0.432 \times 0.568 = 0.245376$. The dispersion coefficient was 1.18: it was thus in Lexis's sense normally random and demonstrated that the single letters of the samples represented mutually independent events with the typical Gaussian random dispersion.[73] A similar result would be expected if 200 batches of 100 balls were to be taken from an urn containing 4319 white and 5681 black balls.[74]

Already in the count of the groups of 100 it was apparent that the direction in which one proceeded made a difference and had considerable consequences. The columns exhibited a greater deviation from the mean value than the rows. By constructing series of each tenth letter, the coherence of the words was lost. In the articulation of language, the necessity of making a vowel follow a sequence of consonants, and vice versa, did not apply here. What remained were disparate and discrete letters, devoid of meaning. This produced exceptional statistical phenomena, such as "myrnnschteme" in the first column with only three vowels among ten letters or "ochaeeaawor" in the second with the amazing number of seven. These nonsensical monsters of words could be taken from Morgenstern's *Lalula* of 1905 or the typographic poems of Raoul Hausmann after 1917: "klekwapufzi", or "fmsbwtözäu".[75]

The comparison material of independent samples, which allowed Markov to highlight the peculiarities of language, he found through one of the oldest known coding operations. Around 450 B.C., in Lysander's Sparta, an instrument called a *scytale* was in usage that Plutarch described as follows:

> When the ephors send an admiral or general on his way, they take two round pieces of wood, both exactly of a length and thickness, and cut even to one another; they keep one themselves, and the other they give to the person they

> send forth; and these pieces of wood they call *scytales*. When, therefore, they have occasion to communicate any secret or important matter, making a scroll of parchment long and narrow like a leathern thong, they roll it about their own staff of wood, leaving no space void between, but covering the surface of the staff with the scroll all over. When they have done this, they write what they please on the scroll, as it is wrapped about the staff; and when they have written, they take off the scroll, and send it to the general without the wood. He, when he has received it, can read nothing of the writing, because the words and letters are not connected, but all broken up; but taking his own staff, he winds the slip of the scroll about it, so that this folding, restoring all the parts into the same order that they were in before, and putting what comes first into connection with what follows, brings the whole consecutive contents to view round the outside. And this scroll is called a *staff*, after the name of the wood, as a thing measured is by the name of the measure.[76]

In cryptology, this type of encryption is called a 'transposition cipher' because only the order of the letters changes, not the letters themselves as in substitution ciphers. It is the connections between the letters that are erased. The variant used by Markov is known as columnar transposition and a double version of this cipher was used, for example, by the revolutionaries who murdered Czar Alexander II in 1881. It was called 'nihilist transposition' after them and was "the most popular cipher of the Russian underground".[77] Modified versions of this cipher were in use until recent times, for example, the ÜBCHI cipher of the German Wehrmacht in the First World War or as a tool of many secret agents after the Second World War.[78] Even today, double columnar transposition is regarded as difficult to decipher.[79] It is safe to assume that Markov knew of this cipher and that it provided inspiration for his study.

Cryptography played a role in the history of the poem *Eugene Onegin*; moreover, precisely this method. In 1904, the encrypted fragments of a tenth chapter were presented to the autograph section of the St Petersburg Academy of Sciences (see Figure 6). The literary historian Piotr Osipovich Morozov, who became a corresponding member of the Academy in 1912, deciphered Pushkin's text three years before Markov's lecture. He published the plaintext in 1910 in the Academy journal *Puskin i ego sovremenniki*. It was a harsh critique of the Czar and a description of the Decabrists' circles. Pushkin, who was already persecuted on political grounds, judged the content so explosive that he destroyed the manuscript on 19 October 1830, five years after the attempted *coup d'état*. A short time later, he re-wrote the chapter from memory, but this time in code (see Figure 5).

It is most likely that Markov knew about the sensational discovery and deciphering of Pushkin's tenth chapter. Nabokov gave this description of the cryptogramme:

> A column of sixteen lines ... representing the first lines of stanzas I–X and XII–XVII. Under this, separated by a horizontal dash, another column ... representing the second lines of stanzas I–IX. Two sets of lines ... in the left-hand margin, the lines parallel to the margin; the lower marginal set represents the second lines of

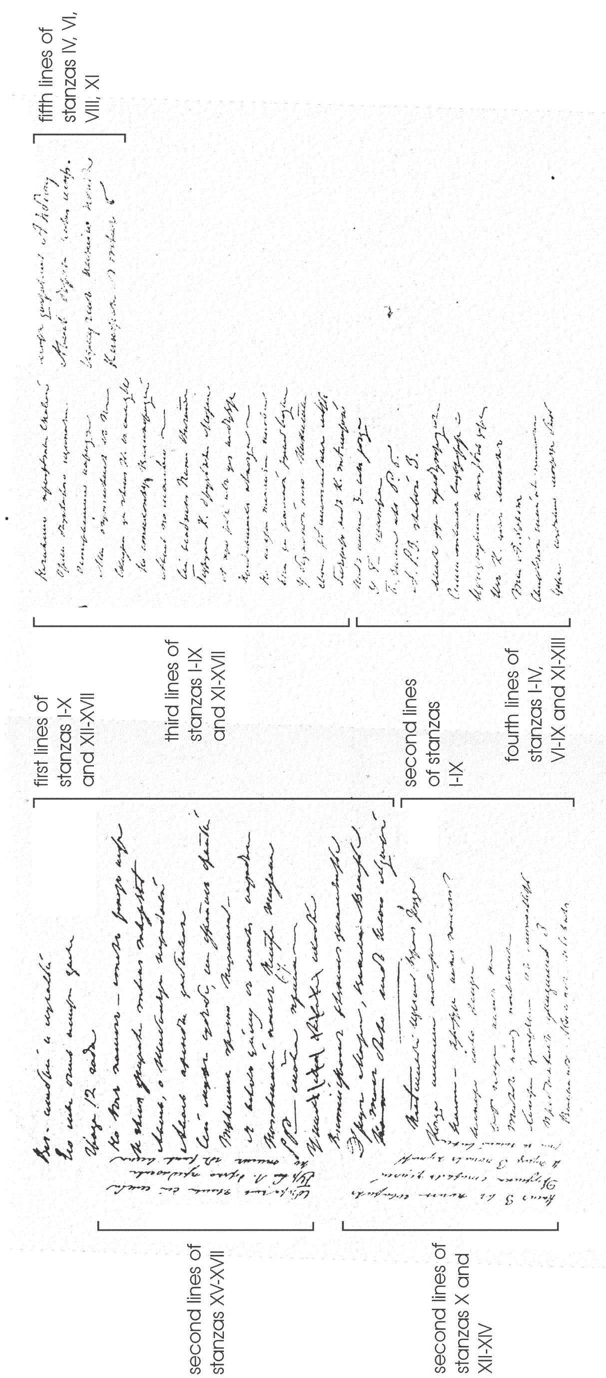


FIG. 6. Pushkin's cryptogram of the tenth chapter.[80]

> stanzas X, XII–XIV, and the upper marginal set represents the second lines of stanzas XV–XVII. A column of twenty-seven lines ... down the left-hand side of the page, representing the third lines of stanzas I–IX and XI–XVII, followed (without any gap or dash) by the fourth lines of stanzas I–IV, VI–IX, and XI–XIII. A column of four lines ... at the top of the right-hand side of the paper, representing what I take to be the fifth lines of stanzas IV, VI, VIII, XI.[81]

Pushkin combined lines of the poem, which occupied the same positions in the stanzas, to conceal their context and meaning. He arranged whole paragraphs in lines underneath each other, in which each verse had its own column, and then read them vertically. He attempted to disguise particularly disparaging passages further by using abbreviations: "O R[ússkiy] glúp[ïy] násh na[ród]" — "O our R[ussian] stup[id] na[tion]".[82] In Nabokov's opinion, he capitulated after encoding the fifth verse: "I also suggest that he soon noticed that something was very wrong with his cipher and in utter disgust gave up the whole matter."[83] The method of the mathematician Markov, which combined letters that were separated in the text by four letters, thus imitated exactly Pushkin's abandoned attempt to encipher his poem, where verses with the same distance from each other were brought together. However, that Markov combined the vertical counts of ten letters each so that each fifth letter was selected, also had a different reason. This compression was necessary because otherwise the results for 20,000 letters would not have fit onto one page. On the other hand, stronger compression would have resulted in part in three-figure sums. This did not affect the mathematical results because the combination of letters, which were separated by four others, was almost as weak as if separated by nine.

SUMMARY

Markov's analysis found in text something that writing supersedes, in the two-fold sense of destroying and preserving: orality, more precisely, in the form of its physical organ — the mouth. The traces of the speech organ influence the written text in two ways. On the one hand, it is free to mutate individual sounds and see if these are accepted by others. On the other, it is constrained in general by the fact that it can only articulate certain combinations. This constraint becomes very apparent when that which is formed in speech is fixed discretely in written text. On the two-dimensional page of printed letters, the potential for re-combination, and thus the reservoir of chains of letters that are possible, but not in use, becomes strikingly obvious. This knowledge can then be used to conceal messages by transposition cyphers, as Pushkin did with the tenth chapter of *Eugene Onegin*.

Markov proved that language can be better grasped when one includes the relationships between the elements. Saussure formulated this in a more radical way:

> *In the language itself, there are only differences.* Even more important than that is the fact that, although in general a difference presupposes positive terms between which the difference holds, in a language there are only differences, *and no positive terms.*[84]

For the first time in mathematics the use of signs received treatment as being differential. It was not single letters that were examined as substantial, as in early cryptanalysis, but their connections. In this way, d'Alembert's problem, mentioned at the beginning of this paper, was at least partially solved. Transition probabilities now allow us to calculate that the word "constantinopolitanensibus" uttered by a person is more probable than "nbsaeptolnoiauostnisnictn". Although both contain the same letters, that which is unsayable is sorted out. The method, therefore, also determines the degree to which text represents orality.

ACKNOWLEDGEMENTS

The author thanks Prof. Dr Siegfried Zielinski for guidance and Gloria Custance for her help with the translation of this article.

REFERENCES

Translations are by the author unless otherwise indicated.

1. Although these works form the basis of many mainly mathematical papers and are of undeniable importance, there are no easily accessible translations into Western European languages. The numerous references to the texts in the scientific literature almost without exception refer to the Russian originals. Thus, for my Ph.D. thesis I made a new translation into German, which is included in the thesis as an appendix: A. A. Markov [1913], "Ein Beispiel statistischer Forschung am Text 'Eugen Onegin' zur Verbindung von Proben in Ketten", in: David Link, "Poesiemaschinen / Maschinenpoesie" (Inaugural Diss., Humboldt-Universität Berlin and Kunsthochschule für Medien Cologne, 2002; http://edoc.hu-berlin.de/docviews/abstract.php?lang=ger&id=25251), 187–203. A rather clumsy translation appeared 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 *Wahrscheinlichkeitsrechnung*. *Cf.* A. A. Markov, "Beispiel einer statistischen Untersuchung über den Text 'Eugen Onegin', das den Zusammenhang von Merkmalen in der Kette veranschaulicht", *Exakte Ästhetik*, vi (1969), 52–59; A. A. Markov, "Sur un cas remarquable des épreuves liées en chaîne", in: *Bicentenaire de la loi des grands nombres: 1713–1913* (St Petersburg, 1913, in French), 56–66. Russian original: A. A. Markov, "Essai d'une recherche statistique sur le texte du roman 'Eugene Onegin', illustrant la liaison des épreuves en chaîne", *Bulletin de l'Académie Impériale des Sciences de St.-Pétersbourg*, vii/3 (1913, in Russian), 153–62. An English translation and an article investigating the propagation of Markov's ideas to Western Europe by the same author is forthcoming in David Link, "Chains to the West: Markov's Theory of Connected Events and its transmission to Western Europe", *Science in context*, xix/4 (2006).
2. His work is discussed in Anders Hald, *A history of mathematical statistics from 1750 to 1930* (New York, 1998). For a good bibliography and a more complete overview, see Oscar B. Sheynin, "A. A. Markov's work on probability", *Archive for history of exact sciences*, xxxix (1988), 337–77.
3. A. A. Markov [1906], "Extension de la loi de grands nombres etc.", in: Yu. V. Linnik (ed.), *A. A. Markov: Selected works* (Leningrad, 1951, in Russian), 339–61. A. A. Markov, "Recherches sur un cas remarquable d'épreuves dépendantes", *Bulletin de l'Académie Impériale des Sciences de St.-Pétersbourg*, i/16 (1907, in Russian), 61–80; a slightly different version was published in French in: *Acta mathematica*, xxxiii (1910), 87–104. A. A. Markov [1910], "Recherches sur le cas général des épreuves liées en chaîne", in: Linnik (ed.), *A. A. Markov: Selected works* (in Russian), 465–507; English translation: "An investigation of the general case of trials connected into a chain", in: Oscar B. Sheynin, *From Markov to Kolmogorov: Russian papers on probability*

and statistics (Engelsbach, 1998, microfiche), 169–93. A. A. Markov, "Sur les valeurs liées qui ne forment pas une chaîne véritable", *Bulletin de l'Académie Impériale des Sciences de St.-Pétersbourg*, v/3 (1911, in Russian), 113–26; German translation: "Über verbundene Größen, die keine eigentlichen Ketten bilden", in: A. A. Markov, *Wahrscheinlichkeitsrechnung* (Leipzig/Berlin, 1912), 299–311. A. A. Markov, "Sur un cas d'épreuves liées en chaîne multiple", *Bulletin de l'Académie Impériale des Sciences de St.-Pétersbourg*, v/2 (1911, in Russian), 171–86. A. A. Markov, "Sur les épreuves liées en chaîne par les événements laissés sans observation", *Bulletin de l'Académie Impériale des Sciences de St.-Pétersbourg*, vi/8 (1912), 551–72.

4. Cited in Anders Hald, *A history of probability and statistics and their applications before 1750* (New York, 1990), 183.
5. Its universality is due to its explanatory weakness. It does not attempt to postulate laws or forces, but only describes quantitative distributions.
6. Francis Galton, *Natural inheritance* (London, 1889), 66. Thus the normal distribution can be seen as an empirical confirmation of Kant's concept of beauty.
7. See François Viète [1591], *Introduction to the analytical art*, in: Jacob Klein, *Greek mathematical thought and the origin of algebra* (Cambridge, 1968), 315–53. The founder of modern algebra was also a cryptologist; see Simon Singh, *The code book* (New York etc., 1999), 45.
8. See Hald, *History of probability* (ref. 4), 33–80. The main works in this connection are Girolamo Cardano's *Liber de ludo aleae* (1564), the correspondence between Blaise Pascal and Pierre de Fermat in 1654, Christiaan Huygens's *De ratiociniis in ludo aleae* (1657), and Jakob Bernoulli's *Ars conjectandi* (1713).
9. Hald, *History of probability* (ref. 4), 144–82, who discusses Tycho Brahe's *De nova stella* (1573), Johannes Kepler's *Astronomia nova* (1609), and Isaac Newton's *De analysi per aequationes numero terminorum infinitas* (1669).
10. See Hald, *History of probability* (ref. 4), 81–143. Most important here are John Graunt's *Natural and political observations made upon the bills of mortality* (1662), Jan de Witt's *Waerdye van lyf-renten naer proportie van los-renten* (1671), and Edmond Halley's *An estimate of the degrees of the mortality of mankind* (1694).
11. See Adolphe Quetelet, *Sur l'homme et le dévéloppement de ses facultés, ou essai de physique sociale* (Paris, 1835); Francis Galton, *Hereditary genius: An inquiry into its laws and consequences* (London, 1869); Wilhelm Lexis [1879], "Über die Theorie der Stabilität statistischer Reihen", in: Wilhelm Lexis, *Abhandlungen zur Theorie der Bevölkerungs- und Moralstatistik* (Jena, 1903), 170–212.
12. Abraham de Moivre, *The doctrine of chances: or, a method of calculating the probability of events in play* (London, 1718), 109.
13. Jean le Rond d'Alembert [1770], "Doutes et questions sur le calcul des probabilités", in: *Oeuvres philosophiques, historiques et littéraires de d'Alembert* (Paris, 1805), iv, 289–315, pp. 305f.
14. Pierre-Simon Laplace [1776], *Recherches sur l'intégration des équations différentielles aux différences finies et sur leur usage dans la théorie des hasards*, in: *Oeuvres complètes* (Paris, 1878–1912), viii, 69–197, cited in Hald, *History of mathematical statistics* (ref. 2), 66.
15. Al-Kindi (*c*. A.D. 850), cited in Singh, *Code book* (ref. 7), 35.
16. Around 1890, Etienne Bazeries used the frequencies of syllables to decrypt documents about the military campaigns of Louis XIV encrypted with the so-called 'Great Cypher'; *cf*. Singh, *Code book* (ref. 7), 57f.
17. This links his study with one undertaken eighteen years later by Kurt Gödel, which triggered a fundamental crisis in mathematics (Kurt Gödel, "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I", *Monatshefte für Mathematik und Physik*, xxxviii (1931), 173–98). Both studies applied numbers to the material basis of signs, but Markov

analysed texts written by others and thus avoided the self-reference that led to irresolvable paradoxes in the latter. Moreover, Markov investigated sequences of letters of the alphabet and not mathematical symbols.

18. Surprisingly, although these calculations were more extensive, Markov did not mention them at all in his correspondence with Chuprov (*cf.* Kh. O. Ondar (ed.), *The correspondence between A. A. Markov and A. A. Chuprov on the theory of probability and mathematical statistics* (New York etc., 1981)).
19. Sergei T. Aksakov [1858], *Years of childhood* (New York, 1960), 3 and 12.
20. On Pushkin, see Rolf-Dietrich Keil, *Puschkin* (Frankfurt am Main, 1999), 101f.; on Aksakov, see Sergei T. Aksakov [1858], *Bagrovs Kinderjahre* (transl. by Erich Müller-Kamp, Zürich, 1978), 506: "Aksakov went blind in his left eye and the right eye began to flicker. By 1846, his sight had deteriorated so badly that he was hardly able to write his name."
21. Aksakov, *Years of childhood* (ref. 19), 53.
22. Aksakov, *Years of childhood* (ref. 19), 24.
23. Aksakov, *Years of childhood* (ref. 19), 30f.
24. Alexander Puschkin [1833], *Gesammelte Werke,* ed. by Harald Raab (Frankfurt am Main, 1973), iii, 24 and 436.
25. See Alexander Pushkin [1833], *Eugene Onegin*, transl. by Vladimir Nabokov (Princeton, 1981), ii, 145.
26. Latin *prorsus* — forward, straight ahead, to sum up, utterly, wholly.
27. For a different point of view, see Jay David Bolter, *Writing space* (London, 1991), 108: "But in most books, as in the papyrus roll, one path dominates all others — the one defined by reading line by line, from first page to last. The paged book has a canonical order."
28. Jacques Derrida [1967], *Writing and difference*, transl. by Alan Brass (London, 1993), 9: "*angustia:* the necessary restricted passageway of speech against which all possible meanings push each other, preventing each other's emergence."
29. *Cf.* Georg Wilhelm Friedrich Hegel [1830], *Encyclopedia of the philosophical sciences in outline and critical writings*, ed. by Ernst Behler (New York, 1990), 233: "But since it [the intuition] exists only as suspended, and the intelligence is its negativity, the true form of the intuition as a sign is its existence in time, — but this existence vanishes in the moment of being, and its tone is the fulfilled manifestation of its self-proclaimed interiority."
30. See Alan M. Turing [1947], "Lecture to the London Mathematical Society on 20 February 1947", in: D. C. Ince (ed.), *Collected works of A. M. Turing* (Amsterdam etc., 1992), i, 106–24, p. 109: "The idea of using acoustic delay lines as memory units.… The idea is to store the information in the form of compression waves travelling along a column of mercury.… [I]t is quite feasible to put as many as 1000 pulses into a single 5′ tube.… A train of pulses or the information which they represent may be regarded as stored in the mercury whilst it is travelling through it. If the information is not required when the train emerges it can be fed back into the column again and again until such time as it *is* required."
31. The first machine of this kind was put on the market at the end of 1958 by Hans Bauer. It was named Echolette NG 51/S.
32. For example, no. 119 of the *Tehilim* (Psalms), in which eight verses each begin with the same letter and 176 lines run through all 22 letters of the Hebrew alphabet.
33. In Joseph Weizenbaum's ELIZA of 1966, one of the first dialogue programmes that linked questions from a human interrogator to answers from a computer with different probabilities, it had the following function: "The user's statement is terminated by a double carriage return, which serves to turn control over to ELIZA" (Joseph Weizenbaum, "ELIZA — a computer program for the study of natural language communication between man and machine", *Communications of the*

ACM, ix (1966), 36–45, p. 36). In the pseudo-oral communication via terminals it set a caesura, but in this case a caesura between two partners in a dialogue. Thus, on the computer keyboard there is not only the ‘return’ or ‘enter’ button, but also a ‘home’ button, which, however, does not make the line advance forward.

34. See Friedrich A. Kittler [1985], *Discourse networks 1800/1900* (Stanford, 1990), 193: “Spatially designated and discrete signs — that, rather than increase in speed, was the real innovation of the typewriter.”
35. The only known example of oral cryptography was used by the Navajo Division of the U.S. Army during the Second World War. See Singh, *Code book* (ref. 7), 196f.: “After three weeks of intense cryptanalysis, the Naval codebreakers were still baffled by the messages. They called the Navajo language a ‘weird succession of guttural, nasal, tongue-twisting sounds … we couldn’t even transcribe it, much less crack it’.”
36. Markov, “Beispiel statistischer Forschung” (ref. 1), 187.
37. In Aksakov’s text, however, p was 0.44898 and was thus slightly higher. This is rather surprising, for one would expect there to be more vowels in poetry than prose.
38. Adrien-Marie Legendre, *Nouvelles méthodes pour la détermination des orbites des comètes* (Paris, 1805); Carl Friedrich Gauss, *Theoria motus corporum coelestium in sectionibus conicis solem ambientum* (Hamburg, 1809).
39. Also when determining the standard error of the mean, Markov calculated its square, by dividing the variance 5.114 by the number of samples, 200. Normally, the mean dispersion was divided by the square root of this value. Laplace discovered in 1812, that the dispersion of means from samples, which were all taken from the same group, with growing number turned into a normal curve. Through the so-called Central Limit Theorem, in dependence on the quantity of samples, it was possible to calculate the expected deviation from the real average. Because the errors were normally distributed, it was possible to define confidence intervals around the found value, within which the actual average of the sum total would certainly lie. 96% of the samples were approximately within the range of two-fold standard error (Pierre-Simon Laplace, *Théorie analytique des probabilités* (Paris, 1812)). With 96% certainty, it could be said that the real average of vowel frequency in Pushkin’s novel lay between 42.87 and 43.5. The standard error was around 0.16.
40. Abraham de Moivre [1733], “Approximatio ad summam terminorum binomii $(a + b)^n$ in seriem expansi”, in: Abraham de Moivre, *The doctrine of chances,* 2nd edn (London, 1738), 243–59.
41. Friedrich Wilhelm Bessel, “Ueber den Ort des Polarsterns”, *Astronomische Jahrbücher für das Jahr 1818* (Berlin, 1815), 233-41.
42. Carl Friedrich Gauss [1821–28], *Theory of the combination of observations least subject to errors* (Philadelphia, 1995), 3.
43. See Detlef B. Linke, “Die Vorläufigkeit der Neurotheologie”, *Lettre international*, lxi (2003), 106–7, p. 106: “The oral space is therefore a space that ... exhibits predictable parameters which are relatively stable, that is, in the case of its own movements, so that sensory feed-back is not necessary.... Instead of feed-back, there can be combinations of sounds and their ordering under new laws, those of grammar. According to this model, the oral space provides the security that the outside world cannot guarantee.”
44. This game, popular all over Japan since the 1920s, bears a remarkable resemblance to Francis Galton’s random generator ‘Quincunx’, which he developed in 1877 to illustrate the normal distribution (*cf.* Francis Galton, “Typical laws of heredity”, *Nature,* xv (1877), 492–5, 512–14, 532–3).
45. Joseph Fashing and Ted Goertzel, “The myth of the normal curve”, *Humanity and society*, v (1981), 14–31, p. 27; see also p. 16: “The bell curve came to be generally accepted, as M. Lippmann remarked to Poincaré ..., because ... the experimenters fancy that it is a theorem in mathematics and the mathematicians that it is an experimental fact.”

46. The list used is the file russian_words.koi8.gz available at http://www.funet.fi/pub/unix/security/passwd/crack/dictionaries/russian/. The source code of the programme can be accessed via the online version of this article at http://alpha60.de/research/markov/.
47. The values deviate slightly from those in Markov's study, because a modern version of Pushkin's text was used for this comparison.
48. According to Paul Bouissac, the Courses were introduced into Moscow only around 1917, when the Russian linguist Serge Karcevski returned from Geneva where he was in contact with Charles Bally, one of the editors of Saussure's lectures. He emigrated to Switzerland in 1907; *cf.* Paul Bouissac, "Perspectives on Saussure" (2003, http://www.semioticon.com/Bouissac/saussurecompanion.rtf.htm, accessed 15 Aug. 2005).
49. Ferdinand de Saussure [1916], *Course in general linguistics*, ed. by C. Bally and A. Sechehaye (London, 1983), 88f., emphasis supplied.
50. Saussure, *Course* (ref. 49), 96f.
51. Saussure, *Course* (ref. 49), 146.
52. Saussure, *Course* (ref. 49), 89. This formulation demonstrates that Saussure was close to the idea of the autonomy of language, as advocated later by Lacan and Burroughs. "These linguistic features were spread by contact" (Saussure, *Course* (ref. 49), 205) is echoed in Burrough's "blind forces" of language which he identifies as a viral process.
53. Saussure, *Course* (ref. 49), 150.
54. Ferdinand de Saussure, *Linguistik und Semiologie* (Frankfurt am Main, 1997), 431.
55. See Louis Bachelier, "Théorie de la spéculation", *Annales scientifiques de l'Ecole Normale Supérieure*, xvii (1900), 21–86. He explained the fluctuation of the value of stocks with the concept of Brownian motion.
56. Saussure, *Course* (ref. 49), 127.
57. Lexis, "Theorie der Stabilität" (ref. 11), 183.
58. *Ibid*.
59. See Markov's works mentioned in ref. 3. The only example of connected samples in probability theory before Markov was the experiment in which balls were taken out of an urn and not put back. It had the sole function of illustrating *a priori* findings afterwards. The calculations were not tested empirically because no-one saw any reason to doubt them.
60. Markov, "Recherches sur un cas remarquable" (ref. 3); the equation used is on p. 100.
61. In Aksakov's text, too, the dispersion coefficient of 0.25 was clearly "subnormal". Here, the theoretical value for the simple chain is 0.29.
62. Around twenty years later, Alan Turing proved that for the hand-written calculations, with which the Russian developed his findings in lengthy and wearisome counting work, the second dimension was not essential. Turing then proceeded to develop his idea of a universal machine, which replaced to a large extent the 'computers' of the time. In this construction, too, each sign was dependent on the preceding one; mediated, however, by the state of the machine and the programme that linked them. Additionally, the connection was no longer probabilistic, but completely defined. The decisive difference, however, is that Turing's machine — in contrast to Markov's, which always moved towards the right and had no return — turned back upon what it had written and interpreted what it found as a command. It re-wrote it and thus changed its own programme. The algorithm of the Russian mathematician wrote readable signs, Turing's wrote executable signs. For this reason, the former did not possess a halting condition. (The halting condition denoted the coincidence of state and read sign, when the universal machine turned off and delivered a result.) *Cf*. Alan M. Turing [1936], "On computable numbers, with an application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society*, lxii (1936–37), 230–65; Emil Post [1936], "Finite combinatory processes, formulation I", in: Martin Davis (ed.),

The undecidable (Hewlett, NY, 1965), 288–91.

63. Markov, "Sur un cas d'épreuves" (ref. 3); the equation used is on p. 179.
64. That the samples examined were independent was also shown by the fact that the dispersion varied only insignificantly when they were combined in twos, fours, or fives. If chaining would have existed, this procedure would have made it decrease because in the groups formed fewer letters would have been mutually dependent. Dependency would have wandered into the sample, so to speak, and would have affected in decreasing measure one letter's relationship to the next. For the calculation in pairs Markov added two sums next to each other in the last row and in this way arrived at the number of vowels in the series of 200 letters of the text and found the sum of the squares of the deviations from the double mean 86.4 to be 827.6, which did not differ significantly from the number for the hundreds, 1022.8. Similarly, when this was calculated for 50 groups of 400 letters or 40 groups of 500 letters, the variance remained stable: the former was 975.2 and the latter 1004.
65. Wilhelm Lexis, *Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft* (Freiburg, 1877), 34.
66. See Lexis, "Theorie der Stabilität" (ref. 11), 177: "Whereas under the presuppositions made here the inequality $R < r$ can never arise because in this case p would become imaginary which points at an impossibility." The equation $Q = 1$ was discussed repeatedly by Markov and Chuprov in their correspondence as of 23 November 1910 (Ondar, *Correspondence* (ref. 18), 38f.). In their view subnormal dispersion was possible also in independent samples.
67. Saussure, *Course* (ref. 49), 131.
68. Saussure, *Course* (ref. 49), 51.
69. *Ibid*. See also Benoit Mandelbrot, "On the theory of word frequencies and on related Markovian models of discourse", in: *Structure of language and its mathematical aspects* (*Proceedings of the Symposia on Applied Mathematics*, xii; Rhode Island, 1961), 190–219, p. 211: "To an extent unrivalled by other classics in the field, Saussure exhibited an 'esprit géométrique', which was most welcome to a mathematician."
70. Saussure, *Course* (ref. 49), 41.
71. Saussure, *Course* (ref. 49), 57.
72. A. A. Markov, *Probability theory* (St Petersburg, 1900, in Russian). The equations used are found in Markov, *Wahrscheinlichkeitsrechnung* (ref. 3), 203 and 209.
73. In the study of the Aksakov text, letters were combined with each other within the groups of ten that in the text were separated by nine letters. The coefficient normalized to 1.05, and thus corresponded again to a random distribution. Markov demonstrated this by making a table containing this count together with the theoretical distribution for 10,000 independent samples. For calculating the latter, he referred to a passage in his *Probability theory*, which gave Newton's binomial formula; *cf*. Markov, *Wahrscheinlichkeitsrechnung* (ref. 3), 27.
74. The calculation of the mathematical expectation of the square of error by dividing the variance again by the number of samples as above was not possible in the case of dependence of the quantities under consideration because the dispersion was artificially increased by the dependence and no longer corresponded to the fluctuation of the arithmetic mean that would be expected if further 20,000 letters of the text were to be examined. Markov therefore proposed using the result of the first arrangement for this, where the different values were not mutually dependent. The connection of the numbers was also clear from the fact that the sum of the squares of their deviations from the mean changed drastically when the values were combined in twos, fours, or fives. Instead of 5788.8, the sums were 3551.6, 3089.2, and 1004. Because of this combining of values, the groups contained progressively fewer letters, which were next to each other in the text. In the groups of 200, which were formed by combining in pairs, only half were mutually dependent, in the groups of 400, only a quarter, until in the groups of 500, there were none at all. The combination in fives,

whether in columns or rows, each contained the values of one small table and produced entirely independent samples. For this reason, the result was the same for both: 1004.

75. See Kittler, *Discourse networks* (ref. 34), 212; also Bernhard Holeczek and Lida von Mengden (eds), *Zufall als Prinzip* (Heidelberg, 1992), 77.
76. Plutarch [*c*. A.D. 75], *The lives of the noble Grecians and Romans* (Chicago etc., 1952), 362 (Lysander).
77. David Kahn, *The codebreakers* (New York, 1967), 619.
78. See Otto Leiberich [1999], "Vom diplomatischen Code zur Falltürfunktion", *Spektrum der Wissenschaft. Dossier Kryptographie,* 2001, 12–18, p. 17: "In the early 1960s, a foreign espionage organization equipped their agents operating in the Federal Republic of Germany with an encryption technique, double columnar transposition, that was regarded as secure. The foreign cryptologists, however, made one mistake. They gave their agents mnemonic sentences, which they had taken from literary works, as keys."
79. *Cf.* Wayne G. Barker, *Cryptanalysis of the double transposition cipher* (Laguna Hills, 1995).
80. Boris Tomashevski, "The 10th chapter of 'Eugene Onegin': The story of its solution", *Literaturnoe nasledstwo*, xvi–xviii (1934, in Russian), 378–420, p. 384f. *Cf.* Piotr Morozov, "Pushkin's coded poem", *Puskin i ego sovremenniki*, iv (1910, in Russian), 1–12.
81. Pushkin, *Eugene Onegin* (ref. 25), 365f.
82. Pushkin, *Eugene Onegin* (ref. 25), 370.
83. Pushkin, *Eugene Onegin* (ref. 25), 374.
84. Saussure, *Course* (ref. 49), 118.