Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

Structural connections between long genetic and literary texts

Sergey V. Petoukhov

Mechanical Engineering Research Institute of Russian Academy of Sciences, Russia, 101990, Moscow, M. Kharitonievskiy pereulok, 4,
http://eng.imash.ru/, info@imash.ru

Comment: Some elements of this article were presented by the author in his keynote speeches at the following conferences: the International Belgrade Bioinformatics Conference 2018 (Belgrade, Serbia, 18-22 June, 2018, http://belbi.bg.ac.rs/); the Second International Conference "Artificial Intelligence, Medical Engineering, Education" (Moscow, Russia, 6-8 October 2018, http://www.ruscnconf.org/aimee2018/index.html); the First Congress of the International Society of Natural Medicine (Samorin, Slovak Republic, 12-14 October 2018, https://www.simn.info/congress-of-natural-medicine-2018/); the Second International Conference on Computer Science, Enginnering and Education Applications (Kiev, Ukraine, 27-28 January 2019, http://www.uacnconf.org/iccseea2019/index.html).

Abstract. Impressive recent discoveries in genetics have borrowed terminology from linguistics and the theory of communications. Leading structural linguistics experts postulate that human languages were influenced by a combination of genetic language and environment, suggesting the compelling possibility that all organisms may utilize their genetic code in communication mechanisms. The article explores this connection between linguistic languages and genetic language. Long genetic and literary texts are considered in special binary forms of their "n-plet representations", where each text is interpreted as a concatenation of fragments of identical length "n". Probabilities of the presence of such n-plets in two types of considered texts are compared: 1) for long DNA sequences of two kinds of hydrogen bonds 2 and 3 of complementary nytrogenous bases in long DNA sequences of different organisms; 2) for long Russian texts represented as sequences of two phonetic groups of Russian letters in novels by L.Tolstoy, F.Dostoevsky, A.Pushkin, etc. Quantum informatics formalisms are used in modeling some of the results. The author's research is in general alignement with P. Jordan's conjecture, published in 1932, in which he claimed that life's missing laws were the rules of chance and probability of the quantum world. These results can be used for further development of quantum biology (this term was proposed by P. Jordan).

Keywords: DNA sequences, literary texts, probability, binary opposition, alphabet, tensor product, quantum biology, quantum informatics, archetypes.

Content

1. Introduction
2. Some properties of long literary texts in Russian
3. Regarding similar properties of long DNA-texts
4. Quantum biology and life's rules of probability postulated by P. Jordan
5. Some concluding remarks
   Acknowledgments
   References

1. Introduction

Impressive recent discoveries in genetics have borrowed terminology from linguistics and the theory of communications. As experts in molecular genetics note, "*the more we understand laws of coding of the genetic information, the more strongly we are surprised by their similarity to principles of linguistics of human and computer languages*" [Ratner, 2002, p. 203].

© 2019 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

Leading experts in the field of structural linguistics have long believed that languages of human dialogue were formed not by random processes but by a continuation of genetic language or, are, at least, closely connected with genetic language, suggesting the compelling possibility that all organisms may utilize their genetic code in communication mechanisms. Analogies between systems of genetic and linguistic information are of wide and important scientific interest, which this article briefly illustrates. Some relevant concepts will be referred to by R. Jakobson [1987, 1999], one of the most famous lingusitics experts and author of an in-depth theory of binary lingustic oppositions. Jointly with F. Jacob, Nobel Prize winner in molecular genetics, and with other linguistic specialists holding the same views, Jakobson proposed that genetic language is the structural basis of linguistic languages [Jacob et al., 1968; Jakobson, 1985]. Among all systems of information transfer, the genetic and linguistic codes are based on use of discrete components, which in itself make no sense, but serve for construction of the minimum units that make sense. In both the genetic and linguistic languages, we deal with separate units that, individually, make no sense, but in studying their special groupings, their sense becomes clear. This similarity between both information systems is an important fact, yet is only one aspect of this relatively new field of study. According to Jakobson, all relations among linguistic phonemes are decomposed into a series of binary oppositions of elementary differential attributes (or traits). By analogy, the set of four letters of the genetic alphabet contains the three binary sub-alphabets, which allow creating new mathematical models in molecular genetics [Petoukhov, 2017, 2018a]. As Jakobson wrote, the genetic code system is the basic simulator, which underlies all verbal codes of human languages. "*The heredity in itself is the fundamental form of communications ... Perhaps, the bases of language structures, which are imposed on molecular communications, have been constructed by its structural principles directly*" [Jakobson, 1985, p. 396]. These questions had arisen to Jakobson as consequence of his long-term research into the connections between linguistics, biology and physics. Such connections were considered at a united seminar of physicists and linguists, organized by Niels Bohr and Roman Jakobson, jointly, at the Massachusetts Institute of Technology.

"*Jakobson reveals distinctly a binary opposition of sound attributes as underlying each system of phonemes... The subject of phonology has changed by him: the phonology considered phonemes (as the main subject) earlier, but now Jakobson has offered that distinctive attributes should be considered as "quantums" (or elementary units of language)... Jakobson was interested especially in the general analogies of language structures with the genetic code, and he considered these analogies as indubitable*" [Ivanov, 1985]. We are reminded also of the title of the monograph "On the Yin and Yang nature of language" [Bailey, 1982], which is characteristic for the theme of binary oppositions in linguistics.

F. Jacob, Nobel Prize winner in molecular genetics, also considered the relationship between genetics and linguistic languages in connection with the principle of binary oppositions, systematically described in the Ancient Chinese book "I-Ching". He wrote: « *C'est peut-être I Ching qu'il faudrait étudier pour saisir les relations entre hérédité et langage*» (In English: *To understand the relationship between genetics and language, perhaps it would be necessary to study the Ancient Chinese "I Ching"*) [Jacob, 1974, p. 205].

This connection between linguistics and the genetic code interests many researchers, and some even perceive linguistic language as a living organism. In his book, "Linguistic Genetics", Makovsky says: "*A look at language as a living organism, subject to the natural laws of nature, ascends to a deep antiquity ... Research of a nature, of disposition and of reasons of isomorphism between genetic and linguistic regularities is one of the most important fundamental problems for linguistics of our time*" [Makovsky, 1992].

The study of linguistic languages can have important implications toward a deeper understanding and modeling of mental functions and for the appropriate expansion of physics. In the work of B. Josephson, Nobel laureate in physics [Josephson, 2018] he says: "*Regular physics is unsatisfactory in that it fails to take into consideration phenomena relating to mind and meaning, whereas on the other side of the cultural divide such constructs have been studied in detail. This paper discusses a possible synthesis of the two perspectives. Crucial is the way systems realising mental function can develop step by step on the basis of the scaffolding mechanisms of Hoffmeyer, in a way that can be clarified by consideration of the phenomenon of language*".

In the next section the author will describe the structural analogies between long DNA-'texts' and long literary works in Russian. The represented analysis of long literary Russian texts (by L.N. Tolstoy, F.M. Dostoevsky, A.S. Pushkin, etc.) is based on binary peculiarities of the Russian alphabet. In addition, some

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

applications of mathematics, concerning quantum informatics for modeling bio-informational structures, are considered. The results described were derived from the author's long-time researches in the field of quantum biology (see details about the history of quantum biology in [McFadden, Al-Khalili, 2018]). P. Jordan, credited with the authoring the first work on quantum biology, postulated that the mechanisms of living organisms are associated with what he referred to as his 'amplifier theory', based on Bohr's notion of the 'irreversible act of amplification', required to bring the fuzzy quantum reality into sharp focus by 'observing' it [Jordan, 1932]. Jordan claimed that «*life's missing laws were the rules of chance and probability (the indeterminism) of the quantum world that were somehow scaled up inside living organisms*» [McFadden, Al-Khalili, 2018]. It is these laws of chance and probability, postulated by Jordan, that we are looking for in our studies of the probabilistic characteristics of long DNA sequences of hydrogen bonds and nitrogenous bases [Petoukhov, 2018a,b; Darvas, 2018]. The article is connected with these results.

2. Some properties of long literary texts in Russian

The Russian alphabet has a binary-oppositional structure since it has two binary-oppositional sub-alphabets: the sub-alphabet of vowels and the sub-alphabet of consonants. Each of these sub-alphabets also has its own binary-oppositional structure: the sub-alphabet of vowels consists of the sub-sub-alphabet of long vowels and the sub-sub-alphabet of short (or iotated) vowels; the sub-alphabet of consonants consists of the sub-sub-alphabet of voiced consonants and the sub-sub-alphabet of deaf consonants (Fig. 1). The soft sign ь and the hard sign ъ in the Russian alphabet do not convey any sound and therefore they are not taken into account in its phonologic structure.

The Russian alphabet can be considered to consist of the following two class of equivalency (in Fig. 1, the first class is marked by yellow and the second class is marked by green):

1. The first class of equivalency combines all short (iotated) vowels and all deaf consonants: е, ё, ю, я, п, ф, к, т, ш, с, х, ц, ч, щ. We can denote the 14 members of this class by the general symbol 0;
2. The second class of equivalency combines all long vowels and all voiced consonants: а, и, о, у, ы, э, б, в, г, д, ж, з, й, л, м, н, р. We can denote the 17 members of this part by the general symbol 1.

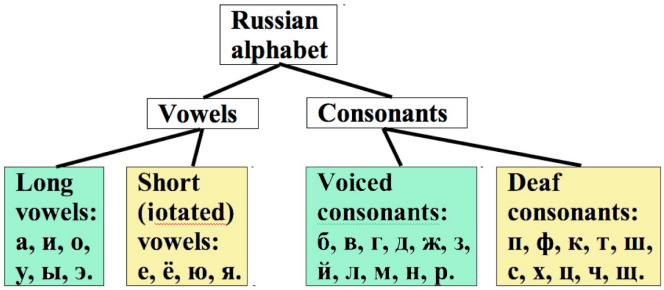


Fig. 1. The binary-oppositional structure of the Russian alphabet.

Adopting this method, any Russian text can be converted into the corresponding binary sequence of these two symbols 0 and 1 (for example, 100110110...) by the following procedure:

1) all punctuation marks and the spacings between words, as well as all soft signs «ь» and hard signs «ъ», are deleted from the text;

2) each of the remaining letters is replaced by the corresponding symbol 0 or 1 depending on which of these two classes of equivalency it belongs.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

For example, as a result of this procedure, the Russian text ”Лев Толстой – великий русский писатель” (in English: «Leo Tolstoy - a great Russian writer») turns into a binary sequence 10101100111011011111000110101001. The computer program for analysis of literary texts was created by graduate student V.I. Svirin, under the guidance of the author.

It is important to mention the role played by the mathematical operation of the tensor product. This well known operation in mathematics, physics and informatics provides a method of joining vector spaces to form larger vector spaces. The tensor product is the crucial operation to understanding the quantum mechanics of multiparticle systems [Nielsen, Chuang, 2010, p. 71] and is one of basic instruments in quantum informatics. The following quotation highlights the importance of the tensor product: «*This construction is crucial to understanding the quantum mechanics of multiparticle systems*» [Nielsen, Chuang, 2010, p. 71] since a postulate of quantum mechanics holds that the state space of a composite system is the tensor product of the state spaces of its components. By definition, under the tensor product of two vectors, each of the components of the first vector is multiplied with all components of the second vector. The expression (1) shows an example of the tensor product (denoted by the symbol $\otimes$) of two 2-dimensional vectors [x, y] and [v, w], which gives in the result one 4-dimensional vector [xv, xw, yv, yw]:

$$[x, y] \otimes [v, w] = [x[v, w], y[v, w]] = [xv, xw, yv, yw] \quad (1)$$

The named binary representations of long Russian literary texts were analyzed [Petoukhov, 2018a] by a method analogous with the analysis of long DNA sequences of numbers of hydrogen bonds 3 and 2 [Petoukhov, 2018a]. In more detail, each of long binary sequences of numbers 0 and 1 (for example the sequence 1-0-1-0-1-1-0-0-1-1-1-0-…) can be represented as a sequence of binary doublets (10-10-11-00-11-10-…), or a sequence of binary triplets (101-011-001-110-…) or, in a general case, as a sequence of binary n-plets (n = 1, 2, 3, 4, 5, …). Such different representations of a long literary text, in the form of sequences of binary n-plets, are termed its “binary n-plet representation”. In this article only long literary texts in Russian are considered, though similar approches are available for long literary texts in some other languages.

A complete alphabet of binary n-plets under fixed value n is termed “the alphabet of binary n-plets” and contains 2^n members. For example, the alphabet of binary monoplets contains 2 members (0 and 1); the alphabet of binary doublets contains 4 members (00, 01, 10, 11), which only exist in binary doublet representations of a considered long literary text; the alphabet of binary triplets contains $2^3 = 8$ members (000, 001, 010, 011, 100, 101, 110, 111), which only exist in the binary triplet representation of a considered long Russian language literary text, etc. It should be mentioned that the tensor family of vectors $[0, 1]^{(n)}$ (where (n) means the tensor power; n = 1, 2, 3, 4, …) contains vectors whose sets of components coincide with corresponding alphabets of binary n-plets. For example, $[0, 1]^{(2)}$ = [00, 01, 10, 11]; $[0, 1]^{(3)}$ = [000, 001, 010, 011, 100, 101, 110, 111], etc. Therefore, it can be seen that, conditionally, all binary n-plet alphabets belong to the same tensor family of vectors.

For each fixed n, we can analyze probabilities (or frequencies or percentage) of each kind of n-plets inside n-plet representations of any long Russian language literary text (or a literary text of any other language translated into Russian). Under a fixed value n, each of these probabilities is equal to the ratio: (the total quantity of a corresponding member of an alphabet of binary n-plets) divided by (the total quantity of these binary n-plets). For example, in the text of the work "Anna Karenina", by Leo Tolstoy, in its binary doublet representation there are 654523 doublets 00, 01, 10 and 11. This number includes 75895 doublets 00, 142504 doublets 01, 142547 doublets 10 and 293577 doublets 11. Correspondingly, the probability of doublets 00 is equal to 75895/654523 = 0,115954672; the probabilty of doublets 01 is equal to 142504/654523 = 0,217721914; the probabilty of doublets 10 is equal to 142547/654523 = 0,21778761; the probability of doublets 11 is equal to 293577/654523 = 0,448535804 (these values are shown rounded to four decimal places in an appropriate row in Fig. 3 in blue color).

In addition to calculating the probabilities of n-plets (n = 1, 2, 3, 4) in n-plet representations of long literary Russian texts, the author has modeled these probabilities by the same quantum-algorithmic method used for modeling probabilities in long DNA sequences of hydrogen bonds 3 and 2, in his article

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

[Petoukhov, 2018a]. This method explores some classical formalism from quantum informatics [Nielsen, Chuang, 2010].

In the author's model approach, phenomenologic probabilities of all members of binary n-plet alphabets $[0, 1]^{(n)}$ (under fixed n, where n = 1, 2, 3, 4, … is much less than the length of the considered literary text) are modeled by numeric values of appropriate coordinates of 2^n-dimensional vector from the tensor family of vectors of probabilities $[0_P, 1_P]^{(n)}$, where 0_P and 1_P denote probabilities of the symbol 0 and of the symbol 1 in the binary monoplet representation of the analyzed long Russian text. These probabilities for the novel «Anna Karenina» are shown in the first column in Fig. 2 and in the first rows of the table in Fig. 3.

Correspondingly, in this model approach, to obtain model values of probabilities in all 2^n members of the binary n-plet alphabet, in a considered binary n-plet representation of the analyzed literary text, the following 2^n coordinates of vectors of probabilities from their tensor family should be calculated:

- in the case of the alphabet of binary doublets, there exist 4 coordinates of the vector $[0_P, 1_P]^{(2)} = [0_P0_P, 0_P1_P, 1_P0_P, 1_P1_P]$; these 4 numeric coordinates correspond to 4 appropriate coordinates of the vector of all members of the alphabet of binary doublets [00, 01, 10, 11];
- in the case of the alphabet of binary triplets, there exist 8 coordinates of the vector $[0_P, 1_P]^{(3)} = [0_P0_P0_P, 0_P0_P1_P, 0_P1_P0_P, 0_P1_P1_P, 1_P0_P0_P, 1_P0_P1_P, 1_P1_P0_P, 1_P1_P1_P]$; these 8 numeric coordinates correspond to 8 appropriate coordinates of the vector of all members of the alphabet of binary triplets [000, 001, 010, 011, 100, 101, 110, 111];
- in the case of the alphabet of binary tetraplets, there exist 16 coordinates of the vector $[0_P, 1_P]^{(4)} = [0_P0_P0_P0_P, 0_P0_P0_P1_P, 0_P0_P1_P0_P, 0_P0_P1_P1_P, 0_P1_P0_P0_P, 0_P1_P0_P1_P, 0_P1_P1_P0_P, 0_P1_P1_P1_P, 1_P0_P0_P0_P, 1_P0_P0_P1_P, 1_P0_P1_P0_P, 1_P0_P1_P1_P, 1_P1_P0_P0_P, 1_P1_P0_P1_P, 1_P1_P1_P0_P, 1_P1_P1_P1_P]$; these 16 numeric coordinates correspond to 16 appropriate coordinates of the vector of all members of the alphabet of binary tetraplets [0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111], etc.

It is important to emphasise that all coordinates of these vectors $[0_P, 1_P]^{(n)}$, where n = 2, 3, 4, … are expressed by only two values, 0_P and 1_P. Summarising, in accordance with the proposed model approach, for an approximate prediction of probabilities of all members of binary n-plet alphabets in a binary n-plet representation of a long Russian literary text - on the basis of knowledge about only two probabilities 0_P and 1_P - the following procedure has been found to be sufficient:

- calculate probabilities 0_P and 1_P of binary monoplets 0 and 1 in the binary sequence;
- calculate the product of these probabilities 0_P and 1_P in coordinates of vectors $[0_P, 1_P]^{(n)}$, where n = 2, 3, 4, … .

The expression (2) shows an example of such calculation of the probability $0_P0_P1_P$ for the member 001 of the alphabet of binary triplets under 0_P = 0,36 and 1_P = 0,64 ($0_P + 1_P = 1$):

$$0_P0_P1_P = 0{,}36*0{,}36*0{,}64 = 0{,}0829 \qquad (2)$$

Figs. 2, 3 represent results – in graphical and tabular forms - of the author's analysis of the novel «Anna Karenina», by Leo Tolstoy, by the above described approach. Significant correspondences can be seen between phenomenologic values (blue points in graphs in Fig. 2) of probabilities of all members of alphabet binary n-plets, and model values (red points in graphs), represented by components of probability vectors $[0_P, 1_P]^{(n)}$, n = 2, 3, 4. These graphs reveal that the model points of red color were almost exactly superimposed on the phenomenologic points of blue color. Fig. 3 shows the proximity of the numerical phenomenological and model values of the studied probabilities. All values are rounded to four decimal places. Therefore, knowing only two probablities 0_P and 1_P of binary monoplets, 0 and 1 in the binary n-plet representation of this well known novel, many probabilties for all members of the alphabets of binary n-plets can be predicted. We presume that a similar model correspondence also holds true for n = 5, 6, ... (if n is much less than the length of the considered literary text) but this should be studied in future research.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

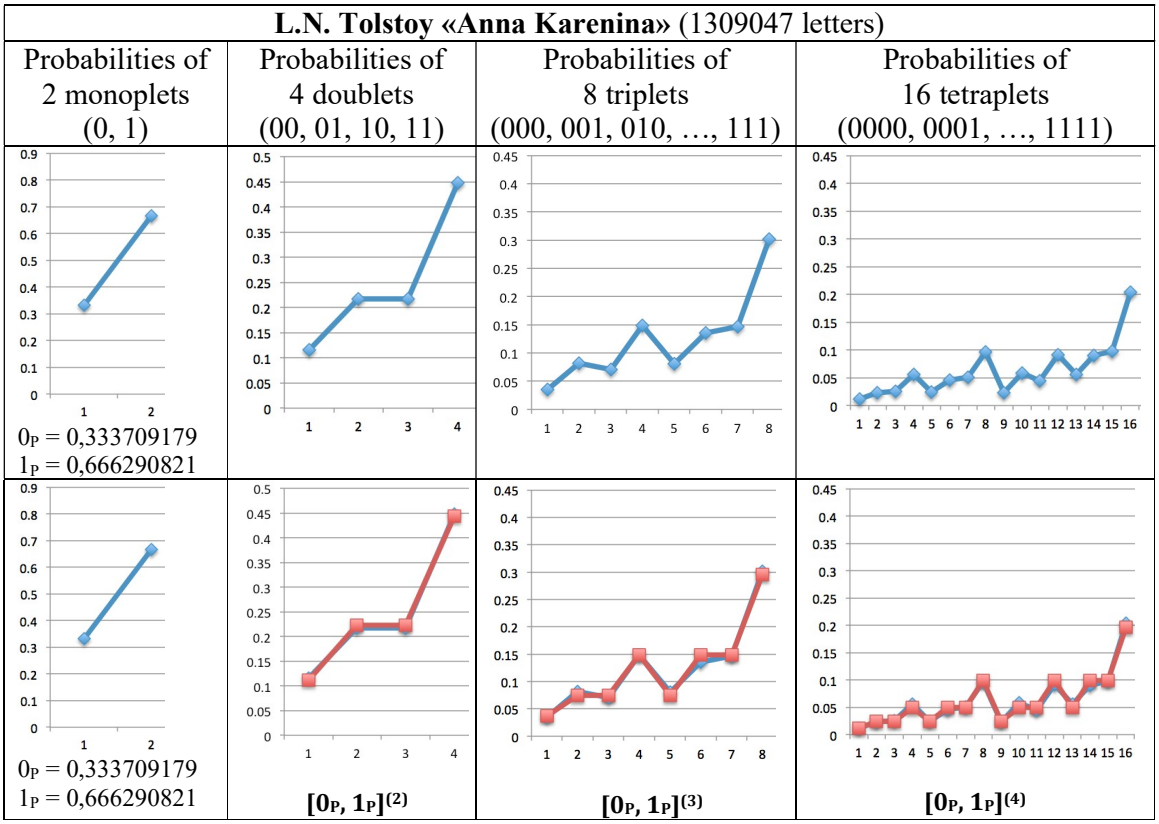


Fig. 2. Graphical analysis results of the novel "Anna Karenina" by Leo Tolstoy (the original literary text was accessed from http://samolit.com/books/62/). Probabilities of members in alphabets of binary *n*-plets (*n* = 1, 2, 3, 4) from the binary *n*-plet representations of this novel are shown. Blue points correspond to phenomenologic values of the probabilities of hydrogen *n*-plets, while red points correspond to model values of the probabilities calculated as components of the vectors $[0_P, 1_P]^{(n)}$, where 0_P and 1_P are probabilities of binary monoplets 0 and 1; (*n*) refers to tensor powers.

L.N. Tolstoy «Anna Karenina» (1309047 letters)	
Probabilities of members in the alphabet of binary monoplets	
Reality:	$0_P = 0{,}3337$; $1_P = 0{,}6663$
Probabilities of members in the alphabet of binary doublets	
Reality:	$0_P0_P=0{,}1160$; $0_P1_P=0{,}2177$; $1_P0_P=0{,}2178$; $1_P1_P=0{,}4485$.
Model:	$0_P0_P=0{,}1114$; $0_P1_P=0{,}2223$; $1_P0_P=0{,}2223$; $1_P1_P=0{,}4439$.
Probabilities of members in the alphabet of binary triplets	
Reality:	$0_P0_P0_P=0{,}0348$; $0_P0_P1_P=0{,}0818$; $0_P1_P0_P=0{,}0708$; $0_P1_P1_P=0{,}1481$; $1_P0_P0_P=0{,}0808$; $1_P0_P1_P=0{,}1353$; $1_P1_P0_P=0{,}1465$; $1_P1_P1_P=0{,}3019$.
Model:	$0_P0_P0_P=0{,}0372$; $0_P0_P1_P=0{,}0742$; $0_P1_P0_P=0{,}0742$; $0_P1_P1_P=0{,}1481$; $1_P0_P0_P=0{,}0742$; $1_P0_P1_P=0{,}1481$; $1_P1_P0_P=0{,}1481$; $1_P1_P1_P=0{,}2958$.
Probabilities of members in the alphabet of binary tetraplets	
Reality:	$0_P0_P0_P0_P=0{,}0114$; $0_P0_P0_P1_P=0{,}0232$; $0_P0_P1_P0_P=0{,}0257$; $0_P0_P1_P1_P=0{,}0559$; $0_P1_P0_P0_P=0{,}0247$; $0_P1_P0_P1_P=0{,}0460$; $0_P1_P1_P0_P=0{,}0505$; $0_P1_P1_P1_P=0{,}0970$; $1_P0_P0_P0_P=0{,}0231$; $1_P0_P0_P1_P=0{,}0582$; $1_P0_P1_P0_P=0{,}0446$; $1_P0_P1_P1_P=0{,}0914$; $1_P1_P0_P0_P=0{,}0565$; $1_P1_P0_P1_P=0{,}0899$; $1_P1_P1_P0_P=0{,}0975$; $1_P1_P1_P1_P=0{,}2045$.
Model:	$0_P0_P0_P0_P=0{,}0124$; $0_P0_P0_P1_P=0{,}0248$; $0_P0_P1_P0_P=0{,}0248$; $0_P0_P1_P1_P=0{,}0494$; $0_P1_P0_P0_P=0{,}0248$; $0_P1_P0_P1_P=0{,}0494$; $0_P1_P1_P0_P=0{,}0494$; $0_P1_P1_P1_P=0{,}0987$; $1_P0_P0_P0_P=0{,}0248$; $1_P0_P0_P1_P=0{,}0494$; $1_P0_P1_P0_P=0{,}0494$; $1_P0_P1_P1_P=0{,}0987$; $1_P1_P0_P0_P=0{,}0494$; $1_P1_P0_P1_P=0{,}0987$; $1_P1_P1_P0_P=0{,}0987$; $1_P1_P1_P1_P=0{,}1971$.

Fig. 3. The numeric representation of the analysis of the novel "Anna Karenina" by Leo Tolstoy (the original literary text was accessed from http://samolit.com/books/62/). Probabilities of members in alphabets of binary *n*-plets (*n* = 1, 2, 3, 4) from the binary n-plet representation of this novel are shown. All values are rounded to four decimal places. Blue numbers correspond to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

phenomenologic values of the probabilities for cases of alphabets named in tabular sections, while red numbers correspond to model values of these probabilities calculated as components of the vectors $[0_P, 1_P]^{(n)}$, where 0_P and 1_P denote phenomenologic values of probabilities of binary monoplets 0 and 1; (n) refers to tensor powers.

Several other examples are shown below of the author's results achieved by similar analysis of famous Russian literary works: L.N. Tolstoy «War and Peace» (Figs. 4, 5); F.M. Dostoevsky «Crime and Punishment» (Figs. 6, 7); «Idiot» (Figs. 8, 9); A.S. Pushkin «Evgenij Onegin» (Figs.10, 11); «Dubrovsky» (Figs. 12, 13); the Russian Bible (Figs. 14, 15). All these results are similar to those described for the novel «Anna Karenina» (Figs, 2, 3): they confirm that probabilities of members studied for different alphabets of binary n-plets (n = 1, 2, 3, 4) are, to some degree, interrelated to each other and that this interrelation can be modeled on the basis of the tensor family of vectors $[0_P, 1_P]^{(n)}$ (n = 1, 2, 3, 4). The computer program for the analysis of literary texts was created by our graduate student V.I. Svirin.

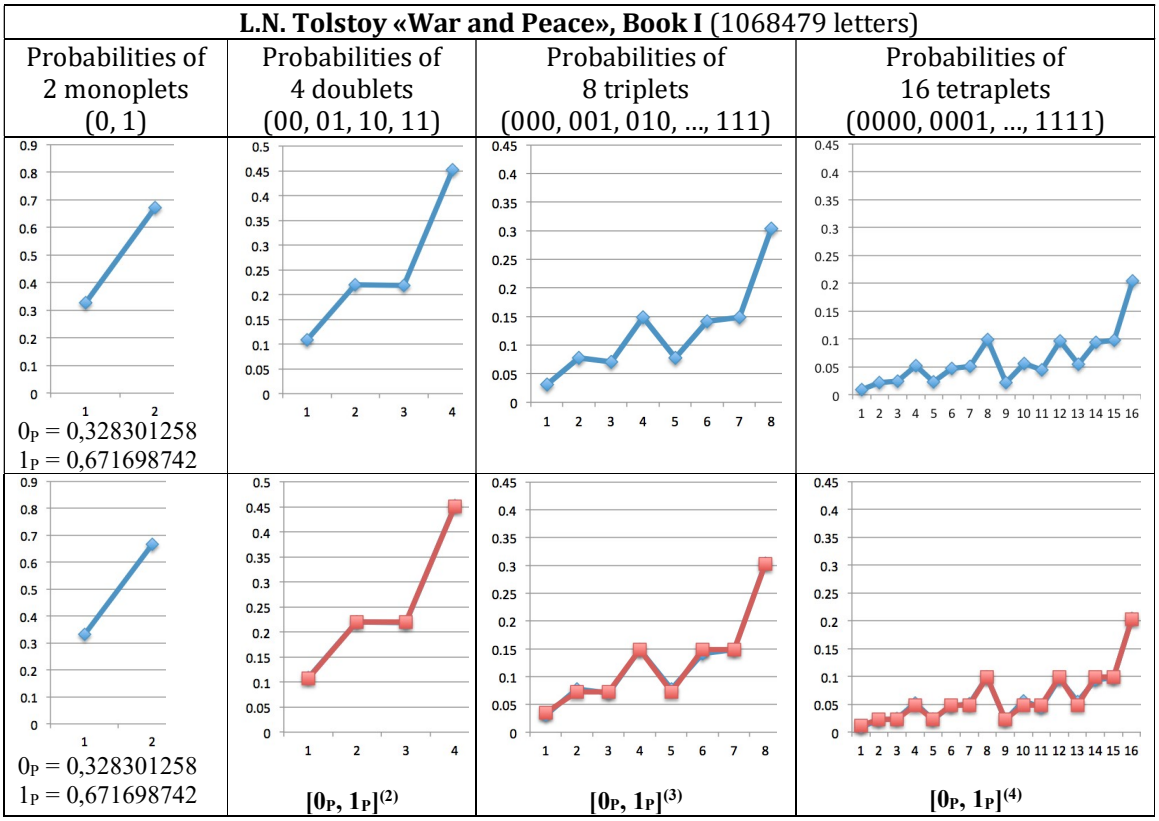


Fig. 4. Graphical analysis results of the novel "War and Peace" (Book 1) by Leo Tolstoy (the original literary text was accessed from http://samolit.com/books/64/). Probabilities of members in alphabets of binary n-plets (n = 1, 2, 3, 4) from the binary n-plet representation of this novel are shown. Blue points correspond to phenomenologic values of the probabilities of n-plets, while red points correspond to model values of the probabilities calculated as components of the vectors $[0_P, 1_P]^{(n)}$, where 0_P and 1_P are probabilities of binary monoplets 0 and 1; (n) refers to tensor powers.

L.N. Tolstoy «War and Peace», Book I (1068479 letters)	
Probabilities of members in the alphabet of binary monoplets	
Reality:	0_P = 0,3283; 1_P = 0,6717
Probabilities of members in the alphabet of binary doublets	
Reality:	0_P0_P=0,1088; 0_P1_P=0,2200; 1_P0_P=0,2190; 1_P1_P=0,4522.
Model:	0_P0_P=0,1078; 0_P1_P=0,2205; 1_P0_P=0,2205; 1_P1_P=0,4512.
Probabilities of members in the alphabet of binary triplets	
Reality:	$0_P0_P0_P$=0,0314; $0_P0_P1_P$=0,0774; $0_P1_P0_P$=0,0705; $0_P1_P1_P$=0,1484; $1_P0_P0_P$=0,0778; $1_P0_P1_P$=0,1419; $1_P1_P0_P$=0,1490; $1_P1_P1_P$=0,3036.

Model:	$0_P0_P0_P0_P=0,0354$; $0_P0_P0_P1_P=0,0724$; $0_P0_P1_P0_P=0,0724$; $0_P0_P1_P1_P=0,1481$; $1_P0_P0_P0_P=0,0724$; $1_P0_P0_P1_P=0,1481$; $1_P0_P1_P0_P=0,1481$; $1_P0_P1_P1_P=0,3031$.
Probabilities of members in the alphabet of binary tetraplets	
Reality:	$0_P0_P0_P0_P=0,0096$; $0_P0_P0_P1_P=0,0220$; $0_P0_P1_P0_P=0,0250$; $0_P0_P1_P1_P=0,0521$; $0_P1_P0_P0_P=0,0234$; $0_P1_P0_P1_P=0,0474$; $0_P1_P1_P0_P=0,0504$; $0_P1_P1_P1_P=0,0988$; $1_P0_P0_P0_P=0,0214$; $1_P0_P0_P1_P=0,0562$; $1_P0_P1_P0_P=0,0451$; $1_P0_P1_P1_P=0,0966$; $1_P1_P0_P0_P=0,0546$; $1_P1_P0_P1_P=0,0943$; $1_P1_P1_P0_P=0,0984$; $1_P1_P1_P1_P=0,2049$.
Model:	$0_P0_P0_P0_P=0,0116$; $0_P0_P0_P1_P=0,0238$; $0_P0_P1_P0_P=0,0238$; $0_P0_P1_P1_P=0,0486$; $0_P1_P0_P0_P=0,0238$; $0_P1_P0_P1_P=0,0486$; $0_P1_P1_P0_P=0,0486$; $0_P1_P1_P1_P=0,0995$; $1_P0_P0_P0_P=0,0238$; $1_P0_P0_P1_P=0,0486$; $1_P0_P1_P0_P=0,0486$; $1_P0_P1_P1_P=0,0995$; $1_P1_P0_P0_P=0,0486$; $1_P1_P0_P1_P=0,0995$; $1_P1_P1_P0_P=0,0995$; $1_P1_P1_P1_P=0,2036$.

Fig. 5. Numeric analysis results of the novel "War and Peace" (Book 1) by Leo Tolstoy (the original literary text was accessed from http://samolit.com/books/64/). Probabilities of members in alphabets of binary *n*-plets (*n* = 1, 2, 3, 4) from the binary n-plet representation of this novel are shown. All values are rounded to four decimal places. Blue numbers correspond to phenomenologic values of the probabilities for cases of alphabets named in tabular sections, while red numbers correspond to model values of these probabilities calculated as components of the vectors $[0_P, 1_P]^{(n)}$, where 0_P and 1_P denote phenomenologic values of probabilities of binary monoplets 0 and 1; (*n*) refers to tensor powers.

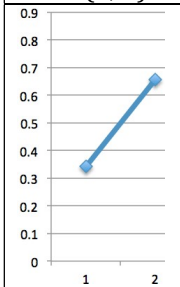
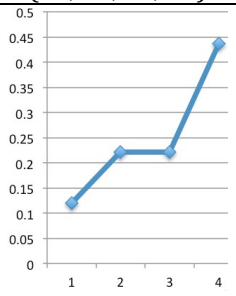
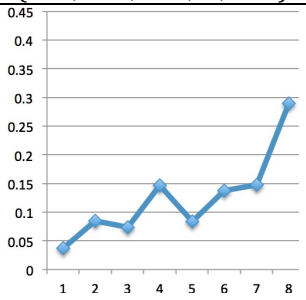
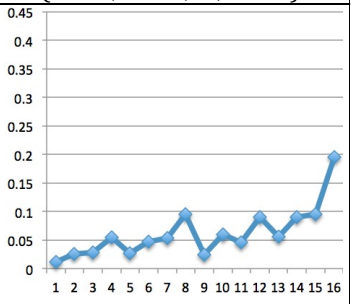
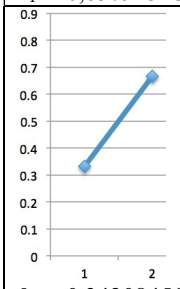
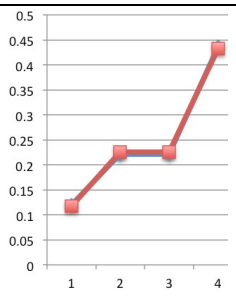
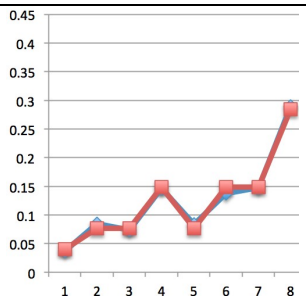
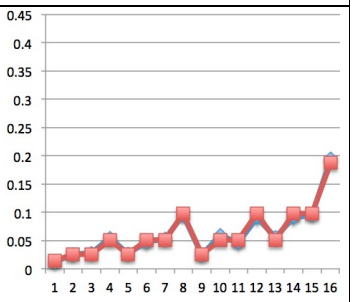
F.M. Dostoevsky «Crime and Punishment» (818099 letters)			
Probabilities of 2 monoplets (0, 1)	Probabilities of 4 doublets (00, 01, 10, 11)	Probabilities of 8 triplets (000, 001, 010, ..., 111)	Probabilities of 16 tetraplets (0000, 0001, ..., 1111)
 $0_P = 0,342084515$ $1_P = 0,657915485$			
 $0_P = 0,342084515$ $1_P = 0,657915485$	 $[0_P, 1_P]^{(2)}$	 $[0_P, 1_P]^{(3)}$	 $[0_P, 1_P]^{(4)}$

Fig. 6. Graphical analysis results of the novel "Crime and Punishment" by F.M. Dostoevsky (the original literary text was accessed from http://samolit.com/books/57/). Probabilities of members in alphabets of binary *n*-plets (*n* = 1, 2, 3, 4) from the binary *n*-plet representation of this novel are shown. Blue points correspond to phenomenologic values of the probabilities of *n*-plets, while red points correspond to model values of the probabilities calculated as components of the vectors $[0_P, 1_P]^{(n)}$, where 0_P and 1_P are probabilities of binary monoplets 0 and 1; (*n*) refers to tensor powers.

F.M. Dostoevsky «Crime and Punishment» (818099 letters)	
Probabilities of members in the alphabet of binary monoplets	
Reality:	$0_P = 0,3421$; $1_P = 0,6579$
Probabilities of members in the alphabet of binary doublets	
Reality:	$0_P0_P = 0,1203$; $0_P1_P = 0,2219$; $1_P0_P = 0,2216$; $1_P1_P = 0,4362$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

Model: $0_P0_P = 0{,}1170$; $0_P1_P = 0{,}2251$; $1_P0_P = 0{,}2251$; $1_P1_P = 0{,}4329$.
Probabilities of members in the alphabet of binary triplets
Reality: $0_P0_P0_P=0{,}0370$; $0_P0_P1_P=0{,}0843$; $0_P1_P0_P=0{,}0738$; $0_P1_P1_P=0{,}1464$; $1_P0_P0_P=0{,}0838$; $1_P0_P1_P=0{,}1377$; $1_P1_P0_P=0{,}1472$; $1_P1_P1_P=0{,}2897$.
Model: $0_P0_P0_P=0{,}0400$; $0_P0_P1_P=0{,}0770$; $0_P1_P0_P=0{,}0770$; $0_P1_P1_P=0{,}1481$; $1_P0_P0_P=0{,}0760$; $1_P0_P1_P=0{,}1481$; $1_P1_P0_P=0{,}1481$; $1_P1_P1_P=0{,}2848$.
Probabilities of members in the alphabet of binary tetraplets
Reality: $0_P0_P0_P0_P=0{,}0119$; $0_P0_P0_P1_P=0{,}0258$; $0_P0_P1_P0_P=0{,}0284$; $0_P0_P1_P1_P=0{,}0551$; $0_P1_P0_P0_P=0{,}0265$; $0_P1_P0_P1_P=0{,}0467$; $0_P1_P1_P0_P=0{,}0535$; $0_P1_P1_P1_P=0{,}0954$; $1_P0_P0_P0_P=0{,}0249$; $1_P0_P0_P1_P=0{,}0594$; $1_P0_P1_P0_P=0{,}0458$; $1_P0_P1_P1_P=0{,}0900$; $1_P1_P0_P0_P=0{,}0560$; $1_P1_P0_P1_P=0{,}0898$; $1_P1_P1_P0_P=0{,}0952$; $1_P1_P1_P1_P=0{,}1954$.
Model: $0_P0_P0_P0_P=0{,}0137$; $0_P0_P0_P1_P=0{,}0263$; $0_P0_P1_P0_P=0{,}0263$; $0_P0_P1_P1_P=0{,}0507$; $0_P1_P0_P0_P=0{,}0263$; $0_P1_P0_P1_P=0{,}0507$; $0_P1_P1_P0_P=0{,}0507$; $0_P1_P1_P1_P=0{,}0974$; $1_P0_P0_P0_P=0{,}0263$; $1_P0_P0_P1_P=0{,}0507$; $1_P0_P1_P0_P=0{,}0507$; $1_P0_P1_P1_P=0{,}0974$; $1_P1_P0_P0_P=0{,}0507$; $1_P1_P0_P1_P=0{,}0974$; $1_P1_P1_P0_P=0{,}0974$; $1_P1_P1_P1_P=0{,}1874$.

Fig. 7. Numeric analysis results of the novel "Crime and Punishment" by F.M. Dostoevsky (the original literary text was accessed from http://samolit.com/books/57/). Probabilities of members in alphabets of binary *n*-plets (n = 1, 2, 3, 4) from the binary n-plet representation of this novel are shown. All values are rounded to four decimal places. Blue numbers correspond to phenomenologic values of the probabilities for cases of alphabets named in tabular sections, while red numbers correspond to model values of these probabilities calculated as components of the vectors $[0_P, 1_P]^{(n)}$, where 0_P and 1_P denote phenomenologic values of probabilities of binary monoplets 0 and 1; (n) refers to tensor powers.

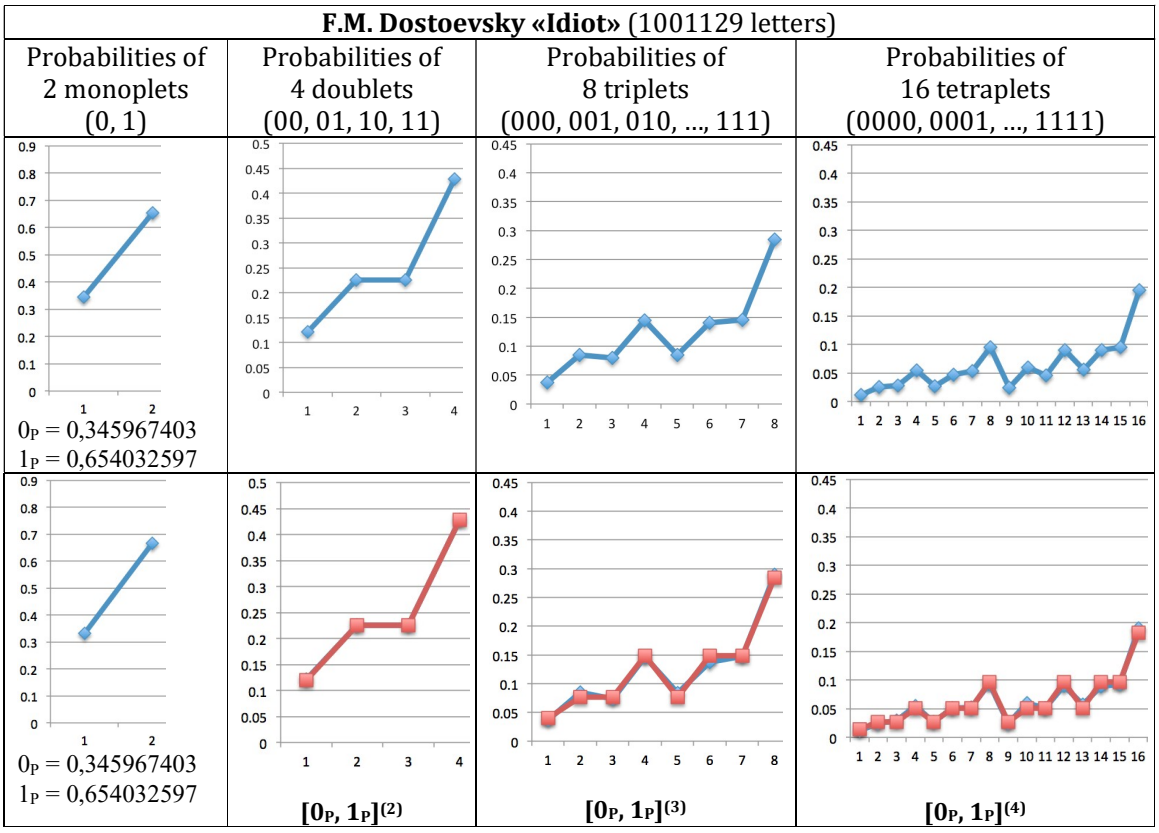


Fig. 8. Graphical analysis results of the novel "Idiot" by F.M. Dostoevsky (the original literary text was accessed from http://samolit.com/books/56/). Probabilities of members in alphabets of binary *n*-plets (n = 1, 2, 3, 4) from the binary *n*-plet representation of this novel are shown. Blue points correspond to phenomenologic values of the probabilities of *n*-plets, while red points correspond to model values of the probabilities calculated as components of the vectors $[0_P, 1_P]^{(n)}$, where 0_P and 1_P are probabilities of binary monoplets 0 and 1; (n) refers to tensor powers.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

F.M. Dostoevsky «Idiot» (1001129 letters)
Probabilities of members in the alphabet of binary monoplets
Reality: $0_P = 0{,}3460$; $1_P = 0{,}6540$
Probabilities of members in the alphabet of binary doublets
Reality: $0_P0_P=0{,}1208$; $0_P1_P=0{,}2251$; $1_P0_P=0{,}2252$; $1_P1_P=0{,}4289$.
Model: $0_P0_P=0{,}1197$; $0_P1_P=0{,}2263$; $1_P0_P=0{,}2263$; $1_P1_P=0{,}4278$.
Probabilities of members in the alphabet of binary triplets
Reality: $0_P0_P0_P=0{,}0367$; $0_P0_P1_P=0{,}0845$; $0_P1_P0_P=0{,}0792$; $0_P1_P1_P=0{,}1449$; $1_P0_P0_P=0{,}0847$; $1_P0_P1_P=0{,}1405$; $1_P1_P0_P=0{,}1405$; $1_P1_P1_P=0{,}2839$.
Model: $0_P0_P0_P=0{,}0414$; $0_P0_P1_P=0{,}0783$; $0_P1_P0_P=0{,}0783$; $0_P1_P1_P=0{,}1480$; $1_P0_P0_P=0{,}0783$; $1_P0_P1_P=0{,}1480$; $1_P1_P0_P=0{,}1480$; $1_P1_P1_P=0{,}2798$.
Probabilities of members in the alphabet of binary tetraplets
Reality: $0_P0_P0_P0_P=0{,}0118$; $0_P0_P0_P1_P=0{,}0251$; $0_P0_P1_P0_P=0{,}0290$; $0_P0_P1_P1_P=0{,}0545$; $0_P1_P0_P0_P=0{,}0276$; $0_P1_P0_P1_P=0{,}0517$; $0_P1_P1_P0_P=0{,}0519$; $0_P1_P1_P1_P=0{,}0941$; $1_P0_P0_P0_P=0{,}0249$; $1_P0_P0_P1_P=0{,}0597$; $1_P0_P1_P0_P=0{,}0505$; $1_P0_P1_P1_P=0{,}0909$; $1_P1_P0_P0_P=0{,}0570$; $1_P1_P0_P1_P=0{,}0885$; $1_P1_P1_P0_P=0{,}0931$; $1_P1_P1_P1_P=0{,}1899$.
Model: $0_P0_P0_P0_P=0{,}0143$; $0_P0_P0_P1_P=0{,}0271$; $0_P0_P1_P0_P=0{,}0271$; $0_P0_P1_P1_P=0{,}0512$; $0_P1_P0_P0_P=0{,}0271$; $0_P1_P0_P1_P=0{,}0512$; $0_P1_P1_P0_P=0{,}0512$; $0_P1_P1_P1_P=0{,}0968$; $1_P0_P0_P0_P=0{,}0271$; $1_P0_P0_P1_P=0{,}0512$; $1_P0_P1_P0_P=0{,}0512$; $1_P0_P1_P1_P=0{,}0968$; $1_P1_P0_P0_P=0{,}0512$; $1_P1_P0_P1_P=0{,}0968$; $1_P1_P1_P0_P=0{,}0968$; $1_P1_P1_P1_P=0{,}1830$.

Fig. 9. Numeric analysis results of the novel "Idiot" by F.M. Dostoevsky (the original literary text was accessed from http://samolit.com/books/56/). Probabilities of members of alphabets of binary n-plets (n = 1, 2, 3, 4) from the binary n-plet representation of this novel are shown. All values are rounded to four decimal places. Blue numbers correspond to phenomenologic values of the probabilities of the named n-plets, and red numbers correspond to model values of these probabilities calculated as components of the vectors $[0_P, 1_P]^{(n)}$, where 0_P and 1_P denote phenomenologic values of probabilities of binary monoplets 0 and 1; (n) refers to tensor powers.

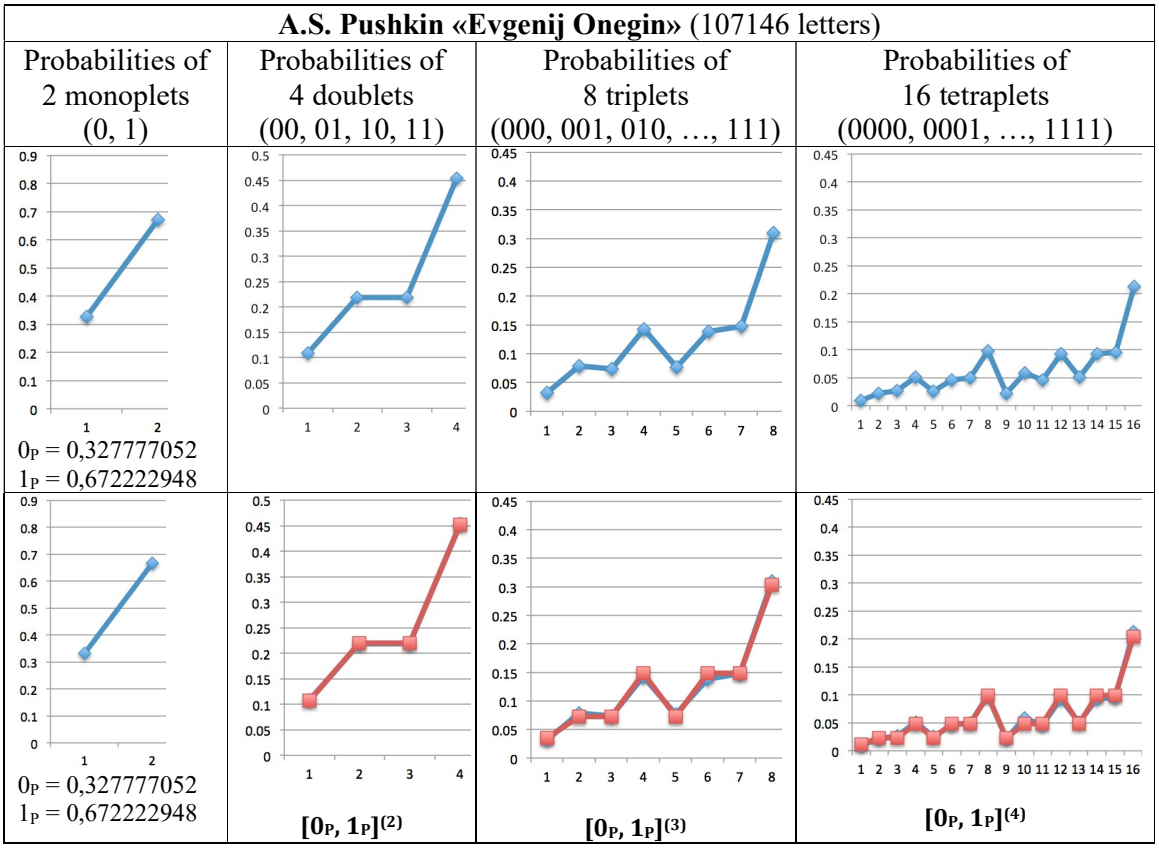


Fig. 10. Graphical analysis results of the novel "Evgenij Onegin" by A.S. Pushkin (the original literary text was accessed from http://tululu.org/b57798/). Probabilities of members of alphabets of binary n-plets (n = 1, 2, 3, 4) from the binary n-plet representation of this novel are shown. Blue points correspond to phenomenologic values of the probabilities of n-plets, while red points correspond to model values of the probabilities calculated as components of the vectors $[0_P, 1_P]^{(n)}$, where 0_P and 1_P are probabilities of binary monoplets 0 and 1; (n) refers to tensor powers.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

A.S. Pushkin «Evgenij Onegin» (107146 letters)
Probabilities of members in the alphabet of binary monoplets
Reality: $0_P = 0{,}3278$; $1_P = 0{,}6722$
Probabilities of members in the alphabet of binary doublets
Reality: $0_P0_P = 0{,}1090$; $0_P1_P = 0{,}2189$; $1_P0_P = 0{,}2187$; $1_P1_P = 0{,}4534$.
Model: $0_P0_P = 0{,}1074$; $0_P1_P = 0{,}2203$; $1_P0_P = 0{,}2203$; $1_P1_P = 0{,}4519$.
Probabilities of members in the alphabet of binary triplets
Reality: $0_P0_P0_P=0{,}0316$; $0_P0_P1_P=0{,}0789$; $0_P1_P0_P=0{,}0738$; $0_P1_P1_P=0{,}1428$; $1_P0_P0_P=0{,}0769$; $1_P0_P1_P=0{,}1389$; $1_P1_P0_P=0{,}1476$; $1_P1_P1_P=0{,}3095$.
Model: $0_P0_P0_P=0{,}0352$; $0_P0_P1_P=0{,}0722$; $0_P1_P0_P=0{,}0722$; $0_P1_P1_P=0{,}1481$; $1_P0_P0_P=0{,}0722$; $1_P0_P1_P=0{,}1481$; $1_P1_P0_P=0{,}1481$; $1_P1_P1_P=0{,}3038$.
Probabilities of members in the alphabet of binary tetraplets
Reality: $0_P0_P0_P0_P=0{,}0098$; $0_P0_P0_P1_P=0{,}0223$; $0_P0_P1_P0_P=0{,}0272$; $0_P0_P1_P1_P=0{,}0509$; $0_P1_P0_P0_P=0{,}0251$; $0_P1_P0_P1_P=0{,}0457$; $0_P1_P1_P0_P=0{,}0500$; $0_P1_P1_P1_P=0{,}0974$; $1_P0_P0_P0_P=0{,}0217$; $1_P0_P0_P1_P=0{,}0588$; $1_P0_P1_P0_P=0{,}0457$; $1_P0_P1_P1_P=0{,}0930$; $1_P1_P0_P0_P=0{,}0513$; $1_P1_P0_P1_P=0{,}0928$; $1_P1_P1_P0_P=0{,}0953$; $1_P1_P1_P1_P=0{,}2131$.
Model: $0_P0_P0_P0_P=0{,}0115$; $0_P0_P0_P1_P=0{,}0237$; $0_P0_P1_P0_P=0{,}0237$; $0_P0_P1_P1_P=0{,}0485$; $0_P1_P0_P0_P=0{,}0237$; $0_P1_P0_P1_P=0{,}0485$; $0_P1_P1_P0_P=0{,}0485$; $0_P1_P1_P1_P=0{,}0996$; $1_P0_P0_P0_P=0{,}0237$; $1_P0_P0_P1_P=0{,}0485$; $1_P0_P1_P0_P=0{,}0485$; $1_P0_P1_P1_P=0{,}0996$; $1_P1_P0_P0_P=0{,}0485$; $1_P1_P0_P1_P=0{,}0996$; $1_P1_P1_P0_P=0{,}0996$; $1_P1_P1_P1_P=0{,}2042$.

Fig. 11. Numeric analysis results of the novel "Evgenij Onegin" by A.S. Pushkin (the original literary text was accessed from http://tululu.org/b57798/). Probabilities of members in alphabets of binary n-plets (n = 1, 2, 3, 4) from the binary n-plet representation of this novel are shown. All values are rounded to four decimal places. Blue numbers correspond to phenomenologic values of the probabilities of the named n-plets, while red numbers correspond to model values of these probabilities calculated as components of the vectors $[0_P, 1_P]^{(n)}$, where 0_P and 1_P denote phenomenologic values of probabilities of binary monoplets 0 and 1; (n) refers to tensor powers.

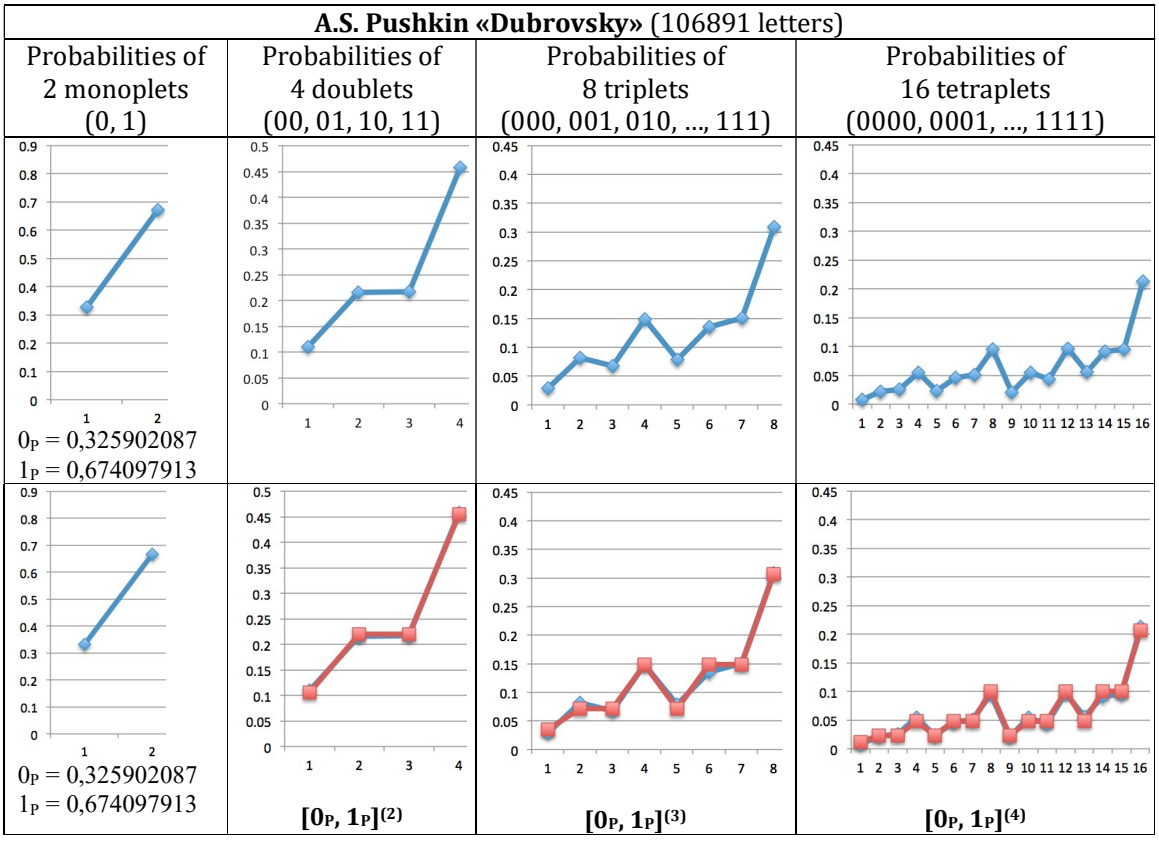


Fig. 12. Graphical analysis results of the novel "Dubrovsky" by A.S. Pushkin (the original literary text was accessed from http://samolit.com/books/61/). Probabilities of members in alphabets of binary n-plets (n = 1, 2, 3, 4) from the binary n-plet representation of this novel are shown. Blue points correspond to phenomenologic values of the probabilities of n-plets, while red

points correspond to model values of the probabilities calculated as components of the vectors $[0_P, 1_P]^{(n)}$, where 0_P and 1_P are probabilities of binary monoplets 0 and 1; (n) refers to tensor powers.

A.S. Pushkin «Dubrovsky» (106891 letters)
Probabilities of members in the alphabet of binary monoplets
Reality: $0_P = 0{,}3259$; $1_P = 0{,}6741$
Probabilities of members in the alphabet of binary doublets
Reality: $0_P0_P = 0{,}1100$; $0_P1_P = 0{,}2152$; $1_P0_P = 0{,}2166$; $1_P1_P = 0{,}4582$.
Model: $0_P0_P = 0{,}1062$; $0_P1_P = 0{,}2197$; $1_P0_P = 0{,}2197$; $1_P1_P = 0{,}4544$.
Probabilities of members in the alphabet of binary triplets
Reality: $0_P0_P0_P=0{,}0295$; $0_P0_P1_P=0{,}0816$; $0_P1_P0_P=0{,}0671$; $0_P1_P1_P=0{,}1491$; $1_P0_P0_P=0{,}0784$; $1_P0_P1_P=0{,}1358$; $1_P1_P0_P=0{,}1503$; $1_P1_P1_P=0{,}3083$.
Model: $0_P0_P0_P=0{,}0346$; $0_P0_P1_P=0{,}0716$; $0_P1_P0_P=0{,}0716$; $0_P1_P1_P=0{,}1481$; $1_P0_P0_P=0{,}0716$; $1_P0_P1_P=0{,}1481$; $1_P1_P0_P=0{,}1481$; $1_P1_P1_P=0{,}3063$.
Probabilities of members in the alphabet of binary tetraplets
Reality: $0_P0_P0_P0_P=0{,}0081$; $0_P0_P0_P1_P=0{,}0223$; $0_P0_P1_P0_P=0{,}0257$; $0_P0_P1_P1_P=0{,}0549$; $0_P1_P0_P0_P=0{,}0234$; $0_P1_P0_P1_P=0{,}0459$; $0_P1_P1_P0_P=0{,}0506$; $0_P1_P1_P1_P=0{,}0952$; $1_P0_P0_P0_P=0{,}0211$; $1_P0_P0_P1_P=0{,}0553$; $1_P0_P1_P0_P=0{,}0437$; $1_P0_P1_P1_P=0{,}0971$; $1_P1_P0_P0_P=0{,}0564$; $1_P1_P0_P1_P=0{,}0916$; $1_P1_P1_P0_P=0{,}0959$; $1_P1_P1_P1_P=0{,}2127$.
Model: $0_P0_P0_P0_P=0{,}0113$; $0_P0_P0_P1_P=0{,}0233$; $0_P0_P1_P0_P=0{,}0233$; $0_P0_P1_P1_P=0{,}0483$; $0_P1_P0_P0_P=0{,}0233$; $0_P1_P0_P1_P=0{,}0483$; $0_P1_P1_P0_P=0{,}0483$; $0_P1_P1_P1_P=0{,}0998$; $1_P0_P0_P0_P=0{,}0233$; $1_P0_P0_P1_P=0{,}0483$; $1_P0_P1_P0_P=0{,}0483$; $1_P0_P1_P1_P=0{,}0998$; $1_P1_P0_P0_P=0{,}0483$; $1_P1_P0_P1_P=0{,}0998$; $1_P1_P1_P0_P=0{,}0998$; $1_P1_P1_P1_P=0{,}2065$.

Fig. 13. Numeric analysis results of the novel "Dubrovsky" by A.S. Pushkin (the original literary text was accessed from http://samolit.com/books/61/). Probabilities of members in alphabets of binary n-plets (n = 1, 2, 3, 4) from the binary n-plet representation of this novel are shown. All values are rounded to four decimal places. Blue numbers correspond to phenomenologic values of the probabilities of the named n-plets, while red numbers correspond to model values of these probabilities calculated as components of the vectors $[0_P, 1_P]^{(n)}$, where 0_P and 1_P denote phenomenologic values of probabilities of binary monoplets 0 and 1; (n) refers to tensor powers.

Russian Bible (3122489 letters)

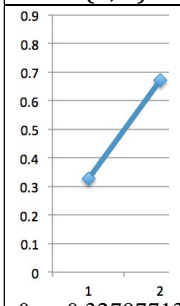
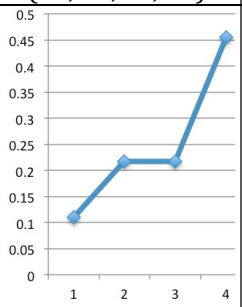
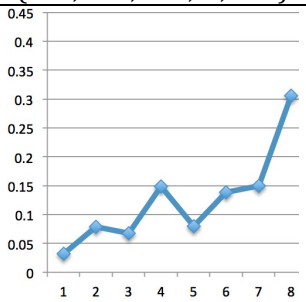
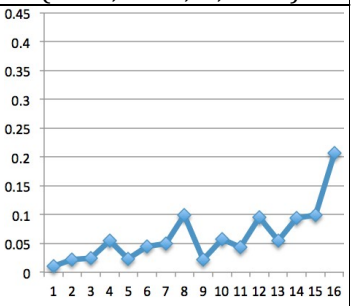
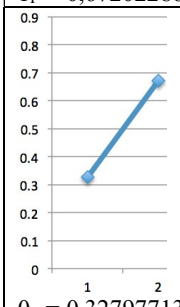
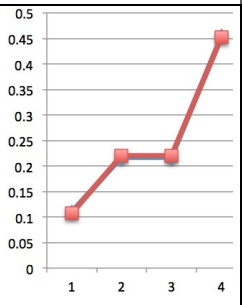
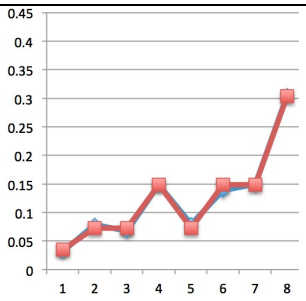
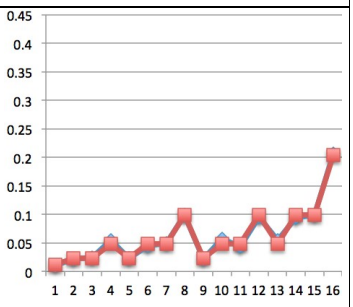
Probabilities of 2 monoplets (0, 1)	Probabilities of 4 doublets (00, 01, 10, 11)	Probabilities of 8 triplets (000, 001, 010, ..., 111)	Probabilities of 16 tetraplets (0000, 0001, ..., 1111)
 $0_P = 0{,}327977136$ $1_P = 0{,}672022864$			
 $0_P = 0{,}327977136$ $1_P = 0{,}672022864$	 **$[0_P, 1_P]^{(2)}$**	 **$[0_P, 1_P]^{(3)}$**	 **$[0_P, 1_P]^{(4)}$**

Fig. 14. Graphical analysis results of the Russian Bible (the original literary text was accessed from http://petoukhov.com/bible.zip). Probabilities of members of alphabets of binary n-plets (n = 1, 2, 3, 4) from the binary n-plet representation of this novel are shown. Blue points correspond to phenomenologic values of the probabilities of n-plets, while red

points correspond to model values of the probabilities calculated as components of the vectors $[0_P, 1_P]^{(n)}$, where 0_P and 1_P are probabilities of binary monoplets 0 and 1; (n) refers to tensor powers.

Russian Bible (3122489 letters)
Probabilities of members in the alphabet of binary monoplets
Reality: $0_P = 0{,}3280$; $1_P = 0{,}6720$
Probabilities of members in the alphabet of binary doublets
Reality: $0_P0_P = 0{,}1107$; $0_P1_P = 0{,}2171$; $1_P0_P = 0{,}2174$; $1_P1_P = 0{,}4548$.
Model: $0_P0_P = 0{,}1076$; $0_P1_P = 0{,}2204$; $1_P0_P = 0{,}2204$; $1_P1_P = 0{,}4516$.
Probabilities of members in the alphabet of binary triplets
Reality: $0_P0_P0_P = 0{,}0318$; $0_P0_P1_P = 0{,}0789$; $0_P1_P0_P = 0{,}0678$; $0_P1_P1_P = 0{,}1484$; $1_P0_P0_P = 0{,}0794$; $1_P0_P1_P = 0{,}1386$; $1_P1_P0_P = 0{,}1493$; $1_P1_P1_P = 0{,}3058$.
Model: $0_P0_P0_P=0{,}0353$; $0_P0_P1_P=0{,}0723$; $0_P1_P0_P=0{,}0723$; $0_P1_P1_P=0{,}1481$; $1_P0_P0_P=0{,}0723$; $1_P0_P1_P=0{,}1481$; $1_P1_P0_P=0{,}1481$; $1_P1_P1_P=0{,}3035$.
Probabilities of members in the alphabet of binary tetraplets
Reality: $0_P0_P0_P0_P=0{,}0099$; $0_P0_P0_P1_P=0{,}0222$; $0_P0_P1_P0_P=0{,}0245$; $0_P0_P1_P1_P=0{,}0543$; $0_P1_P0_P0_P=0{,}0232$; $0_P1_P0_P1_P=0{,}0448$; $0_P1_P1_P0_P=0{,}0493$; $0_P1_P1_P1_P=0{,}0995$; $1_P0_P0_P0_P=0{,}0222$; $1_P0_P0_P1_P=0{,}0568$; $1_P0_P1_P0_P=0{,}0438$; $1_P0_P1_P1_P=0{,}0948$; $1_P1_P0_P0_P=0{,}0552$; $1_P1_P0_P1_P=0{,}0937$; $1_P1_P1_P0_P=0{,}0997$; $1_P1_P1_P1_P=0{,}2063$.
Model: $0_P0_P0_P0_P=0{,}0116$; $0_P0_P0_P1_P=0{,}0237$; $0_P0_P1_P0_P=0{,}0237$; $0_P0_P1_P1_P=0{,}0486$; $0_P1_P0_P0_P=0{,}0237$; $0_P1_P0_P1_P=0{,}0486$; $0_P1_P1_P0_P=0{,}0486$; $0_P1_P1_P1_P=0{,}0995$; $1_P0_P0_P0_P=0{,}0237$; $1_P0_P0_P1_P=0{,}0486$; $1_P0_P1_P0_P=0{,}0486$; $1_P0_P1_P1_P=0{,}0995$; $1_P1_P0_P0_P=0{,}0486$; $1_P1_P0_P1_P=0{,}0995$; $1_P1_P1_P0_P=0{,}0995$; $1_P1_P1_P1_P=0{,}2040$.

Fig. 15. Numeric analysis results of the Russian Bible (the original literary text was accessed from http://petoukhov.com/bible.zip). Probabilities of members in alphabets of binary n-plets (n = 1, 2, 3, 4) from the binary n-plet representation of this novel are shown. All values are rounded to four decimal places. Blue numbers correspond to phenomenologic values of the probabilities of the named n-plets, while red numbers correspond to model values of these probabilities calculated as components of the vectors $[0_P, 1_P]^{(n)}$, where 0_P and 1_P denote phenomenologic values of probabilities of binary monoplets 0 and 1; (n) refers to tensor powers.

Our results show that the described properties of long Russian literary texts reflect the deep specifics of that language and not the particular literary style of a particular writer. It can be assumed that any long literary text in a foreign language, translated into Russian, will demonstrate similar properties.

It will be important to discover if there are similar patterns in the texts of other languages with differing alphabets and differing phonetic features. The author has begun to conduct these studies for literary texts in English, German, French and many other languages. It is hoped to publish the results of these studies in the future.

The next Section shows that analogical properties exist in many long DNA-texts, from data analyzed by the author, and that the described model approach, using the tensor product of vectors of probabilities, is effective not only in the case of long Russian literary texts but also in the case of long DNA-texts. Rules of these probabilities in long DNA-texts and in long literary Russian texts have mathematical analogies with the main law of population genetics - the Hardy–Weinberg law [Petoukhov, 2018a].

3. Regarding similar properties of long DNA-texts

In DNA molecules, genetic information is written in very long sequential texts using only 4 letters: adenine A, cytosine C, guanine G, thymine T. For example, the human genome consists of several billion such letters. In the double helix of DNA, complementary letters C-G and A-T are always located opposite each other and form complementary pairs by means of 3 hydrogen bonds for the pair C-G and by means of 2 hydrogen bonds for the pair A-T (it can be denoted as C=G=3 and A=T=2). Correspondingly, any DNA sequence contains a chain of 2 and 3 hydrogen bonds, for example, 33223223233… . We term such number

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

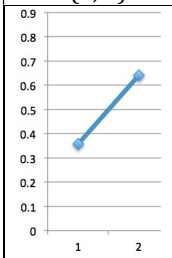
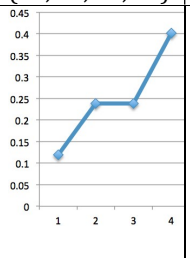
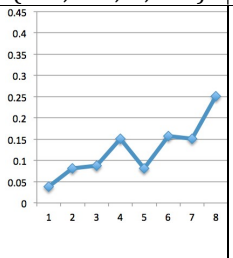
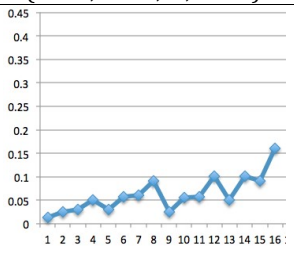
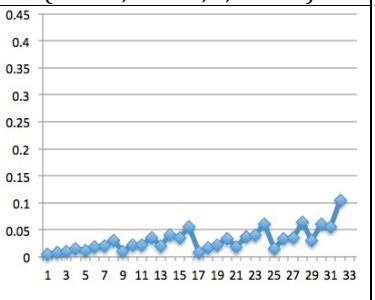
chains of hydrogen bonds: “hydrogen bond sequences” (briefly, “H-sequences” or “H-texts”). The author analysed the properties of such long hydrogen bond sequences for many different organisms (here the term “long” means DNA sequences containing ≥ 100000 letters). The results of the author’s study are briefly presented, below, which confirm the existence of structural connections between long genetic texts and long literary texts.

Each of the long chains (or sequences) of numbers 2 and 3 of hydrogen bonds (for example, the H-sequence 3-2-3-3-3-2-2-2-3-2-3-3-…) can also be represented as a sequence of H-doublets (32-33-32-22-32-33-…), or a sequence of H-triplets (323-332-223-233-…) or, in a general case, as a sequence of H-*n*-plets (n = 1, 2, 3, 4, 5,…). We term these different representations of long sequences of nucleotides A, T, C, G, as forms of H-n-plets sequence, referred to as its “H-*n*-plet representations”.

A complete set of H-*n*-plets under fixed value n is called “the alphabet of H-*n*-plets” and contains 2^n members: for example, the alphabet of H-monoplets contains 2 members (3 and 2); the alphabet of H-doublets contains 4 members (33, 32, 23, 22) and only these members exist in the H-doublet representation of any DNA sequence; the alphabet of H-triplets contains $2^3 = 8$ members (333, 332, 323, 322, 233, 232, 223, 222); etc. It should be mentioned that the tensor family of vectors $[3, 2]^{(n)}$ (where (n) refers to the tensor power, n = 1, 2, 3, 4, …) contains vectors, whose sets of components coincide with corresponding alphabets of H-*n*-plets. For example, $[3, 2]^{(2)} = [33, 32, 23, 22]$; $[3, 2]^{(3)} = [333, 332, 323, 322, 233, 232, 223, 222]$, etc. Therefore, it can be seen that, conditionally, all H-*n*-plets alphabets belong to the same tensor family of vectors and are interconnected with each other.

The article (Petoukhov, 2018a) describes in detail these hidden regularities in the long DNA H-sequences, which are connected with members of the alphabets of H-*n*-plets of many different organisms. These regularities concern probabilities (or frequencies or percentage) of each of hydrogen *n*-plets (briefly, H-*n*-plets) in long DNA H-sequences of the complete set of chromosomes. The author has revealed that for a concrete long DNA sequence, complete sets of individual probabilities of H-*n*-plets (under different values n) are interrelated with each other. This interrelation can be modeled by the same method described above for the analysis of literary texts: if q and p are probabilities of numbers 3 and 2 of hydrogen bonds in the considered DNA, then tensor powers (n) of the 2-dimensional vector [q, p] gives 2^n-dimensional vectors $[q, p]^{(n)}$, components of which are model values of probabilities of members of the corresponding alphabet of H-*n*-plets in this DNA (n = 1, 2, 3, 4, … is much less than the length of such sequences).

For an illustration of this statement, Fig. 16 shows - in a graphical form - an example of phenomenological values of probabilities of all members of alphabets of H-*n*-plets (n = 1, 2, 3, 4, 5) in the case of the DNA sequence of the first chromosome of the plant *Arabidopsis thaliana*, which contains 30427671 nucleotide pairs. Simultaneously, Fig. 16 shows model values of these probabilities as components of 2^n-dimensional vectors $[q, p]^{(n)}$, where q=0,35873552 and p=0,64126448.

Probabilities of 2 monoplets (3, 2)	Probabilities of 4 doublets (33, 32, 23, 22)	Probabilities of 8 triplets (333, 332, ..., 222)	Probabilities of 16 tetraplets (3333, 3332, ..., 2222)	Probabilities of 32 pentaplets (33333, 33332,..., 22222)
 q=0,35873552 p=0,64126448				

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

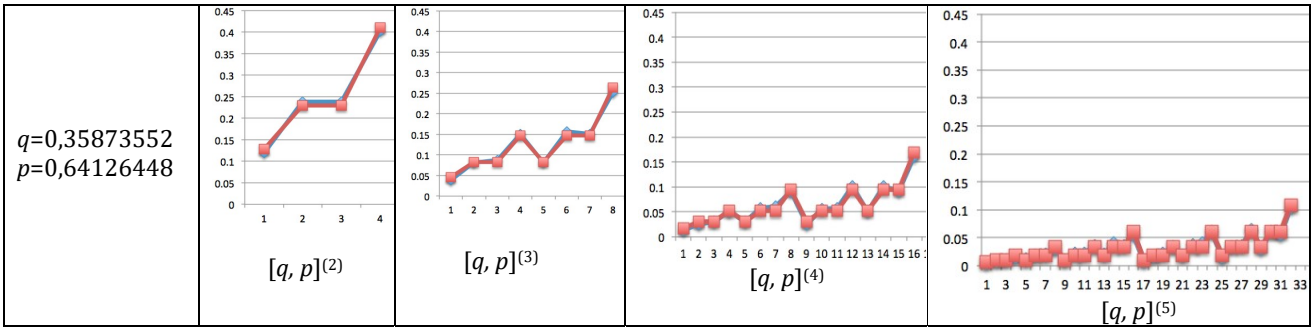


Fig. 16. The graphic representation of probabilities in members of hydrogen *n*-plet alphabets (n = 1, 2, 3, 4, 5) in the DNA sequence of the first chromosome of the plant *Arabidopsis thaliana* (initial data relating to this chromosome were accessed from https://www.ncbi.nlm.nih.gov/nuccore/NC_003070.9). Blue points in the graphs show phenomenological probabilities of n-plets of numbers of hydrogen bonds, while red points show model values of these probabilities as components of 2^n-dimensional vectors $[q, p]^{(n)}$, where q and p are probabilities of hydrogen bonds 3 and 2 in this DNA.

Fig. 17 shows phenomenological and model values of probabilities of all members of hydrogen *n*-plet alphabets (n = 1, 2, 3, 4, 5) for the same DNA sequence as in Fig. 16. It can be seen that these model values reproduce phenomenological values with the level of accuracy, which one can see in Fig. 17.

Probabilities of members in the alphabet of H-monoplets (numbers 3 and 2)
Reality: q = [3] = 0,3587; p = [2] = 0,6413
Probabilities of members in the alphabet of H-doublets (4 numbers: 33, 32, 23, 22)
Reality: [33] = 0,1198; [32] = 0,2390; [23] = 0,2389; [22] = 0,4023.
Model: [33] = 0,1287; [32] = 0,2300; [23] = 0,2300; [22] = 0,4112.
Probabilities of members in the alphabet of H-triplets (8 numbers: 333, 332, 323, 322, 233, 232, 223, 222)
Reality: [333]= 0,0385; [332]= 0,0812; [323]= 0,0880; [322]= 0,1507; [233]= 0,0812; [232]= 0,1577; [223]= 0,1514; [222]= 0,2512.
Model: [333]=0,0462; [332]=0,0825; [323]=0,0825; [322]=0,1475; [233]=0,0825; [232]=0,1475; [223]=0,1475; [222]=0,2637.
Probabilities of members in the alphabet of H-tetraplets (16 numbers: 3333, 3332, 3323, 3322, 3233, 3232, 3223, 3222, 2333, 2332, 2323, 2322, 2233, 2232, 2223, 2222)
Reality: [3333]=0,0132; [3332]=0,0253; [3323]=0,0310; [3322]=0,0502; [3233]=0,0311; [3232]=0,0570; [3223]=0,0601; [3222]=0,0906; [2333]=0,0253; [2332]=0,0560; [2323]=0,0570; [2322]=0,1007; [2233]=0,0502; [2232]=0,1008; [2223]=0,0907; [2222]=0,1607.
Model: [3333]=0,0166; [3332]=0,0296; [3323]=0,0296; [3322]=0,0529; [3233]=0,0296; [3232]=0,0529; [3223]=0,0529; [3222]=0,0946; [2333]=0,0296; [2332]=0,0529; [2323]=0,0529; [2322]=0,0946; [2233]=0,0529; [2232]=0,0946; [2223]=0,0946; [2222]=0,1691.

Fig. 17. Phenomenological values (in blue color) and model values (in red color) of probabilities of all members of the alphabets of hydrogen *n*-plets (n = 1, 2, 3, 4) in the DNA sequence of the first chromosome of *Arabidopsis thaliana* (appropriate graphs are shown in Fig. 16; initial data relating to this chromosome were accessed from https://www.ncbi.nlm.nih.gov/nuccore/NC_003070.9). Numbers in square brackets refer to probabilities of corresponding hydrogen *n*-plets (for example, the symbol [323] refers to the probability of the H-triplet 323 in the triplet-representation of the hydrogen sequence of this DNA like 322-232-233-…). All values are rounded to the fourth decimal place.

Similar results have been obtained in our analysis of the plant, *Arabidopsis thaliana*; nematode, *Caenorhabditis elegans*; fruit fly, *Drosophila melanogaster*; house mouse, *Mus musculus;* and Homo Sapiens, drawing on nuclear chromosome data and DNA sequence data obtained from GenBank.

These results assume the existence of the following general rule for eucariots about probabilities (or frequencies) of hydrogen bonds in complete sets of nuclear chromosomes:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

- in complete sets of nuclear chromosomes of different organisms, a probability of any member of the H-*n*-alphabets has approximately the same value in all chromosomes.

However, further research is needed to define a degree of universality and precision of this rule.

The author has also calculated probabilities of members of H-*n*-alphabets (n = 1, 2, 3, 4, 5) in 19 genomes of bacteria and archaea from the full list in the article [Rapoport, Trifonov, 2012, p. 2]: "*Aquifex aeolicus, Acidobacteria bacterium, Bradyrhizobium japonicum, Bacillus subtilis, Chlamydia trachomatis, Chromobacterium violaceum, Dehalococcoides ethenogenes, Escherichia coli, Flavobacterium psychrophilum, Gloeobacter violaceus, Helicobacter pilory, Methanosarcina acetivorans, Nanoarchaeum equitans, Syntrophus aciditrophicus, Streptomyces coelicolor, Sulfolobus solfataricus, Treponema denticola, Thermotoga maritima and Thermus thermophiles*". The calculated sets of these probabilities were also modelled on the basis of sets of coordinates of appropriate vectors $[q, p]^{(n)}$. These results confirm that the proposed model approach on the basis of vectors $[q, p]^{(n)}$ can be used to obtain idealized models of probabilities of all members of H-*n*-alphabets in actual long DNA sequences (n = 1, 2, 3, 4, … is much less than the length of such sequences).

The proposed model allows the prediction of probabilities of many members of H-*n*-alphabets in long DNA sequences with a high level of accuracy, on the basis of knowledge of probabilities of only two numbers - 3 and 2 – of hydrogen bonds in the DNA.

It is important to remember that if a quantum state can be represented as a vector of a Hilbert space, such a state is termed a pure quantum state. If a pure state $|\psi\rangle$ can be written in the form $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$, where $|\psi_i\rangle$ is a pure state of the i-th subsystem, it is said to be *separable* [Nielsen, Chuang, 2010]. Otherwise it is termed *entangled*. Correspondingly, the present model approach, using the tensor product, represents long DNA sequences of n-plets of hydrogen bonds (n = 2, 3, 4,…) as quantum systems in their separable pure state. The observed difference between the real and model values of the probabilities under study can be interpreted as a violation of a separable pure state, due to the presence of an entangled state. In the case of human chromosomes, this difference between the real and model values of the probabilities studied is more significant than that of the plant *Arabidopsis thaliana* shown in Fig. 16. This infers that, from the model point of view, entangled states are more pronounced in the case of human chromosomes. On the basis of this model approach it is interesting to compare characteristics of such entangled states in genomes of many organisms having different levels of consciousness. Whether the human genome possesses the highest level of such entangled state or not?

4. Quantum biology and life's rules of probability postulated by P. Jordan

Let us return to the historically first work on quantum biology [Jordan, 1932] written by one of creators of quantum mechanics. He asked: Are the laws of atomic and quantum physics of essential importance for life? In fact, Jordan had been thinking about this question for over a decade and had been using the term «Quantumbiologie» since the late 1930s. Jordan believed that the specifics of living organisms are based on special laws that should be discovered in the future. Moreover, he claimed that «*life's missing laws were the rules of chance and probability (the indeterminism) of the quantum world that were somehow scaled up inside living organisms*» [McFadden, Al-Khalili, 2018]. Here, the difference between biological and inanimate objects should be explained. Jordan correctly pointed out that inanimate objects were governed by the average random motion of millions of particles, such that the motion of a single molecule has no influence whatsoever on the whole object. This insight is usually credited to Erwin Schrödinger, who later claimed that life was different from inorganic chemistry because of its dependence on the dynamics of a small number of molecules. Jordan similarly argued that the few molecules that control the dynamics of living cells within the control center have a dictatorial influence, such that quantum-level events that govern their motion, such as Heisenberg's uncertainty principle, are amplified to influence the entire organism. Jordan called this his 'amplifier theory' and based it on Bohr's notion of the 'irreversible act of amplification' that is required in order to bring the fuzzy quantum reality into sharp focus by 'observing' it. Jordan believed that living organisms were uniquely able to carry out this amplification in a way that was conspicuously different from inanimate matter, such as a Geiger counter. Jordan was convinced he could extend quantum indeterminism from the subatomic world to macroscopic biology. He even made a connection with free will by suggesting a link between quantum mechanics and psychology. Jordan's

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

insistence that living organisms have a unique ability to amplify the quantum into the macroscopic world has a lot of resonance with modern views of quantum biology [McFadden, Al-Khalili, 2018].

In accordance with these pioneering ideas of Jordan on the hidden laws of probability in living organisms, the author conducted a systematic study of the rules of probability in long DNA sequences as it has been described in this article and in [2018b]. DNA sequences are a very important and convenient for such computerized research, which in Jordan's day was impossible.

In addition, the author pays special attention to DNA sequences of hydrogen bonds, taking into account the well-known data on the exceptional role of hydrogen bonds in biological organization and in water. For example, L. Pauling believed and wrote that «*the significance of the hydrogen bond for physiology is greater that of any other single structural feature*" [Pauling, 1940, p. 450]. The author's book [Petoukhov, 2001] describes the regular proton groupings in the molecular genetic system showing phenomenological data on the structural and "arithmetic" role of protons (nuclei of hydrogen atoms) in genetics; these phenomenological data are connected in particular with Yin-Yang dichotomous schemes, digrams, trigrams and hexagrams of the Ancient Chinese book «I Ching» written a few thousand years ago and including the famous table of 64 hexagrams in Fu-Xi's order. "I Ching" declares a universality of a cyclic principle of organization in nature; traditional Oriental medicine is based on positions of this book. As known, the famous expert in molecular genetics G. Stent was the first author published a hypothesis about a possible connection between genetic code structures and a symbolic system of the "I Ching" [Stent, 1969, p. 64]. A few authors have supported him and his hypothesis later.

The ancient Chinese claimed that this table of 64 hexagrams is a universal archetype of nature, a universal classification system. They knew nothing about the genetic code, but the genetic code with its 64 codons and other important features is strikingly similar to this ancient table and the system of the "I-Ching". It should also be noted that the numbers 2 and 3 of hydrogen bonds in DNA, played the basic role in Ancient Chinese arithmetic. According to the creator of analytical psychology, Carl Jung [Fordham, 1978; Jung, 1947; Samuels, 2005], trigrams and hexagrams of the "I Ching" fix a universal set of archetypes (innate psychic structures). Jung was a connoisseur of this book and called it a "great and unique" work. On the basis of the concept of archetypes with using schemes of the "I Ching", he developed his amplification method that he successfully applied to the treatment of patients. Wolfgang Pauli made a prominent contribution to the development of the concept of archetypes as a result of his many years of cooperation on this topic with Jung [Pauli, Jung, 2001]. Pauli introduced Jung to the ideas and concepts of quantum mechanics and quantum biology. It seems not to be accidental that Jung used the same term «amplification» in his method that Jordan used in his work on quantum biology. The materials on structural connections between the genetic code and schemes of the "I Ching" were additionally described in detail in [Petoukhov, 1999, 2001, 2008; Petoukhov, He, 2009; Hu, Petoukhov, Petukhova, 2017b].

Taking into account the structural connection of the genetic molecular systems and Jung's archetypes of the unconscious with schemes of the «I Ching», the concept of "biological archetypes" in an expanded sense, should be borne in mind, including universal biological properties and phenomena. One such biological archetype or universal biological property is the main psychophysiologic law by Weber-Fechner (http://en.wikipedia.org/wiki/Weber–Fechner_law): the intensity of the perception is proportional to the logarithm of stimulus intensity. It is known that different types of inherited sensory perception are subordinated to this law: sight, hearing, smell, touch, taste, etc. Because of this law, the power of sound in music and technology is measured on a logarithmic scale in decibels. One can suppose that the innate Weber–Fechner law (WF-law) is a law pertaining to the human nervous system. However, its meaning is actually much wider because it holds true in many kinds of lower organisms without a nervous system: "*this law is applicable to chemo-tropical, helio-tropical and geo-tropical movements of bacteria, fungi and antherozoids of ferns, mosses and phanerogams ... . The Weber-Fechner law, therefore, is not the law of the nervous system and its centers, but the law of protoplasm in general and its ability to respond to stimuli*" [Shultz, 1916, p. 126]. The conception of multi-resonant genetics proposes the mathematical model of the Weber-Fechner law on the base of natural resonant frequences of a particular class of vibrational systems with 2 degrees of freedom [Petoukhov, 2015a, 2016a]. It seems that the structural connections described in this article between long genetic and literary texts represent one of the biological archetypes in bioinformatics.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

5. Some concluding remarks

It is worthy of note that the probabilities of alternation of vowels and consonant letters have been the focus of studies in Russian literary texts by professional mathematicians. Their results are important for many scientific problems including problems of artificial intelligence [Domingos, 2015]. For example, Russian mathematician A.A.Markov, who is a creator of the famous "Markov's chains", developed his theory (of chains) using analysis of the poem "Evgeniy Onegin", where he manually calculated probabilities of vowels and consonants more than 100 years ago [Markov, 1924]. Analysis of probabilities of different n-plets of vowels and consonants in the poem "Evgeniy Onegin" was also made in the work devoted to the mentioned Markov's research [Petrenko, 2018]. The great Russian mathematician, A.N. Kolmogorov, studied rhythms of verses to reveal the mechanism of developing internal goals of self-organizing systems [Kolmogorov, 1997; Nikolaev, 1999]. In contrast to all known research concerning vowels and consonants and their concatenations in doublets, triplets, etc., in Russian literary texts, the author is studying the following:

- First, the author studies the total probabilities of all representatives of two described classes of equivalency of Russian letters (in Fig. 1 the first class is denoted by yellow and the second class is denoted by green) including the probabilities of different n-plets of these representatives (n = 1, 2, 3, 4, …); (as distict from the probabilities of alternation of separate vowels and consonant letters);
- Second, from this standpoint the author studies analogies between long literary texts and hydrogen sequences of DNA to test the concept that linguistic languages are continuations of genetic languages, which are basic for all biological organisms;
- Third, the author studies possibilities of applications of formalisms of quantum informatics to model long literary texts and DNA texts for developing the known hypotheses of some authors concerning quantum-informational organization of informatics in living bodies.

Science has led to a new understanding of life itself: «*Life is a partnership between genes and mathematics*» [Stewart I., 1999]. The author's results, published in various editions, favourably testifies that the mathematics of quantum informatics and of the theory of resonances of oscillatory systems with 2^n degrees of freedom, could be such a partner of the genetic system [Petoukhov, 2015a,b,c; Petoukhov, 2016a; Petoukhov, Petukhova, 2017b; Hu, Petoukhov, Petukhova, 2017a, 2018].

The results described above concerning the parallels between long genetic and literary texts were obtained by the author on the basis of his thoughts regarding the deep connections that exist between molecular genetics and quantum informatics, and of his hypothesis that living organisms are quantum-algorithmic essences [Petoukhov, 2018a,b; Petoukhov, Petukhova, Svirin, 2018; Petoukhov, Svirin, 2018].

Quantum computers use mainly the following mathematical formalisms:

- unitary (or orthogonal) operators;
- the tensor (or Kronecker) multiplication of matrices;
- the logical operation of modulo-2 addition.

The author's research has revealed that different families of structured DNA alphabets of n-plets of nucleotides C, G, A, T, and also of hydrogen bonds, correspond to these formalisms in a high degree [Petoukhov, 2018a,b; Petoukhov, Petukhova, Svirin, 2018]. The research described in this paper also points to the possibility that applications of mathematics of quantum informatics for modeling biological phenomema can help the understanding of many biological phenomena. For example, an adult human organism has around 10 trillion (10^{14}) cells and each cell contains an identical copy of DNA, genetic information used for the physiological functioning of the organism as a holistic system of cells. It is important to ask how such a huge number of cells reliably function as a cooperative whole? Quantum informatics and associations with quantum computers can help to model and understand holistic biological systems and can assist in creating new and effective systems of artificial intelligence on the basis of biological prototypes. (For more information on perspectives of using quantum informatics for artificial intelligence see [Biamonte et al, 2017]).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

Many authors have supposed that living organisms use principles of quantum informatics [Altaisky, 2001; Igamberdiev, 2013; Josephson, 2018; Matsuno, 1999, 1999, 2003, 2015; Matsuno, Paton, 2000; Mikheenko, 2018; Patel, 2001a,b,c; Penrose,1996; Quantum Aspects of Life, 2008]. For example, in his thoughts concerning quantum computers in living organisms, R. Penrose appeals to the fact that tubulin proteins exist in two different configurations, and they can switch between these configurations like triggers, providing bio-computer functions [Penrose, 1996]. By contrast to this "protein standpoint", results of the model approach of resonance genetics testify that the molecular-genetic level, which is the deepest level of living organisms, is connected with the principles of quantum computers [Petoukhov, 2018a,b; Petoukhov, Petukhova, 2017a; Petoukhov, Petukhova, Svirin, 2018].

It should be noted that in his researches the author is purposefully looking for the intersection of bioinformational structures with the formalisms of quantum informatics. For example, the data given in this article concerning the probabilities of *n*-plets in long literary texts and long DNA sequences can also be interpreted on the basis of the known theorem: the probability of the product of two independent events is equal to the product of the probabilities of these events. But the consideration of DNA *n*-plets as combinations of independent events poorly corresponds to the phenomenology of the genetic coding system, in which, for example, in single-stranded DNA the probabilities of *n*-plets combined from DNA letters C, G, A, T depends on the letter order in *n*-plets of the same letter composition. The individual probability of doublet CG can differ by a factor 6 from the individual probability of the doublet GC in the long DNA sequence (see such example in [Petoukhov, 2018b]). This means that the probabilities of the doublet CG and the doublet GC in the long single-stranded DNA cannot be simultaneously represented as the product of the probabilities of the monoplets C and G; in other words, the probabilities of douplets CG and GC cannot be considered as the product of probabilities of independent events C and G. In the case of DNA chains of numbers 3 and 2 of hydrogen bonds, we are referring to the total probability for the set of doublets CG + GC because both these doublets of DNA letters belong to one class of equivalence "33" doublets of hydrogen bonds. Since members CG and GC of this equivalence class of hydrogen bonds are composed of dependent events, it seems to be inappropriate to model the probabilities of hydrogen n-plets as the product of two independent events; therefore the use of quantum algorithmic modeling is preferred, with the tensor product of the probability vectors and with other formalisms of quantum informatics, which have proven themselves in other genetic models, including the concepts of multi-resonance genetics and of geno-logic coding [Petoukhov, 2008, 2011, 2016b, 2017, 2018b; Petoukhov, He, 2009; Hu, Petoukhov, Petukhova, 2017b; Petoukhov et al., 2017; Petoukhov, Petukhova, Svirin, 2018].

Referring to the connection between quantum informatics and resonances, E. Schrodinger emphasised the basic meaning of resonances in quantum mechanics: "*The one thing which one has to accept and which is the inalienable consequence of the wave-equation as it is used in every problem, under the most various forms, is this: that the interaction between two microscopic physical systems is controlled by a peculiar law of resonance*» [Schrodinger, 1952, p.115]. In considering an exact balance in nature between bundles of energy, lost by one system and gained by another, he noted: «*I maintain that it can in all cases be understood as a resonance phenomenon*» [ibid, p.114].

L.Pauling used ideas of resonances in quantum mechanical systems in his theory of resonance in structural chemistry. His book [Pauling, 1940] about this theory is the most quoted among scientific books of the 20th century. The actual molecule, as Pauling proposed, is a sort of hybrid, a structure that resonates between the two alternative extremes; and whenever there is a resonance between the two forms, the structure is stabilized. He wrote: «*Among the most interesting problems of science are those of the structure and properties of substances of biological importance. I have little doubt that in this field resonance and the hydrogen bond are of great significance, and that these two structural features will be found to play an* important *part in such physiological phenomena as the contraction of muscle and the trasmission of impulses along nerves and in the brain»* [Pauling, 1940, p. 570].

In the 1940's Pauling was unaware of the existence of the double helix of DNA, but the author's results, concerning the requlrities of hydrogen bonds in long DNA sequences, confirm in some degree his prediction regarding the important role of hydrogen bonds in the transmission of biological information. These results also seem to have relations with research into the unusual properties of water in living bodies (see for example [Pollack, 2013] and materials of annual conferences on the physics, chemistry and biology of water, http://www.waterconf.org/). Many authors believe that hydrogen bonds play an important role in the transmission of information through water and other water-based liquids. Biological matter contains mainly structured water. For example, jellyfish–consist of 99% water, yet are vibrant, energetic

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

creatures. Hydrogen is a component of almost all organic matter and is present in all living cells, where the number of hydrogen atoms is almost 63% [Hornak, 1996]. Taking into account the existing hypotheses about the cosmic origins of life, it should be recalled that hydrogen is the main component of stars and interstellar gas. A hydrogen atom, devoid of an electron, is a proton; the participation of numbers of protons in the structural organization of the molecular genetic system is described in the book [Petoukhov, 2001]. It may be reasonable to consider that living bodies are, in essence, a construction of hydrogen atoms and hydrogen bonds. Science, in the future, should answer in what degree this thought is true.

Pauling's theory uses the fundamental principle of a minimal energy because – in resonant combining of parts into a single unit – each of the members of the ensemble requires less energy for performing their own work than when working individually. Of course, this fundamental principle can be used in many other cases of resonances in different systems. In particular, in the article [Petoukhov, Petukhova, 2017b], we used the theory of oscillators with many degrees of freedom to model some phenomena of Mendelian genetics and to analyze structures of genetic-molecular alphabets; also to explain the phenomena of segregation in these molecular alphabets, the existence of dominant and recessive resonances in nitrogenous bases of DNA and RNA that have been postulated by analogy with dominant and recessive alleles in Mendelian genetics. A hypothesis concerning resonance in genetic systems as binary computers was formulated in [Petoukhov, 2016a; Petoukhov, Petukhova, 2017b]. The concept of the multi-resonance genetics was put forward taking into account deep analogies between molecular-genetic structures and the mathematical theory of resonances of oscillatory systems with many degrees of freedom [Petoukhov, 2016a].

In this paper regarding genetic and literary structural interconnections it is important to recall that our speech communication and singing are also based on the inherited ability to use resonances when generating and perceiving sounds. Complex structured systems of resonances of vibrational biosystems with many degrees of freedom can be assumed as a deep reason of described structural analogies. Resonances participate in our perception of music, which has some inherited aspects. The hypothesis concerning the close parallels between genes and music was formulated in [Josephson, Carpenter, 1996]. In our works [Petoukhov, 2008, 2015c; Koblyakov, Petoukhov, Stepanyan, 2015; Petoukhov, He, 2009; Hu, Petoukhov, Petukhova, 2017a] we described a connection between the molecular structure of DNA, including its sequences of hydrogen bonds, with ratios of musical harmony in the Pythagorean musical scale and in a new musical system of so-termed "Fibonacci-stages musical scales", resembling the known phyllotaxis laws of inherited biological morphogenesis. Figuratively speaking, taking into account this connection, living bodies can be considered as complex musical instruments that are developed in phylogenesis and onthogenesis.

Another example of the theoretical approaches in biology relating to the principle of resonances is given by works devoted to cell language theory [Ji, 2015, 2017]. This author postulated an analogy between enzymic catalysis and blackbody radiation, which was modeled by Planck due to his thoughts regarding the resonances of many oscillators. He has noted that some important biological phenomena are described by histograms, which are analogical to the histograms of blackbody radiations. Ji has proposed a generalization of the Planck equation for modeling many biological phenomena that embody long-tailed histograms. By analogy with the principle of quantization of energy in quantum mechanics, Ji has postulated a quantization of free energy levels in enzymes.

E.Schrodinger (1944) noted: "*from all we have learnt about the structure of living matter, we must be prepared to find it working in a manner that cannot be reduced to the ordinary laws of physics… because the construction is different from an anything we have yet tested in the physical laboratory*». For comparison, the enzymes in biological organisms work a million times more effectively than catalysts in the laboratory [Varfolomeev, 2005]. We believe that such ultra-efficiency of enzymes in biological bodies is defined not only by laws of physics, but also by mathematics of quantum informatics and quantum logic, and therefore - in accordance with Schrodinger - this ultra-efficiency cannot be reduced to the ordinary laws of physics.

Acknowledgments: Some results of this paper have been possible due to a long-term cooperation between Russian and Hungarian Academies of Sciences on the topic "Non-linear models and symmetrologic analysis in biomechanics, bioinformatics, and the theory of self-organizing systems", where S.V. Petoukhov was a scientific chief from the Russian Academy of Sciences. The author is grateful to G. Darvas, E. Fimmel, M. He, Z.B. Hu and I.V. Stepanyan for their collaboration. Special thanks to V.I. Svirin who created computer programs for described researches on the basis of author's technical tasks. Special thanks also to the German Academic Exchange Service (DAAD) for providing the very useful internship for the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

author in autumn 2017 at the Institute of Mathematical Biology of the Mannheim University of Applied Sciences (Germany) where the host was Prof. E. Fimmel. I am deeply grateful to J. S. Reid for his edits to the English grammar of this article.

References

Altaisky M.V., Filatov F.P. (2001). Genetic information and quantum gas. arXiv:quant-ph/0106123v1, submitted on 22.06.2001.

Bailey Ch.-J.N. (1982). On the Ying and Yang Nature of Language. Ann Arbor: Karoma.

Biamonte J., Wittek P., Pancotti N., Rebentrost P., Wiebe N., Lloud S. (2017). Quantum machine learning. – *Nature,* v. 549, pages195–202 (14 September 2017), doi:10.1038/nature23474.

Darvas G. (2018). Petoukhov's rules on symmetries in long DNA-texts. *Symmetry: Culture and Science*, v. 29, # 2, pages 318-320, https://doi.org/10.26830/symmetry_2018_2_318, http://journal-scs.symmetry.hu/abstract/?pid=673

Domingos P. (2015).The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. New York: Basic Books.

Fordham M. (1978). *Jungian psychotherapy: A study in analytical psychology*. Chichester, U.K.: Wiley.

Hornak J.P. (1996). *The Basics of MRI*. https://www.cis.rit.edu/htbooks/mri/

Hu Z.B., Petoukhov S.V., **Petukhova E.S**. (2017a). Generalized crystallography, the genetic system and biochemical esthetics. - *Structural Chemistry*, v. 28, №1, pp. 354-368. doi:10.1007/s11224-016-0880-0, http://link.springer.com/journal/11224/28/1/page/2

Hu Z.B., Petoukhov S.V., Petukhova E.S. (2017b). I-Ching, dyadic groups of binary numbers and the geno-logic coding in living bodies. - *Progress in Biophysics and Molecular Biology,* v. 131, December 2017, pp. 354-368. (https://doi.org/10.1016/j.pbiomolbio.2017.08.018).

Hu Z.B., Petoukhov S.V., Petukhova E.S. (2018). On symmetries, resonances and photonic crystals in morphogenesis. *Biosystems*, online 14.09.2018, https://doi.org/10.1016/j.biosystems.2018.09.004

Igamberdiev A.U. (1993). Quantum mechanical properties of biosystems: a framework for complexity, structural stability, and transformations. *Biosystems*, 31 (1), pp. 65–73.

Ivanov Viach.Vs. (1985). Linguistic way of Roman Jakobson. – In: Jakobson R.O. Selected works. – M., Progress (in Russian).

Jacob, F. (1974). Le modele linguistique en biologie. *Critique, Mars*, *30*(322), pp.197-205.

Jacob F., Jakobson R., Levi-Strauss C., Heritier P.L. (1968). Vivre et parler. – «Lettre francaises», p. 1221-1222, Feb. 1968.

Jakobson, R.O. (1985). Selected works. M., Progress (1985, in Russian).

Jakobson, R. (1987). *Language in Literature.* Cambridge: MIT Press (1987).

Jakobson, R.O. (1999). *Texts, Documents, Studies*. Moscow: RGGU (1999, in Russian).

Ji, S. (2015). Planckian distributions in molecular machines, living cells, and brains: The wave-particle duality in biomedical sciences. - *Proc. Intern. Conf. on Biology and Biomedical Engineering*. Vienna, March 15-17, pp. 115-137, http://www.conformon.net/wp-content/uploads/2016/09/PDE_Vienna_2015.pdf

Ji S. (2017). *The Cell Language Theory: Connecting Mind and Matter*. - World Scientific Publishing Europe Ltd. , 450 p. ISBN-13: 978-1848166608.

Jordan P. (1932). Die Quantenmechanik und die Grundprobleme der Biologie und Psychologie. Naturwissenschaften 20, 815–821. (doi:10.1007/BF01494844).

Josephson B.D. (2018). The Physics of Mind and Thought. Preprint, DOI: 10.13140/RG.2.2.36516.32640/2 (https://www.researchgate.net/publication/328968105)

Josephson B.D., Carpenter T. (1996). What can music tell us about the nature of the mind? A Platonic Model. Stuart R. Hameroff, Alfred W. Kaszniak & Alwyn C. Scott (eds.), Toward a Science of Consciousness, MIT Press. 1996. https://philpapers.org/rec/JOSWCM

Jung C.G. (1947). On the nature of the psyche. *Coll. Works* (Vol. VIII). London: Routledge & Kegan Paul.

Koblyakov A.A., Petoukhov S.V., Stepanyan I.V. (2015). The Genetic Code, the Golden Section and Genetic Music. *Presentation at the conference «IS4IS Summit Vienna 2015. The information society at the crossroads»,* 3-7 June 2015, Vienna, Austria, http://sciforum.net/conference/70/paper/2776.

Kolmogorov A.N. (1997). Semiotic messages. – Novoie literaturnoie obozrenie, v. 24, p.216-244 (in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

Russian). http://magazines.russ.ru/nlo/1997/24/kholmog.html;
Makovsky M.M. (1992). Linguistic genetics. M., Nauka (in Russian).
Markov A.A. (1924). *Probability calculus*. Moscow, Gosudarstvennoie izdatelstvo (in Russian)
Matsuno K. (1999). Cell motility as an entangled quantum coherence. – *BioSystems*, 51, pp. 15–19.
Matsuno K. (2003). Quantum mechanics in first, second and third person descriptions. – *BioSystems*, 68, pp. 107-118.
Matsuno K. (2015). On the physics of the emergence of sensorimotor control in the absence of the brain. - *Progress in Biophysics and Molecular Biology*, 119, pp. 313-323.
Matsuno K., Paton R.C. (2000). Is there a biology of quantum information? – *BioSystems*, 55, p.39–46
McFadden J., Al-Khalili J. (2018). The origins of quantum biology. *Proceedings of the Royal Society A*: *Mathematical, Physical and Engineering Sciences*. 12 December 2018. https://royalsocietypublishing.org/doi/full/10.1098/rspa.2018.0674.
Mikheenko P. (2018). Possible superconductivity in brain. https://arxiv.org/abs/1812.05602 , submitted on 13.12.2018.
Nielsen M.A., Chuang I.L. (2010). *Quantum Computation and Quantum Information*. New York: Cambridge University Press.
Nikolaev A. (1999). The musical probability of the verse. – Report at the conference "Pushkin as a symbol of Russian culture" (MIPT, November 19, 1999), http://www.biometrica.tomsk.ru/kolmogorov/kolmogor11.htm
Patel A. (2001a). Quantum algorithms and the genetic code. Pramana – Journal of Physics, v. 56, 2-3, pp. 367-381, arXiv:quant-ph/0002037.
Patel A. (2001b). Testing quantum dynamics in genetic information processing. – Journal of Genetics, v. 80, 1, pp. 39-43.
Patel A. (2001c). Why genetic information processing could have a quantum basis. – Journal of Biosciences, v. 26, 2, pp. 145-151.
Pauli W., Jung C. (2001). C. A. Meier, ed. *Atom and Archetype, The Pauli/Jung Letters*, 1932–1958. Princeton, New Jersey: Princeton University Press. ISBN 978-0-691012-07-0.
Pauling L. (1940). The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An I introduction to Modern Structural Chemistry, 2nd ed. Oxford University Press, London, 664 p.
Penrose R. (1996). *Shadows of the Mind: A Search for the Missing Science of Consciousness*. - Oxford University Press, USA, 480 p.
Petoukhov S.V. (1999). Genetic code and the Ancient Chinese «Book of Changes». – Symmetry: Culture and Science, v. 10, № 3-4, p. 211-226.
Petoukhov S.V. (2001). *The Biperiodic Table of the Genetic Code and Numbers of Protons*. - Moscow, MKC, 258 p. ISBN 5-7012-0203-8. http://petoukhov.com/biperiodic_table_of_the_genetic_code-Petoukhov.pdf .
Petoukhov S.V. (2008) *Matrix genetics, algebras of the genetic code, noise immunity*. M., RCD, 316 p. (in Russian).
Petoukhov S.V. (2011). Matrix genetics and algebraic properties of the multi-level system of genetic alphabets. - *Neuroquantology,* 2011, Vol 9, No 4, p. 60-81, http://www.neuroquantology.com/index.php/journal/article/view/501
Petoukhov S.V. (2015a). Resonances and genetic biomechanics. - *Symmetry: Culture and Science*, т. 26, №3, p. 379-397, http://petoukhov.com/PETOUKHOV_IN_SCS_2015.pdf
Petoukhov S.V. (2015b). The concept of resonances in genetics ("resonant bioinformatics"). - Selected Works of the VII International Congress "Weak and Superweak Fields and Radiation in Biology and Medicine", September 07-11, 2015, St. Petersburg, pp. 1-41, http://www.biophys.ru/lib/sci/resonance/460-resonance-00001 (in Russian).
Petoukhov S. V. (2015c). Music and the Modeling Approach to Genetic Systems of Biological Resonances. *Presentation at the conference «**IS4IS Summit Vienna 2015.** The information society at the crossroads*», 3-7 June 2015, Vienna, Austria, http://sciforum.net/conference/70/paper/2812
Petoukhov S.V. (2016a). The system-resonance approach in modeling genetic structures. – *Biosystems*, January 2016, v. 139, p. 1-11, http://petoukhov.com/PETOUKHOV_ARTICLE_IN_BIOSYSTEMS.pdf .
Petoukhov S.V. (2016b). Symmetries of the genetic code, Walsh functions and the theory of genetic logical holography. - *Symmetry: Culture and Science*, v.27, №2, p. 95-98.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 February 2019 doi:10.20944/preprints201812.0142.v2

Petoukhov S.V. (2017). Genetic coding and united-hypercomplex systems in the models of algebraic biology. - *Biosystems*, v. 158, August 2017, pp. 31–46.

Petoukhov S.V. (2018a). The Genetic Coding System and Unitary Matrices. *Preprints* 2018, 2018040131 (doi: 10.20944/preprints201804.0131.v2).

Petoukhov S.V. (2018b). The rules of long DNA-sequences and tetra-groups of oligonucleotides. - arXiv:1709.04943v5, 5th version (8 October 2018), 159 pages.

Petoukhov S.V., He M. (2009). *Symmetrical Analysis Techniques for Genetic Systems and Bioinformatics: Advanced Patterns and Applications*. Hershey, USA: IGI Global. 271 p.

Petoukhov S.V., Petukhova E.S. (2017a). Symmetries in genetic systems and the concept of geno-logical coding. - *Information*, 2017, *8*(1), 2; doi:10.3390/info8010002, http://www.mdpi.com/2078-2489/8/1/2/htm

Petoukhov S.V., Petukhova E.S. (2017b). Resonances and the quest for transdisciplinarity. – *Information Studies and the Quest for Transdisciplinarity*. Editors M. Burgin, W.Hofkirchner, World Scientific, p. p. 467-487.

Petoukhov S.V., Petukhova E.S., Hazina L.V., Svirin V.I., Stepanyan I.V., Silova T.V. (2017). Geno-logical coding, united-hypercomplex numbers and systems of artificial intelligence. – In the book "*Advances in Artificial Systems for Medicine and Education*", Editors Z.B. Hu, S.V. Petoukhov and M.He. Part of the "Advances in Intelligent Systems and Computing" book series (AISAC, volume 658), p. 2-13, Springer, https://link.springer.com/search?query=978-3-319-67348-6.

Petoukhov S.V., Petukhova E.S., Svirin V.I. (2018). New Symmetries and Fractal-Like Structures in the Genetic Coding System. In: Hu Z., Petoukhov S., Dychka I., He M. (eds) Advances in Computer Science for Engineering and Education. ICCSEEA 2018. P. 588-600. Advances in Intelligent Systems and Computing, vol 754. Springer, Cham, DOI: https://doi.org/10.1007/978-3-319-91008-6_59

Petoukhov S.V., Svirin V.I. (2018). The New Wide Class of Symmetries in Long DNA-Texts. Elements of Quantum-Algorithmic Genetics. - Biologia Serbica, v. 40, №1, p. 51, Special Edition, ISSN 2334 6590, UDK 57(051). Book of Abstracts, Belgrade Bioinformatics Conference 2018, 18-22 June, 2018, Belgrade, Serbia, page 51.

Petrenko Yu.I. (2018). Separation and Ranking of Vovels in Russian texts for cryptanalysis tasks. – Open education, v. 22, № 1, pp. 59-69 (in Russian). DOI: http://dx.doi.org/10.21686/1818-4243- 2018-1-59-69

Pollack G.H. (2013). *The Fourth Phase of Water: Beyond Solid, Liquid, and Vapor*. USA: Ebner & Sons, ISBN-10: 096268954, ISBN-13: 978-0962689543

Quantum Aspects of Life (2008). / editors Abbott D., Davies P.C.W., Pati A.K.; foreword by Sir Roger Penrose. ISBN-13: 978-1-84816-253-2.

Rapoport A.E., Trifonov E.N. (2012): Compensatory nature of Chargaff's second parity rule. *Journal of Biomolecular Structure and Dynamics*, November 1-13, DOI:10.1080/07391102.2012.736757

Ratner V.A. (2002). *Genetics, Molecular Cybernetics.* – Novosibirsk: Nauka, 272 p. (in Russian).

Reid J.S. (2017). Secrets of Cymatics. https://youtu.be/uMK3OVBjx2Q.

Reid J.S. (2018). Secrets of Cymatics II. https://www.youtube.com/watch?v=K8w_fr0vfGo .

Samuels A. (2005). Amplification (Analytical Psychology). International Dictionary of Psychoanalysis. *Encyclopedia.com.* 22. https://www.encyclopedia.com/psychology/dictionaries-thesauruses-pictures-and-press-releases/amplification-analytical-psychology

Schrödinger E. (1944). *What is life?* - Cambridge Press.

Schrödinger E. (1952). Are there quantum jumps? Part I. - *The British Journal for the Philosophy of Science*, Vol. 3, No. 10, pp. 109-123. Published by: Oxford University Press on behalf of The British Society for the Philosophy of Science Stable, URL: http://www.jstor.org/stable/685552 .

Shults E. (1916) The organism as a creativity. - In the book "*Questions of theory and psychology of creativity*", Russia, Kharkov, vol.7, p. 108-190 (in Russian, "Voprosy teorii i psikhologii tvorchestva").

Stent G.S. (1969). *The coming of the golden age*. New York: The Natural History Press.

Stewart I. (1999). *Life's Other Secret: The New Mathematics of the Living World*. Penguin, NY.

Varfolomeev S.D. (2005). *Chemical Enzymology* (textbook of Moscow State University named after M.V. Lomonosov). - Publishing Center "Academy", 480 p. ISBN 5-7695-1953-3 (in Russian).