Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

Structural Connections between Long Genetic and Literary Texts

Sergey V. Petoukhov[1]

[1]-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 materials of this article were presented by the author in his keynote speech at the International Belgrade Bioinformatics Conference 2018 (Belgrade, Serbia, 18-22 June, 2018, http://belbi.bg.ac.rs/), in his keynote speech at the Second International Conference "Artificial Intelligence, Medical Engineering, Education" (Moscow, Russia, 6-8 October 2018, http://www.ruscnconf.org/aimee2018/index.html) and also in his presentation at 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/).

Abstract. Impressing discoveries in the field of the genetic code have been described by its researchers by means of the terminology borrowed from linguistics and the theory of communications. Leading experts on structural linguistics believe for a long time already that languages of human dialogue were formed not from an empty place, but they are continuation of genetic language or, anyhow, are closely connected with it, confirming the idea of information commonality of organisms. The aticle continues the theme about a connection of linquistic languages with the genetic language. It describes results of comparative study of long Russian literary texts (novels by L.Tolstoy, F.Dostoevsky, A.Pushkin, etc.) and long sequences of hydrogen bonds in double helixes of DNA of different organisms. Formalisms of quantum informatics are used in modeling some of these results taking into account thoughts of many researches about possible using principles of quantum informatics in organisation of living bodies.

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

1. Introduction

Impressing discoveries in the field of the genetic code have been described by its researchers by means of the terminology borrowed from linguistics and the theory of communications. As experts in molecular genetics mark, "*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].

Leading experts on structural linguistics believe for a long time already that languages of human dialogue were formed not from an empty place, but they are continuation of genetic language or, anyhow, are closely connected with it, confirming the idea of information commonality of organisms. Analogies between systems of genetic and linguistic information are contents of a wide and important scientific sphere, which can be illustrated here in short only. One can reproduce here some thematic thoughts by R. Jakobson [1987, 1999], who is one of the most famous experts and who is the author of a deep theory of binary oppositions in linguistics. He and other persons holding the same views suppose that we possess a language, which is old as a life and which is the most live among all languages. Among all systems of an information transfer, the genetic code and linguistic codes only are based on use of the discrete components, which in itself have no sense, but serve for construction of the minimum units, which are making sense. In the both cases of the genetic language and of a linguistic language, we deal with separate units which, taken in itself, have no sense, but they get a sense after their special grouping. A similarity between the both information systems does not be exhausted by this fact at all. 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 the four letters of the genetic alphabet contains the three binary sub-alphabets (Fig. 1), which allows to number columns and rows of the genetic matrices $[C, A; T, G]^{(n)}$ (Fig. 2). As Jakobson

© 2018 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

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 have arisen at Jakobson as consequence of its long-term researches of connections of linguistics, biology and physics. Such connections were considered at a united seminar of physicists and linguists, which was organized by Niels Bohr and Roman Jakobson jointly in 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]. One can remind 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. Linguistics gives one of significant examples of existence and importance of ensembles of binary oppositions in information physiology.

The Nobel Prize winner in molecular genetics F. Jacob was also thinking about deep relations between genetics and linguistic languages in a 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*» (it means in English: *perhaps, for revealing of relations between genetics and language it would be necessary to study them through the Ancient Chinese "I Ching"*) [Jacob, 1974, p.205].

Similar questions about a connection of linguistics with the genetic code excite many researchers. In addition many researchers perceive a linguistic language as a living organism. The book "Linguistic genetics" [Makovsky, 1992] says: "*The opinion about language as about a living organism, which is submitted to the laws of a 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*".

The importance of studying the phenomenon of linguistic languages for a deep understanding and modeling of mental functions and for the approproiate expansion of physics is noted in the work of the Nobel laureate in physics B. Josephson [Josephson, 2018]: "*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*".

This article describes author's results about structural analogies between long DNA-texts and long literary works in Russian. The represented analysis of long literary Russian texts (written by L.N. Tolstoy, F.M. Dostoevsky, A.S. Pushkin, etc.) is based on binary peculiarities of the Russian alphabet. In addition, some applications of mathematics of quantum informatics for modeling bioinformational structures are considered.

2. Some properties of long literary texts in Russian

As known, 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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

The Russian alphabet can be considered as consisting 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. Any of 14 members of this class we denote 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. Any of 17 members of this part we denote by the general symbol 1.

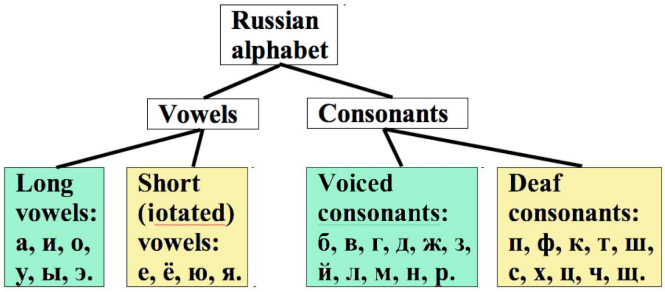


Fig. 1. The binary-oppositional structure of the Russian alphabet.

Any Russian text can be converted into the corresponding binary sequence of these two symbols 0 and 1 (like as 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 its belonging to one of these two classes of equivalency.

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 10101100111011011110001101010001. The computer program for our analysis of literary texts was created by graduate student V.I.Svirin in line with a technical task formulated by the author.

For further description one should remind about the mathematical operation of the tensor product. This operation is well known in mathematics, physics and informatics, where it gives a way of putting vector spaces together 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 speaks about the tensor product: «*This construction is crucial to understanding the quantum mechanics of multiparticle systems*» [Nielsen, Chuang, 2010, p. 71] since in line with the postulate of quantum mechanics: 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 componets 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]:

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

We analize the named binary representations of Russian long literary texts by means of the method [Petoukhov, 2018a], which was used for analyzing long DNA sequences of numbers of hydrogen bonds 3 and 2. More precisely, 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 also 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, …). We call such different representations of a long literary text in the form of sequences of binary n-plets as its “ binary n-plet representations”.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

A complete alphabet of binary n-plets under fixed value n is called “the alphabet of binary n-plets” and it 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 representation of any 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 any long literary text, etc. One can mention that the tensor family of vectors $[0, 1]^{(n)}$ (where (n) means the tensor power; $n = 1, 2, 3, 4, \ldots$) 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. By this reason, one can note conditionally that all binary n-plet alphabets belong to the same tensor family of vectors.

In binary n-plet representations of long literary texts in Russian, we analyze probabilities (or frequencies or percentage) of each of members of concrete alphabets of binary n-plets in concrete long texts in 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 exist the total number 654523 of doublets 00, 01, 10 and 11, including 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 calculation of probabilities of all members of alphabets of n-plets ($n = 1, 2, 3, 4$) in long literary Russian texts, the author models all these probabilities with a good level of accuracy by means of the same quantum-algorithmic method, which was used for modeling probabilities in long DNA sequences of hydrogen bonds 3 and 2 in his article [Petoukhov, 2018a]. This method explores some classical formalism from quantum informatics [Nielsen, Chuang, 2010].

In our 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 not too large) 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 get model values of probabilities of all 2^n members of the binary n-plet alphabet in a considered binary n-plet representation of the analyzed literary text, one should calculate the following 2^n coordinates of vectors of probabilities from their tensor family:

- 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.

One should emphasise here that all coordinates of these vectors $[0_P, 1_P]^{(n)}$, where $n = 2, 3, 4, \ldots$ are expressed by means of only two values 0_P and 1_P. Briefly speaking, in line with the proposed model approach, for an approximate prediction of values 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 - it is enough to do the following:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

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

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 our analysis of the novel «Anna Karenina» by Leo Tolstoy by means of the described approach. One can see nice correspondences between phenomenologic values (blue points in graphs in Fig. 2) of probabilities of all members of alphabets of binary *n*-plets and model values (red points in graphs) represented by components of probability vectors $[0_P, 1_P]^{(n)}$, $n = 2, 3, 4$. One can see in these graphs that the model points of red color turned out to be almost exactly superimposed on the phenomenologic points of blue color. Fig. 3 shows the proximity of the numerical phenomenological and model values of the studyed probabilities. All values are rounded to four decimal places. So, knowing only two probablities 0_P and 1_P of binary monoplets 0 and 1 in the binary *n*-plet representation of this novel, one can predict - with good level of accuracy - dozens of probabilties of all members of the alphabets of binary n-plets for this Russian novel with a qood level of accuracy. We presume that a similar model correspondence holds true also for $n = 5, 6, \ldots$ (if n is not too large) but this should be studied in future researches.

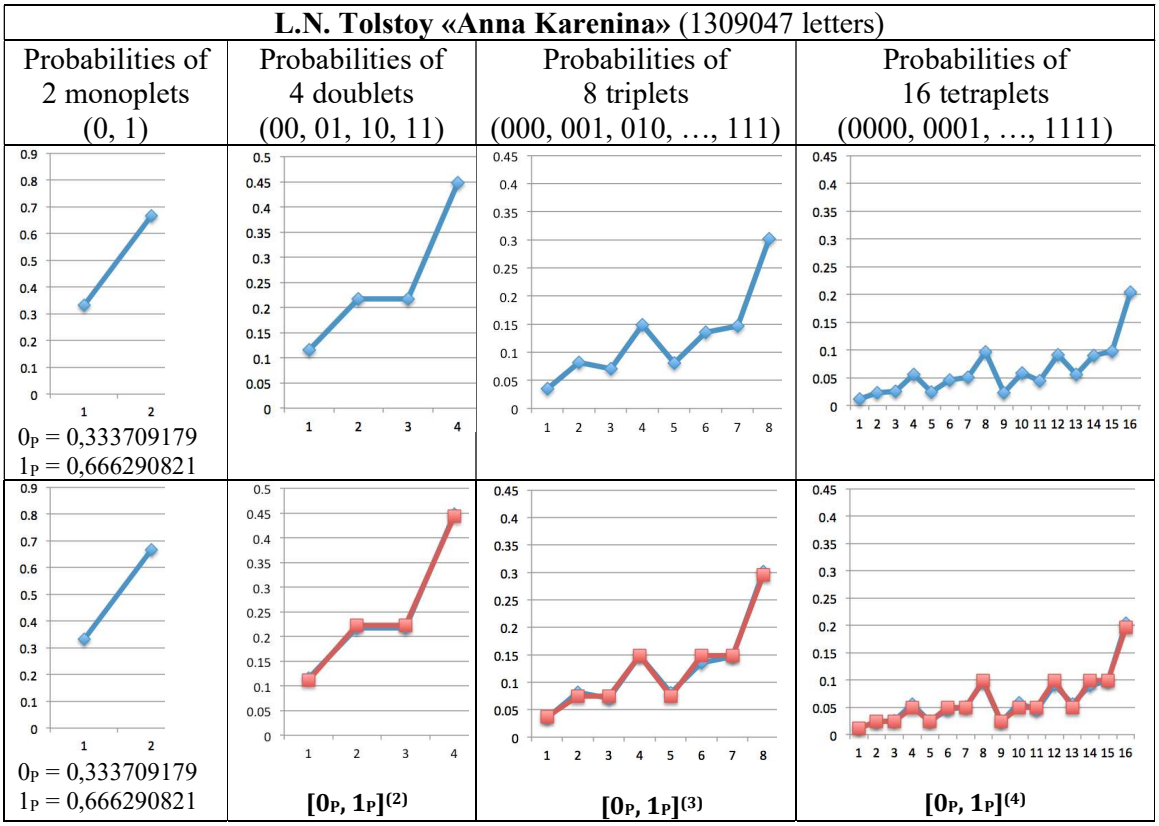


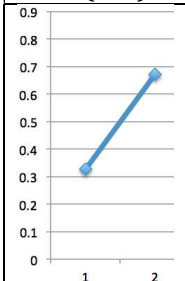
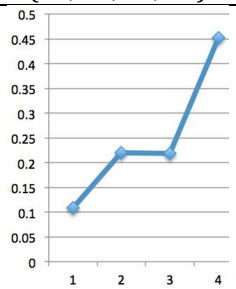
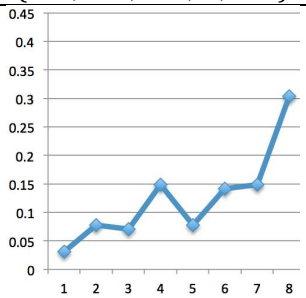
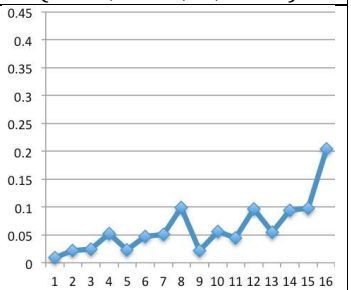
Fig. 2. Graphical representations of results of the analysis of the novel "Anna Karenina" by Leo Tolstoy (the original literary text is taken from http://samolit.com/books/62/). Probabilities of members of alphabets of binary *n*-plets ($n = 1, 2, 3, 4$) from the binary *n*-plet representation of this novel are shown. Blue points correspond to phenomenologic values of the probabilities of hydrogen *n*-plets, and 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) means tensor powers.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

L.N. Tolstoy «Anna Karenina» (1309047 letters)	
Probabilities of members of the alphabet of binary monoplets	
Reality:	$0_P = 0{,}3337$; $1_P = 0{,}6663$
Probabilities of members of 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 of 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 of 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 results of the analysis of the novel "Anna Karenina" by Leo Tolstoy (the original literary text is taken from http://samolit.com/books/62/). 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 for cases of alphabets named in tabular sections, 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) means tensor powers.

Below we represent other examples of our results received by similar analysis of famous literary works in Russian: L.N. Tolstoy «War and Peace» (Figs. 4, 5); F.M. Dostoevsky «Crime and Punishment» (Figs. 6, 7) and «Idiot» (Figs. 8, 9); A.S. Pushkin «Evgenij Onegin» (Figs.10, 11) and «Dubrovsky» (Figs. 12, 13); the Biblie in Russian (Figs. 14, 15). All these results are similar to the described results for the case of the novel «Anna Karenina» (Figs, 2, 3): they confirm that studyed probabilities of members of different alphabets of binary n-plets (n = 1, 2, 3, 4) are interrelated each other in a certain degree 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.

L.N. Tolstoy «War and Peace», Book I (1068479 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)
 $0_P = 0{,}328301258$ $1_P = 0{,}671698742$			

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

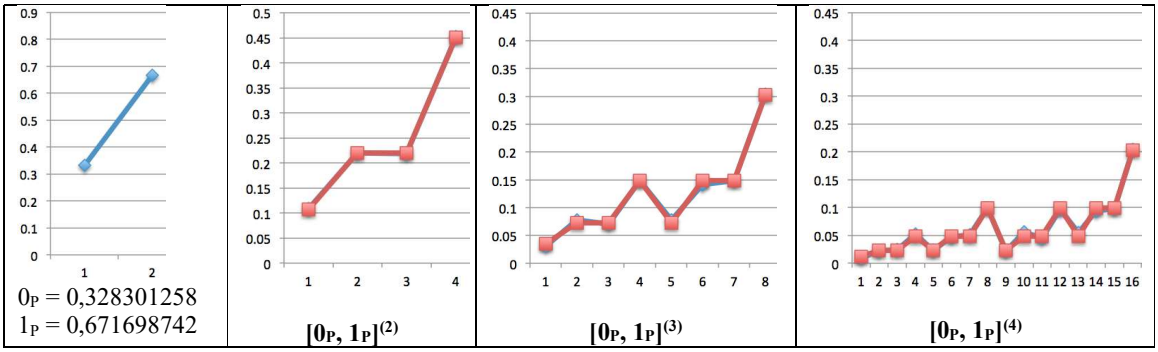


Fig. 4. Graphical representations of results of the analysis of the novel "War and Peace" (Book 1) by Leo Tolstoy (the original literary text is taken from http://samolit.com/books/64/). Probabilities of members of alphabets of binary *n*-plets (*n* = 1, 2, 3, 4) from the binary *n*-plet representation of this novel are shown. Blue points correspond to phenomenologic values of the probabilities of *n*-plets, and 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) means tensor powers.

L.N. Tolstoy «War and Peace», Book I (1068479 letters)
Probabilities of members of the alphabet of binary monoplets
Reality: 0_P = 0,3283; 1_P = 0,6717
Probabilities of members of 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 of 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: $0_P0_P0_P$=0,0354; $0_P0_P1_P$=0,0724; $0_P1_P0_P$=0,0724; $0_P1_P1_P$=0,1481; $1_P0_P0_P$=0,0724; $1_P0_P1_P$=0,1481; $1_P1_P0_P$=0,1481; $1_P1_P1_P$=0,3031.
Probabilities of members of 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 representations of results of the analysis of the novel "War and Peace" (Book 1) by Leo Tolstoy (the original literary text is taken from http://samolit.com/books/64/). 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 for cases of alphabets named in tabular sections, 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) means tensor powers.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

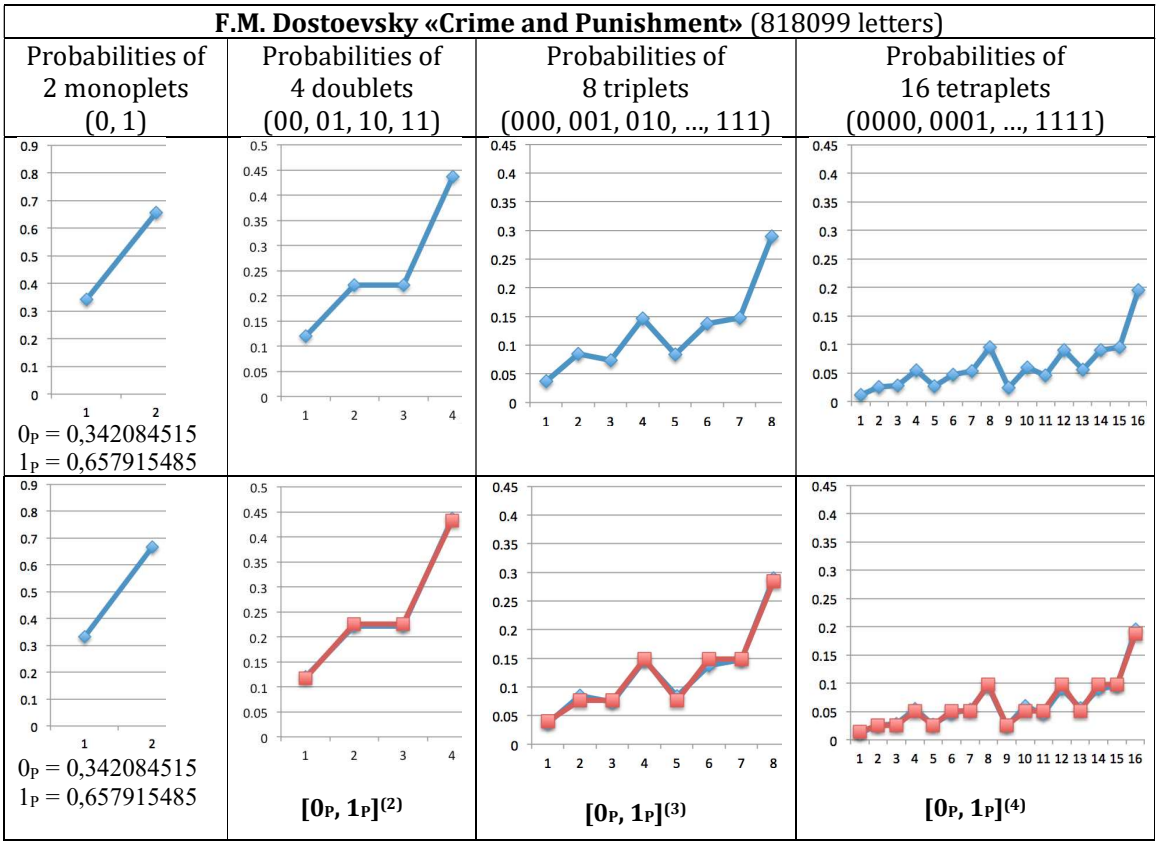


Fig. 6. Graphical representations of results of the analysis of the novel "Crime and Punishment" by F.M. Dostoevsky (the original literary text is taken from http://samolit.com/books/57/). Probabilities of members of alphabets of binary n-plets (n = 1, 2, 3, 4) from the binary n-plet representation of this novel are shown. Blue points correspond to phenomenologic values of the probabilities of n-plets, and 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) means tensor powers.

F.M. Dostoevsky «Crime and Punishment» (818099 letters)
Probabilities of members of the alphabet of binary monoplets
Reality: $0_P = 0{,}3421$; $1_P = 0{,}6579$
Probabilities of members of the alphabet of binary doublets
Reality: $0_P0_P = 0{,}1203$; $0_P1_P = 0{,}2219$; $1_P0_P = 0{,}2216$; $1_P1_P = 0{,}4362$.
Model: $0_P0_P = 0{,}1170$; $0_P1_P = 0{,}2251$; $1_P0_P = 0{,}2251$; $1_P1_P = 0{,}4329$.
Probabilities of members of 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 of 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: $0_P0_P0_P0_P=0{,}0137$; $0_P0_P0_P1_P=0{,}0263$; $0_P0_P1_P0_P=0{,}0263$; $0_P0_P1_P1_P=0{,}0507$; $0_P1_P0_P0_P=0{,}0263$; $0_P1_P0_P1_P=0{,}0507$; $0_P1_P1_P0_P=0{,}0507$; $0_P1_P1_P1_P=0{,}0974$; $1_P0_P0_P0_P=0{,}0263$; $1_P0_P0_P1_P=0{,}0507$; $1_P0_P1_P0_P=0{,}0507$; $1_P0_P1_P1_P=0{,}0974$; $1_P1_P0_P0_P=0{,}0507$; $1_P1_P0_P1_P=0{,}0974$; $1_P1_P1_P0_P=0{,}0974$; $1_P1_P1_P1_P=0{,}1874$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

Fig. 7. Numeric representations of results of the analysis of the novel "Crime and Punishment" by F.M. Dostoevsky (the original literary text is taken from http://samolit.com/books/57/). 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 for cases of alphabets named in tabular sections, 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) means tensor powers.

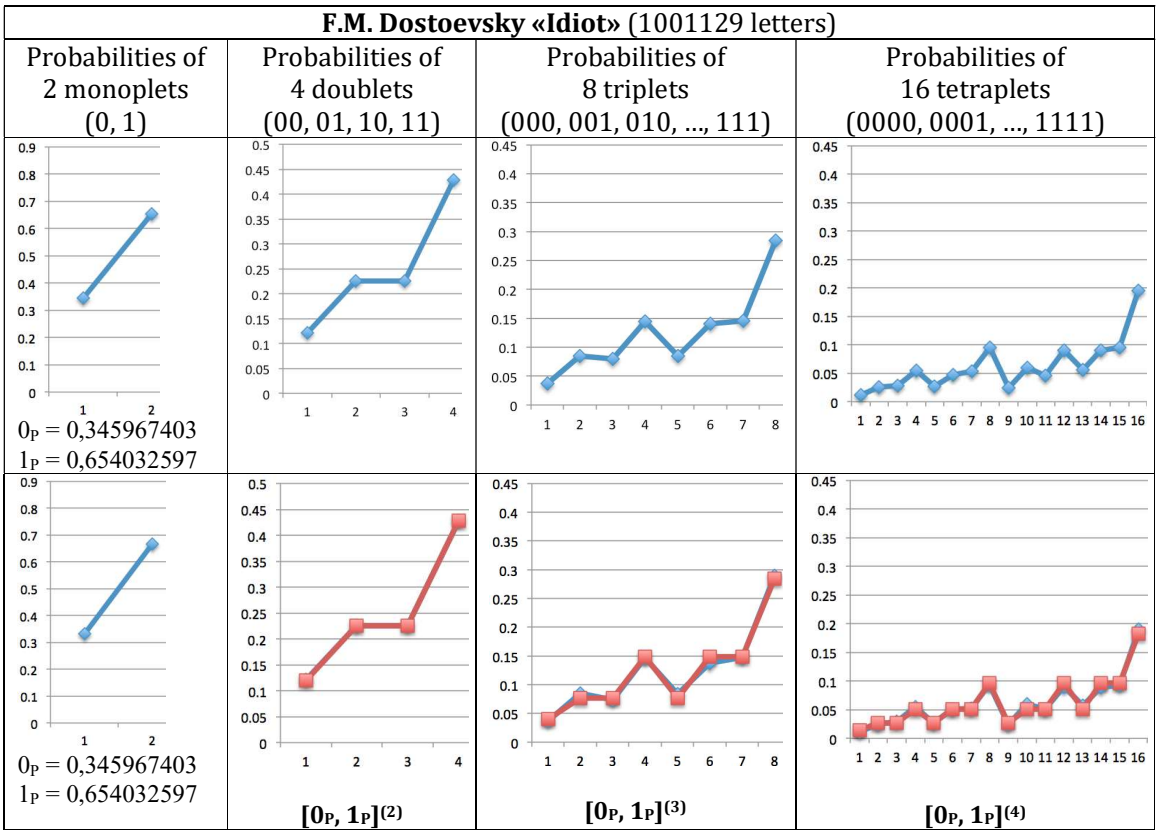


Fig. 8. Graphical representations of results of the analysis of the novel "Idiot" by F.M. Dostoevsky (the original literary text is taken from http://samolit.com/books/56/). Probabilities of members of alphabets of binary *n*-plets (*n* = 1, 2, 3, 4) from the binary *n*-plet representation of this novel are shown. Blue points correspond to phenomenologic values of the probabilities of *n*-plets, and 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) means tensor powers.

F.M. Dostoevsky «Idiot» (1001129 letters)	
Probabilities of members of the alphabet of binary monoplets	
Reality:	$0_P = 0{,}3460$; $1_P = 0{,}6540$
Probabilities of members of 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 of 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:	$0_P0_P0_P=0{,}0414$; $0_P0_P1_P=0{,}0783$; $0_P1_P0_P=0{,}0783$; $0_P1_P1_P=0{,}1480$; $1_P0_P0_P=0{,}0783$; $1_P0_P1_P=0{,}1480$; $1_P1_P0_P=0{,}1480$; $1_P1_P1_P=0{,}2798$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

Probabilities of members of 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 representations of results of the analysis of the novel "Idiot" by F.M. Dostoevsky (the original literary text is taken 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) means tensor powers.

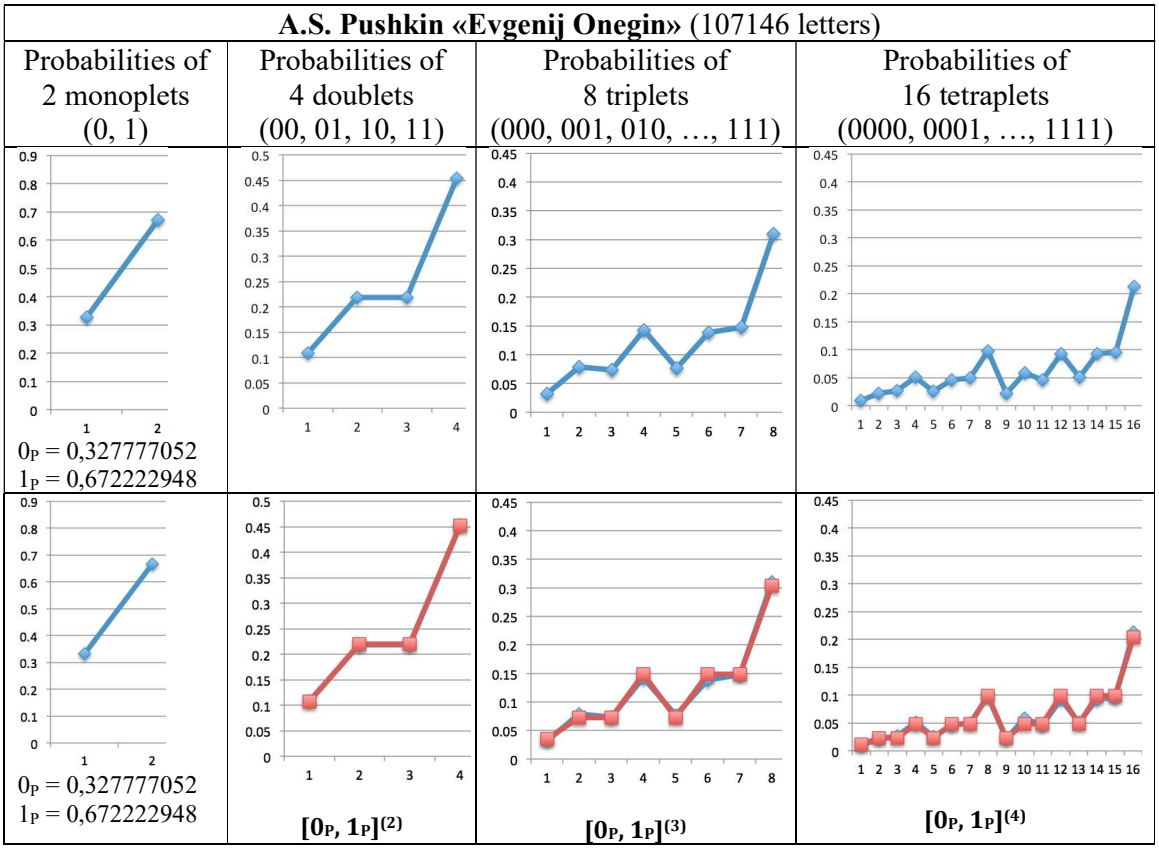
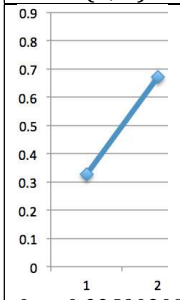
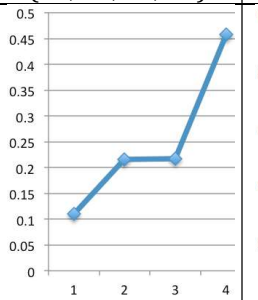
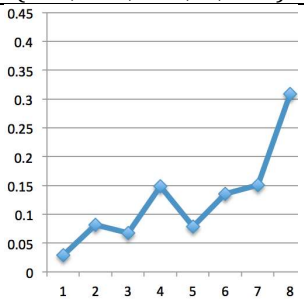
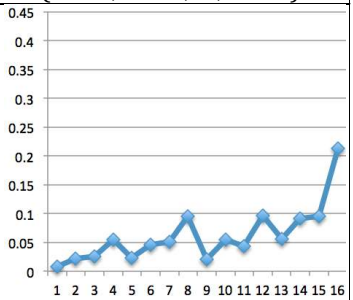
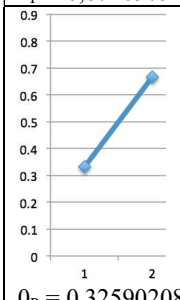
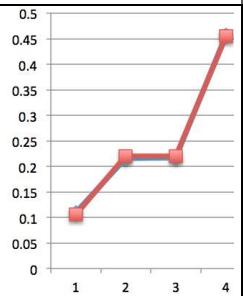
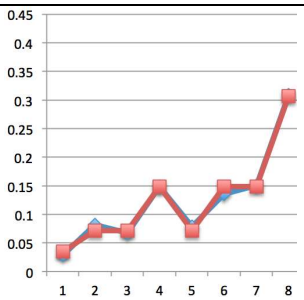
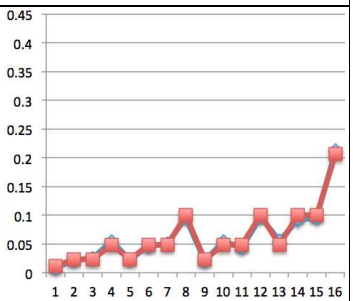


Fig. 10. Graphical representations of results of the analysis of the novel "Evgenij Onegin" by A.S. Pushkin (the original literary text is taken from http://tululu.org/b57798/). Probabilities of members of alphabets of binary n-plets (n = 1, 2, 3, 4) from the binary n-plet representation of this novel are shown. Blue points correspond to phenomenologic values of the probabilities of n-plets, and 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) means tensor powers.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

A.S. Pushkin «Evgenij Onegin» (107146 letters)
Probabilities of members of the alphabet of binary monoplets
Reality: $0_P = 0{,}3278$; $1_P = 0{,}6722$
Probabilities of members of 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 of 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 of 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 representations of results of the analysis of the novel "Evgenij Onegin" by A.S. Pushkin (the original literary text is taken 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. 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) means tensor powers.

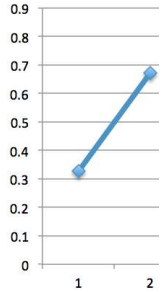
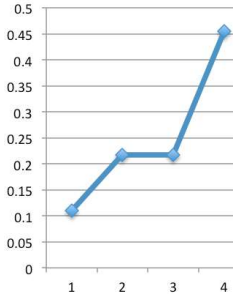
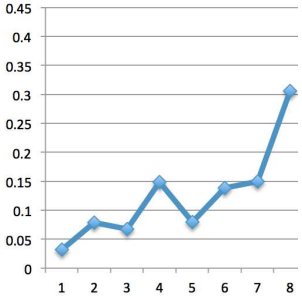
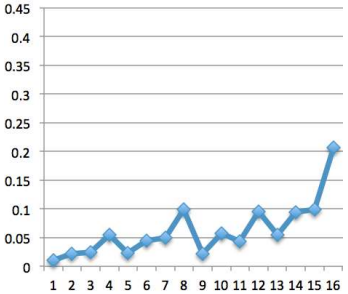
A.S. Pushkin «Dubrovsky» (106891 letters)			
Probabilities of 2 monoplets (0, 1)	Probabilities of 4 doublets (00, 01, 10, 11)	Probabilities of 8 triplets (000, 001, 010, ..., 111)	Probabilities of 16 tetraplets (0000, 0001, ..., 1111)
 $0_P = 0{,}325902087$ $1_P = 0{,}674097913$			
 $0_P = 0{,}325902087$ $1_P = 0{,}674097913$	 $[0_P, 1_P]^{(2)}$	 $[0_P, 1_P]^{(3)}$	 $[0_P, 1_P]^{(4)}$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

Fig. 12. Graphical representations of results of the analysis of the novel "Dubrovsky" by A.S. Pushkin (the original literary text is taken from http://samolit.com/books/61/). 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, and 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) means tensor powers.

A.S. Pushkin «Dubrovsky» (106891 letters)
Probabilities of members of the alphabet of binary monoplets
Reality: $0_P = 0{,}3259$; $1_P = 0{,}6741$
Probabilities of members of 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 of 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 of 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 representations of results of the analysis of the novel "Dubrovsky" by A.S. Pushkin (the original literary text is taken from http://samolit.com/books/61/). 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) means tensor powers.

Bible (in Russian, 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)
			
$0_P = 0{,}327977136$ $1_P = 0{,}672022864$			

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

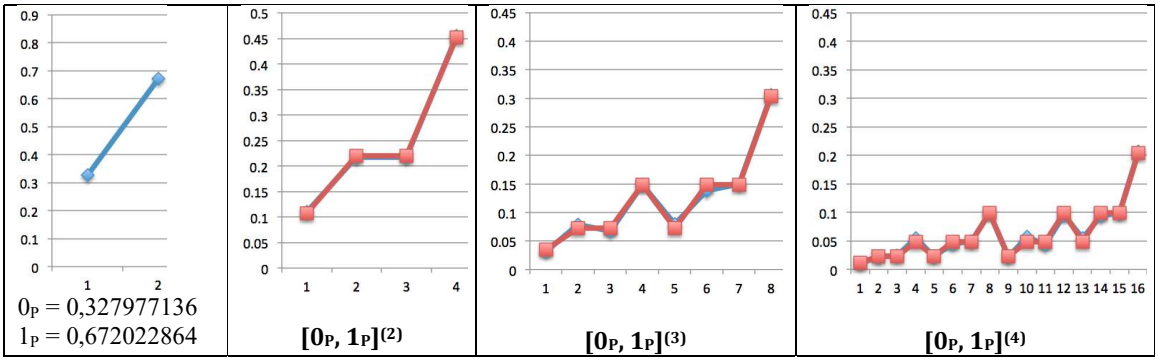


Fig. 14. Graphical representations of results of the analysis of the Bible (the original literary text in Russian is taken from http://petoukhov.com/bible.zip). Probabilities of members of alphabets of binary *n*-plets (*n* = 1, 2, 3, 4) from the binary *n*-plet representation of this novel are shown. Blue points correspond to phenomenologic values of the probabilities of *n*-plets, and 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*) means tensor powers.

Bible (in Russian, 3122489 letters)
Probabilities of members of the alphabet of binary monoplets
Reality: $0_P = 0{,}3280$; $1_P = 0{,}6720$
Probabilities of members of 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 of 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 of 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 representations of results of the analysis of the Bible (the original literary text in Russian is taken 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. 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*) means tensor powers.

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

Are there similar patterns not only in long Russian literary texts, but also in texts in many other languages that have other alphabets and other phonetic features? We conduct these studies for literary texts in English, German, French and many other languages. We hope to publish the results of this study later.

The next Section shows that analogical properties exist in many long DNA-texts, which was tested 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 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 has mathematical analogies with the main law of population genetics - the Hardy–Weinberg law [Petoukhov, 2018a].

3. About similar properties of long DNA-texts

In DNA molecules, genetic information is written in very long DNA-texts in a form of sequences of only 4 letters: adenine A, cytosine C, guanine G, thymine T. For example, the human genome is a text with several billions of genetic letters A, T, C and G. In 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 call such chains of numbers of hydrogen bonds as "hydrogen bonds sequences" (briefly, "H-sequences" or "H-texts"). We analysed properties of such hydrogen bonds sequences for many long DNA of different organisms (here the term "long" means that DNA contains ≥ 100000 letters). Below results of our study are briefly presented, which confirm the existence of structural connections between long genetic and literary texts.

Each of 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 be represented also 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 call such different representations of a long DNA sequence of nucleotides A, T, C, G in forms of sequences of H-n-plets 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 it 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. One can mention that the tensor family of vectors $[3, 2]^{(n)}$ (where (n) means the tensor power, $n = 1, 2, 3, 4, \ldots$) 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. By this reason, one can note conditionally that all H-*n*-plets alphabets belong to the same tensor family of vectors and are interconnected each other.

The article (Petoukhov, 2018a) describes in details those hidden regularities in many tested H-sequences of long DNA of different organisms, which are connected with members of the alphabets of H-*n*-plets. These regularities concern probabilities (or frequencies or percentage) of members of the alphabets 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, probabilities of members of different alphabets of H-*n*-plets (under different values *n*) are interrelated each other. This interrelation can be modeled by the same method, which was 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.

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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

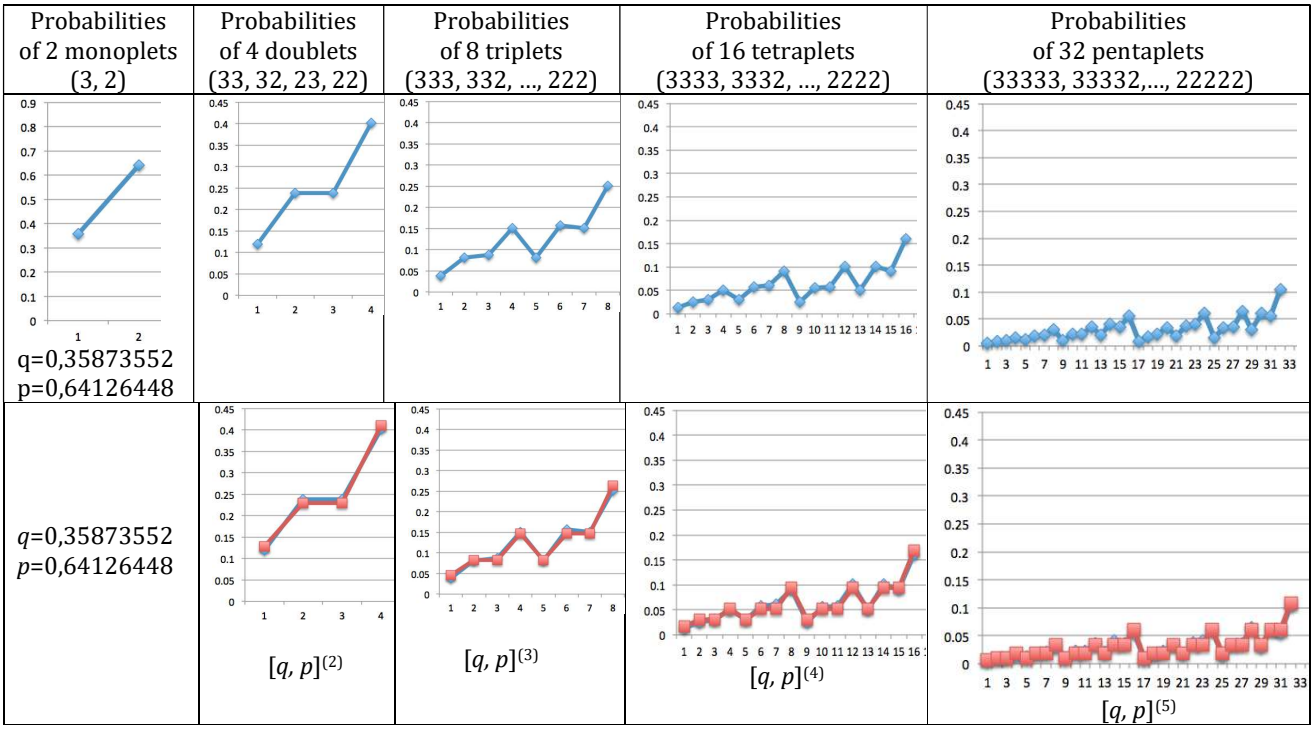


Fig. 16. The graphic representation of probabilities of members of hydrogen *n*-alphabets (n = 1, 2, 3, 4, 5) in the DNA sequence of the first chromosome of a plant *Arabidopsis thaliana* (initial data about this chromosome were taken from https://www.ncbi.nlm.nih.gov/nuccore/NC_003070.9). Blue points in graphs show phenomenological probabilities of n-plets of numbers of hydrogen bonds. 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*-alphabets (n = 1, 2, 3, 4, 5) for the same DNA sequence as in Fig. 16. One can see that these model values reproduce phenomenological values with good level of accuracy.

Probabilities of members of the alphabet of H-monoplets (numbers 3 and 2)
Reality: q = [3] = 0,3587; p = [2] = 0,6413
Probabilities of members of 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 of 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 of 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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

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 about this chromosome were taken from https://www.ncbi.nlm.nih.gov/nuccore/NC_003070.9). Numbers in square brackets denote probabilities of corresponding members of alphabets of hydrogen *n*-plets (for example, the symbol [323] denotes the probability of the number 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 (with some other levels of accuracy) have been obtained in our analysis of the complete sets of nuclear chromosomes on the basis of initial data about their DNA sequences from the GenBank: a plant *Arabidopsis thaliana*, a nematode *Caenorhabditis elegans*, fruit fly *Drosophila melanogaster*, house mouse *Mus musculus*, Homo sapiens.

These results allow assuming existence of the following general rule for eucariots about frequencies of hydrogen bonds in complete sets of nuclear chromosomes:

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

Of course, further researches are needed to define a degree of universality and precision of this rule.

We have calculated also 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 $[q, p]^{(n)}$ with good results (with different levels of accuracy in different genomes). These results confirm that the proposed model approach on the basis of vectors $[q, p]^{(n)}$ can be used to get idealized models of probabilities of all members of H-*n*-alphabets in real long DNA sequences (*n* = 1, 2, 3, 4, … is not too large).

The proposed model allows predicting probabilities of dozens of members of H-*n*-alphabets in a long DNA sequence with a high level of accuracy on the base of knowledge of probabilities of only two numbers - 3 and 2 – of hydrogen bonds in this DNA.

One should recall that if a quantum state can be represented as a vector of a Hilbert space, such state is called a pure quantum state. If a pure state $|\psi>$ can be written in the form $|\psi> = |\psi_1> \otimes |\psi_2>$, where $|\psi_i>$ is a pure state of the i-th subsystem, it is said to be separable [Nielsen, Chuang, 2010]. Otherwise it is called entangled. Correspondingly our 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 a plant *Arabidopsis thaliana* shown in Fig. 16; from the model point of view, this means that in the case of human chromosomes, entangled states are more pronounced.

Some concluding remarks

One can remind that the probabilities of alternation of vowels and consonants letters were specially been studying 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 famous "Markov's chains", developed his theory of these 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 letters in the poem "Evgeniy Onegin" was also made in the work devoted to the mentioned Markov's research [Petrenko, 2018]. Great Russian mathematician A.N. Kolmogorov studied

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

rythms of verses for revealing the mechanism of developing internal goals of self-organizing systems [Kolmogorov, 1997; Nikolaev, 1999]. In contrast to all known researches about vowels and consonants letters and their concatenations in doublets, triplets, etc. in Russian literary texts, we study the following:

- Firstly, we study not the probabilities of alternation of separate vowels and consonants letters but 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 of different n-plets of these representatives (n = 1, 2, 3, 4, …);
- Secondly, from this standpoint we study analogies between long literary texts and hydrogen sequences of DNA to check the known idea that linguistic languages are continuations of genetic languages, which are basic for all biological organisms;
- Thirdly, we study possibilities of applications of formalisms of quantum informatics to model long literary texts and DNA texts for developing the known hypothesis of some authors about 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 set of our results published in various editions testifies in favor that mathematics of quantum informatics and of theory of resonances of oscillatory systems with 2^n degrees of freedom could be such partner of the genetic system [Petoukhov, 2015a,b,c; Petoukhov, 2016a; Petoukhov, Petukhova, 2017b; Hu, Petoukhov, Petukhova, 2017a, 2018].

The described above results about parallelisms between long genetic and literary texts were obtained by the author on the basis of his thoughts about deep connections of molecular genetics with 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.

In a series of our works, we have found 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]. In our opinion, applications of mathematics of quantum informatics for modeling biological phenomema can help to understand many biological phenomena. For example, an adult human organism has around 10 trillion (10^{14}) human cells and each of cells containts an identical complect of DNA, genetic information from which is used for physiological functioning organism as the holistic system of cells. How such huge number of cells can reliably functioning as a cooperative whole? Quantum informatics and associations with quantum computers can help to model and understand holistic biological systems and also can help to create new effective systems of artificial intelligence on the basis of biological prototypes (about perspectives of using quantum informatics for artificial intelligence see [Biamonte et al, 2017]).

Many authors supposed that living organisms use principles of quantum informatics [Igamberdiev, 2013; Josephson, 2018; Matsuno, 1999, 1999, 2003, 2015; Matsuno, Paton, 2000; Patel, 2001a,b,c; Penrose,1996; Quantum Aspects of Life, 2008]. For example, in his thoughts about quantum computers in living organisms, R. Penrose appeals to the known fact that tubulin proteins exist in two different configurations, and they can switch between these configurations like triggers to provide bio-computer functions [Penrose, 1996]. By contrast to this “protein standpoint”, results of the model approach of the resonance genetics testify that already 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].

One should note that in our researches we are purposefully looking for the intersection of bioinformational structures with the formalisms of quantum informatics. For example, the data given in this article about the probabilities of the members of the alphabets 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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

independent events is equal to the product of the probabilities of these events. But the consideration of DNA *n*-plets as combinations of independence events as a whole badly 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: for example, 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 talking here about the total probability for the set of doublets CG + GC because both these doublets of DNA letters belong one class of equivalence "33" of 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 not appropriate to model the probabilities of hydrogen n-plets as the product of two independent events; therefore we prefer using quantum algorithmic modeling with the tensor product of the probability vectors and with other formalisms of quantum informatics, which have already proven themselves in some 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].

What one can say about a connection of 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 that time Pauling didn't know about the double helix of DNA, but our results about requlrities of hydrogen bonds in long DNA sequences confirm in some degree his prediction about an important role of hydrogen bonds in the transmittion of biologic information. Our results seem to have relations with researches of 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/). We hope that the invention of the high-precision device «Cymascope» [Reid, 2017] for studying resonances phenomena in water will help to clear additionally questions about the role of hydrogen bonds in the transmittion of information through water substances. Biological bodies contain many structural water. For example, jellyfishes consist of 99% water, but despite this they live perfectly. Hydrogen is a part of almost all organic matter and is present in all living cells, where the number of atoms to hydrogen is almost 63% [Hornak, 1996]. Taking into account the existing hypotheses about the cosmic origin of life, one can 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]. Is it possible to say that living bodies are created on the basis of a constructor of hydrogen atoms and hydrogen bonds? Science in 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 members of the ensemble requires less energy for performing own work than when working individually. Of course, this fundamental principle can be used in many other cases of resonances in different systems as the physical base. In particular, in the article [Petoukhov, Petukhova,

2017b], we used theory of oscillators with many degrees of freedom to model some phenomena of Mendelian genetics and to analyze structures of genetic-molecular alphabets; to explain phenomena of segregation in these molecular alphabets, the existence of dominant and recessive resonances in nitrogenous bases of DNA and RNA have been postulated by analogy with dominant and recessive alleles in Mendelian genetics. The hypothesis about genetic systems as binary computers on resonances 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 mathematical theory of resonances of oscillatory systems with many degrees of freedom [Petoukhov, 2016a].

One can add here that our speech communication and singing are based on the inherited ability to use acoustic resonances in speech communication and singing when generating and perception of sounds. Resonances participate in our perception of musics, which has some inherited aspects. The hypothesis about close parallels between the concepts of gene and musical idea was formulated in [Josephson, Carpenter, 1996]. In our works [Petoukhov, 2008, 2015d; Koblyakov, Petoukhov, Stepanyan, 2015; Petoukhov, He, 2009; Hu, Petoukhov, Petukhova, 2017a] we described a connection of molecular structure of DNA, including its sequences of hydrogen bonds, with ratios of musical harmony in the Pythagorean musical scale and in so called Fibonacci-stages scales resembling the known phyllotaxis laws of inherited biological morphogenesis. Figuratively speaking, taking into account this connection, one can consider living bodies as complex musical instruments, which are developed in phylo- and onthogenesis.

Another example of theoretical approaches in biology on the basis of the principle of resonances is given by works devoted to the 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 idea about huge number of resonances. He has noted that some important biological phenomena are described by histograms, which are analogical to histograms of blackbody radiations. Ji has proposed a generalization of the Planck equation for modeling many biological phenomena, having long-tailed histograms. By analogy with the principle of quantization of energy in quantum mechanics, Ji 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 in a million times more effectively than catalysts in the laboratory. What makes the enzyme in the body for 1 second, a catalyst in the laboratory can make only for 100 thousand years. 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 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.

References

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.

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

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

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.

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.

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

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 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

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.

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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

Petoukhov S.V. (2001). The Biperiodic Table of the Genetic Code and Numbers of Protons. – Moscow, MKC, 258 p. ISBN 5-7012-0203-8.

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). Resonances and genetic biomechanics. - *Symmetry: Culture and Science*, v. 26, №3, 2015, p. 379-397, http://petoukhov.com/PETOUKHOV_IN_SCS_2015.pdf

Petoukhov S.V. (2015c). 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. (2015d). 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.

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 December 2018 doi:10.20944/preprints201812.0142.v1

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]

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 .

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