The Organization of Alphabets of Nucleic Acids and Proteins as a Structural Prototype of Human Language Alphabets

Yury Figatner
South Mountains University, Oaxaca (Mexico)

Pavel Makagonov
Mixtec Technological University, Oaxaca (Mexico)

Abstract
An alphabet analysis of English, Spanish, Russian, and Vietnamese texts, which belong to different language groups, shows that among vowels there is a predominant group of four letters that is equal to the number of nucleotides of NA (nucleic acids) in the biological language alphabets. The quantity of consonants in these texts is equal to (in English and Russian) or near to (in Spanish) 20, which is the number of triplet groups of mRNA and the same number of protein amino acids. As is well known, DNA and RNA nucleotides combine the functions of energy carrier with the functions of information carrier. It was ascertained by a functional analysis of vowels that they perform the same functions. At the same time the consonants carry only information. Therefore consonants are similar not to nucleotides but to triplets of mRNA and to protein amino acid. We believe that the optimal number of letters in the alphabet of biogenetic languages was brought about by bio-evolution through the process of natural selection. The analysis of the alphabets of human languages leads us to believe that this achievement of bio-evolution was realized naturally in the evolution of human languages alphabets as well. It could therefore be concluded that the development of this approach could enlarge the spectrum of linguistic analysis methods of human languages.

Proceedings of the 10th World Congress of the International Association for Semiotic Studies (IASS/AIS)
Universidade da Coruña (España / Spain), 2012. ISBN: 978-84-9749-522-6 Pp. 739-750

The main aim of interdisciplinary investigation in linguistics and genetics is a search for general laws of the structural and functional dynamics of biogenetical and human languages. Knowledge of laws of any science permits us to create techniques of control of their objects. In our case, it means control of language organization. Without this there is no possibility to solve the key problem of modern computer linguistics which is the creation of natural language for human-computer communication.

The use of linguistic lexis by geneticists appears as a result of their joint research with linguists. The following facts come to mind:

For decades scientists have explained to students that the main elements of living beings – nucleotides and amino acids – are the letters of a biogenetic alphabet. In «The biochemistry foundation» (Leninger 1982: 15, 107) wrote that amino acids can be linked in varying sequences to form a vast variety of proteins, just as the 26 letters of the English alphabet can be combined to form an almost endless variety of words, sentences and even books. Four nucleotides form all nucleic acids and twenty amino acids are the basic structural building units of proteins. These are the same in all organisms including animals, plants, and micro-organisms. This group of 20 amino acids can be considered to be an alphabet of *the protein structure language* and the four nucleotides form the alphabet of the *DNA language*.

With a high degree of certainty specialists say that genome genetic information is realized according to the laws of language dynamics. Data on the constancy of number of letter groups in alphabets of proteins and nucleic acids draws special interest.

1. The nucleic acids DNA, mRNA, and rRNA are consisting of four letters-nucleotides - adenine (A), guanine (G), cytosine (C), and thimine (T). In RNA uracyl (U) replaces thimine. Besides, in DNA and RNA 20 groups of nucleotide triplets are marked out (which are codons and anticodons). And namely they contain information about amino acid structure of proteine.
2. The proteins consist of 20 amino acids and each of them codes with one of 20 groups of triplets mRNA.

The constancy of structure of the bio-genetic alphabets of all living beings is generally recognized and can be a sign of the existence of a general law.

1. GOALS AND OBJECTIVES

The goal of this work is a search for a law of language organization. It has not been found by methods of genetics and that is why we turn to linguistics. Furthermore we use interdisciplinary systems analysis of data about human language alphabets, and the data about the evolution of social institutions (Figatner 1998) because languages change essentially in the process of technological evolution.

The main problem was to develop a ***similarity method*** for formalization and operationalization a definition of law. But we searched for ***definition of sense (semantic)*** of the object in behavioral studies because it must give way to formalizing in the exact sciences (Hilbert, Bernays 1934: 23-25).

2. METHODS OF RESEARCH: ANALOGY SEARCH AND ANALYSIS OF SIMILARITY LEVEL

The fundamental principal of general systems theory is the similarity principle (Gigch 1978: 55, 336). A good example of this is mathematics itself. It is based on a postulate about the similarity of characteristics and relations of different natural objects, which mathematicians describe with the same numerical or character expressions.

The principle of similarity is interesting in that it open the possibility of a purposeful search for the known structural-functional characteristics of one branch of learning object (X) in the object (Y) of another branch. If we have data, which recognize X and Y on the same class of action with a high grade of probability (in our case, classes of languages) then, in spite of belonging to a different level of organization (in our case, biological and social), then very likely they will can reveal functions known for X, in object Y, for previously unknown functions or functions which researchers did not use to correspond them to each other. As a result we can obtain not only a new classification of objects, but a possibility to construct for Y previously unknown functional models. And these models always serve as a source of new methods and techniques of researching.

The principal value for transformation of the similarity principle to a ***similarity method*** is the operationalization of a definition. In accordance with opinions of logicians: the conception «algorithm» included partly to definition of conception «operationalization». The sense of an operational definition as a transfer from theoretical structure to experimental data through step-by-step instructions and to concrete actions and measurable results is an important point for revealing in principle the method of similarity.

Detecting some characteristic the similarity of objects X and Y, that belong to different branches of studying we can transfer theoretical knowledge about operationally determined object X to still non-operationalized characteristics of object Y. And then we can apply to Y the known for X series of instructions, which describe activity that the researcher has to realize for the discovery of the variable value. If the examination of values for Y is confirmed by data obtained by the study method of its branch of knowledge, then Y obtains an operational definition.

So, the similarity method is the method of discovering the operational definition of a law by transfer of the known structure of measuring for one object to another object, which beforehand defied measurement.

2.1. Limitations.

– We studied only alphabetic languages, and only alphabets of texts, but not speeches.

– We limited ourselves by setting a goal, and we exclude the mathematical linguistics methods.

– An analysis of peptides signal functions was not run; we studied only the role of amino acids as the letters of bio-genetic alphabets.

3. OBJECTS AND MATERIALS OF RESEARCH

The frequency of letters in the English alphabet was measured in the electronic data base of digitalized scientific books and articles of sociological contents with the total volume of 4138 pages (7907275 letters).

The frequency of letters in Russian texts was measured in an electronic data base of 1000 scientific social-humanitarian books: the common quantity of pages is 330733, or 336 716 362 letters (see table 3).

The frequency content of the alphabet in Spanish texts was studied on the sampling of newspaper articles of 1995. This is the part of the text corpora which is used in the Department of Frances and Romanic philology of the Autonomous University of Barcelona for revealing all possible regularities related with Spanish. The sampling contains of 400 pages of texts (192220 words) without eliminated overhead information.

The frequency content of the alphabet in Vietnamese texts was studied on the data of a representative sampling formed for 80000 syllables or near the 250000 letters (Remarchuk, Makagonov 1976: 76-88). The sampling was prepared from the works of 50 Vietnamese writers. First a frequency list of Vietnamese syllables was formed and on the basis of this list a frequency calculation was made.

4. RESULTS AND DISCUSSION. COMPARISON OF FREQUENCY GROUP FOR ALPHABETS OF BIOGENETIC AND HUMAN LANGUAGES

The alphabets form the primary level of organization of genome and human languages. We are interested in them because the letters in human texts, as well as the amino acids in the texts of polypeptide chains, are substantial language carriers. Their properties as information carriers, for example spacial form carriers, are similar to the state of an electron, or state of «holes», photons, or their groups, used as letters for coding and generating computer texts. It is likely that in creating new computer languages it is necessary to take into consideration the nature and organization of these elementary standard carriers.

4. 1. Alphabets of nucleic acids and proteins

Letters of DNA alphabets are represented by 4 nucleotides in the first row of table 1. They are used for record-keeping about all proteins of organism in ten thousand genes. The information is rewritten from DNA to molecules of informational mRNA.

They «exit» from the nucleus to the cytoplasm of cells and are used for protein synthesis. The molecules of rRNA and tRNA also are synthesized in DNA. The first (rRNA) is composed of a skeleton of ribosome, the protein manufacturing machinery of all proteins. They bind to mRNA to receive proteins' «drawings». Transport tRNAs are substantially smaller (usually about 74-95 nucleotides) found in cytoplasm their «own» «letter» - amino acid, and transfer them to ribosome as a raw material for synthesis of the polypeptide chains of proteins.

Let us emphasize the following: the information is encoded in DNA, mRNA, and rRNA by four single letters – nucleotides (tab.1). At the evolution process more complex code appears. These are genes that are constructed of triplets and are assigned to protein synthesis' control.

The sense of triplets is the encoding of amino acids. For their turn, tRNAs contain triplets-*anti*codons (tab 1), which recognize triplets-codons in mRNA and by this way find in it the place for transporting to ribosome amino acids to synthesize the protein's polypeptides.

We work out in detail these characteristics of nuclide acids, because geneticist call the triplet a «word». But linguistic semantics shows that a meaningful word always bears (is charged) complex sense, revealed by other several meaningful words. For example, the sense of the word «apple» is described by a lot of special terms in dictionary entries of botanists. So, the triplets encode only the other ***letters.***

4.2. Why geneticists call amino acids letters

In the polypeptide chains of proteins the amino acids are grouped into stretches (spirals and strands) less or about 20 amino acids. The structures of stretches and the whole chains take different conformations, which became apparent through the type or effectiveness of protein functions, such as catalyze or other activities. I.e., the protein activity fulfills no single amino acid, but the groups and chain stretches of them. On the analogy of human language, these stretches - the groups of amino acids are appropriately called «the words» and the polypeptide chain itself – as «the sentence».

TABLE 1.

ALPHABETS OF LANGUAGES: DNA, MRNA, RRNA, TRNA, AND PROTEINS (AT THE LEFT), AND MATRIX PRESENTATION OF TRIPLET CODONS BY GENETICIST (AT THE RIGHT) *(THE FIGURE BORROWED FROM: LEWIN 1983: 57-60, 336).*

Alphabet DNA – Nucleotides				
1	A (adenine)	G (guanine)	C (cytosine)	**T (thimine)**
Alphabet mRNA and rRNA – Nucleotides				
2	A (adenine)	G (guanine)	C (cytosine)	**U (uracyl)**[1]
Alphabet Proteins – Amino acids[2]**; и alphabet triplets of DNA, mRNA and rRNA – Triplet-codones**				
3	Serine	Glycine	Alanine	Leucine
	UCU, UCC, UCA, UCG, AGU, AGC	**GGU, GGC GGA, GGG**	**GCU, GCC GCA, GCG**	**CUU, CUC, CUA, UUA, CUG, UUG**
4	Lysine	Valine	Threonine	Aspartic acid
	AAA, AAG	**GUU, GUC GUA, GUG**	**ACU, ACC ACA, ACG**	**GAU, GAC**
5	Glutamic acid	Proline	Asparagine	Arginine
	GAA, GAG	**CCU, CCC CCA, CCG**	**AAU, AAC**	**CGU, CGC, CGA, AGA, CGG, AGG**
6	Phenylalanine	Isoleucine	Glutamine	Tyrosine
	UUU, UUC	**AUU, AUC AUA**	**CAA, CAG**	**UAU, UAC**
7	Cysteine	Histidine	Methionine	Tryptophan
	UGU UGC	**CAU CAC**	**AUG**	**UGG**
Alphabet tRNA – nucleotide triplet-*anti*codones*				

1st base	2nd base: U	2nd base: C	2nd base: A	2nd base: G
U	UUU (Phe/F) UUC (Phe/F) UUA (Leu/L) UUG (Leu/L)	UCU (Ser/S) UCC (Ser/S) UCA (Ser/S) UCG (Ser/S)	UAU (Tyr/Y) UAC (Tyr/Y) UAA Ochre (*Stop*) UAG Amber (*Stop*)	UGU (Cys/C) UGC (Cys/C) UGA Opal (*Stop*) UGG (Trp/W)
C	CUU (Leu/L) CUC (Leu/L) CUA (Leu/L) CUG (Leu/L)	CCU (Pro/P) CCC (Pro/P) CCA (Pro/P) CCG (Pro/P)	CAU (His/H) CAC (His/H) CAA (Gln/Q) CAG (Gln/Q)	CGU (Arg/R) CGC (Arg/R) CGA (Arg/R) CGG (Arg/R)
A	AUU (Ile/I) AUC (Ile/I) AUA (Ile/I) AUG (Met/M)	ACU (Thr/T) ACC (Thr/T) ACA (Thr/T) ACG (Thr/T)	AAU (Asn/N) AAC (Asn/N) AAA (Lys/K) AAG (Lys/K)	AGU (Ser/S) AGC (Ser/S) AGA (Arg/R) AGG (Arg/R)
G	GUU (Val/V) GUC (Val/V) GUA (Val/V) GUG (Val/V)	GCU (Ala/A) GCC (Ala/A) GCA (Ala/A) GCG (Ala/A)	GAU (Asp/D) GAC (Asp/D) GAA (Glu/E) GAG (Glu/E)	GGU (Gly/G) GGC (Gly/G) GGA (Gly/G) GGG (Gly/G)

Fig. 1. All triplet codons are included. 61 codon correspond to amino acids, and all amino acids with the exception of tryptophan and methionine encoded by more than one codon. Usually, codons - synonymous form groups in which the first two bases are common and the third varries. Three codons provoke termination. The base reading frame written by convention: 5′ to 3′.

* Alphabet tRNA consist of triplets –anticodons complementary to the codons.
For example, codon in mRNA - ACG, and its' anticodon in tRNA is a complementary group UGC.

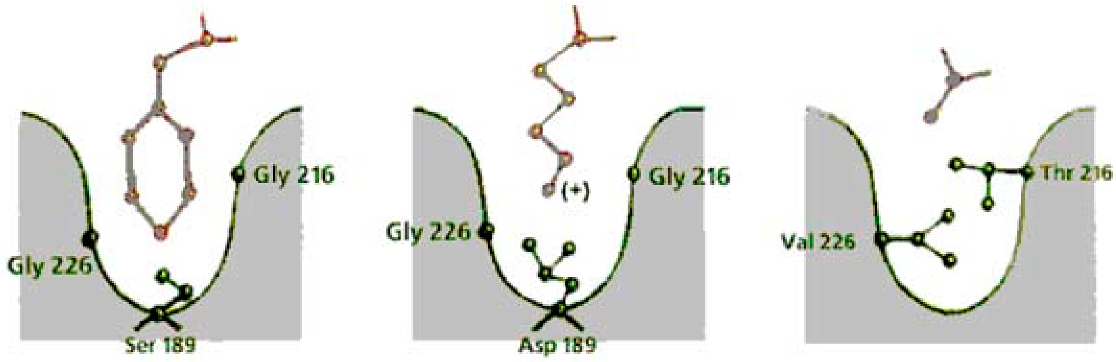

FIG 2. ORGANIZING OF SUBSTRATE-BINDING SITE OF DIFFERENT SERINE PROTEASES (*THE FIGURE BORROWED FROM*: BRANDEN, TOOZE 1991).

Calling the amino acid a «letter», geneticists denote an important sense of language units — control of their carrier behavior. An analysis of role of amino acids in protein helps to understand this sense. A spatial pattern of amino acids is geometrical information. Dynamics of forms, conformations determine physicochemical characteristics of amino acids, peptide chains, and the whole protein. Thereby they change a protein activity. If one amino acid in the chain replace by the other or change its conformation, then it will be changed the type or effectiveness of protein activity. (See figure.2).

Speaking in the technical terms of linguistics: if the ***letter*** - amino acid is changed then the new ***word***- stretches will appear. This word will change the sense of chain-***sentence,*** and with it the sense of whole ***text*** in a protein molecule. I.e., the geometrical information of structural units of protein controls its (protein) functions. These structural units form (constitute) simultaneously the substantial body and the unit of text language of protein.

This analysis confirms the competence of the definition of an amino acid as a letter in a protein language given by geneticists. As the main goups of protein language's elements appear to be similar to the elements of human languages, it is of interest to analyse a human language alphabet.

4.3. The groups of vowels and consonants in alphabets of human languages

Let us turn to alphabets of texts of human languages. They are important especially because of completeness of sense expression, in particular, professional-technical and scientific languages. In everyday oral language, unlike the texts, the most part of semantic is replaced by material context. In the texts a language realizes its lexical- and semantic characteristics and it is more adequate to goal of searching the general law of language and the models of structure of natural language for Human-Computer Interaction (or Human-Computer Interface).

Frequencies of letters of English, Russian, and Spanish texts are presented in the table 2 in decreasing order from left to right. Frequencies of Vietnamese letters in the same order are presented in the table 2. The logarithms of frequencies are given in tables plotted in figures 3-4. It can be seen that the English and Russian texts contain the groups of vowels ***E, I, O, A*** and (in the Cyrillic alphabet) ***О, Е, И, А***. They precede three times more in frequencies the other vowels about 1.57 times in the logarithmic scale).

Unlike vowels, the first groups of four consonants in English and Russian texts show of less sharp distinction in frequencies from the other consonants. Nevertheless it concentrates our attention to accentuation (protuberance) in both languages the identical in phonetic first groups of consonants. T, N, S, R and (in the Cyrillic) H, T, C, P though they are arranged in different order. (Table.2)

As one can see in the Table 2c and in the Table.4 in the Spanish and Vietnamese languages also we can see an inclination for separation the first four more frequent letters among all their vowels. But it is only a tendency. (See figure 3a.) We pay attention to the fact that Spanish first group of vowels consist of the same letters as the English one: (E, A, O, I). This fact can be explained by simultaneity of development of alphabet in text with speech. And the vowel pattern in the alphabetical human languages was formed in early Indo-European language.

Table 2. Letters' frequencies in the English, Russian, and Spanish texts

English %				Russian %				Spanish %			
Vowels				Vowels				Vowels			
e	**i**	**o**	**a**	**о**	**е**	**и**	**а**	**e**	**a**	**o**	**i**
12.45	**8.53**	**8.12**	**7.63**	**11.08**	**9.16**	**8.66**	**7.26**	**13.40**	**12.73**	**9.34**	**6.97**
u	y			ё	э	ы	я	u	y		
2.55	1.71			0.25	0.31	1.92	2.15	4.17	0.81		
				у							
				2.29							
				ю	й						
				0.70	1.28						
Consonants				Consonants				Consonants			
t	**n**	**s**	**r**	**н**	**т**	**с**	**р**	**n**	**s**	**r**	**d**
9.68	**7.29**	**6.87**	**5.89**	**7.03**	**6.69**	**5.96**	**4.94**	**7.37**	**7.33**	**6.86**	**5.23**
h	l	c	d	в	л	к	м	l	t	c	p
4.66	4.21	4.09	3.03	4.52	3.92	3.32	3.21	5.22	4.43	4.29	2.55
f	m	p	g	д	п	з	г	m	b	g	v
2.55	2.47	2.17	1.50	2.75	2.57	1.67	1.58	2.48	1.22	1.09	0.94
b	w	v	k	ч	б	х	ж	q	h	f	z
1.28	1.24	1.16	0.33	1.49	1.43	1.01	0.88	0.92	0.79	0.71	0.46
x	z	q	j	ц	щ	ш	ф	j	x	k	w
0.23	0.15	0.11	0.09	0.64	0.48	0.47	0.37	0.43	0.17	0.07	0.02

** Letters Ñ, RR, LL, and CH covering low frequencies group (~ 30000-80 000) do not represented in the table for the technical difficulties, but they do not have an influence on the overall picture of consonant frequencies distribution.

So, we can establish a fact that in spite of all studying languages belong to different groups of languages, in two studied texts' alphabets there is clearly defined dominating by frequency group of vowels. This tendency presents also in the Spanish.

The group of four most frequent vowels of Vietnamese texts also exhibits a tendency to dominating. It is interesting by the fact that pattern of its letters (See tab.2) approximates to pattern of vowels general for three Indo-European languages, in spite of the fact that the Vietnamese language belong to Austro-Asian family.

The Spanish-speaking Americas and differ from the region of Madrid parts of Spain have a lot of variant linguistic usages. There have been several initiatives to reform Spanish. It attracted our attention in table 2c, that W and K are low frequency consonants of non-Spanish origin, and two other consonants: Z and RR can be an object of discussion of Spanish language reform. So in the future the number of consonants could be reduced to 20.

The modern Vietnamese texts presentation based on Latin graphical characters was created by Catholic missionary Alexandre de Rhodes in XVII century and was adopted officially in 1910. In the Vietnamese languages there are 11 vowels, but together with bivocals (vowel diphthongs) or vowel clusters it comes to 26 vocal units.

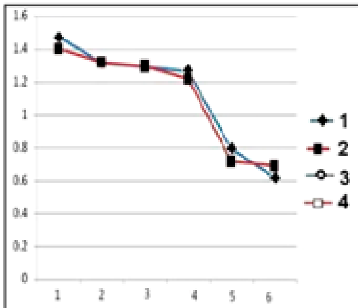

Figure 3. Logarithm of frequencies for English (1), and for first six vowels of Russian (2)

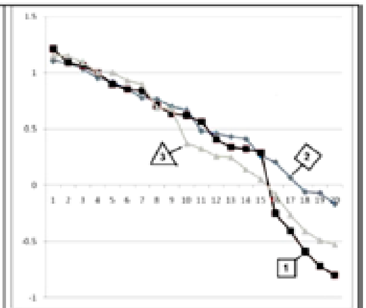

Figure 4. Logarithm of frequencies for twenty consonants of Russian (1), English (2), and Spanish (3) languages.

Table 4. Vietnamese vowel

1	A	O^	I	O
	15675	10830	10014	8817
2	U	U*	O*	E^
	5345	5103	4555	3677
3	A^	A(	E	
	3657	1555	1161	

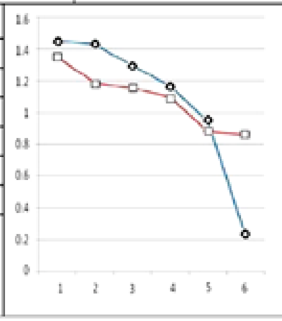

Figure 3a. Logarithm of frequencies for Spanish (3), and Vietnamese (4) vowels.

The Vietnamese language has the six tones system which modulates vocal units. Jointly it gives 122 «tonals»: vocal and so-called semivowels. They are more numerous the number of consonants, which are 23. This redundancy enables to give Vietnamese an opportunity to understand text, represented only by tonals without consonants. (That was verified by experiment with 10 Vietnamese students of Moscow State University). So, the Vietnamese consonants carry only auxiliary information; and redundancy of «full Vietnamese alphabet» (considering all tonals) exceed much the redundancy of «alphabetical units» of Indo-European languages. But namely in the part of consonants the phonetic the spoken Vietnamese language changes greatly even during the life of one generation. As an alphabetic language it is very young. That is why we limit ourselves to demonstrating vowels only. (Tab. 4).

4.4. The functional similarities of the main groups of letters of biogenetic languages and the frequency groups of human languages

The main conception of this work is composed of assumptions about numerical equality between two main groups of letters of human languages and of genome; namely – between the number of nucleotides and vowels (4), and between the number of amino acids and consonants (20).

In the first approximation the submitted data confirm this correspondence. Thereby it is proved be correct the working hypothesis, that human languages are evolving in direction of optimal structure of their alphabets. To pass to formalizing, it is necessary to operationalize the supposed regularity. For that we need to reveal the structures and functions of language elements, for which it can exist the optimizing action of natural selection. This analysis could give possibility to answer the question about the distinction of nucleotides and triplets as the elements of genome languages.

An alphabet of verbal language was formed in the co-evolution of speach and text, that activated each other. In according to V.A. Bogoroditsky, the consonants are «mouth-closers», and the vowels are «mouth-openers», and the first without the second are not effective in speech reproduction (Bogoroditsky 1909: 5). For us the following is important:

Producing the vowels by resonators and air stream energy, and its function to transmit this energy to consonants, is similar to relations of single nucleotides and triplets. All four nucleotides of DNA and Uracyl of RNA (see tab. 1) in their phosphate forms fulfill the functions of carriers of quantified (exactly measured) portion of energy for metabolic reactions in the bodies of all living being (Lehninger 1982: 413, 433-434). DNA is synthesized from the energy forms of its predecessors, which is why the energy transport costs are reduced to minimum. This mechanism of energy supply is used in synthesis – record of RNA- texts.

So, the letters – nucleotides of DNA and RNA alphabets (Tab.1) simultaneously fulfill functions of carriers of energy and information in DNA self-replication process and in synthesis of RNA texts on it. For the criteria «transfer an energy and information» the nucleotides are similar to vowels.

During the DNA and RNA synthesis letters – nucleotides spend energy for binding with NA chain. From that moment they form the part of triplets. The triplets also do not transfer energy and, as we saw, they fulfill only the information function – encoding the letters of other alphabets. The same manner ***consonants*** in a human speech fulfill information functions without energy functions.

Similarly the vowels spend their energy to produce consonants in a speech. In other words, having lost energy for «writing» DNA- and RNA- texts, nucleotides lose the energy sense of vowels. According to this reasoning both nucleotides and triplets are similar to consonants. Does it mean that triplets are similar to letters?

4.5. Similarity of triplet texts to some types of texts of human languages

Having disclosed the similarity of triplets and consonants, let us pay attention to vowels in human texts. In technical and scientific texts, and particularly in computer programs, vowels do not carry that sensory and vocal meaning, which is peculiar to speech. It is a canonical claim of professions and disciplines, excluding an influence of emotional speech intonations in their semantic. In these texts vocalism of vowels have no sense, and they are assimilated with *consonants*. i.e., not only in a human speech but in the mentioned types of texts vowels lose their energy sense, fulfilling only informational functions, as the consonants do. As we saw above, the similar reduction of vowel functions takes place in the nucleic acids' language. This implies that from the similarity of triplets (which are a signal group in PNA text) to consonants it does not follow their similarity to letters. It is useful to keep in mind, that:

– human languages are constructed by a system of codes but not by only one of them;

– triplet-codons correspond to only one of the codes and only as a translation agent, and DNA with RNA also have other functions.

To understand the role of triplet code in the system of genom language code we have formed a working hypothesis of full set of units of nuclein acids language: Letter - nucleotide; Word - a group of triplets, which are the coding of the word - fragment of the protein chain; Sentence - exon, usually separated from other exones (particularly by other introns); Text - the totality of gene exones, which are coding protein.

Form this point of view it can be assumed that the triplets are similar to some «parts of words» (like syllables or morphemes) or to pronominal, interjectional, or service word wich does not have there own sense. For example, a triplet with amino-acid (when it is joined to tRNA) can become the word pointing to the place in the chain of matrix's RNA.

5. CONCLUSIONS. FOUNDATION FOR OPTIMIZING ACTION OF NATURAL SELECTION IN THE EVOLUTION OF GENOME AND HUMAN LANGUAGES

The combination of nucleotide functions as carriers of information and energy probably was necessary in biogenetic language evolution because it reduced energy and information loss and also time spending and damage risks in transference to DNA and RNA.

This statement remains hypothetical because the early stages of bio-evolution still have not been studied. But in the same way, verbal speech and text appeared to be more effective than para-languages (gesture and others) and rock paintings in the human evolution, because they freed the body and the hands for joint actions, needed less energy for transporting, and were more exact.

Forming human speech with vowels only brings with it: 1. Heavy energy expenses and communication slowness. 2. Diminution of its informative capacity and exactness of signs and sense transmission.

Consonants, in spite of complicating the alphabet, improved essentially these parameters.

Likewise the 61 «consonant» of triplet- codons in DNA and mRNA, 31 «consonants» of triplet anticodons tRNA and 20 «consonants» amino acids increase incommensurably exactness and informational capacity in communication of parts of hereditary apparatus, if compare them hypothetically by these characteristics with the language which consists only of four nucleotides- energy-carriers.

To see the functional dependence of informational processes in a cell on these parameters state, we give values of its velocities (rates). «Velocity of DNA- polymerase motion in vivo ~ 1000 nucleotides per second.» (Lewin 1983: 428). «At 37 ºC transcription mRNA fullfill with the velocity 2500 nucleotides in minute» (Lewin 1983: 116). The examples show the processes in single proteins, but in a cell thousands of them are working simultaneously. For one second in a cell millions of communication acts take place. In a society the quantity of communication acts is much more, especially in our time. Probably, an increase or decrease in the alphabet by one letter can provoke an over-expenditure of energy as well as a worsening of communication parameters. And it can be in a greater degree with increasing of metabolism velocity (speaking about the cell) or velocity of social life.

6. THE MAIN RESULTS

1. In alphabets of natural human languages, two groups of vowels and consonants are revealed which can be compared numerically with two groups of letters of biogenetic languages: nucleotides DNA and RNA with group of vowels in human language texts (4); amino acids with group of consonants (20).

2. The method of structural- functional similarity confirms the relevancy of denomination of amino acids in proteins, and nucleotides in DNA and RNA as letters. This made it possible to formulate a hypothesis about entire a composition of units of nuclein acids and proteins language, and also about similitude between triplets and syllabis or to words without their own meaning. The method, seemingly, can be used for studying of elements of more complex levels of language organization.

3. Accepting the biogenetical languages as pattern of optimal number of letters, similar to vowels (nucleotides), and number of letters, similar to consonants (amino acids), it can be put forward the working hypothesis, that (at least technical) human languages evolve in the direction of optimal number of vowels and consonants (letters) in their alphabets. In turn, the relationships between elements of human languages can serve as standards for analysis of functions of the biogenetics languages' elements. It does not mean, that they «aim» to repeat the structure of alphabets biogenetic languages. More likely, it would be well to say that the evolutional achievement of biogenetical languages – admittedly optimal structure of their alphabets, - could be incorporated as an elemental basement in alphabetical level of languages more complex, social level of organization.

REFERENCES

Bogoroditsky, Vasiliy (1909): *Study of physiology of All-Russian pronunciation,* Kazan.

Branden, Carl; Tooze, John (1991): *Introduction to Protein Structure,* New York: Garland Publ., Inc.

Figatner, Yury (1998): *On the problem of the forecast of postindustrial technological revolution: elaboration of systemic criteria,* Moscow: INION RAN.

Gigch, John van (1978): *Applied General Systems Theory,* New York: Harper & Row. [Cit.: Russian Edition. Vol. 1. Moscow: Mir. 1981.]

Hilbert, David; Bernays, Paul (1934): *Grundlagen der Mathematik,* Vol. I: Berlin: Springer. [Cit.: Russian Edition. Moscow: Nauka. 1979.]

Lehninger, Albert Lester (1982): *Principles of biochemistry,* Vol. 1, 2: New York: Worth Publishers, Inc.. [Cit.: Russian Edition. Moscow: Mir. 1985]

Lewin Benjamin (1983): Genes: New York: John Wiley and Sons Inc. [Russian Edition. Moscow: Mir. 1987.]

Remarchuk, Vladislav**;** Makagonov, Pavel (1976): «Study of compiling of frequency list for Vietnamese syllabus». *Vietnamese linguistic articles,* Mocsow: Nauka.