Macromolecular Mechanism Of Language

Pavel E. Moroz
Hudson Health Spa
353 West 57th Street
New York, New York 10019

Received January 4, 1979; October 20, 1981

ABSTRACT. On the basis of families of Indo-European words, a comparison has been made between the principles of word changes in the evolution of language and chromosomal mutations and recombination. Deletion, addition, inversion, translocation, transposition and other phenomena known for chromosomes can be followed in the evolution of words. Principles of biological evolution, such as divergence, convergence, change in function, irreversibility, increasing complexity of organization, and accelerating rate are also true for the evolution of language. This paper expresses the idea that mental activity and its superstructure -- language, knowledge, culture -- are based on the same macromolecular mechanism as heredity. Information in general is considered to be a function of macromolecules.

* * *

Language is a window into the brain. Its highly quantitative and discrete nature, its immense but strict combinations of phonemes, words, and sentences could make it an instrument of brain research, potentially as precise as electronencephalography or neuropharmacology. This paper suggests the possibility of quantitative analysis of the linguistic activity of mind and discusses some biological aspects of the evolution of language.

Evolution of language and principles of biological evolution

The evolution of speech is a sequence of anatomo-physiological changes in organs of articulation. The evolution of language is a sequence of inventions organizing speech into a logical system, expressing the increasing complexity of civilization. The difference between speech and language is the difference between physiological function and cultural tradition. The ability of speech can express itself in the form of many different languages. The phylogenesis of speech left scanty palaeobiological and comparative-anatomical traces. Comparative linguistics could probably illuminate somewhat the evolution of speech, since those phonemes or their combinations which are common in most unrelated languages, for example, Chinese and Indo-European, obviously belong to most ancient forms of speech. The similarities in unrelated languages are most likely connected with speech; the differences -- with culture. The history of language, at least within the Indo-European group, of which a proto-language was spoken possibly in the fifth millenium B.C. (Watkins, 1968), exhibits many traits of biological evolution and follows many of its principles.

Divergence. Presented in the form of a tree, divergence is a graphic emblem of evolution. The tree of the Indo-European family of languages (Watkins, 1968) is similar to any other tree of biological evolution.

* * * * * * *

Evolutionary Theory 6: 33-42 (March 1982)
The editors thank T. Tamari, F.S. Szalay, and another referee for help in evaluating this paper.

In the process of divergent evolution the accumulation of genetic differences leads to a genetic incompatibility (with the substitution of about 10^3 genes), which means the emergence of a new species (Dobzhansky, 1970). In the evolution of language, the accumulation of a few hundreds of different words forms two linguistically incompatible dialects when people poorly understand each other. Species and dialect appear to be analogous.

Convergence. A similarity of function often produces a similarity of structure. The streamlined form of the body of a submarine, a torpedo, a shark, an ichthyosaur, a penguin and a dolphin are examples of intratechnological, intrabiological and bio-technological convergence.

The solution of an arithmetic problem by different methods or the formulation of the same diagnosis from different symptoms, are the simplest examples of cognitive convergence.

In language one may distringuish the convergence of synonymns and homographs. Synonyms, such as way and route or intense and strong re-represent the convergent approach to the same semantic function of words with different origins. This is the same convergence of ideas, given in a semiotic form, as the expression 1 radian = 57^o, reflecting the equivalence on two views on the nature of the angle. Homographic convergence is the identity of signs with a profound difference in meaning. Italian ancore has two meanings -- "anchor" and "still, yet". English sound means acoustical "sound" and "strong, healthy". Poets profit from these chance coincidences by playing on words. Natural selection profits from the originally accidental resemblance of the caterpillar and a twig ("biological homographs"), turning it into the phenomenon of mimicry.

Increasing complexity of organization. It suffices to compare a primitive unicellular organism with the higher animals, to see how much the increasing complexity of organization is inseparable from the evolutionary process. This can also be well-traced in the evolution of such old products of technology as the ship or the loom, whose path of greater complexity from simple wooden constructions to huge polytechnical automatons is very "biological".

The growth in the complexity of language appears in the increase in the number of the "lexical population" (dictionary growth) and in the increase of the number of letters in the word, in the increasing complexity of the morphological structure of the word and its etymological dismemberment (instead of oak and beat -- transcendentalism and radio-engineering), in the increasing complexity of the behavior of each individual word (the expansion of the semantic spectrum of the word, the increase in semantic load and degree of abstraction).

The rate of evolution. Beginning with the origin of life, the rate of evolution has continuously increased. The ascent from protobiont to primary vertebrates took about two billion years, the evolution of vertebrates lasted 700 million years, of birds -- 200 million years; of mammals -- 70-100 million years; and of man -- about 2 million years (Astafiev, 1972). The higher and more complex the organization, the faster the evolution.

E. Mayr (1976) indicates that there are blue-green algae going back to one or two billion years ago that are virtually identical with living forms. Marine invertebrates like Lingula and Limulus exist for periods of between 300 and 600 million years. On the other hand, many vertebrates show visible evolutionary change in a period of 100 years. The possible explanation of why the pace of evolution increases with higher organization is mainly a statistical one. Higher forms have an increased number of specialized cellular formations, organs and functions, and an evolutionary change in each of them increases the probability of the change of the whole organism in comparison with primitive forms.

Publishing can serve as a rough mirror of the evolution of ideas and knowledge. The astonishing growth in published items from the beginning of bookprinting gives students of science a good reason for speaking of the exponential growth of technical and scientific knowledge. It is possible that the exponent of development of life and the exponent of civilization only differ in temporal dimensions.

The pace of evolutionary changes on the level of species, of class and type is roughly comparable to the progress of the second, minute and hour hands of a clock. This is also true of language. It evolves as an organism of words and, just as cells change more slowly than the whole organism, words change more slowly than language as a whole, e.g. Russian _syt_, German _satt_, English _sat_ and Latin _satis_ are very similar in these dissimilar languages. Exactly in the same way the growth in the "dictionary" of mechanical engineering (screws, nuts, rivets, gears, bearings) lags behind the expanding stream of constructions.

The change in functions. Examples of change of function in biology are the transformation of the swimming bladder as a flotation regulator in the fish into an organ of breathing, the transformation of fins into walking extremities and the latter into flippers.

A shift in the meaning of words is characteristic of the evolution of language. Now _Herostrat_ and _Maecenas_ are no longer names, but only aesthetic symbols. The word _drummer_ ironically came to designate a noisy, traveling salesman to the extent that real drummers, as armies became industrialized, became merely a memory with a heroic aura.

The history of the word _vokzal_ (Ushakov, 1925) is interesting. In the XVIII century in the village of Vauxhall near London, there was a public amusement garden. Borrowing "Vauxhall" the French began to call similar amusement centers in France by this name. With the beginning of the first railroad in Russia, a concert hall near Petersburg was called _Vokzal_ (Russian phonetical equivalent of Vauxhall). Subsequently, railroad stations throughout Russia came to be called _vokzal_ and nothing pleasurable remains in this word. So its evolution was accompanied by a radical shift in function: the name of the village -- a common noun for entertainment parks -- railroad stations.

In cognate languages, words with the same root can acquire completely different and frequently even opposed functions. In Russian _cerstvy_ means "dried-up" or "stale", while in Czech it means "fresh".

The irreversibility of evolution is evident also in language. The divergence of languages from common Indo-European has increased their mutual differences and alienation. Then the expansion of cultural ties, the spread of Greek and Latin and the common fund of religious, scientific and technical terminology began to show a pan-European linguistic convergence. Finally, the emergence of synthetic languages, such as Esperanto and computer languages, heralds a future common language; the clockhand of evolution will make a complete revolution, but the two common languages will be as different as 12 midnight and 12 noon.

It is difficult to believe that all these cases of close similarity in the evolution of living nature and of language are examples of fortuitous analogy, not of homology. Below, an attempt is made to show that a fundamental identity exists in the internal mechanisms of biological evolution and the evolution of language.

Pavel E. Moroz

The Chromosome and the Word

The word possesses a very clear structure of phonemes and a very diffuse semantic structure. A word is a concept with a minimal content and a maximal degree of freedom. A sentence limits the freedom of a word. For example: *He lies with his back on the ground*; but *No lies were in his background*. The chromosome possesses the same clear semiotic structure (the sequence of genes) and the same reticence, reservation of the genetic content, which clarifies itself through the gene and enzymatic interaction during onthogenesis. Cells having originally the same chromosome set become specialized in various directions during differentiation, producing finally erythrocytes, muscles or neurons depending on the developmental context, the "sentence".

Words are characterized by the same basic types of mutations as chromosomes. Deletion is a loss of a letter or a syllable: *korona* -- *krona* (Russ.). The word *mineralogy* is the result of a deltion of a syllable in *mineralology* (Ushakov, 1925). Duplication: *dig* -- *digging*; Italian *dottore* was formed from *doctor* by means of duplication through substitution of *c* for *t* and the addition of *e*. Inversion: the Russian family name *Silvestrov* is often interpreted as *Silverstov*; Russian *futljar* and *tarelka* descend from German *Futteral* and *Teller*, which turned into Polish *talerka* (Ushakov, 1925). In both Russian words the central event is the inversion of *ral* into *lar* and *ler* into *rel*.

A translocation is an exchange of parts between nonhomologous chromosomes, and transposition is a movement of a block of genes to a new position within a chromosome (Hexter and Yost, 1976; Dobzhansky, 1970). From *leucocytosis* and *erythropoesis* two other medical terms may be combined by means of reciprocal translocation or recombination: *leucopoesis* and *erythrocytosis*. Crossing of Russian words *rastorjenie* and *vstraivat* gives *vtorjenie* and *rasstrativat*; *christoprodavets* and *knigoljub* gives *knigoprodavets* and *christoljub*. With English words *senseless* and *meaningful*, translocation produces *meaningless* and nonexisting "*senseful*". Such nonsense may also be produced from chromosome parts.

Two cases of transposition may be distinguished on the base of linguistic material: an equivalent and an inequivalent one. *Autoradiographia* and *radioautographia* is an equivalent transposition. *Cytoplasm* (a part of a cell) and *plasmacyte* (a type of cells) is an inequivalent transposition, becuase of the resulting change of sense.

In the chromosome, some genes are more mutable than others. In words, vowels also change with special ease. This is why in comparative linguistics the skeleton of the word is seen in the consonants.

Mutation in chromosomes is multiple, serial (multiple allelomorph series) with the formation of a number of characteristic (and nor arbitrary) mutants of a given gene. The laws of alternation of consonants and vowels correspond to this in words. The alternation of (b) and (v) is typical for many Indo-European languages, as for example, *fevrier* and and *februar*, *fiber* and *fever*, Latin *caballum* and French *cheval*, but the alternation of (v) and (s) or (s) and (a) is unthinkable. Such a mutation is improbable insofar as these letters are too different, non-homologous, non-allelomorphic.

The word *electrocution* may be considered as the sewing together of the fragments of two words -- *electrical execution* -- with the following "chromosomal" mechanism: tearing off *ical* and *exe* (deletion), then the uniting of the remaining parts in one string with the insertion of the linking letter (o). In exactly the same way (the reduction and the union in a new string) one may explain such words as cyborg (cybernetic organism), Euratom, UN (United Nations) and other abbreviations.

Irreversibility of evolution does not exclude reversibility of mutations (only an aggregate of mutations is irreversible). The mutated eye color in Drosophila can be restored, as a result of reverse mutation. A microbial mutant which has acquired resistance to an antibiotic can, by means of a reverse mutation, become sensitive anew (Dobzhansky, 1970). Here is an example of a reverse letter mutation. A presumable root ruk (hand) existed in Slavic parent language. Then in ancient Russian, k turned into ts -- rutse: and in contemporary Russian ts anew returned to k in ruka (Ushakov, 1925).

The comparison of the word with the chromosome may seem inexact insofar as a word usually consists of 3-10 letters while the eukaryotic chromosome consists of thousands of genes. But in the process of crossing-over, chromosomes can exchange shorter segments which are analogous to a word, and the sentences, paragraphs and chapters, formed with the aid of words, can approach in quantify of signs and, possibly, in quantity of information, a chromosome or set of chromosomes. Furthermore, in some Indian languages, words are not separated at all; also in German, not rare are whole complexes of a number of words (e.g. gallenblasenwandverletzung). Finally, the spaces between the words are not informationally empty because text without spaces would be equivocal or senseless (which is the case with such words as gynecology, which could be interpreted as gyn-ecology, ecology of a woman, or developmental, instead of development -al). Thus, the spaces are in their essense punctuation marks, and the flow of speech and the chain of text are as uninerrupted as the nucleotide chain containing terminating and initiating codons, which are also punctuation marks.

Tabulation of Mutations

Now the views set forth in this paper will be applied to an analysis of etymological transformations in a genealogy of a specfic Indo-European family of words. Table 1, which is compiled on the basis of Walter Porzig's book (1954; Chapter 7), summarizes words tracing back to the conjectural ancestor root rudhro (red) of the parent common language. All phonetic mutations are related to this word. The diphthongs (column 3) are equated with vowels. The words are grouped according to their length. It is noteworthy that the longest words (6-7 phonemes) belong to most ancient languages, the shortest (3 phonemes) belong to relatively new languages (Irish and English; also German -- rot). Deletions (see Table 1) seem to be an important type of progressive mutatitions leading to reduction of words. In genetics, too, deletions can account for the reduction and loss of organs in phylogenesis.

While some words of a developing language may shorten, some others become longer. The English noun inconceivableness (Webster, 1967) gives a good example of the evolutionary lengthening. It stems from the Latin capere -- to seize or to take; capere + com = concipere (Latin) ⟶ conceivre (Old French) ⟶ conceive (English) + able = conceivable. Negation (un) leades to unconceivable, of which the noun is unconceivableness. Thus, a simple root, capere, has become overgrown by a set of extensions (un, con, able, ness). Another interested example is the word indefeasible. It stems from the French fair (to do) ⟶ feasible + negation de = defeasible. Negation (in) of the negation (de) gives **indefeasible** which formally is a return to feasible, but this return, because of the

Table 1. Etymological mutations in the family of the Indo-European word <u>rudhro</u> (red).

Number of phonemes per word	Language	1	2	3	4	5	6	7	8	Type of Mutation
6	Common		r	u	d	h	r	o		ancestor form, "wild type"
	Greek	e	r	y	t	-	r	o	s	1,8 -A, 3,4 -S, 5 - D
	Tocharian		r	-	t	ä	r	y	e	3 - D, 4,7 -S,5 -T to 3,8 -A
	Tocharian		r	ä	t	-	r	a	m	3,4 7 -S,5-D, 8-A, Inv <u>tä</u> to <u>ät</u>.
	Indic		r	u	d	hi	r	a		7 - S
	Umbrian		r	ou	d	h	-	o	s	6 - D, 8 - A
5	Umbrian		r	u	f	-	r	u		4,7 -S, 5-D
	Latin		r	u	f	-	-	u	s	4,7 -S, 6-D, 8-A
	Latin		r	o	b	-	-	u	s	3,4,7 -S, 5,6, -D, 8-A
	Latin		r	u	b	e	r	-		4,5 -S, 7 - D
	Celtic		r	ou	d	-	-	u	s	5,6 - D,7 - S, 8-A
	Germanic		r	hu	d	d	h	-		5 - Du to 4, 6 -T to 5
	Church Slavonic		r	ъ	d	-	r	ъ		3,7 -S, 5 - D
4	Umbrian		r	o	f	-	-	u		3,4,7 - S, 5,6, -D
	Germanic		r	au	d	-	r	-		5,7 - D
	Gothic		r	au	b	-	-	-	s	4 - S, 5,6,7 -D, 8 - A
	Russian	k	r	o	v	-	-	-		1 - A, 4 - S, 5, 6, 7 -D
3	Old Irish		r	ua	d	-	-	-		5,6,7 - D
	Old English		r	ea	d	-	-	-		3 - S, 5,6,7 - D
	English		r	e	d	-	-	-		3 - S, 5,6,7 - D
Percentage of mutations			0	45	50	85	55	85		
Result of mutations			0	3o 1ä 1e 1ea 1y 1ъ 1D	3b 3f 3t 1v	14D 1ä 1e 1d	10D 1h	8D 5u 2a 1y 1ъ		

Designation of mutations: A - addition, D - deletion, Du - duplication, Inv - inversion, T - transposition, S - substitution irreversibility of evolution was attained not by means of simplification of defeasible to feasible, but by its complication to indefeasible.

Progress through lengthening and through shortening is characteristic of both a words and a chromosome. In chromosomes, the acquisition of new genes during phylogenesis leads to expansion of the genome which, in higher forms contains more genes than in lower ones. According to this an increased number of prefixes, suffixes and their combinations leads in a developing language to an increased vocabulary -- genone of language.

The phoneme r (Table 1, column 2) is very stable (no mutations). It dominates in the word and its change would apparently mean the loss of the whole word. The second r (column 6) which closes the root, as the first opens it, is less stable. Being resistant to substitutions, it is liable to deletions (languages of groups 5,4,3). Whether the stability of r is a physiological property of this phoneme, which does not have close "alleles", or whether this stability of r is determined by its important position cannot be clarified by the example of only one word, but the possible role of the place of a phoneme is very reminiscent of the "position effect" known in chromosome genetics (Roberts and Pembrey, 1978). Other phonemes (letters) of rudhro, including the consonant d, are quite variable. For each initial phoneme of rudhro a total percentage of mutations is given. Most mutable are h and o (85%). A trend exists for the consonant h to be replaced by vowels (a and e) in contrast to d. The genes, as is well known, also differ significantly in their rates of mutation and in their most probable alleles. Among the phonemic mutations those of point type (one phoneme mutation) as single deletion, substitution, addition prevail over more complex mutation, as transposition, inversion, duplication (both Tocharian, Germanic), which is also true for bilogical mutations (gene, point mutations are more frequent than chromosome mutations). The table shows that both structural and statistical principles of biological and linguistic mutability are alike.

Table2 summarizes the etymological changes of the Indo-European word snusos. In accord with Table 1, the first and especially the second consonant (s and n) are stable. They are the stronghold of the word. Table 2 confirms the "position effect": the first s is subject to deletions, but is stable to other mutations, whereas the second s (position 4) shows not only deletions but also substitutions (kh and r). This means that the

Table 2: Eytomological mutatitions in the family of the Indo-European word snusos (daughter-in-law)

Language	Phonemes 1	2	3	4	5	6	Type of mutation
Indo-European	s	n	u	s	o	s	
Sanskrit	s	n	u	s	a	-	5-S, 6-D
Greek	-	n	u	-	o	s	1,4-D
Latin	-	n	u	r	u	s	1-D, 4,5-S
Armenian	-	n	u	-	-	-	1,4,5,6-D
Albanian	-	n	u	s	e	-	1,6-D, 5-S
Slavonic	s	n	u	kh	a	-	4,5-S, 6-D

Russian	s	n	o	kh	a	–	3,4,5-S, 6-D
Old English	s	n	o	r	u	–	3,4,5-S, 6-D
Percentage of mutations	44	0	22	77	77	66	
Result of mutations	4D		2o	2D	3a	6D	
				2r	2u		
				2kh	1D		
					le		

Designations are the same as in Table 1.

stability of a consonant is not so much its immanent property but depends on the phonemic environment. However, the percentage of mutations in both Table 1 and 2 shows that some consonants (n, s, r) are more stable than others (d, h). This resistance or liability of phonemes to mutations may be accounted for by the phylogenetic age of a phoneme. If Haeckel's law that onthogenesis recapitulates the phylogenesis is true for speech, then the consonants mastered at the earlier age are those acquired at the earlier stage of phylogenesis, and, like any other phylogenetically ancient traits, they will be more resistant to mutation than more recent consonants. Unfortunately, the estimation about which consonants are mastered earlier by children is controversial (De Vito, 1970). Probably, phonemes, especially consonants, which are common in most unrelated languages, are most ancient phylogenetically and, supposedly, most resistant to mutations.

The reasons why some phonemes, regardless of their absolute rate of mutability, easily turn into some relating phonemes and do not turn into others is easier to understand from the point of view of the number of distinctive features, the number of contrasts (De Vito, 1970), as a measure of the evolutionary distance between phonemes. For example, b and p are distinguishable from each other by the feature of voice. More removed from one another are d and p: d is voiced and apical, p is voiceless and labial. One can assume that the resistance of phonemes to reciprocal mutations is directly proportional to the number of contrasts between them, i.e., to the number of anatomical and functional distinctions between corresponding organs of articulation. One of the biological equivalents to this is the phenomenon of metaplasia -- the local transformation of one tissue into another, which occurs more easily the less the phylogenetic distance between the tissues.

The similar symptomatology of linguistic and genetic phenomena does not necessarily mean that linguistic transformations occur on the level of chromosomes. Due to mechanisms of transcription and translation, copies of a chromosome or its parts are present in the cytoplasm in the form of nucleotide and polypeptide chains and they are more available for fast and fluent reactions connected with speech than the chromosome itself. Macromolecular chains show all kinds of recombinations and nutations known for chromosomes. Therefore, linguistic mutations are compatible with them. If the neural information is encoded in the polypeptide chains of the surface of a neuron and is retrieved in the form of nerve impulses (Moroz, 1980), this may be also the way of the storage and processing of verbal information including the chromosome-like changes of words. According to hypothesis the basic dictionary may be stored in chromosomes of neurons, responsible for speech, but its part -- the vocabulary of the everyday language -- is written most likely in the cell surface polypeptides.

Biology of Information. The nature of information is that ultimate level on which the striking similarity of forms and mechanisms of heredity and language, as well as other expressions of brain activity, appears as a homology, not an analogy (Moroz, 1977). Information appears to be a function of macromolecular structure of nucleic acids and proteins, characterized by stability of content, its incidental or programmed change and mutual exchange of parts of content. These characteristics of macromolecular information express themselves variably in genetic and mental activities. Stability of content results in heredity or its mental derivative -- memory; changeability of the content is mutation and recombination or their mental derivative -- informational associations, recombinative thinking. The mechanism of the expression of information is the genetic code or its possible nervous modification -- the code of nerve impulses (Moroz, 1980). There are no separate worlds of somatic life, of thinking, of language, of culture. There is a world of macromolecular information in whcih stability, recombination and selection express themselves in somatic, instinctive and mental life. Linguistic phenomena can be ultimately reduced to few main macromolecular mechanisms. This probably is even more true for the language of music, with its 8-letter compact alphabet, based on frequency of sound.

Since basic macromolecular mechanisms in heredity, immunology, thinking and nervous regulation have much in common, cognitive-like mechanisms are present in non-nervous and non-mental processes (recognition of cell identity, meiotic synapsis of pair chromosomes with the recognition of identical loci), and typically genetic phenomena like storage of programs or recombination are present in the thinking process. Hence, there is a certain way of processing macromolecular information in living nature which appears in the form of reproduction, immunological reactions, thinking, language, etc. Therefore, when part of living nature -- the brain -- seeks to memorize, compare, recombine and select uncontroversial information about other living functions, e.g. about language as a part of biology, the mind uses the same principles which it investigates.

BIBLIOGRAPHY

Dobzhansky, Th.: Genetics of the evolutionary process. Columbia Univ., New York - London (1970).

Hexter, H., Yost, H.T.: The science of genetics. Prentice-Hall, N.J. (1976).

Mayr, E.: Evolution and the diversity of life. Harvard Univ. Press. Mass. (1976).

Moroz, P. E.: Physical limits of biological and technical evolution. Biosci, Communicat. 3, 93-412 (1977).

Moroz, P. E.: A hypothesis of the code of nerve impulses. Acta Biotheoretica, 29, 101-109 (1980).

Porzig, W.: Die Gliederung des indogermanischen Sprachgebiets. Heidelberg (1954).

Language

Roberts, J.A.F. and M.E. Pembrey: An introduction to medical genetics. Oxford (1978).

Ushakow, D.N.: Concise introduction into linguistics. Moscow (1925) (In Russian).

De Vito, J.A.: The physiology of speech and language. Random House, N.Y. (1970).

Watkins, C.: Indo-European and the Indo-Europeans. The American Heritage Dictionary. Houghton Mifflin Co. 1496-1497 (1968).

Webster's Dictionary. Merriam Company, Mass. 1967

* * * * * * *

The author gratefully appreciates the important advice, suggestions and corrections of Dr. Tal Tamari and Dr. Frederic Warburton, as well as editorial assistance and the discussions of Dr. David Vilbur.