საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 3, №2, 2009
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 3, no. 2, 2009

Linguistics & Semiotics

«Paradigms» in Linguistics and the Problem of the Isomorphism between the Genetic Code & Semiotic Systems

Thomas V. Gamkrelidze

Academy Member, Georgian National Academy of Sciences

Dedicated to the 80th Birthday of my friend & colleague Vyacheslav V. Ivanov, an eminent scholar of our time – Linguist, Semiotician, Culture Historian.

ABSTRACT. The problem of validity of the notion of «Paradigm» in science (in *Thomas Kuhn's* sense of the term) is being much discussed in connection with its application to Humanities and Social Sciences. Notwithstanding some objections to this effect, the notion of «Paradigm» should be considered effective in application to Humanities, especially to Linguistics, enabling us to propose a reasonable historical classification of diverse directions in the Science of Language. © *2009 Bull. Georg. Natl. Acad. Sci.*

Key words: *Paradigm, Paradigms in Linguistics, Biosemiotic Paradigm.*

We may envisage the history of European (and American) Linguistics as an interchange of certain «Paradigms» existing parallelly, or replacing one another, in space and time that may be presented globally as follows:

Paradigm I:

Port-Royal Grammar (Grammaire générale et raisonnée) [Antoine Arnaud & Claude Lancelot]

Paradigm II:

Comparative Historical Linguistics (Franz Bopp; Karl Brugmann & Neogrammarians; Antoine Meillet et al.)

Paradigm III:

Japhetic Linguistics (Nicholas Marr)

Paradigm IV:

Linguistique synchronique (Ferdinand de Saussure):
a) Phonology & Structural Linguistics (Nicholas Trubetzkoy, Roman Jakobson)
b) Descriptive Linguistics (Leonard Bloomfield et al.)
c) Structural Typology & Universals of Language (Joseph Greenberg, Irine Melikishvili; Georgi Klimov)

Paradigm V:

Transformational Generative Grammar (Noam Chomsky)

Paradigm VI:

Nostratic Theory & Language Macrofamilies (Holger Pedersen; Vladislav Illich-Svitich, Sergey Starostin, Aaron Dolgopolsky; Allan Bomhard; Joseph Greenberg, Merrit Ruhlen)

Paradigm VII:

Typological Comparative Historical Linguistics (Roman Jakobson, Oswald Szemerényi, Winfred Lehmann, Paul Hopper, Vyacheslav Ivanov/Thomas Gamkrelidze)

A particularly insightful *«Paradigm»* revealing the isomorphism existing between the genetic code and different semiotic systems was advanced in the second half of the 20th century; it may be called a *«Biosemiotic Paradigm»*.

As is known, a great discovery was made in the 50's of the past century in molecular biology, shedding light on the hereditary mechanism. Heredity was found to correspond to information recorded along the chromosomes by means of a definite type of chemical alphabet. *Four* nucleotides or "chemical radicals" serve as the initial elements of this alphabet – its "letters"; combining with each other in infinite linear sequences of nucleic acids, these elements create, as it were, a chemical text of genetic information. Similarly to a phrase constituting a segment of a definite linguistic text formed with the aid of a linear sequence of a small number of initial discrete units – letters or phonemes – an individual gene corresponds to a definite segment in a long chain of nucleic acids presenting the *four* initial chemical radicals. And similarly to the linguistic code, in which the initial units – phonemes – are per se devoid of meaning but serve for the building through definite combinations of minimal sequences expressing a definite content within the given system, precisely in the same way, in the genetic code it is not a separate element of the system, not an individual chemical radical that is informative, but special combinations of these initial *four nucleotides* of *three elements* each, forming so-called "triplets". Since in all $4^3 = 64$ *combinations of three* can be formed by these *four initial elements,* the genetic vocabulary comprises *64 "words",* of which three triplets represent "punctuation marks", denoting in a long sequence of nucleic acids the beginning and end of a "phrase", while the rest correlate with one of the 20 amino-acids, and among such "triplets" further "synonymous words" are identifiable, i. e. several sequences that correlate with one and the same amino-acid.

The establishment of such correlations between "triplets" of the *four initial elements* and 20 amino-acids, and the conversion of a long chain of "triplets" into a protein sequence of amino-acids – into a peptide chain – is precisely the deciphering or decoding of the hereditary information contained in the genetic code, similarly to the decipherment of a Morse code message when translating it into some other language.

Obviously enough, all living beings on earth possess "knowledge" of the genetic code in the sense of being capable of correctly deciphering the genetic "words" forming the content of the genetic information, and accordingly of synthesizing protein sequences. In this sense, the genetic code is universal, every living thing on earth possessing a key to it (Jacob 1977).

Thus, the infinite variety of living beings is reducible in the final analysis to long genetic "messages" formed according to the rules of linear combination of the elements of the genetic code, exhibiting striking features of structural similarity with the linguistic code. It is not fortuitous that from the very moment of the decipherment of the genetic code molecular genetics began to borrow extensively linguistic concepts and terminology in its further research into the mechanism of heredity.[1]

However, the linguistic code underlying natural languages has a much greater number (than four) of initial

[1] For its part, linguistics also has cases of borrowing some concepts and terms from molecular genetics. Thus, e. g., it is suggested that, in the theory of markedness, the members of the hierarchical relation of "markedness" – called earlier "unmarked" ~ "marked" (ultimately traceable to the terminology of the Prague Linguistic School according to which the members of this binary relation were characterised as *merkmallos ~ merkmalhaltig)* – may now be called – in conformity with their content – as "dominant" vs. "recessive", and the "relation of markedness" be reformulated as "relation of dominance" (cf. Gamkrelidze 1979).

units-phonemes – whose combinations constitute the minimal meaningful elements of a sound language, this being one of structural features distinguishing it from the genetic code. This creates a redundancy in a language system, permitting the correction or the reconstruction of the established sequences of initial units and correction of distortions in messages that result from "noise" under the impact of external factors. The genetic code lacks such a property; hence any permutation or elimination of individual elements in the linear sequence of nucleotides inevitably leads to a distortion of the originally recorded genetic information.

The structural isomorphism evidenced by the two different information-carrying systems – genetic and linguistic – built on the principle of a linear combination of initial discrete units, raises a phenomenological question as to the nature of these systems and to the causes of such structural isomorphism. Various points of view are being advanced.

Most characteristic in this respect was the controversy between the two famous scientists representing different fields of knowledge: The linguist *Roman Jakobson* and the biologist-geneticist *François Jacob*: Is the observed structural isomorphism between the two codes – genetic and verbal – purely external, resulting from a mere convergence induced by similar information needs, or does this isomorphism stem from the phylogenetic principle of construing the linguistic code according to the structural patterns of the genetic code; is it perhaps due to the fact that the foundations of the linguistic patterns superimposed upon molecular communication have been modelled directly upon its structural principles? The latter assumption was upheld by *Roman Jakobson*, whereas *François Jacob* assumes rather an analogous structuration of different information-carrying systems with analogous functions.

The Jakobsonian conception of structural isomorphism between the genetic and linguistic codes presupposes an evolutionary process of superposition of the linguistic code immediately on the genetic and copying its structural principles, this having been effected under conditions of an unconscious possession by the living organism of knowledge on the character and structure of the latter. This fully refers to the sphere of the "unconscious" – the unconscious possession by the organism of information on the structure of its essential mechanisms. And all this was manifested not only in the phylogenetic process of shaping the language mechanism according to the model of the genetic code, but also in the various creative acts of outstanding individuals who build special information-carrying (semiotic) systems according to the model and principles of the genetic code, apparently without explicit familiarity with the structure of the latter.[2]

In this connection it is appropriate to recall the "Theory of the glottogonic process" advanced by the outstanding linguist and philologist of his time *Nicholas Marr*, who possessed a peculiar scholarly intuition, at times leading him to logically unfounded, but quite unexpected solutions of certain phenomena in a right direction.

Thus, for example, *Nicholas Marr* reduced the historically existing diversity of languages to precisely *four* (sic!) initial elements consisting, strange as it may seem, of peculiar sound "triplets" – meaningless sequences of three sounds: *sal, ber, yon, rosh.* According to *Nicholas Marr*, any text of arbitrary length in any language of the world is, in the final analysis, the result of a phonetic transformation of only these initial *four elements* – per se signifying nothing – combined into definite linear sequences. This, in *Marr*'s view, made for the unity of the glottogonic process.

Nicholas Marr's glottogonic theory has no rational basis whatsoever; it contradicts also the logic of modern theoretical linguistics and general linguistic methodology, and in this sense it is irrational and irrelevant to Linguistics proper. However, this theory – representing a peculiar structural model of language, very close to the genetic code – is not irrelevant to Science and Psychology, in general, and may serve to illustrate the occurrence in an outstanding personality of intuitive ideas on the structure of genetic code, evidently copied by him subconsciously in developing an original model of language as an information-carrying system.

Of course, *Nicholas Marr* could not have had explicit and conscious knowledge of such a structure of the genetic information system, as neither could those ancient Chinese philosophers who, about three thousand years earlier, compiled the book *I Ching* (Chinese "*Book of Transformations*"), in which they developed a special system of "transformations" *of four* binary elements formed of the "masculine principle" *yang* and the "feminine principle" *yin,* and grouped into "triplets", yielding a total of 64 triple sequences, analogous to the genetic "triplets". It is with the help of such "triplets" that the diversity of the living world is described in this

[2] The very scientific penetration into the structure of the genetic code essentially amounts to the human beings becoming explicitly aware of the structure of their own genetic mechanism, to the transfer from the "unconscious" to the sphere of conscious of the knowledge on the ultimate structure of all living beings on earth, implicitly built into each living organism.

ancient Chinese symbolic system. In this context, especially significant seem to be also the systems with *four* elements of the world in the cosmogony of the *Ionians*, with *four* humours of the human body in *Hippocrates*, and others, supporting the idea that a strict relation is imposed by a sort of unconscious filiation between the discussed systems and the genetic code.

All these symbolic semiotic systems (in particular the ancient Chinese "*Book of Transformations*" of the binary elements *yin* and *yang,* and *Nicholas Marr's* model of language) strikingly correspond – down to quantitative parameters – to the structure of the genetic code that evidently served in the creators of these systems as an unconscious modelling substratum.

I assume that *Nicholas Marr*'s so-called *"Japhetic Theory"* is a special *«Paradigm»* that came into being on ideological grounds as a counterpart to *«Comparative Historical Paradigm»*.

ლინგვისტიკა და სემიოტიკა

«პარადიგმები» ენათმეცნიერებაში და იზომორფიზმის პრობლემა გენეტურ კოდსა და სემიოტიკურ სისტემებს შორის

თ. გამყრელიძე

აკადემიკოსი, საქართველოს მეცნიერებათა ეროვნული აკადემია

ეძღვნება მეგობრისა და კოლეგის, თანამედროვეობის გამოჩენილი მეცნიერის — ვიაჩესლავ ივანოვის დაბადებიდან 80 წლისთავს.

„პარადიგმის" ცნება (ამ ტერმინის *თომას კუნისეული* გაგებით) შემოტანილია ენათმეცნიერებაში და მოცემულია საენათმეცნიერო მიმდინარეობების კლასიფიკაცია ლინგვისტური „პარადიგმების" მიხედვით.

REFERENCES

Gamkrelidze, Thomas V. 1979. "Hierarchical relations among phonemic units as phonological universals". *Proceedings of the 9th International Congress of Phonetic Sciences* (Copenhagen) II, 9-15.

Jacob, François. 1977. The linguistic model in biology. In: D. Armstrong and C. H. van Schooneveld (eds) *Roman Jakobson, Echoes of his Scholarship.* Lisse, de Ridder, 186-192.