

Studying Text Structure in Linguistic Synergy

Bikesh Revovna Ospanova

Karaganda State Technical University, Karaganda, Kazakhstan

Abstract: The article is devoted to studying the text structure from the position of the linguistic-synergetic approach. In the article there are reflected the results of the study, whose purpose was using the information amount measure permitting to analyze the general mechanisms of information-entropic text characteristics, underlying all spontaneously running in the surrounding world processes of information accumulation which lead to the systems structure self-organization. The text is considered as a complicated multilevel synergetic system. There is analyzed its structure within the frames of multidimensionality. In the article there are presented some aspects of the experimental approach to computing of Kazakh and English texts, as well as experimental data demonstrating the results. The urgency of studying the text structure from the position of linguistic synergy is defined by the need of the deeper studying of the language which is not only one of the main means of the impact on the public consciousness, but also the integrate, unstable, changing evolutionary world which gives the right to speak of it as of a self-organizing system; insufficient degree of studying the laws of linguistic synergy and at last, insufficiency of the existing research approaches for studying the text at a specific sight angle permitting to put to the front plan the studying of the text inner life synergetic nature. These problems in linguistics and other scientific disciplines enter the number of little studied ones and require their deep studying from the side of linguists and representatives of other sciences.

Key words: Linguistic-synergetic approach • Self-organizing system • Information

INTRODUCTION

The humanity and the processes of globalization taking place all over the world entered the stage of their development which is called the information society. In these conditions appearing new paradigms of cognition is quite well-formed and synergy becomes the most integral in them.

Synergy is a new trend in the man's cognition of the nature, himself and the sense of his existing. Besides, a new quality in cognition is achieved due to using nonlinear thinking and synthesizing the achievements of various sciences when designing the universe image. Today synergy is developed by scientists of various scientific disciplines, so it can be considered an interdisciplinary trend of scientific studies.

The basic ideas of synergy are presented in the works by R. Kohler [1], G. Haken [2], G. Altman [3], I.R. Prigozhin [4], W. Wildgen W. [5], Brown C., G. Yule [6], Schiffrin D. [7] and in the studies of the home scientists V.I. Arshinov, V.G. Budanov, R.G. Piotrovskiy, S.P. Kapitsa, S.P. Kurdyumov, V.S. Ratnikov, A.P. Rudenko and others.

The main propositions of synergy relating immediately to linguistic applications are at present actively studied by linguists. They develop new ideas of synergy, new world vision, connected with studying the principles of self-organization, nonlinearity, non-equilibrium, bifurcation changes, time irreversibility, instability as the underlying characteristics of the evolution processes.

Today synergy in linguistics presents a new unifying trend whose aim is to reveal the general ideas, methods, laws of the language transition from one level of organization to another. The language, being in constant development and movement, is a complicated, dynamic, self-organizing system. In this connection it is worth noting that linguistics is closely connected with synergy as a new interdisciplinary trend of studying the systems consisting of a lot of components or subsystems which interact between each other in a complicated manner. Alongside with this there appear new aspects of discussing and considering the synergetic approach to the language analysis. Thus, there appears a new trend of studies, i.e. linguistic synergy that is based on synergy presentations of the complicated self-organizing systems

development. The text is one of the most complicated objects of linguistic studies. At present there are made attempts to cast a new glance at the text and its structure from the position of linguistic synergy. The diversity of approaches to the text gives birth to multiple variants of its understanding.

Methodology of Studying: In the process of studying on the complex system base there were used the following methods: general scientific methods of system analysis, modeling, linguistic synergetic approach, methods of quantitative analysis; information analysis, comparative method for revealing information-entropic characteristics of texts. There were revealed the main criteria of linguistic-mathematical paradigm based on methods of self-organization theory. There were considered texts of various genre-stylistic nature, carried out within the frames of the aspect of the text general structure.

The Main Part: The works of a lot of scientists are devoted to studying the synergetic trend in linguistics, namely: R.G. Piotrovskiy, I.A. German, V.A. Pishchalnikova, L.S. Pikhtovnikova, S.K. Gural, I.Yu. Moiseyeva, G.G. Moskalchuk, K.I. Belousov and others. The processes taking place in the language in the works of these scientists began to be considered from the point of view of complexity, nonlinearity, chaos, attractor, synergy; there is observed the transition from the static vision of the text to the dynamic one. There is a noteworthy fact that the studies of linguistic synergy scientists permit to have a new glance at the linguistic problems and to comprehend them in a new way.

So, well-known linguist R.G. Piotrovskiy, generalizing the experience of linguistic statistics, engineering linguistics, studies in the field of methodology of language teaching, notes that the synergetic approach in linguistics is necessary:

- In the structural-typological comparison of different languages;
- In comparing entropic characteristics of various texts, their style belonging;
- In constructing and realizing algorithms of the text automatic reprocessing, particularly, in texts computer translation and annotation [8].

Linguistic synergy, by the words of K.I. Belousov, permits to consider the text in the “text-environment” aspect as an open nonlinear self-organizing system, whose processes of self-organization are connected with

the probabilistic laws of natural objects forming. In the researcher’s opinion, “the text is a self-organizing functional metasystem, whose individual systems tend to establish between each other the unified temp-rhythmic evolution parameters based on the order parameter, presented by the attractor (creative, cyclic), as well as on the basis of the governing parameters of the whole metasystem, i.e. motivation-purposeful guidelines of the man speech activity and universal phenomena of any natural object form-building” [9].

G.G. Moskalchuk considers the laws of the text forming as a natural object on the material of a wide corps of various styles texts. In this researcher’s opinion, the text form is a result of the man’s sense-generative activity and the language representative function. As the text structural organization is conditioned by the nature laws, the issues related to its self-organization are to be considered in close relation to the man. Consequently, in a text as in an open non-equilibrium system, the order exists only due to the man’s energy flow. G.G. Moskalchuk considers that the text is a product of the man’s intellectual activity presenting an information clot and invested with a certain form. By the scientist’s words, in the text perception the man perceives its information and structural negentropy, accumulates and uses the information which, in turn, excites the organism’s and intellect’s energy pulses, i.e. “the energy put in the text structure by its author transforms in the text into a system of language symbols and signs, that possesses a certain structural organization and comprises a certain information amount” [10].

Well-known linguist N.S. Valgina speaks of such an aspect in the text characteristic, as “a measure of probability of pragmatic information projected to a potential reader”. Pragmatic information is a fixed in the language expression the speakers’ attitude to the realities, to the report content, to the addressee. This measure defines the information usefulness (value) for the user’s achieving of the put purpose. By the scientist’s words, if this measure of pragmatic (new, useful, perceived by the reader) information is called the term “entropy”, then in relation to functionally different texts this measure will be defined as follows: for official and business texts the entropy adequacy is principal, for scientific ones it is limited by a special readers’ circle and therefore quite predictable; for journalistic and paper texts the entropy adequacy is principal but unpredictable; for fiction texts it is not principal and unpredictable” [11]. Indeed, the complete understanding, the achieving of the information effect thought of by the author, cannot be achieved

because information is a state reflecting structural relations in the world, but not a thing or an object which can be passed.

Thus, generalizing the abovementioned, we come to a conclusion that the linguistic synergetic approach to the text is the most substantiated, as within this approach the text is studied as an open non-equilibrium nonlinear self-organizing system possessing an ability to exchange with the surroundings and information.

Within the frames of our study there was carried out an information-entropic analysis of the texts belonging to five genre-stylistic types of the modern Russian and Kazakh languages. There were developed methods of information analysis of written texts with the unified quantitative evaluation of their perfection based on use of the law of preserving the sum of information and entropy.

Our studies were conditioned by the necessity to study various genres text material with the aim of its perfection. By the way of using mathematical computations we obtained the values of a letter entropy taking into account one, two, three, four, five and six letters of the text in Russian and Kazakh. There was developed a linguistic mathematical model for analyzing the text structure built on the basis of the fundamental law of preserving the sum of information and entropy using Shannon's formula. In the general characteristic of the entropic-information (entropy is the measure of disorder, information is the measure of disorder removal) analysis of the texts we used Shannon's statistical formula to determine the text perfection, harmony:

$$H = -\sum_{i=1}^N p_i \log_2 p_i \quad [12]$$

where p_i is probability of revealing some uniform element of the system in their set N ;

$$\sum_{i=1}^N p_i = 1, p_i \geq 0, i = 1, 2, \dots, N.$$

For linguistics the language entropy is an important measure. The language entropy is a general measure of the probabilistic-linguistic relations in the language text.

$$H_1 = H(\alpha_1) = -0,142 \cdot \log_2(0,142) - 0,094 \cdot \log_2(0,094) - \dots - 0 \cdot \log_2(0) \approx 4,3742$$

If we compare value H_1 with value $H_0 = \log 32 = 5$ bit, we'll see that the irregularity of the alphabet different letters occurring leads to the reducing of information contained in the text letter by about 0.63 bit.

In this connection we carry out a comparison between the data characterizing the numeric evaluation of these measures in Kazakh and Russian.

As the Russian alphabet consists of 32 letters (31 letters, 1 blank), so, according to this,

$$H_0 = \log 32 = 5 \text{ bit.}$$

H_0 is the maximum value of the text entropy consisting in the receiving one letter of the Russian text (information contained in one letter) under the condition that all the letters are considered to be equally probable; bit is the unit of measuring information.

The Kazakh alphabet consists of 43 letters (42 letters, one blank), so, according to this, $M \frac{\log 43}{\log m} = M \frac{H_0}{\log m}$. Here

$$H_0 = \log 43 = 5,4 \text{ bit.}$$

- entropy of the experiment consisting in receiving one letter of the Kazakh text (information contained in one letter) under the condition that all the letters are considered to be equally probable.

The computations show that value H max in Russian (32 letters in the alphabet, letters *с и ё, ь* and *ъ* are expressed by the same combination and the blank (-) as an empty space between words) does not practically differ from H max of the Kazakh alphabet content (42 letters and a blank):

$$\text{In Russian } H_0 = \log 32 = 5 \text{ bit}$$

$$\text{In Kazakh } H_0 = \log 43 = 5,4 \text{ bit}$$

Let's consider the use of the information-entropic analysis on the example of a Russian text of the journalistic style of speech. The material for the experiment will serve a piece of a paper article [13]. For computing the text information we calculated the probabilities of one letter occurring, two-, three-, four-, five and six-letter combinations in this text of 500 signs volume, 429 signs without blanks.

By equalizing the frequencies to the probabilities of corresponding letters occurring, we'll obtain an approximate value for one-letter entropy:

For the further characteristic of the information amount and the text entropy we'll transit to computing the conditional entropy H_2, H_3, H_4, H_5, H_6 .

As we see, the conditional entropy of the text accounting for two-letter combinations is equal to:

$$H_2 = H\alpha_1(\alpha_2) = H(\alpha_1\alpha_2) - H(\alpha_1) = -0,0199 \cdot \log_2(0,0199) - \dots - 0,0159 \cdot \log_2(0,0159) - 0,0139 \cdot \log_2(0,0139) - 0,0020 \cdot \log_2(0,0020) + \dots + 0,142 \cdot \log_2(0,142) + 0,094 \cdot \log_2(0,094) + 0 \cdot \log_2(0) \approx 3,0423$$

The text entropy accounting for three-letter combinations:

$$H_3 = H\alpha_1\alpha_2(\alpha_3) = H(\alpha_1\alpha_2\alpha_3) - H(\alpha_1\alpha_2) = -0,0081 \cdot \log_2(0,0081) - \dots - 0,0081 \cdot \log_2(0,0081) - 0,0020 \cdot \log_2(0,0020) + 0,0199 \cdot \log_2(0,0199) + \dots + 0,0159 \cdot \log_2(0,0159) + 0,0020 \cdot \log_2(0,0020) \approx 0,7895$$

The text entropy accounting for four-letter combinations:

$$H_4 = H\alpha_1\alpha_2\alpha_3(\alpha_4) = H(\alpha_1\alpha_2\alpha_3\alpha_4) - H(\alpha_1\alpha_2\alpha_3) = -0,0060 \cdot \log_2(0,0060) - \dots - 0,0060 \cdot \log_2(0,0060) - 0,0020 \cdot \log_2(0,0020) + 0,0081 \cdot \log_2(0,0081) + \dots + 0,0081 \cdot \log_2(0,0081) + 0,0020 \cdot \log_2(0,0020) \approx 0,5605$$

The text entropy accounting for five-letter combinations:

$$H_5 = H\alpha_1\alpha_2\alpha_3\alpha_4(\alpha_5) = H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5) - H(\alpha_1\alpha_2\alpha_3\alpha_4) = -0,0060 \cdot \log_2(0,0060) - 0,0040 \cdot \log_2(0,0040) - \dots - 0,0020 \cdot \log_2(0,0020) + \dots + 0,0060 \cdot \log_2(0,0060) + \dots + 0,0020 \cdot \log_2(0,0020) \approx 0,0451$$

At last, we transit to computing the value H_6 , which is to be minimum, i.e. is practically reduced to. So, the text entropy accounting for six-letter combinations is equal to:

$$H_6 = H\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5(\alpha_6) = H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\alpha_6) - H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5) = -0,0040 \cdot \log_2(0,0040) - 0,0040 \cdot \log_2(0,0040) - 0,0040 \cdot \log_2(0,0040) - \dots - 0,0020 \cdot \log_2(0,0020) + 0,0060 \cdot \log_2(0,0060) + \dots + 0,0020 \cdot \log_2(0,0020) \approx 0,0108$$

In the process of our study, when computing the number of repeating different letter combinations in the journalist text, we came to the following indicators (in bits):

H_1	H_2	H_3	H_4	H_5	H_6
4,3742	3,0423	0,7895	0,5605	0,0451	0,0108

These figures confirm the fact that in Russian journalistic texts with increasing the information there takes place the reducing of the uncertainty degree (entropy).

Now let's carry out the analysis of the Kazakh journalistic text. The material for the experiment served a paper article from the Republican paper "Yegemen Kazakhstan" [14], a daily governmental paper that publishes official information of the state authoritative bodies work, normative legislatives documents and information of the authorized state bodies on issues of home and foreign policy.

The text contained 500 signs with blanks and 438 signs without blanks. As a result of computing there were obtained the following data.

The text entropy accounting for one letter:

$$H_1 = H(\alpha_1) = -0,124 \cdot \log_2(0,124) - 0,086 \cdot \log_2(0,086) - \dots - 0,002 \cdot \log_2(0,002) \approx 4,4253$$

The text entropy accounting for two-letter combinations:

$$H_2 = H\alpha_1(\alpha_2) = H(\alpha_1\alpha_2) - H(\alpha_1) = -0,028 \cdot \log_2(0,028) - 0,026 \cdot \log_2(0,026) - \dots - (0,002) \cdot \log_2(0,002) + 0,124 \cdot \log_2(0,124) + 0,086 \cdot \log_2(0,086) + \dots + 0,002 \cdot \log_2(0,002) \approx 2,7267$$

The text entropy accounting for three-letter combinations:

$$H_3 = H\alpha_1\alpha_2(\alpha_3) = H(\alpha_1\alpha_2\alpha_3) - H(\alpha_1\alpha_2) = -0,01 \cdot \log_2(0,01) - 0,01 \cdot \log_2(0,01) - \dots - 0,002 \cdot \log_2(0,002) + 0,028 \cdot \log_2(0,028) + 0,026 \cdot \log_2(0,026) + \dots + 0,002 \cdot \log_2(0,002) \approx 1,0687$$

The text entropy accounting for four-letter combinations:

$$H_4 = H\alpha_1\alpha_2\alpha_3(\alpha_4) = H(\alpha_1\alpha_2\alpha_3\alpha_4) - H(\alpha_1\alpha_2\alpha_3) = -0,008 \cdot \log_2(0,008) - 0,008 \cdot \log_2(0,008) - \dots - 0,002 \cdot \log_2(0,002) + 0,01 \cdot \log_2(0,01) + 0,01 \cdot \log_2(0,01) + \dots + 0,002 \cdot \log_2(0,002) \approx 0,3301$$

The text entropy accounting for five-letter combinations:

$$H_5 = H\alpha_1\alpha_2\alpha_3\alpha_4(a_5) = H(\alpha_1\alpha_2\alpha_3\alpha_4a_5) - H(\alpha_1\alpha_2\alpha_3\alpha_4) = -0,008 \cdot \log_2(0,008) - 0,008 \cdot \log_2(0,008) - \dots - 0,002 \cdot \log_2(0,002) + 0,008 \cdot \log_2(0,008) + 0,008 \cdot \log_2(0,008) + \dots + 0,002 \cdot \log_2(0,002) \approx 0,1198$$

The text entropy accounting for six-letter combinations:

$$H_6 = H\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5(a_6) = H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5a_6) - H(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5) = -0,008 \cdot \log_2(0,008) - 0,006 \cdot \log_2(0,006) - \dots - 0,002 \cdot \log_2(0,002) + 0,008 \cdot \log_2(0,008) + \dots + 0,002 \cdot \log_2(0,002) \approx 0,0657$$

In the process of our study, when computing the number of repeating different letter combinations in the Kazakh journalist text, we came to the following indicators (in bits):

$$H_1 \quad H_2 \quad H_3 \quad H_4 \quad H_5 \quad H_6 \\ 4,4253 \quad 2,7267 \quad 1,0687 \quad 0,3301 \quad 0,1198 \quad 0,0657.$$

Thus, the further computations of the texts from one- to six-letter combinations for Kazakh and Russian are not identical. Based on the computations carried out we can suppose that in the journalistic texts of both languages with increasing information there takes place the reducing of the uncertainty degree (entropy). Entropy in Kazakh and Russian is equal (in bits):

In Kazakh:

$$H_1 \quad H_2 \quad H_3 \quad H_4 \quad H_5 \quad H_6 \\ 4,4253 \quad 2,7267 \quad 1,0687 \quad 0,3301 \quad 0,1198 \quad 0,0657$$

In Russian:

$$H_1 \quad H_2 \quad H_3 \quad H_4 \quad H_5 \quad H_6 \\ 4,3742 \quad 3,0423 \quad 0,7895 \quad 0,5605 \quad 0,0451 \quad 0,0108$$

Thus, an important result of the study is the fact that the information-entropic analysis of the text structure was carried out based on Shannon's information entropy using the classical definition of probability. As a result of the studies we established that any language text, from a

unit word to a large literature work, can be presented as a system whose elements are individual letters and its parts present the same letters combinations. Respectively, with the help of the synergy theory of information there can be carried out a structural analysis of arbitrary texts on the side of their randomness and order by the number and frequency of the letters. We established that with the transition to a higher level of organization there takes place the reducing of the text entropy.

Using a text as a universal model, we can establish the limits of changeability in which there can be performed linguistic systems self-organization and development.

CONCLUSION

Thus, resting upon the own studies of the text and the studies of modern linguists, such as R.G. Piotrovskiy, I.A. German, V.A. Pishchalnikova, G.G. Moskalchuk and others, we came to the conclusion that the text can be analyzed from the position of linguistic synergy. The linguistic synergic approach to the text structure modeling as a self-organizing object is necessary for the further studying of philology. Based on using the arsenal of linguistic methods and conceptions, it is necessary to find algorithms permitting, with the help of theory of information and mathematical computations, to carry out deep analytical studies in the field of language, particularly, in organizing the text structure in its various aspects.

ACKNOWLEDGEMENT

The article has been prepared within the frames of the state grant for budget program No 1013.

REFERENCES

1. Kohler, R., 2004. Synergetic Linguistics // The Encyclopedia of Language and Linguistics / Ascher R.E. (Ed.). – Oxford: Pergamon Press, 8: 4454-4455.
2. Haken G. and M. Synergy, 2005. Mir, pp: 404.
3. Prigozhin, I., 1985. From the existing to the occurring: Time and complexity in physical science. M.: Nauka, pp: 327.
4. Altmann, G., 2005. On the Dynamic Approach to Language // Linguistic Dynamics: Discourses, Procedures and Evolution / Ballmer T.T. (Ed.). – B.; N. Y.: Mouton de Gruyter, pp: 181-189.
5. Wildgen, W., 1990. Basic Principles of Self-Organization in Language // Synergetics of Cognition. L.P., Tokyo, Hong Kong: Springer-Verlag Berlin Heidelberg, pp: 415-426.
6. Brown, C. and G. Yule, 2008. Discourse Analysis. Cambridge.
7. Schiffrin, D., 2004. Approaches to Discourse.- Oxford; Cambridge, MA.
8. Piotrovskiy, R.G., 1996. Linguistic synergy // Scientific-technical information. Ser. 2. Information Processes and Systems, pp: 12.
9. Belousov, K.I., 2002. Text form in activity expression (theoretical-experim.study): Diss.... cand. of Philology. Kemerovo, pp: 19.
10. Moskalchuk G.G., 2010. Text structure as a synergetic process. 2-d Ed.-M.: URSS, pp: 296.
11. Valgina, N.S., 2003. Theory of text. - M.: Logos, pp: 279.
12. Shannon, K.E., 1963. Mathematical theory of communication // Works on theory of information and cybernetics. – M.: IL, pp: 243-332.
13. Ulyankina Yelena, 2012. The paper is closer for the official// Novy vestnik, 23(625): 13.
14. Zharymbetova, N., 2012. There are needed correct decisions for agriculture // Yegemen Kazakhstan, pp: 22.