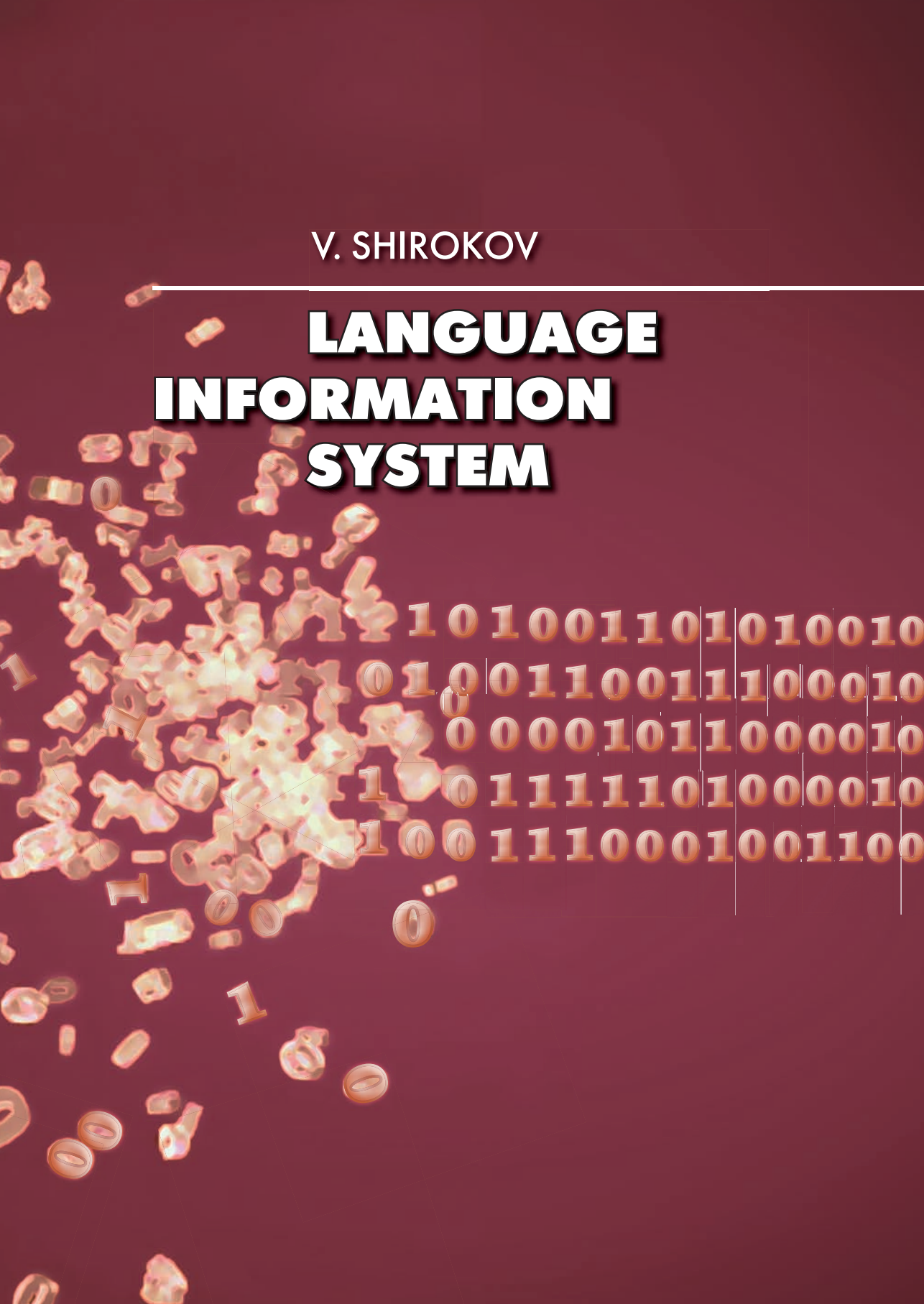
V. SHIROKOV

# LANGUAGE INFORMATION SYSTEM



NATIONAL ACADEMY OF SCIENCES OF UKRAINE
UKRAINIAN LINGUA-INFORMATION FUND
OF THE NAS OF UKRAINE

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
УКРАЇНСЬКИЙ МОВНО-ІНФОРМАЦІЙНИЙ ФОНД
НАН УКРАЇНИ

В.А. ШИРОКОВ

# МОВА ІНФОРМАЦІЯ СИСТЕМА

*ПРОЄКТ*
*«УКРАЇНСЬКА НАУКОВА КНИГА*
*ІНОЗЕМНОЮ МОВОЮ»*

КИЇВ
АКАДЕМПЕРІОДИКА
2021

V. SHIROKOV

# LANGUAGE INFORMATION SYSTEM

*PROJECT*
*«UKRAINIAN SCIENTIFIC BOOK*
*IN A FOREIGN LANGUAGE»*

KYIV
AKADEMPERIODYKA
2021

https://doi.org/10.15407/akademperiodyka.451.160
UDK 81:004
S54

**Reviewers:**

*Anatolii P. ZAHNITKO*, Corresponding Member of the NAS of Ukraine, Professor, Dr. Sci. (Phil.)

*Oleksandr V. PALAGIN*, Academician of the NAS of Ukraine, Professor, Dr. Sci. (Engin.)

*Approved for publication by Ukrainian Lingua-Information Fund of the NAS of Ukraine (September, 3, 2020, Protocol No. 6)*

*The publication was funded within the framework of the Targeted Complex Program of the NAS of Ukraine "Scientific Bases of Functioning and Providing for Conditions for the Development of the Scientific and Publishing Complex of the NAS of Ukraine"*

**Shirokov V.**

S54 Language. Information. System / V. Shirokov; Ukrainian Lingua-Information Fund of the NAS of Ukraine. — Kyiv: Akademperiodyka, 2021. — 160 p.

ISBN 978-966-360-451-0

The monograph deals with the presentation of fundamental information problems of linguistics and is the result of many years of research and observations of the author in the field of information properties of natural human language. These studies have led the author to the need to penetrate deeply into the phenomenology of cognitive-communicative processes based on natural language, study their information properties and build on this basis formalized linguistic models focused on very specific applications.

The development of these ideas in the modern civilizational and evolutionary context stimulates the development of the ideas of quantum linguistics, which is currently in the field of scientific forecasts.

**UDK 81:004**



**ISBN 978-966-360-451-0**

# CONTENT

# PREFACE

The world is evolving. It joined the network or, as it is called, the network-centric phase of evolution. World Network has become an environment where the cognitive-communicative scenario of the development of civilization is playing. An economy based on knowledge, every hour, and every second requires a tremendous amount of new scientific-technical, business, operational and other information the production of which is also included in the overall cognitive-communicative process.

According to the well-known company IDC, the total global volume of created and replicated data amounted to more than 1.8 zettabytes (18 trillion GB) in 2011. According to IDC forecasts the amount of data on the planet will be doubled (at least) every two years until 2020. If this trend is considered fair, and at the moment there is no serious reason to doubt it, then now the world data adapted the web of data has exceeded the level of $10^{14}$ or even $10^{15}$ GB. If, following A.N. Kolmogorov (1987)[1], we would consider information as the periphrasis of complexity concept, then given above fact clearly confirms the basic empirical rule (necessary condition) of the general evolution theory: *the system evolves when its complexity increases*. In other words, *the complexity of the evolving system is always growing*.

However, there is a legitimate doubt that such large amounts of information can be effectively processed, and most importantly — adequately perceived and understood by their recipients in real-time. After all, the psycho-mental-physiological human nature, including the ability to perceive and process information is unlikely to have changed significantly since the days of Adam. The basic cognitive tract of a human, namely: «*perception → sensation → experience → awareness → understanding → reflection → reaction*»[2] (Palagin, Shirokov, 2000) continues to act like thousands of years ago, though, so to speak, civilization information sur-

[1] Kolmogorov A.N. Tri podhoda k opredeleniyu ponyatiya «kolichestvo informacii». *Teoriya informacii i teoriya algoritmov*. Moscow: Nauka, 1987. S. 220.

[2] Palagin A.V., Shirokov V.A. Principles of cognitive lexicography. *Informational theories & application*. 2000. Vol. 9, No. 2, P. 43—51.

roundings have changed radically. The exponential growth of information[3] is not accompanied by an adequate increase of human capabilities of its assimilation and the more effective use.

Thus, there is the main contradiction of the modern era: *evolution law requires increasing amounts of information produced and perceived by humanity, and society, considered as the sum of the individuals and the relationships between them, due to the peculiarities of human nature is unable to take advantage of these volumes and is forced to restrict its production, processing, and usage*. Evolution as if by itself «includes» the mechanisms of its inhibition, which sometimes show themselves in very dramatic forms. Catastrophic scenarios of the development of civilization, the hypothesis of a technological singularity, etc. become increasingly popular[4].

The urgency of those challenges has stimulated new approaches to handling large amounts of information and led to the concept of «*Big Data*», which is symbolically characterized as the «three *V*», namely: *Volume* (volume — petabytes of stored data), *Velocity* («speed» — data receiving, transformation, loading, analysis, processing and reaction in real-time) and *Variety* («diversity» — the processing of structured, semi-structured and unstructured data of different types). Recently to «three *V*» two more were added: *Veracity* — the accuracy of the data, which users have begun consider to be of greater importance, and *Value* — the value of the extracted and stored information. Now Big Data is recognized as the second most important (after virtualization) trend in information technology infrastructure, which is considered more powerful than even energy efficiency and monitoring.

These features of the present stage of a Network development set up a problem of creating tools that can «take» a considerable part of the primary functions of the cognitive human tract. Thus the problem of intellectualization of the Network steadily moved to the forefront of relevance.

Among the many aspects of intellectualization, we select the language. Indeed, among a large number of definitions of artificial intelligence, there is also the following one: «*Artificial intelligence — a form of individualization of*

---

[3] Muzyakov S. Informacionnaya sreda i usloviya eksponencial'nogo rosta ob'ema znanij v sovremennom obshchestve.*Vlast'*. 2012. No. 4. S. 42—46.

[4] Vinge Vernor. The Coming Technological Singularity:How to survive in the post-human era. *Vision-21. Interdisciplinary Science and Engineering in the Era of Cyberspace. Proceedings of a symposium consposored by the NASA Lewis Research Centre and the Ohio Aerospace Institute and held in Westlake.* (March 30—31 1993, Ohio); Kurzweil Ray. TheSingularityIsNear: When Humans Transcend Biology. Viking. 2006. 652p.; Panov A.D. Tekhnologicheskaya singulyarnost', teorema Penrnouza ob iskusstvennom intellekte i kvantovaya priroda soznaniya. URL: http://www.intelros.ru/pdf/metafizika/2013_3/7.pdf; Medvedev D.A. Sverhtekhnologii i obshchestvo v XXI veke. Kurs lekcij [Videolekciya 10. Tekhnologicheskaya singulyarnost']. 2008. URL: http://univertv.ru/video/sociologiya/

*technical systems that have linguistic status*»[5], and therefore *intelligence itself should be considered as a form of individualization of systems that have a linguistic status*.

Linguistic support of the Network is a leading factor and primary interface for Human-Network and Human-Human interactions *through* the Network. This situation will occur soon. So, the problem of combining («integrating») ideas and technologies of virtualization, Big data, and intellectualization (mainly through the mechanisms of natural language) is becoming increasingly important. But the peculiarity of linguistic means is that languages exist only in the form of separate national languages, whose speakers, in fact, able to develop linguistic software components at the appropriate level and acceptable quality. The basis of such developments are well codified, annotated and representative models and arrays of linguistic data that represent all (ideally) aspects of the functioning of a language, both in cognitive and communicative terms. Such means are lexicographic systems provides a far-reaching generalization of the dictionary concept[6], and linguistic corpora[7], which operates the large digital arrays of linguistically qualified and annotated texts.

It should be noted thatthe creation of the lexicographic systems and linguistic corpora, which are the basis for the further development of high-quality means of linguistic support itself is a quite complex and extremely time-consuming task requiring a significant number of qualified professionals. However, many aspects of these works at this stage is not subject to automation and must be performed, so to speak, manually. Moreover, the problems of developing theoretical tools that can form the necessary basis for the formalization of intellectual properties of language and the creation of effective models of intellectualization are actualized. However, the solution to this problem, as it often happens in science, meets the need of revision even the original principles of the description of the studied objects to bring them into line with those problems, which should be solved at this new stage.

A quarter of a century ago author has distinguished a class of objects that he qualified as «language-information systems». The name of this class includes the main concepts put into the title of the book, namely: language, information, and system. We at various times and on various occasions explored many aspects of this triad. At the same time, in a compact monographic form, they have not yet been collected and submitted in a unified form. This book aims to fill this gap.

---

[5] Shirokov V.A. Fenomenolohiia leksykohrafichnykh system. Kyiv: Nauk. dumka, 2004. S. 20. (Admittedly, however, this definition remains the concept of «language status» unexplained. As always, the devil is in the details!).

[6] Shirokov V.A., Buhakov O.V., Hriaznukhina T.O., Kostyshyn O.M., Kryhin M.Yu. Korpusna linhvistyka. Kyiv: Dovira, 2005. 472 s.

[7] Shirokov V.A. Informatsiina teoriia leksykohrafichnykh system. Kyiv: Dovira, 1998. 331 s.

The basis of our research from the very beginning was based on an information approach to the description of the linguistic phenomenology. Considering the language as a system for implementing a special kind of information processes, we applied the information-informational and linguistic ideas of the outstanding Russian mathematician A.N. Kolmogorov to develop our concept of the scientific description of a natural language.This approach led us to formulate a fairly universal phenomenological principle, which we called *the lexicographic effect in information systems*, which clarifies the mechanism of generation in the language continuum discrete complexes of linguistic units representing certain aspects of the interaction of linguistic forms and corresponding contents.

The evolution of ideas about the nature of language led to certain agreements that we took as a basis for the presentation. These agreements are a set of postulates, peculiar linguistic «presumptions», which the author (explicitly or implicitly) is supposed to follow in this book.

The list of marked presumptions is assumed to be open, and, therefore, the book itself represents only certain «prolegomena» to the regular presentation of the information concept of constructing the theory of language. At the moment in the list of linguistic presumptions, the author has 7 sentences, namely, the following[8].

## List of linguistic presumptions

1. Any thought can be expressed in a natural language form.

2. A natural language is a tool that provides a variety of information transformations.

3. Transformations of information in a language are carried out in two main aspects — cognitive and communicative.

4. Between any two languages, it is possible to establish equivalence in the meaning of propositions 1—3.

5. Mathematics is a kind of language.

6. Language forms a system in the sense that will be clarified below.

7. The main system-forming relationsof language are the following: «subject — object» and «form — content».

More broadly, the essence of the above presumptions is set out below. However, it should be noted, that even the approach to their formulation crystallized during attempts to answer the following questions:

1. What is a «system» in linguistics, and how it relates to the concept of «system» in general?

[8] We are deeply grateful to Professor Violetta Koseskqa-Tosheva from the Institute of Slavic Studies of the Polish Academy of Sciences for the informative and fruitful discussions of these linguistic presumptions, thanks to which they acquired the published form.

2. What is the relationship between information and language?

3. How «Form» and «Content» should be interpreted in language? What are the linguistic (extra-linguistic) form and linguistic (extra-linguistic) content?

4. What are the boundaries of the formalization of the language system?

5. What is the possible expansion of the boundaries of the conceptual description of the language system?

6. What is the relationship between the «grammatical» and «lexicographic» way of describing the language system? Is it possible to combine them and if so, how to do it?

The author must admit that at the moment he does not have complete and, perhaps, even quite distinct answers to these questions, answers that could completely satisfy the readers. Deliberately referring to some internal inconsistency of the text, the author turns to the reader with a request for condescension, especially since it (the contradiction), if desired, can very well be attributed to factors that stimulate the development of scientific ideas about the nature of language.

# INTRODUCTION TO INFORMATION DESCRIPTION OF LINGUISTICS: INFORMATION AND LANGUAGE

## Presumption No. 1

The first of our linguistic presumptions we formulated as follows: «Any thought can be expressed in the natural language form. «This not irreproachable, and possibly doubtful from the general philosophical point of view, a statement we ask to perceive as an invitation to consider the following topics: what is thought?, what is language (specifically, natural)?, what is the form of the language and how does it display the content (particularly, the «thought»)?

It is possible to formulate the questions a little differently, in particular. The «substance» of thought, as we believe, *can be* mapped to the «substance» of the language, i. e. the «substance» of thought has a fairly adequate expression in linguistic forms. This means that the language has potencies and corresponding internal system resources to generate reliable and confident forms in which the thought is expressed.

On the other hand, it (the language) discloses its content: a) in extralinguistic reality as a display of mental (or psychomental, and also sensory-perceptual) processes and b) the actual linguistic reality or linguistic system, as will be discussed below.

## About information doctrine, which the author considers close to his ideas about the world and language. Or, in other words, information about his world outlook

The author had announced an informational approach to his conception of language in the title of the book. Moreover, Presumption No. 2 states that «A *natural language is a tool that provides a variety of information conversions*». The author, sharing a belief about the informational nature of a natural language, considers that it is necessary to speak out about the main concepts of the nature of the information which the author considers relevant to the subject matter.

Information (from lat. informatio — awareness, explanation, description) — in a broad sense is an abstract concept with many meanings depending on the context. In the narrower sense, it is data (messages, data) regardless of the presentation form.

Of course, the concept of «information» is interpreted and used differently in different scientific disciplines. In each discipline, the concept of «information» is connected to various worldviews and different conceptual paradigms. At the same time this concept extensively used in the traditional, everyday, common sense[9] — it is information, knowledge, reports about the situation that a person perceives from the outside through senses (sense of vision, hearing, taste, smell, touch, ...), and, of course, the instruments through which he makes external world observations and monitoring.

Information can be stored, transferred, and processed in various forms, including symbolic (sign) form. The same information could have different forms of representation that are implemented through certain sign systems, constructed from some basic elements («*alphabets*»), with rules to perform operations over them.

The information process requires at least three elements: the **source** of the information, its **recipient,** and the **communication device** providing the **delivery** of information from **source** to the**recipient**.

Thus, information is always information about something, i.e. about the corresponding object, which is its source. Therefore it would appear reasonable that the information is generic, a universal feature of all things and processes — there is no «entity», not possessing the information. According to such authority as A.N. Kolmogorov[10], information exists objectively, regardless of whether someone perceives it or not, although it appears only in the process of its perception. Information reflects such universal properties of things and processes as *structure, complexity, heterogeneity, patterns of change, states of an object*, etc.

Note that in our previous works[11] an overview of various opinions, views, approaches, concepts, and information theories is presented. So, here we will not analyze this subject. We will try to outline some general natural parameters and laws governing the informational properties of systems, especially in connection with such fundamental concepts as matter, energy, and knowledge — it turns out that information is very closely connected with these concepts. For this, we need more or less clear, correct, and metaphor-free definitions of the noted concepts.

---

[9] Google on request «information» (February 24, 2016, when this text was written), gave no less than 8 560 000 000 links, the number greater than the number of the inhabitants of the Earth at that time. But on September 18, 2020, when this text was being translated into English, Google's response to the request for «information» had already reached an incredible 16 180 000 000 (!).

[10] Kolmogorov A.N. Teoriya peredachi informacii.*Sessiya Akademii nauk SSSR po nauchnym problemam avtomatizacii proizvodstva* (15—20 oktyabrya 1956): Plenar. zasedaniya. Moscow: Izd-vo AN SSSR, 1957. S. 66—99.

[11] Shirokov V.A. Informatsiina teoriia leksykohrafichnykh system. S. 19—50; Shirokov V.A., Buhakov O.V., Hriaznukhina T.O., Kostyshyn O.M., Kryhin M.Yu. Korpusna linhvistyka. S. 105—120.

This is necessary because recently, as a result of the general tendency to freely use scientific knowledge, these concepts are operated quite arbitrarily, especially by people who are far from science. Therefore, we first give basic definitions, that we will follow. Let us make a few comments, so to speak, of a physical and metaphysical nature, which will help to enter the circle of concepts relating to the subject of consideration.

Firstly, it is intuitively clear that the concept of knowledge corresponds to the concept of information, but they are quite different from each other.

Information, as noted above, is an objective characteristic of objective phenomena and processes. In our opinion, there is a deep analogy between the definitions of the concept of information, on the one hand, and energy, which is also some kind of objective characteristic of things, on the other. And although linguists are not directly involved in the study of the energetic properties of things, we consider this analogy to be very useful, based on general methodological and general cultural prerequisites. The noted analogy is very easy to trace, referring to the Table 1.

So, we can see certain parallelism in the determination of the properties of the concepts of energy and information. At the same time, we emphasize that these concepts manifest completely different properties and characteristics of objects, which can be seen from paragraph 1. given in the comparison Table. However, despite the noted essential differences, there are very significant connections between them. For example, a remarkable property of information is that to obtain it, it is necessary to expend a certain amount of energy. It turns out that the minimum energy required to obtain one bit of information is calculated by the formula[12]:

$$kT \cdot ln\ 2 \text{ erg}, \tag{1}$$

where $k = 1.38 \cdot 10^{-16}$ erg/K° is the Boltzmann constant, and $T$ is the absolute temperature of the system in which information is generated. At 300 K°, that is, at normal temperature, this value is $4 \cdot 10^{-14}$ erg or approximately $10^{-2}$ electron volts. This is a very small quantity. For example, the minimum amount of energy required to obtain 100 terabytes of information (approximately the same amount is contained in 50 billion pages of printed text) is approximately 1 erg. We still do not know the real information processes working with such tremendous efficiency!

In a certain sense, there are reverse processes, namely those where, figuratively speaking, information can be converted into energy. The idea of the possibility of such processes arose a long time ago in connection with a discussion of the so-called «Maxwell demon» paradox, which is an example of a mechanism that, in a sense, converts information into energy, i.e. uses information as a «fuel».

[12] Vol'kenshtejn M.V. Teoriya informacii i evolyuciya. *Kibernetika zhivogo: biologiya i informaciya*.Moscow: Nauka, 1984. S. 45—53. Vol'kenshtejn M.V. Entropiya i informaciya. Moscow: Nauka, 1986.190 s.

*Table 1.* **Comparison of the properties of the concepts «ENERGY» and «INFORMATION»**

| ENERGY | INFORMATION |
|---|---|
| 1. Energy is a universal quantitative characteristic of physical systems. There are no physical systems that are not characterized by energy | 1. Information is a universal quantitative characteristic of any system. There are no systems that are not characterized by information |
| 2. The indicated value characterizes the intensity of the processes occurring in physical systems | 2. The indicated value characterizes the complexity of the processes occurring in the systems and the complexity (heterogeneity, structuredness) of the systems themselves |
| 3. The amount (quantity) of energy (both measured in an experiment, and theoretically calculated) is represented by numbers | 3. The amount (quantity) of information (both measured, and calculated theoretically) is represented by numbers |
| 4. Energy dimension: [mass] [length]$^2$ [time]$^{-2}$ | 4. Dimension of information: [bit] or [energy] × [temperature]$^{-1}$ |
| 5. In systems of various nature, energy has different manifestations and characteristics. They are reflected in the methods of experimental observation and measurement of energy effects, as well as in the methods (models) of their theoretical description | 5. In systems of different nature, information has various manifestations and character. They are reflected in the methods of experimental observation and measurement of information effects, as well as in the methods (models) of their theoretical description |
| 6. Various manifestations of energy for different systems and levels of consideration are usually called forms of energy (mechanical, electrical, magnetic, etc.). The fundamental property is the conservation law: in a closed system, all processes are occurred so that energy, turning from one form to another, remains constant | 6. Various manifestations of information for various systems and levels of consideration are usually called forms of information. Information exists in the form of data, texts, knowledge, models, etc. Information processes are accompanied by transformations of information from one form to another |
| 7. Different levels of matter are characterized by their specific methods of energy description, represented by the appropriate theoretical models and mathematical formalisms | 7. Different types of systems are characterized by their specific methods of information description, that are presented by the appropriate theoretical models and mathematical formalisms (for example, the Hartley, Shannon, and Kolmogorov models) |

An example of such a process is described in the book[13]. This effect gives reason to believe that similar processes occur in sociotechnical systems; they are described in detail in the works[14].

[13] Stratonovich R.L. Teoriya informacii. Moscow: Sov. radio, 1975. 423 s., gl. 12.

[14] Shirokov V.A. Informatsiino-enerhetychni transformatsii ta informatsiine suspilstvo. *Nauka, innovatsiia, informatsiia*. 1996. No. 1. S. 48—66; Shirokov V. A. Informatsiina teoriia leksykohrafichnykh system; Shirokov V.A. Fenomenolohiia leksykohrafichnykh system.

Thus, we note that information processes do not constitute a closed system and very quickly lead to another area of describing the properties of objects — an energy one. This prompts specialists involved in the informational properties of the language to gradually realize the insufficiency of the informational approach alone, even in their own, predominantly informational field. And, in our opinion, as we get deeper into the properties of language, the ideas, facts, and methods of other sciences will be used more and more often in linguistics. This shows the meaning of transdisciplinarity — the main defining feature of the modern development of science.

Note that not any information can be converted into a useful resource. In the particular case, this process, figuratively speaking, can implement even one of Maxwell's information demon. In real situations, for this purpose, whole complexes of organizations and institutions are created (following Selfridge[15], we will use the name «pandemoniums» for them), and they ensure the production, repeated transformation of information and its transactions until it enters the socio-technical system at the right time, and in the right form, adapted to the perception of the said system. It is the processes of production and targeted multiple conversion and transportation of information that gives it a new quality that allows us to qualify it as *knowledge*. Therefore, now we can give such a working definition of the concept of knowledge:

*knowledge is information, the form of which is the carrier of transformations that adapt its reception by the social system.*

We emphasize that this definition does not claim to be absolute and it is working, even technical. Note that it completely lacks constructions such as «the correct display of the laws of the world in the head of a person», characteristic of philosophical definitions of knowledge. Our definition, we believe, is as objective as possible, since it contains an objective concept of information and a fairly clear idea of the role of its transformations in the social system. The absence of references to «correctness» and «laws of the world» we consider completely justified, because, as is it known, there are false knowledge, as well as the knowledge that does not reflect the laws of the world. Despite such lapidary of our definition, quite concrete conclusions follow from it.

First, *if information* as such represents certain *objective* properties of things, then *knowledge* carries the potential of the *subjective*. Indeed, before getting into the production and economic system, the information should become a fact of consciousness: first individual, then collective, or at least group. Moreover, it, of course, undergoes some transformations following the specifics of the functioning of consciousness. And this specificity is such (and such is the construction of

[15] Selfridge O.G. Pandemonium: a paradigm for learning / Mechanisation of thought processes. London. HMSO, 1959. P. 511—531.

the human intellectual apparatusaccordingly), that numerous transformations and interactions between mental and linguistic structures take place in it.

Studies of great psychologists and linguists, among which we mention Wilhelm von Humboldt, V. Bekhterev, L. Vygotsky, N. Chomsky, R. Shenk, and many others, have convincingly confirmed that any mental process in a person has its reflection in the linguistic sphere. So it's correct to speak of a single *thought-speech* process. We also emphasize the important role of specialization of the linguistic subsystem as the main communicative tool in human society.

From the foregoing follows one more rather simple definition of knowledge *as information verbalized according to the laws of the language system*. Indeed, in a linguistic form, such important aspects of human life as the cultural code, the scientific and linguistic worldviews, the linguistic consciousness and subconscious (collective and individual), etc. are compressed.

And here, a scientific or contextual-objective picture of the world, which is usually primary in information, turning into socially significant knowledge, actively interacts with the language system and «picks up» from it the properties inherent in this system. In general, linguistic and contextual-subject components comprise two completely different aspects of the functioning of knowledge. Moreover, it is the linguistic form that prevails on a sufficiently large segment of the knowledge functioning, up to the act of its direct consumption by society. It should be noted that the linguistic form, acquired by knowledge, is quite convenient and flexible in itself.

Firstly, it is universal: we are convinced that almost any information can be verbalized — this is reflected in our first Presumption.

Secondly, each person is quite qualified to use this form, assimilating it from early childhood.

Thirdly, the linguistic form is linear, as a result of which it is easily amenable to various information operations — coding, transformation, storage, transportation, etc.

But the flip side of this convenience and flexibility is that in the natural-language form of information only elements of the *form* of the language system are more or less explicitly presented, while *ontological* or *semantic* components are presented only implicitly. It logically follows from this that the «extraction» of semantic information from verbalized knowledge, that is, a natural language text, inevitably leads to the application of the principles of the language system, so to speak, in the «opposite» direction. Namely, if we symbolically depict the process of verbalization of information, as:

$$L\, I = T, \tag{2}$$

where $I$ is information, $L$ is the «converter» of information into a language object, $T$ is ф text, then the extraction of semantic information will look like this:

$$L^{-1}T = I^{\,\mathrm{sem}}. \tag{3}$$

This last process we call the extraction of knowledge from the text.

Formulas (2)—(3) are much easier to write than to organize the corresponding processes represented by these formulas. Recently, there has been a significant increase of scientific publications on this topic. Many of them apply to methods of automated construction of knowledge models by text in a natural language. However, there are no particularly outstanding successes in these areas.

Firstly, it should be noted that all methods are very dependent on the design of the language system of a particular national language in which the input information is presented. Therefore, to switch to text analysis for each language, it is necessary to carry out certain stages of the analysis of language structures that are implicitly contained in the constructions of the $L$ and $L^{-1}$ operators. Such steps, in particular, are analysis of sign systems, grammatical analysis of language structures, contextual, syntactic, semantic, pragmatic, statistical analysis, etc. The creation of effective automatic procedures focused on the implementation of the indicated varieties of analysis is a rather complex scientific and technical task.

The set forth above approach to information and knowledge was formulated in our work «Noospheric dimensions of information and knowledge»[16].

As of now, there are software tools that can to some extent perform operations with texts described above. The most common among them are search engines, in particular, Internet ones, capable of performing several natural language functions. However, these tools do not yet bring complete satisfaction to users. On the contrary, everyone knows how «noisy» the search tools of the Internet are. Our experience shows that such tools and, accordingly, technologies should be comprehensive and combine many conceptual paradigms because it is hardly worth hoping that one scheme can cover the whole variety of cognitive situations that arise during the intellectual processing of texts for the extraction of knowledge from them.

The information approach to the language requires the involvement of the most formalized ideas about the nature of information in its relation to the nature of the language. Among such ideas, an important place is occupied by theories that make it possible to qualify and evaluate information from a quantitative side. Currently, the most common are measures of the amount of information, the introduction of which is associated with the names of Hartley, Shannon, and Kolmogorov.

The formal object of R. Hartley's information measure is a finite set consisting of $N$ elements. In 1928, R. Hartley proposed a logarithmic measure for measuring the amount of information of discrete messages:

$$I = log_2 N, \tag{4}$$

where $I$ is the amount of information received, and $N$ is the number of possible (equally probable) outcomes or various messages that can be received from a

[16] Shirokov V.A. Noosfernye izmereniya informacii i znaniya. *Noosferologiya: Nauka, obrazovanie, praktika* / Red. O.A. Gabrielyan. Simferopol': TNU im. V.I. Vernadskogo. 2008. 463 s.

source of discrete messages with an alphabet of $M$ letters with a message length of $n$ letters.

In 1948, an American engineer and mathematician C. Shannon suggested a formula for calculating the amount of information for events with different probabilities. According to Shannon, the amount of information $I$ for a system of $N$ possible events, where $p_i$ are the probabilities of individual events, the amount of information for events with different probabilities is determined by the formula:

$$I = -\sum_{k=1}^{N} p_k \log_2 p_k. \quad (5)$$

Here we will not consider examples and interpretations of the application of information measures of Hartley and Shannon, sending the reader to numerous publications on this topic. We will quote only a fragment from the book of Umberto Eco «The Open Work»[17], in which the author tried to apply the ideas of information theory to the study of literary works, in particular musical and literary ones. In this book, there is a chapter «Openness, Information, Communication», which contains sections «Information Theory», «The Concept of Information in Wiener Theory», «The Difference between Meaning and Information», «Meaning and Information in Poetic Communication», «Information Transfer», «Poetic Speech and Information», «Correlation of Information Theory With Musical Speech», «Information, Order and Disorder», «Information and Psychological Transaction», «Transaction and Openness», «Information and Perception».

Let us cite a rather long excerpt from this chapter, namely, from the fragment «*Note of 1966*», placed after the section «Information, Order, and Disorder». It seems to us that in this fragment U. Eco, who used information concepts close to the views of R. Hartley, N. Wiener, and C. Shannon, most clearly expressed his attitude to the possibilities and prospects of applying information approaches to the study of texts (literary and music). So (quote):

«It's quite simple to show that this theory[18] did not appear to explain the nature of the poetic message and that it does not apply to processes in which denotative and connotative meanings come into play: it is so simple to show that no one can object.

However, precisely because the theory of information is simply *not applicable* to an aesthetic phenomenon, many researchers are trying to use it in this area. Precisely because it does*not apply* to significative processes, they are trying to use it to explain linguistic phenomena. And finally, precisely because these concepts, perceived in their original meaning, have nothing to do with works of art, in this essay we seek to find out to what extent they can be applied to it. If they were ap-

[17] Eco Umberto. Opera aperta. Forma e indeterninazione nelle poetiche contemporanee. Milano: Bompiani, 1967.

[18] Information Theory (auth.).

plicable immediately, it would not be worth the time to find out the possibilities of this application. On the contrary, we proceeded from the fact that in the framework of the theory of communication, when its mechanism (and verification is necessary here) *must* be reduced to actions common to any communicative mechanism, even to those that involve the simple transmission of signals without connotative meaning through one channel and perceived by a machine that understands them as instructions for further actions based on a pre-set code and can fix a one-to-one relationship between a given signal and a given mechanical or electronic action.

On the one hand, the above objection would have been fully valid if the following points had not been clarified:

1) The application of concepts related to information theory to the field of aesthetics is not the reason for designing the idea of an open, polysemous, ambiguous work. In fact, on the contrary, it is the presence of a certain share of ambiguity and polysemy in each work of art that makes us believe that the categories of computer science are to some extent suitable for recognizing this phenomenon.

2) The application of informatics categories to the phenomena of communication is now perceived as something universally recognized, and many researchers, starting from Jacobson, who connects the idea of binary with the phenomena of a language, to Piaget and his students, who apply the concept of information to the phenomenon of perception, Levy-Strauss, Lacan, Russian semiologists, Max Benze, the new Brazilian criticism, etc. When such an interdisciplinary meeting takes place, so rich in representatives of various directions from many regions of the world, you feel that there is more to it than a skillfully distributed fashion or thoughtless extrapolation. There is a categorical device, which is perceived as a key that can open many doors.

3) Even if we were dealing with an irrepressible search for analogies with uncontrolled extrapolations, we still have to admit that cognition is also made possible through the efforts of the imagination of a building hypothesis, that imagination that dares to go in the shortest and, perhaps, not entirely reliable ways. Excessive severity and the most justified caution can lead away from obviously dangerous roads, but these roads could lead to the plateaus, from which the overall picture would look clearer with all its interchanges and highways, which went unnoticed during the first topographical study.

4) The categorical apparatus of information theory seems methodologically cost-effective only when it is included in the context (however, researchers understand this only gradually, and only recently) of general semasiology. Before rejecting concepts related to computer science, we must test them in the light of semasiology».

As you can see, in this fragment, U. Eco gives some skeptical arguments regarding the usage of information theory to study the subtle effects of meaning the carriers of which are fiction texts. At the same time, his reference to semasiology

and the installation on verification of precisely the seismological properties and potentials of information categories seems very valuable. Here, his insight was revealed by the researcher, who intuitively felt that the cognitive potential of such a fundamental concept, which is information, far exceeds the capabilities and limits of those formal models and information theories that were built at the time of writing of his book.

## A. Kolmogorov's algorithmic complexity and information measure associated with it

It now seems to us that many fears of U. Eco and the prejudices of other researchers regarding the possibilities of using information ideas and methods are largely overcome in A. Kolmogorov's algorithmic theory of information, based on the formalization of the concept of complexity[19], when considering the latter from a somewhat non-standard perspective. Since our approach is close to the information concept of A. Kolmogorov, we will proceed immediately to his elementary and popular presentation. We believe that it is in Kolmogorov's approach that the informational properties of objects are presented in a compact form, which we perceive as the closest to the properties of a natural language, considered by our Presumptions (second and third) as an information transformer.

The introduction of the Kolmogorov informational measure aims to clarify the concept of information, firstly, without involving the probabilistic approach characteristic of C. Shannon's informational measure, and secondly, by providing the possibility of applying this measure to individual objects, and not just their ensembles.

The fundamental concepts on which this approach is based are the concepts of a ***constructive object*** and its ***description***. We consider these concepts first in a non-rigorous, intuitively accessible wording, «in simple terms» as they say.

By a constructive object, we will mean a certain «construction» constructed (assembled, arranged) from certain predetermined elements that belong to some finite set. Elements of this set are not supposed to have a more detailed structure, but of course, there is a way to distinguish them. They can be included in a constructive object an unlimited number of times. By this circumstance, we do not impose the condition of finiteness on the constructive object itself.

An illustrative model of constructive objects are chains of symbols that belong to some finite completely ordered set, which we will call the *alphabet*. Let us write examples of various chains depicting various structural objects:

$$000000000000000000000000000000000000000000000000 \quad (6)$$

$$\begin{gathered}\text{abababababababababababababababababababababababababa} \\ \text{bababababababababababababababababababababababababab}\end{gathered} \quad (7)$$

$$4c1j5b\int 2p0w1x8rx2\aleph y39mg5q85s7\forall ra\surd jfdp \quad aq7ni\oplus i \quad qe9\xi vz \quad (8)$$

[19] Kolmogorov A.N. Tri podhoda...

Under the description of a constructive object, we will understand its mapping into a certain «language». Moreover, this mapping is such that in its image in this «language» we can unambiguously reconstruct the displayed constructive object. Such a mapping itself is not unique — a lot (possibly infinite) of such descriptions can be associated with the same object. The object itself can also serve as a description of itself. For example, we can write the object (6) quite unambiguously as «***0 50 times***», and object (7) as «***ab 50 times***». We see that, even though the second object is twice as long as the first one, the volumes of their descriptions are almost equal. True, we cannot say that these descriptions are minimal, i.e. there are no shorter descriptions of these objects. As for the third object, it is «difficult» to come up with a shorter description than his presentation (we will use the term *auto-presentation*).

From these examples, one can conclude an intuitive idea of the complexity of a constructive object as a method of constructing its shortest description, that is, one whose length is minimal. By the length of the description, we mean the number of characters that make up the description text. If we dwell on this understanding volume of complexity, it becomes clear that the complexity measure, in principle and under certain refinements, is measurable and quantitative. It is also intuitively clear that the measure of the complexity of a constructive object is primarily related to its properties, such as heterogeneity and structure.

Considerations of a similar nature were used as the basis for Kolmogorov's approach. So, according to Kolmogorov, information about a certain object is considered obtained when it is possible to reproduce (reconstruct) this object (an adequate model of the object) from some finite description (a set of features). To strictly define this concept and construct an information measure Kolmogorov had had to apply such fundamental concepts as an algorithm, Turing machine, and recursive function. We also add that Kolmogorov's information theory goes back to the ideas of the theory of computational complexity (complexity of algorithms), which, is the source of interpreting information as a measure of the complexity, heterogeneity, and structure of systems, as well as confidence in the universality of such a feature as complexity, since any system, regardless of its nature, is characterized by a certain complexity, heterogeneity and has a certain structure, even if it is trivial.

The author allowed himself, for rigor and order, to give a mathematical formalization of the concept of complexity and information that is not burdened with details, which we set out here following A. Kolmogorov's work «Three Approaches to Defining the Concept of Information Amount»[20]. At the same time, we assume that the reader has at least a superficial familiarity with the concepts of a set, the mapping of sets, their Cartesian product, as well as the concepts of functions defined on sets (including recursive functions). These are minimal mathe-

[20] Kolmogorov A.N. Tri podhoda...

matical tools that form a language, in which facts that provide a link between the informational properties of the world and the properties of linguistic objects can be quite organically presented. On the other hand, this is confirmed by the *fifth of the linguistic presumptions, treating mathematics as a kind of language*.

Thus, we consider a countable set $X = \{x\}$. A countable set is a set consisting of elements that can be numbered with integers so that different numbers correspond to different elements. As mathematicians say, to establish a one-to-one correspondence (*isomorphism* or *bijection*, *bijective mapping*) between elements of the set $X$ and some infinite subset (or the whole set) of integers. Any countable set has the cardinality of the set of integers or, that is the same, the countable set that is isomorphic to the set of integers. A finite set is isomorphic to some proper finite subset of the set of integers. Assume that there is a one-to-one correspondence between the set $X$ and the set $D$ of binary words starting with «one», in other words[21], let a bijective map (bijection) be given:

$$n: X \to D, \tag{9}$$

such that each $x \in X$ uniquely corresponds to a certain $d = n(x)$, $d \in D$, and vice versa. We assume that:

*1. $n(x)$* is a general recursive function on $D$. Denote the length of the binary word $d \in D$ by $l(d)$, that is, the number of «zeros» and «ones» contained in it. Then $l(n(x)) = l(x) + C$, where $C$ is a constant.

2. There is a unique mapping $\chi: X^2 = X \times X \to X$, so that for $\forall x \in X, y \in X\ \exists z \in X$, and for $z = \chi(x, y) \equiv (x, y)$ and $n(z) = n(x, y)$ is a general recursive function of $n(x)$ and $n(y)$,moreover:

$$l(x, y) \le C_x + l(y),$$

where the constant $C_x$ depends only on $x$. We assume that the isomorphism (9) is established so that the set $X$ is also considered as a set of binary words. Assume that there exists a partially recursive function $\phi(p, x)$ that maps the binary word $x$ to a binary word $y$, and $p$, $p \in D$ is interpreted as an algorithm (or program) that «processes» $x$ into $y$:

$$p: x \to y, \tag{10}$$

while $\phi$ represents the programming method (language). Without loss of generality, we assume that $p$ for a given $x$ is given by a certain binary word. Denote:

$$K_\phi(y|x) = \begin{cases} \min|_p l(p), \text{ if } \phi(p, x) = y \\ \infty, \text{ if there is no finite } p \text{ such that } \phi(p, x) = y. \end{cases} \tag{11}$$

Thus, $K_\phi(y|x)$ is the length of the minimal program $p$ that render $x$ into $y$ by a given programming method $\phi$. This quantity is called the ***complexity*** $y$ to $x$ for

[21] The set $D$ of binary words starting with «one» consists of the following elements: 1, 10, 11, 100, 101, 110, 111, etc. This is a countable set and, therefore, it is isomorphic to the set of integers.

a given $\phi$. On an intuitive level, it is clear that «complexity *y*concerning *x* for a given $\phi$» is associated with some heterogeneity of the object *x*, i.e. with how irregularly placed «zeros» and «ones» in it, how heterogeneous this distribution can be, and cannot be significantly simplified and, therefore, does not allow a shorter description $\phi\ (p, x) = y$. Of course, the dependence of the complexity value on the programming method $\phi$ is a drawback of the described method, but there is a theorem [22] that states the existence of the «best» programming method *A* such that for any partially recursive function $\phi$ the inequality holds:

$$K_A\ (y|x) \leq K_\phi\ (y|x) + C_\phi, \tag{12}$$

where the constant $C_\phi$ depends only on $\phi$ and does not depend on *x* and *y*.

The quantity $K_A\ (y) \equiv K_A\ (y|1)$, «normalized» with the respect to the single element $x = 1$, can naturally be considered as the complexity of the element *y*. The amount of information in the object *x* relative to the object *y* is defined as the difference:

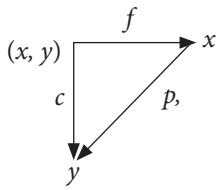
$$I_A\ (x|y) = K_A\ (y) - K_A\ (y|x). \tag{13}$$

The last formula determines the measure of information — so-called algorithmic measure of Kolmogorov's information.

Kolmogorov's information measure is very instructive. Its defining property is relativity. Indeed, we can determine the amount of information that an object contains only by comparing its properties with the properties of another object that we are quite familiar with. The property of «subjectivity» of Kolmogorov's measure is noteworthy. After all, it explicitly contains a certain operator, the program — «subject», which provides a comparison of the properties of objects *x* and *y*. We will return to the discussion of this property.

For the sake of exposition simplicity, as is customary in mathematics, both objects are represented by binary numbers (i.e. «substantially» homogeneous objects), the formulation of the problem of comparing the properties of *x* and *y* does not seem to be too «complicated». This, however, can change dramatically if objects of different substantial nature are chosen as *x* and *y*.

Formula (13) and all thisdescribed approach to determining information through algorithmic complexity can be interpreted in a somewhat different way. Consider a triangular diagram:



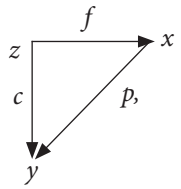

(14)

where $x$ and $y \in X$; the element $(x, y) \in X^2$, and due to the existence of themapping $\chi\ X^2 \to X$, $\chi\ (x, y) = z$, $z \in X$. In the diagram (14), mappings *f* and *c* project the

[22] Kolmogorov A.N. Tri podhoda...

element $(x, y)$ onto first and second factors, respectively, and formulas (11)—(13) and the interpretation of complexity and information measure remain valid.

Assume that the object $z$, which until then represented the image of the Cartesian product $(x, y)$ under the mapping $\chi$, actually is the object of the outside world that is, to some extent, independent of $x$ and $y$. This assumption allows us to construct this diagram:




(15)

where $z \in Z$, $Z$is the source set of objects $z$; with $x \in X$, $y \in Y$. Then the mappings $f$ and $c$ determine certain interpretations of the object $z$, and, also, the mapping$p$ interprets $x$ through $y$. It is natural to assume that object $x$ represents the«formal» properties of the object $z$, provided that the mapping$f$contains some analog of the «communication medium» and «perceptual-sensory apparatus», and $y$ contains its «substantial» properties, and the connection between the «form» and the «content» takes on $p$. The requirement ofminimizing $p$ is completely natural here because the «interpretation» of the form (and its result is the content!) should not contain any «extra» elements random with a respect to $x$ (and also $z$). The construction thus defined, based on the triple of objects $(z, x, y)$ and the triple of mappings $(f, c, p)$ forming the diagram (15), allows a natural interpretation as a construction that is a carrier and, in fact, realizer of the «form-content» relationship.

So we see that the «form-content» relationship in such a peculiar way is «encrypted» in Kolmogorov's algorithmic theory of complexity and information, which means that this relationship is closely connected with information processes; and even more, we are convinced that it is their system-forming relationship. At the same time, it is also one of the main system-forming relations of the natural language, which the well-known linguistic theories persistently and convincingly confirm.

Based on the above comparison, it is quite logical to try to derive the basics of linguistic theory, using ideas, concepts and, so to speak, the «mode» of Kolmogorov's information theory.

The most fundamental, in our opinion, is the idea of a constructive object and its description, which is expressed in formula (10): $p$: $x \to y$. Since language objects are real entities that have their ontological nature, then, of course, the property of constructiveness should be understood from the very beginning as some simplification, coarsening of reality, some only approximately correct idea of these objects. This situation is typical for any scientific research: science does not study its objects in their guises, but only on more or less adequate models. And the property of constructiveness, which we attribute to linguistic objects and phenomena, is the very first assumption, a basic assumption, allowing the further development of the theory.

The next step is to build an adequate description of the studied object, allowing its identification and reproduction of its properties. We will try to establish what properties an operator (reflection) *p* must have, that implements the construction of a description of the language object to which we prescribed the property of constructiveness.

Firstly, this operator must have in its composition a certain analog of the sensory apparatus, which allows it to perceive the properties of the observed object and to distinguish it from other objects. The properties taken in this way should then be fixed in the environment of the internal states of the given operator and undergo some analytical and synthetic procedures, the result of which is a constructive (in fact, formal) object, which represents an expression of some completely defined combination of the operator's internal states. This last expression is identified with the description of the original observed object. In Kolmogorov's information theory, the operator itself, which processes the properties of the observed object into its internal representation (description), is also represented in the form of some constructive object. As such, it also takes the form of some description, and specifically, words in the alphabet of internal states (here they are marked with binary words). Among the possible descriptions, the minimum one should be chosen, taking into account some additional conditions, and in this way, an expression is obtained for the complexity of the original object regarding its description. It serves as the source material for constructing the information measure.

An analysis of the relationships in the process of constructing a description of the source object led us to diagrams (14) and (15), which already allow interpretation of Kolmogorov's information process in terms of the «form-content» relationship. However, from the construction of the operator that implements the reflection (10) described by us, the relation «subject-object» also follows. Indeed, the operator *p*, appearing in formula (10), has a «sensory» apparatus that allows it to interact with the «external» medium, memorize signals from this medium, and carry out their «analytical-synthetic» processing, i.e. to produce analogs of «mental speech» processes, since their final product is speech expressions — «words», but only in the binary alphabet. Thus, in the structure of Kolmogorov's information process in an implicit form, not only the «form-content» relationship appears, but also the «subject-object» one. Both of these relations are system-forming for the natural language, in which the *seventh of our linguistic presumptions are confirmed and substantiated.* In the next section, «Phenomenology of the Language and the Picture of the Language World», we will consider these relations in more detail. However, the reference to these properties, as the properties of system formation, compels us to make some refinement of the concept of a system, since in this book this concept is used in a sense slightly different from what is usually accepted in linguistics. We will analyze this refinement in the section «Linguistics and Systems Approach».

# PHENOMENOLOGY OF LANGUAGE AND PICTURE OF LANGUAGE WORLD [23]

## Phenomenological grounds for describing the language

The experience of constructing any theories shows that the success of the formalization of the description of a particular subject area is directly associated with the choice of adequate objects of the conceptual presentation of its phenomena. This entirely applies to language. It should be recognized, however, that when creating linguistic schemes, their authors often tend to narrow the field of linguistic phenomenology, so these constructions often become quite chamber-like. Of course, this restriction greatly simplifies the task of researchers, for which simplification one has to pay with the adequacy of the description itself, its incomplete, insufficient for the application, correspondence to the phenomenology of the language and, finally, the efficiency of the final products — linguistic technologies. The latter has become increasingly important as we move into the depths of the so-called society of knowledge, which the author considers the second and highest phase of the information society.

The boundaries of the field of linguistic phenomenology are not permanent and should be clarified in connection with the tasks to be solved in the research process. Moreover, it is possible and very different approaches to the definition of the principles themselves, based on which the said clarification is supposed to be made.

In this regard, I wish to quote the words of Isaac Newton of his first, written at the age of eighteen, scientific work «The Universal Language», *devoted to fundamental grammatical problems* [24]*: «The dialects of individual languages are so dif-*

[23] Inthissection, theresultsof our monograph are partially used: Shirokov V.A. Elementy leksykohrafii. Kyiv: Dovira, 2005. 304 s., as well as the results of following works: Shirokov V.A. Grammatika kak fenomenologicheskaya problema. *Bionika intellekta*. 2013. 1 (180). S. 3—14; Shirokov V.A., Shevchenko I.V. Hramatyka u fenomenolohichnomu vymiri. *Movoznavstvo*. 2014. No. 4. S. 3—27; Shirokov V., Shevchenko I. On the phenomenological approach to grammar. *Cognitive studies. Etudes cognitives*. Warsaw: SOW Publishing House, 2015. 15, P. 3—34.

[24] The universal language (Newton I. Of an Universall Language.Modern Language Review). This work was published only in 1957. The translation into Russian language, made L.V. Knorina (with its note and afterword and a preface by V.A. Uspenskiy), was published in «Semiotics and Informatics». Moscow, 1986. vol. 28.

*ferent that a universal language cannot be deduced from them as faithfully as from the nature of things themselves, which is the same for all peoples and based onwhich the entire language had been created at the beginning. The system is constructed from individual substances (spirits and bodies* 25*), each of which is in a given position, state, or condition. The use of language consists in the fact that one person can designate for another in what state one or another substance is, was, will be, may be, should be, desired to be, prompted to be, etc. That's enough of an introduction for what will follow».*

From the above brief fragment, it is clear that Isaac Newton, formulating the initial principles of his theoretical-linguistic concept, suggested the insufficiency of what is called «linguistic material» nowadays, and expanded the boundaries of the phenomenology of language to the «nature of things themselves, which for all nations, and based on which the whole language was created at the beginning».It is noteworthy that Newton proceeds from the existence of a certain «universal language», the «dialects» of which are all «separate languages». In the XXI century, the whole enormous complexity of I. Newton's attempt to deduce the language system from the «nature of things themselves» is understandable, but now we cannot but admire the intellectual athleticism of the young genius who made such a bold attempt more than 350 years ago.

On the other hand, this attempt can be described as an apotheosis of the phenomenological approach to the study of language. After all, the nature of things «is» the subject, i.e. «is given» to it through mastering the totality of its phenomena, which, in the end, become the facts of language. The methodological setting for deduction of the language system from the nature of things seems all the more valuable because it introduces language into the general circle of phenomena and entities as one of the manifestations of the general nature of things.

Of course, in defining the boundaries of linguistic phenomenology, it is necessary to observe a certain moderation, so that theoretical constructions, if possible, end not only with the formulation of general principles but also with results that are useful for practical applications. And although that the self-esteem of many linguists should flatter the saying of the Apostle John: «In the beginning was the Word, and the God has Word, and the Word was God», which introduces the language means in the original arsenal of the Creator, the study of human languages should be perhaps attributed to the later stages of the evolution of the world.

However, speaking of the evolution of language, we should not forget about creationist views on its nature. Here is how Umberto Eco, although somewhat ironically, describes the act of creating the language through William of Baskerville, one of the main characters of the novel «The Name of the Rose» [26]: «*God warned*

[25] A modern scientist would say «from fields and particles», which, however, would be no more than a terminological difference. Welldone, 18-year-oldNewton!

[26] EkoUmberto. Imya rozy / Per. s ital. E. Kostyukovich. SPb.: Simpozium, 2005.

*Adam to not eat from the Tree of Knowledge of good and evil, and it was God's law. But at the same time, he allowed Adam, moreover, authorized him to give names to things in the civilian world. And in this respect, in a carnal way, he granted complete freedom to his homager. Yes, exactly, even though some of our contemporaries claim that the nomina sunt consequential rerum*[27]*. However, the Book of Genesis on this account states quite clearly: The Lord led all creatures to the man, to see what name he would give them, and as the man named the living creature, so is it to be called aeonian. And although there is no doubt that the first man approached the matter strictly and responsibly, naming in his Eden language every thing and every creature, he was guided by the nature of named, yet the fact that while doing this, he took on some kind of sovereign power is not canceled: decide which of the many names, at his discretion, best corresponds to the nature of the object being named. For, it has been established that the names that different people use to describe the same concepts are different, but only the concepts, that is, the signs of things, are constant and the same for everyone. And the word nomen (name) undoubtedly comes from nomos, that is, in Greek means «law», precisely because the nomina are created by groups of people ad placitum, i.e. by a free joint decision».* [28]

Thus, it can be assumed that Isaac Newton made an attempt, in fact, already carried out by Adam. After all, according to William of Baskerville, Adam «*naming every thing and every creature in his language of Eden was guided by the nature of the named*». And although it is not quite clear where this nature was known to the first man (perhaps the Creator who was present at this was advising him?!), the methodological setting on managing the nature of things when naming them becomes completely understandable. Just as it becomes clear why Newton, without sufficient knowledge of the nature of things at the age of eighteen, but wanting to embody such knowledge in his theory of language, devoted most of his scientific life precisely to the knowledge of the nature of things.

Speaking of evolutionism and creationism concerning language, it is impossible not to recall the judgment on this topic of another Wilhelm, namely, such authority as Wilhelm von Humboldt, who noted: «No matter how natural the assumption of the gradual formation of a language might seem to us, it could arise only at once. <...> For a person to understand at least one single word, the entire language and all its interrelationships should already be inherent by him[29]». Our observations on the systemic relationships of language[30] confirm Humboldt's stated idea, although nowadays we are not ready to give preference in this subject to either evolutionism or creationism.

---

[27] Names are derived from things (*lat*.).

[28] Eko Umberto. Imya rozy.

[29] Gumboldt V. Ob otlichiyah v stroenii chelovecheskih yazykov i ih vliyanii na duhovnoe razvitie chelovechestva. SPb., 1859. 366 s.

[30] Shirokov V. System relations in explanatory dictionaries. *Proceedings of the VI Int. Scientific and Technical Conf*. CSIT 2011. Lviv, 2011, P. 260—264.

In connection with the question of the phenomenology of language, let us refer to the views of the most prominent representative of the phenomenological trend in the philosophy of the XX century, Edmund Husserl. In the third volume of his Logical Studies [31], he writes: «*Modern grammar believes that it should be based solely on psychology or other empirical sciences. In contrast, we see here that the old idea of a* ***general*** *and, in particular,* ***a priori grammar*** *acquires (since we discover the laws defining possible forms of meanings) the foundation that cannot be doubted and, in any case, some definite sphere of significance (Gültigkeit). To what extent other areas belonging to the Apriori grammar can be discovered is beyond the scope of our interests. Within pure logic, there is a sphere of laws, distracted from any objectivity. These laws, in contrast to logical laws in the usual and exact sense, could reasonably be called {pure logical grammatical (reinlogisch grammatische)}*[238]*. {And it would be even better to contrast the pure doctrine of value forms with its pure doctrine of significance (reine Geltungsiehre)*»}.

At first glance, it seems that in this case, the phenomenologist E. Husserl chooses in favor of apriorism. However, the last century with its great discoveries in the field of logic and mathematics (associated with the names of G. Frege, G. Kantor, B. Russell, A. Whitehead, D. Gilbert, K. Godel, L. Wittgenstein, A. Tarsky, L. Brouwer, L. Zade, and others), as well as quantum-relativistic physics (A. Einstein, A. Poincaré, M. Planck, N. Bor, L. de Broglie, V. Heisenberg, E. Schrödinger, P. A. M. Dirac, M. Born, J. von Neumann, R. Feynman, and others) had expanded the concept of logic so much, at the same time bringing it closer to substance and subject, that the combination of logic and grammar according to E. Husserl now looks not at all the way it was seen, for example, by the rationalist R. Descartes in XVII century. So, Husserl's remark, as we see it, is quite interpretable in the phenomenological spirit, and Husserl's «pure logical-grammatical (reinlogisch grammatische) laws» are quite comparable in their conceptual epistemology with Newton's «nature of things».

What are the general features of the latter, which are reflected in linguistic forms? For the answer let us turn to what we call the picture of the language world. As aphoristically expressed by L. Wittgenstein: «*The boundaries of my language define the boundaries of my world*», and earlier V. von Humboldt in a letter to F. Wolf in 1804 wrote: «*I managed to discover — and I become more and more aware of this, that through the language one can cast a glance over the highest and deepest spheres and all the diversity of the world*».

## The linguistic picture of the world and its conceptual representation

The linguistic picture of the world can be schematically represented as follows Fig. 1. In this scheme, the external impressions of the world through the perceptual-sensory apparatus of a person (the subject of the thought-speech process), when

[31] Gusserl E. Sobr. soch. T. III (1). Logicheskie issledovaniya. Issledovaniya po fenomenologii i teorii poznaniya / Per. s nem. V.I. Molchanova. Moscow: Dom intellekt. knigi, 2001.

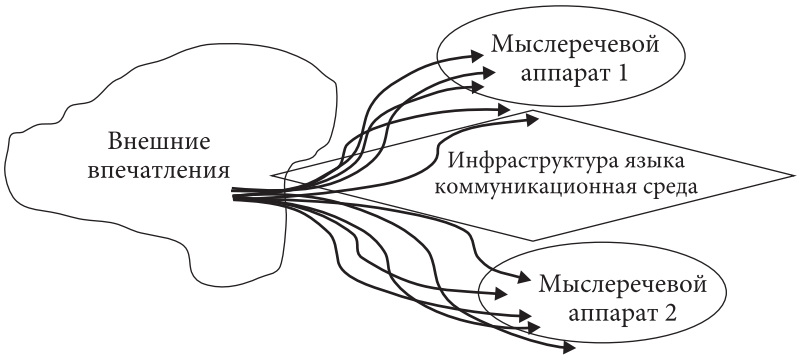



***Fig. 1.*** The linguistic picture of the world *(Thought-speech apparatus 1, Thought-speech apparatus 2, External impressions, Language infrastructure, Communication field)*

transformed, get into his thought-speech apparatus, where they undergo processing, acquiring a language form. Through the communication system — the language infrastructure — the exchange of verbally expressed information with another subject of the thought-speech process is carried out.

Even from such a lapidary scheme, quite definite conclusions follow regarding the structure of the language system. First, it is a conclusion concerning the relative stability of the forms acquired by language structures in the process of evolution. After all, communicators (participants in the communication process) are obliged to exchange information in agreed «formats» and «protocols» of data, because without this communication is impossible at all. The mentioned means of «standardization» of data exchange by definition should not change (or change little) in the process of communication, which ensures the required stability of the forms of language structures. The second relates to the specificity, uniqueness, and diversity of forms in which language data is implemented in the contours and elements of the paths of the thought-speech and communication processes. As noted by Pitts and McCulloch[32]: «*The language through which information is transmitted (in the brain)... does not correspond and should not correspond to the language that people use to communicate with each other*». This is understandable, since the thought-speech and communication processes, per se, are different tasks that require different means for their resolution. And if the function of the communication system is merely the transmission of language data without or with minimal distortion, the thought-speech apparatus solved much more diverse tasks: coding-decoding, analysis, synthesis, transformation, interpretation, con-

[32] Quoted from the book: Pribram K. Yazyki mozga. Eksperimental'nye paradoksy i principy nejropsihologii / Per. s angl. Ya.N. Danilovoj i E.D. Homskoj; Red. i predislovie A.R. Luriya. Moscow: Progress. Red. lit-ry po filosofii. 1975.

ceptualization, comparison, memorization, short-term and long-term storage, etc. Marked processes occur in different environments and are carried out in different, so to speak, hardware and software implementations. We believe that the properties of information and features of the flow of information processes, generally speaking, depend on the physical substrate in which they are implemented. The last statement finds its rationale in quantum information, the research, and the development of which has been particularly intensified recently [33].

The first question that arises in formulating the principles of modeling language substance is the question of the modeling objects, namely: what are the objects of the language, and what we are going to model. We take as a starting point in this work that *our objects of the language are certain psychophysical states and processes occurring in the human thought-speech apparatus, and its oral and written forms serve as elements of the speech process infrastructure*. This statement is somewhat different from the customary for traditional linguistics view, according to which the substance of the language is sound substance, which forms the language substrate to be studied and modeled. We proceed from the obvious fact that speech does not arise in the vocal cords — form already belongs to the infrastructure of the linguistic, or rather, thought-speech process.

It is obvious that the thought-speech process itself is integrated, i.e. it contains both linguistic and mental components. In the thought-speech apparatus, it is expressed in the form of a dynamic system of interconnected reflexes, the content, and character of which is explored, for example, in the book by V.M. Bekhterev [34], published in 1909, but still not lost its relevance, as well as in the previously mentioned monograph K. Pribram and many other works. According to V.M. Bekhterev, natural language is one of the so-called connecting reflexes that occur in the human brain. Consequently, the separation of language processes from mental processes, accepted by many linguists, as well as attempts to study the language «by itself» seems to us an unjustified and methodologically incorrect simplification. The linguistic system should be considered as open and transparent, which implies both a significant expansion of the phenomenological base of the language and a modification of the appropriate conceptual tools.

Oral and written forms of speech, in this sense, play the role of models of thought-speech processes and their communicative environment (infrastructure) at the same time. With such a factorization it can be argued that they represent the language periphery.

Let us, however, make a reservation about the possible underestimation of the infrastructural components of the language, which may arise since we recog-

[33] See, for example, the book: Hrennikov A.Yu. Vvedenie v kvantovuyu teoriyu informacii. Moscow: FIZMATLIT, 2008. 284 p., as well as numerous references to it and articles in journals: *Europhysics Letters. J. Applied Physics, J. Optical Society of America*, *Nature Photonics*, *Nature Physics,* New *J. Physics* et al.

[34] Bekhterev V.M. Ob'ektivnaya psihologiya. Moscow: Nauka, 1991. 480 s.

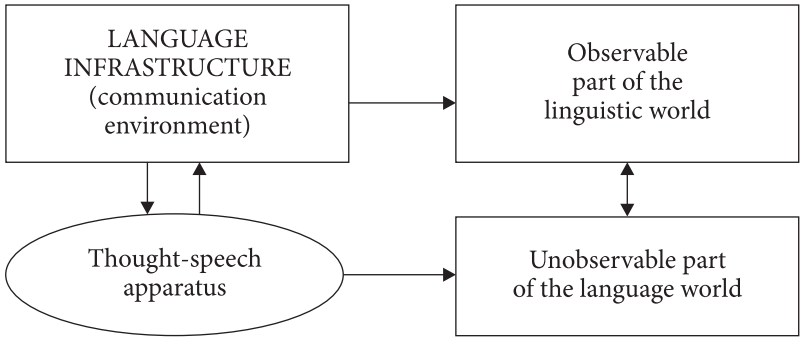



***Fig. 2.*** Picture of the language world

nize psychophysical states and processes of the thought-speech apparatus as the «core», primary language objects. The fact is that modern data show that knowledge of the language and its proficiency are not innate properties of a person. Only the language ability is innate, for which there are certain areas responsible for speech in the human brain from birth. And the process of «installing» a language in a person necessarily requires the presence of such infrastructural elements as so-called «external» and «egocentric» speech, which already operate in the early stages of the ontogenetic development of the child's language system and end with the formation of an «internal» speech, which bookended the process of creating a full-fledged speech apparatus [35]. Thus, the language periphery is an integral element of the language system. Also, it is one of the elements that ensure its informational transparency.

It should be noted that the psycho-physical condition and processes (including thought-speech), usually not fully accessible to direct observation, and even more so for objective fixation. Oral and written forms of speech, serve as representatives of the observed components of the states of language objects and processes occurring in the thought-speech apparatus. As such, they are used as the main objects in the conceptual modeling of a language. This fact further emphasizes the importance of the language periphery in the study of the language system. Thus, in the linguistic picture of the world, one can distinguish the observable and not directly observable components of the thought-speech process (Fig. 2).

We recognize the conventionality of such factorization because some elements of thought-speech processes are already observable (but only some) nowadays. However, in what form, form, and format language data is stored, what models, structures, and types of this data are — all this cannot be directly observed yet, and it is not quite clear how this is possible in principle.

[35] Vygotskij L.S. Myshlenie i rech. Izd. 5-e, ispr. Moscow: Labirint, 1999. 352 s.

## The lexicographic effect in information systems

The fundamental properties of subjectness can be clarified by analyzing the mechanisms of the lexicographic effect in information systems that we formulated in the 90s of the last century [36]. To do this, we need the concept of a lexicographic description of systems, the presentation of which follows the works [37]. Under the lexicographic description, we understand a much more general type of system description than that adopted in the classical linguistic lexicography and dealing with dictionary-making. In short, lexicographic description of the object under study consists in:

a) selection in it a discrete spectrum of certain units relating to its ontological nature, namely, such discrete, relatively stable entities that, when combined, explicate the phenomenal manifestations of the object under study;

b) construction of interpretations of the mentioned spectra both from the side of the forms of their manifestation and from the side of the inner content associated with these forms and manifested in them.

The basic principles of the lexicographic description in our understanding are based on the information theory of systems [38], so that this description, based on information interpretation, acquires much more general features than directly related to linguistic phenomenology, and, applies to any objects where information processes occur with systemic signs, only to some extent similar to the corresponding signs of language substance.

According to his typology, the approach we have adopted is described as phenomenological, since it is based on the general informational manifestations of objects and is not a specific consequence of their particular structure. At the same time, we will emphasize our conscious striving, when considering the informational aspects of reality, to be as close as possible to that circle of phenomena that are close to natural language or that can qualify as such.

A common feature of all processes of information exchange is the transformation of information from one form to another, and modern natural science theories quite clearly confirm the fundamentally discrete («quantum») nature of interaction and exchange processes and, therefore, basic discretization of the processes of perception and interpretation (description) of reality. The noted discretization has at least one feature common to all known processes, which we believe is of a fundamental nature. Namely: observing and generalizing the behavior of various systems, we conclude that *in the process of evolution (dynamics,*

[36] The first regular exposition of the theory of the lexicographic effect in information systems was carried out in our book: The Informative Theory of Lexicographic Systems.

[37] Shirokov V.A. Fenomenolohiia leksykohrafichnykh system; Shirokov V.A. Elementy leksykohrafii; Shirokov V.A. Komp'iuterna leksykohrafiia. Kyiv: Nauk. dumka, 2011.

[38] Shileiko A.V., Kochnev V.F., Khymushyn F.F. Vvedenye v ynformatsyonnuiu teoryiu system. Mosow: Radyo y sviaz, 1985. 278 s.

*self-development...) of a system* of any nature *in its structure, its interaction (observation) with some subjectmanifests a subsystem of relatively stable discrete entities* («order subsystem») that play the role of its *elementary information units*, so that all other phenomena of the system are in a certain way organized combinations of these elementary information units.

The marked subsystem has properties that are in some way related to the properties of the lexical system of a natural language: it «generates» in its structure something like a thesaurus and grammar with the properties of signify, meaning, content, polymorphism, etc., inherent in these constructions; it is the carrier of both the «plane of expression» and the «plan of content». This circumstance explains our use of the term «lexicographic effect».

Sets of elementary information units are characterized by the property of «substantiality», as well as other aggregations caused by objective processes, as a result of which the marked aggregates tend to have relative stability of their characteristics, ensuring their localization in the respective areas of the system parameters. Similarly, we conclude that any lexicographic effect develops in the environment of subject-to-object relations, and manifestations of subjectivity vary within extremely broad boundaries, starting with the universal reflection property inherent in all things and ending with mental and cognitive reactions and movements of intellectual entities.

The above-described set of phenomena is the content of the lexicographic effect in information systems.

It can be affirmed that in the study of any subject areas, experts investigate the lexicographic effects occurring in these areas. Thus, the lexicographic effect has not only a phenomenological component but also a methodological one, since it has a certain «operational potential», stimulating in the process of modeling certain systems to establish and define the corresponding complexes of elementary information units, taking into account, specifying and representing their properties. In this incarnation, the concept of a lexicographic effect acts as a method of abstracting data.

Since the lexicographic effect is expressed in the representation of some, in essence, a continual universum through discrete sets, there is an attractive possibility of its formalization using the Löwenheim-Skolem theorem known in model theory [39]. The latter consists of the fact that under certain restrictions it is possible to establish some isomorphism between uncountable and countable sets; in a sense, the potential infinity can have a finite interpretation a finite infinity model is constructed. The Löwenheim-Skolem theorem states that any consistent first-order solvable theory that has an uncountable model also has a countable model. This means that if a certain structured set is given by a countable set of rules, then there is a countable set (that is, a proper subset of the set of natural

[39] Puaza Bruno. Kurs teorii modelej. Almaty, 2001. 460 s. (ebook in PDF).

numbers) on which to construct an accurate model of this structured set, where all the original axioms will be fulfilled. Consequently, there is such a representation of an infinite description of an object through a finite one, which contains all the information about an infinite object.

Thus, the lexicographic effect can be considered as a phenomenological substantiation of the theory of complexity and the corresponding theory of information, and vice versa: complexity theory, Kolmogorov information, and the Löwenheim-Skolem theorem can naturally be considered as formal correlates of the lexicographic effect in information systems [40].

Hereinafter, the complex of elementary information units of object $D$, induced as a result of the development of the lexicographic effect $Q$, we denote by $I^Q(D)$ or simply $I(D)$, if the reference to the type of lexicographic is not significant. The system of elementary information units has a certain structure. In particular, it is possible to dedicate a core in it — a certain subsystem $I_0^Q(D) \subset I^Q(D)$ and define the generating procedure $\pi$:

$$\pi: I_0^Q(D) \to I^Q(D). \tag{16}$$

The triplet $(I^Q(D), I_0^Q(D), \pi)$ will also be identified with the system of elementary information units and used along with $I^Q(D)$, $I(D)$, $I_0^Q(D)$, $I_0(D)$ as equivalents, assuming that the generating procedure is defined, known and understood from the context.

Summing up the content of this phenomenology, we state that processes similar to those described occur in all fairly complexly constructed natural and socio-technical systems and, in general, in systems of any nature, in which sources, transducers and consumers of information operate, and, therefore, analogs of perceptual-sensory acts and intellectual processes take place, implemented in the environment of the systemic «structure — substance — subject» triad relations.

What has been said provides an additional basis for the validity of the conclusion about the existence and universality of the lexicographic effect in information systems with the characteristics of the*sign, meaning, content, and polymorphism*, similar to those in natural language, which also has these properties [41]. These factors reinforce the tendency to construct as formally defined structures as possible, which are representatives of lexicographic effects, and their application to the description of phenomena of any nature, in particular, natural-language ones.

We see numerous confirmations of the above in several natural phenomena. Even though according to modern concepts, the entire universe is placed in a

[40] Golovko G.G., Shirokov V.A. Teorema Levengejma — Skolema kak formal'nyj korrelyat leksikograficheskogo effekta v informacionnyh sistemah. Mega Ling' 2006. *Gorizonti prikladnoï lingvistiki ta lingvistichnih tekhnologij: Dop. Drugoï mizhnar. nauk. konf.* (20—27 veresnya, 2006, Ukraïna, Krim, Partenit). S. 201—202.

[41] Solomonik A. Semiotika i lingvistika. Moscow: Molodaya gvardiya, 1995. 352 s.

four-dimensional space-time continuum, the observed quantities depend on a small number of so-called «world constants» (speed of light, electron charge, and mass, Planck constant, etc.), which together with a set of integers constitute a kind of physics «alphabet», which are expressed all meaningful statements about the behavior of physical systems (the values of observable quantities). The situation is similar with the scientific description of other systems: all chemicals are certain «words» in the alphabet of chemical elements, and their interconversions are the «sentences» of this language; proteins mainly consist of 20 amino acids, DNA molecules are built based on four types of nucleotides, etc. The chemical elements themselves are elementary information units relative to the lexicographic effect, determined by the Pauli principle, which allows no more than one electron to exist in one quantum state, thus regulating the order of filling the electron orbits of atoms and, therefore, determining the possible set of chemical elements. Examples of this kind can result in tens.Significantly,this behavior is typical not only for, so to speak, real objects of the world — natural and technical — but also for conceptual level formations operating with ideal objects, abstractions, and mental constructions.

## The subject in the structure of the lexicographic effect

From the construction of the Kolmogorov information measure and diagram (15), it follows that the subject has a dualistic nature: it has a mechanism for direct («perceptual-sensory», reflecting the form of things) perception of objective reality, and is also endowed with an apparatus for «intelligent» processing of the information obtained by interpreting its «content». Also, it is endowed with a kind of «interface» between these two ways of information processing. It is necessary to emphasize that although information characteristics (properties associated with the complexity of the object) manifest (actualize, explicate...) in the process of subject — object» interaction, this can happen only when these properties are in «collapsed» form hidden in the structures of both subject and object.

Thus, «subjectivity» is not only an external feature concerning an object but, in fact, an internal ability of an object to «reflect», its immanent property. The stated understanding of subjectivity has a close connection with the quantum principles of the description of reality, according to which the fundamental characteristic of an object is its state, which in theory, i.e. in the process of scientific description, acquires the features of a basic conceptual object. In this regard, we present some general considerations regarding the concept of system states. This concept, which is used in many natural, socio-humanitarian, and technical disciplines, is, in our opinion, the most deeply theoretically and practically developed in quantum mechanics, where it is fundamental.

According to the canonical doctrine of quantum mechanics, each system at a certain point in time is in a certain state. The system state is formalized as the

solution of the Schrödinger equation for this system. Since the Schrödinger equation is a particular type of differential equation in partial derivatives, the set of solutions associated with the states of the system, forms an infinite-dimensional Hilbert space. Consequently, the number of states of a quantum mechanical system is theoretically infinite.

The system state represents the most complete description of the theory and determines the probabilistic interpretation, but it is, generally speaking, not a directly observable quantity. Observed quantities are represented in quantum mechanics by Hermitian operators that act in the Hilbert space of states, and the possible values of observable quantities are calculated as matrix elements of these operators in the state space. However, in some other theories, the system states are observable quantities. For example, in classical mechanics, the state of a material point is defined by a pair of coordinate-momentum at a specific point in time: (*x (t), p (t)*), which are observed, both separately and together. In quantum mechanics, there is a fundamental limit to the simultaneous measurement of the coordinates and momenta, which is determined by Heisenberg's uncertainty relation.

Therefore, the concept and status of the monitored quantities and noninvariant are defined in different scientific (and other) theories. It could be required for the theory to operate only on the observed values, but this question is not simple. It was greatly debated during the formation of quantum theory and has not lost relevance nowadays. Achievements of this field of theoretical knowledge contain such general methodological lessons and attitudes that they can and should be learned by any science that has ambitions for the theoretical understanding of the nature of the things it studies.

The first and most important of them is, perhaps, the fact that both observable and directly unobservable quantities are used to characterize the states of objects. Moreover, according to the conviction of most scientists, it is impossible to build a theory from only one observable quantity. In this regard, let us turn to the analysis of the idea or principle of observability.

This principle, laid by A. Einstein as the basis for the special theory of relativity and developed in sufficient detail in the substantiation of quantum mechanics, turned out to be very methodologically productive to the description of other complex processes. As A. Einstein noted in his discussion with V. Heisenberg in 1926 when discussing the fundamentals of quantum mechanics [42], it is impossible to build a theory from observable quantities alone (although without them no scientific theory is unthinkable at all). Only the theory itself should determine which of its quantities are directly observable and which are not. In particular, in quantum mechanics, states, generally speaking, are not directly observable quan-

[42] This discussion is described in detail by V. Heisenberg in his book: Part and Whole: Gejzenberg V. Fizika i filosofiya. Chast' i celoe / Per. s nem. Moscow: Nauka, 1989. S. 191—196.

tities. The latter, as was said, corresponds to a certain type of operators acting in the state space; their values represent the values of the observables.

So, observable and not directly observable quantities have different ontological and logical status in different theories; however, as far as we know, a general view on this subject has not yet been developed in detail. In the light of the foregoing, such a simple interpretation of the relations between the observable and not directly observable quantities of the theory is suggested: *they represent, respectively, the «formal» and «substantial» sides of the object under study.*

When applied to language objects, such an interpretation can be detailed in the sense that the state of any language unit can be decomposed into the formal part (achievable for direct perception by the subject, be it a sound or a graphic image), and the content is represented by a combination of «all contexts» in which this linguistic unit can function — this circumstance makes the indicated part of the state directly unobservable.

In a scientific discussion about the logical and psychological foundations of the phenomenon of observability, it is worth mentioning such a philosophical attitude as the Mach principle [43], according to which sensory impressions are ordered in human thinking in a way that provides for the most economical arrangement of these impressions in stable complexes. Characteristically, that A. Einstein [44], considering this principle too banal for it to be able to play the role of universal epistemological law, noted the special role of language in the ontological-logical-psychological development of the cognition process. He considers linguistic constructions not only as a way of fixing sensory complexes but also as a reflection of what exists (or even can only exist) outside these complexes and without connecting with them. In our opinion, Einstein's comments (and he was very sensitive to questions of the philosophy of knowledge) about the role of language is not accidental— they confirm our belief in the universality of cultural and informational processes at all levels of reality. We recall in this connection the conclusion formulated above that even the very possibility of the existence of such a phenomenon as language is a consequence of the property of «being a subject».

Thus, in the world, and therefore in cognition, there is an «invisible», that is, something that cannot be directly observed. This gives us reason to recognize the role of faith in cognitive-intellectual processes, which, in the words of the Apostle Paul, is «certainty in the invisible» [45].

However, it should be recognized that the deeper epistemological causes of observability-non-observability are not disclosed in quantum theory. In our opinion, these reasons lie in the phenomenology of the complexity of the mani-

[43] Mach principle — the principle of «saving thinking».

[44] Gejzenberg V. S. 191—196.

[45] Faith is the exercise of confidence and confidence in the invisible. St. Ap. Paul, Hebrews 11, 1.

fested being. The above presentation of the theory of complexity allows us to consider this question in more detail.

It is generally accepted that the simplicity of a scientific theory is aesthetically pleasing to most researchers — the simplicity and beauty of the mathematical scheme suggested by nature have great convincing power for them. Note that at the time of the construction of quantum theory, the concept of simplicity (and the antonym, and therefore the related concept of complexity) had belonged to general linguistic. Then the theory of complexity was not formulated yet, as you know, it appeared only in the fifties of the last century. The connection of such a characteristic as the complexity of objects and their descriptions (and hence their simplicity!) with information was also not clarified, and quantitative measures were not known for assessing these quantities and their relationships. What was said about the concept of complexity developed by A.N. Kolmogorov and other scientists, its connection with the informational aspects of describing reality, and with the concept of information and its quantitative measure, has a deep connection with the criterion of simplicity and beauty of a scientific theory.

The minimality of the description of the studied object, which, according to Kolmogorov, is an objective measure of the amount of information about this object, encourages scientists (at least at the level of the subconscious mind) to find descriptions of such type, although it does not indicate ways and does not give recipes, since, generally speaking, it refers to the class of algorithmically insoluble problems. However, the lack of ways and recipes does not deny the objectivity of the existence of a minimal description, it is just evidence that there is no formula or algorithm for obtaining new scientific truths. And when such a description is found, then, obviously, it should look like the simplest one — in fact, it is. Consequently, the criterion of simplicity (or beauty) of a scientific theory, in our opinion, is not so much a consequence of Mach's principle of economy of thinking (which Einstein qualifies as «suspiciously commercial» and which has only a very indirect relation to the essence of the matter, since it rather refers to the fundamental informational property of objectively existing things, rather than the line of thinking as a subjective process), but follows from the general nature of information and corresponds to the formal definition of the measure of its quantity according to Kolmogorov.

Indeed, when a description of the studied object (process, system, etc.) is received, which most adequately corresponds to its essence, then this description must be minimal since it provides only essential information about the studied object and does not contain a description of random, non-essential details, which «clutter up» the essential «extra» elements. The scientist, so to speak, instinctively seeks to get just such a description of the object under study that is consistent with the definition of Kolmogorov's information measure, based on the minimization of the description; this, in our opinion, also explains the psychological confidence that the researcher feels when he manages to get a simple (beautiful!) formula, equation, derivation, etc.

The formalism of the theory of complexity is both transparent and deep, it should be perceived ontologically, as an objective property of things. One of the non-trivial manifestations of this feature is that the complexity of a composite entity, generally speaking, is not equal to the sum of the complexities of its entities. More precisely, complexity is not an additive function of the system. In other words, if there is a certain system consisting of other, «smaller» subsystems that are its constituents, i.e. if:

$$D = \bigcup_i D_i; D_i \cap D_j = \varnothing, i \neq j,$$

where the symbol $D$ denotes the system in question and $D_i$ its components, then:

$$K(D) \neq \sum_i K(D_i), \tag{17}$$

where $K(D)$ is a quantitative measure of the complexity of the system $D$, and $K(D_i)$, respectively, are quantitative measures of the complexity of its constituents $D_i$ (as a rule $K(D) < \Sigma\, K(D_i)$). These representations, of course, apply to individual $K(D_i)$, as well as to their components.

In the process of formation, functioning, and interaction of composite systems, a phenomenon occurs, and we qualify it as «complexity self-compensation».

The content of this phenomenon is as follows.

Any reality, being an object of observation (in other words, entering into a «subject-object» relationships), manifests its essence in some limited forms, achievable for perception by the subject [46]. The manifested complexity of the noted forms (potentially it is infinite, because matter, according to modern concepts, has no boundaries of divisibility, so to speak, «in-depth» and does not allow control of all its interconnections and mediations) is compensated, consistent with the «perceptual-sensory» apparatus of the perceiving subject or device (which is the same).

The nature of the interaction of constituents forming a certain unity (integrity), identified as a composite object, is such that they manifest in a «connected» state only a certain part of their full, «immanent» complexity. The need for such behavior can be interpreted as a property that provides the fundamental possibility of knowing the «manifest» being, and maybe even its existence. Otherwise, the complexity of any object would be infinite (and potentially it is), but the complexities of the individual components are «self-compensating» in the process of forming the whole. So it can be argued that the potential complexity of any thing is infinite, because, as noted, for now, we do not see the boundaries of the divisibility of matter, and each lower structural level has its non-zero complexity. But all varieties of component complexity do not «manifest» as a whole «simultaneously», they are revealed only «on a leveled» basis. Therefore, the complexity in each case is subject to «renormalization», if we go from the analogy with quan-

[46] In this regard, we recall the famous saying of A. Einstein: «God is inventive, but not evil», which emphasizes that at each stage of cognition, the world manifests itself in the finite forms of complexity that make this knowledge possible.

tum electrodynamics, where the procedure of «subtraction of infinities» must also be applied to eliminate divergences.

A clear example of the self-compensation of complexity is provided by the language. For example, the length of the one dictionary entry of the explanatory dictionary, which takes into account the effects of grammatical and lexical semantics, including the multiplicity of grammatical meanings, lexical polysemy, phraseological structure of the lexeme, etc., can be considered as the measure of the complexity of a particular word. Meanwhile, a word in a sentence (specific context) functions only in a certain meaning — one or a kind of «mixture» of several possible meanings for polysemic items and, therefore, the measure of its complexity in a specific context is determined only by a part of the dictionary entry, and in some cases, it can be only tenths or even hundredths of the total complexity of the item. Thus, the complexity of the whole sentence may be less than the total complexity of a single word, which is the part of it.

The construction of being is paradoxical! The phenomenological approach suggests: complex things can consist of even more complex ones. In this sense, «more» is smaller than «minimal». In our opinion, a non-trivial confirmation of this thesis is the well-known effect, which has both an ontological, epistemological, and even psychological dimension — it concerns the complexity of scientific theories: atomic theory, for example, does not seem simpler than molecular theory, the nuclear theory does not seem simpler than atomic theory, elementary particle theory is not simpler than nuclear theory, etc. In linguistics, for example, word theory («lexicology») also does not seem simpler than sentence theory («syntax»). In light of the foregoing, the principle of reductionism, according to which complex things should consist of simpler ones, seems not only not obvious, but even dubious, which leads to some revision of the basics of standard system analysis, which will be discussed in the next section. At this level, we propose to take into account the effects described by the theory of complexity, so that it takes on the features and status of a natural and general science, and not just a purely mathematical doctrine.

So, the mechanism of self-compensation of complexity seems to us to be such a universal «universal» of the «subject-object» relationship that it should perhaps be attributed to the basic principles of system analysis. We suggest to considering the relation «form-content» as another universal.

## Information essence of the relation «form — content»

To analyze the details of the unfolding of the relations «form — content» (RFC), consider a diagram that symbolically depicts the process of perception of some object by a certain subject:

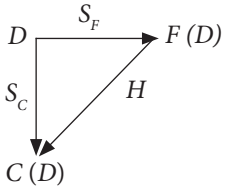
$$S : D \longrightarrow V(D), \tag{18}$$

where the letter *D* denotes «something» from the real (or imaginary) world, acting as an *object* of the process of perception (observation, study, attention, experience...) from some *S*, which we consider to be the *subject* of this process; *V*(*D*) we denote the result of this process. Note that *S* can be a person or a device designed by a person, or a man-machine system, or anything else endowed with the properties of perception and sensation («display»); *S* can also be a «collective subject» — a group of people, a social community, an ethnos, a nation, a people, an aggregate of nations or even humanity as a whole.

Here we do not detail the principles and methods of arranging the subject *S*, except for one: *S* has two defining properties: a) the perception and sensation of external signals («perceptual-sensory» apparatus) and b) the ability to process them internally (awareness and interpretation). Thus, as a result of the physical, mental, intellectual, and other limitations of the subject *S*, the entire set of properties of the object *D* for its perception is divided into two, not very clear, ambiguous, changing, and indistinctly detachable parts. To the first of them, we attribute those properties of *D*, which are more directly perceived by the «perceptual-sensory» apparatus *S* — we denote this part by *F (D)* and treat it as a combination of properties of *D* related to its form from the point of view of subject *S* perceiving *D*. To the second part, we attribute the properties of *D*, which are not directly perceived by the perceptual-sensory apparatus *S* but reflected in it indirectly through the means of internal processing. We denote this part by *C (D)* and will consider it as a combination of meaningful properties of *D*, again, from the point of view of the perception of the subject *S*. Thus, formula (18) takes the following form:

$$D \xrightarrow{S_F} F(D) \xrightarrow{H} C(D), \tag{19}$$

where the symbol $S_F$ is designated the action of the «perceptive-sensory» apparatus of the subject *S*, the result of which is a set of formal (from the point of view of *S*) properties of *D*; the symbol *N* indicated a mechanism which interprets (the relationship between form and content) and ensuring the integrity of perception of the object *D* subject to *S* (if it is indeed possible to provide the specified integrity). At the same time, suggesting the existence of a mechanism that allows the transition from *D* to *C (D)* — let's denote the mechanism through $S_C$ — and we get the following transformation of the diagram (19):




(20)

where, as we see, there was a «decomposition» of the subject *S* into its «constituents» $S_F$ and $S_C$, which are responsible for the reconstruction of the formal and

substantive properties of *D*, respectively. So, in the above decomposition, the subject appears to have a dualistic nature: it has a mechanism for direct («perceptual-sensory», taking into account the shape of things) perception of objective reality *D*, as well as an apparatus for «intelligent» processing of the information received by interpreting its «content». Besides, it is provided with a certain «interface» between these two methods of information processing, which is represented in the diagram by the element *H*.

We are not inclined to absolutize the scheme described above. There is no clear boundary between *F* (*D*) and *C* (*D*), as there is none at all between form and content. The properties of *S* were also almost not detailed, although for general reasons it was decomposed into $S_F$ and $S_C$. Consequently, this approach may indeed qualify as phenomenological, since it does not rely on the assumption of a possible «construction»*S* and the mechanisms of its functioning. From these considerations it can be affirmed that the scheme outlined is quite general — it does not contain any specific «anzatans».

The only characteristic of the language, we believe the assumption that *F (D)* should have a linear character, that is, be represented by linear sequences of discrete objects, the source of which is a certain finite set. Considering the above, even the very possibility of the existence of such a phenomenon as a language is, as noted above, a consequence of the fundamental property of *S*«to be a subject», i.e. those for which an object appears to have its own external (form) and internal (content) side. The relationship between these various aspects of perception, symbolically depicted by the quantities of $S_F$, $S_C$, *H*, is distinguished by a great variety, the source of which is fundamentally inherent (i.e. those that in principle cannot be eliminated) properties of the perceiving subject *S*: variability, irregularity, diversity, limitation, vagueness, etc.

Let us note one more feature of the process of deploying RFC. Modern cultural studies tend to interpret it as a symbol of postmodernism, namely, the influence of the subject on the object, i.e. the possibility of changing the state of object *D* in the process of its perception (observation, research...) by subject *S*. The fact is that in for the process to occur, symbolically depicted in diagrams (18)—(20), in many cases it is necessary to «activate» object *D* for it to «manifest» its properties that *S* «is interested» in. In the classical scientific paradigm, it was believed that suchexcitation of an object can be made as small as desired, and neglect it, believing that it does not significantly change the state of the object. However, the development of science has found that this is not generally the case. Historically, the first scientific discipline, taking into account the influence of the subject (device) on the object of study, was quantum mechanics.

The stated corresponds with the information theory of A. Kolmogorov. The obvious similarity, the «affinity» of the diagram (19) with the definition of information according to Kolmogorov, as well as the further reasoning that led to the

construction of the diagram (20), suggests that the basis of both of them is the same pattern. The very form of the presentation of information measures indicates a certain process, the result of which was the generation of the «alphabet» — the sign system of object representation. Deploying mappings (*f, c, p*) in diagrams (14)—(15) that match the elements of the RFC from the diagram (20), leads to a comparison of the latter with the components of information processes, which in Kolmogorov's theory are reduced to mathematical relationships (algorithms, recursive functions, ...), given on discrete sets.

Relations «subject-object» and «form-content», being the backbone of the language, literally permeates all levels and elements of the language system. Thus, for linguistic research, the issues of identification and delimitation of the elements of these relations are of particular importance. To do this, we have to clarify the concept of a system.

# LINGUISTICS AND SYSTEMATIC APPROACH

## Structure, substance, and subject in the definition[47] of a system

We figured that the word «*system*» and its derivatives occur 149 times in the previous text of our book. Of autonomous words, it is one of the most frequent, and among the elements of the term system, perhaps it is the most common. There is no doubt that this term carries a very large semantic, conceptual burden in scientific texts. Linguistics in this set is no exception. After all, the systematic nature of the language is one of the creeds of modern linguistics and confirms its understanding, which takes into account not only the properties of individual elements of the language but also the relationships (relations) between them, as well as the properties of these relations themselves. It is this circumstance, in the opinion of most linguists, that transfer to the language the property of systemic.

However, as a rule, linguistics does not extend further than the stated statement (which, however, is very useful in practice). Thus, the property of systemic essentially boils down to the concept of structure, because the structure of an object, as it is known, is determined by the set of its elements and connections (relations, operations...) between them.

Regarding the possible other, additional properties that can and, as we believe, should be the basis for the definition of the concept of a system, different opinions were also expressed at one time — both in the field of linguistics and beyond.

In particular, they discussed the possibility of taking into account the substantial properties in the definition of the concept of a system, although in the sense it is rather negative. As noted by G.P. Melnikov in his famous work «Systematic Approach in Linguistics» [48]: «In works on the systemic-structural methodology and purely structuralist works, the problem of «substance», «materiality», «substrate» is either not discussed at all, or is mentioned only to show that it

[47] The author is fully aware that there can be no exact definition of the concept of a system due to the fundamental nature of this concept, therefore, here the word «definition», that applied to the concept of a system, is used in a somewhat «Pickwick» sense.

[48] Mel'nikov G.P. Sistemnyj podhod v lingvistike. *Sistemnye issledovaniya. Ezhegodnik*. 1972. Moscow: Nauka, 1973. S. 183—204.

is not significant in the study of an object as a system. The well-known thesis of F.de Saussure «language is a form, not a substance» is fully coincides with the statement by W. Ross Ashby about what the general theory of systems should do: «It is necessary to exclude from consideration two factors that are not relevant. The first of these is «materiality» — the idea that the machine should be made of real materials... In the same way, any reference to energy is irrelevant...». Now, after a considerable time has passed since the publication of such views, it has become almost obvious that they are idealizations, methodological simplifications that do not adequately reflect reality.

And in fact, let's take, for example, the concept of information and its real embodiment in computer systems. For a long time, it was believed that the fundamental properties of computing systems do not depend on the «substance» from which information converters are composed, whether it be radio tubes, semiconductors, or microelectronic chips. It was believed that only the quantitative parameters of the calculators — memory, and speed — depend on this substance. However, the situation has changed dramatically with the discovery of quantum information, not in the sense of a banal «transition of quantity into quality», but because quantum information has properties that are fundamentally different from classical ones (we are talking about quantum superposition and quantum entanglement [49]), so computing systems created based on such a «substance» will also have system properties that are fundamentally different from classical computers. Given example is just one of the illustrations of the substantial definiteness of systems.

More than 40 years ago, G.P. Melnikov had suggested expanding the definition of the concept of a system to include substantial properties [50]. Thus, according to Melnikov, structure, and substance are «system-forming» concepts of the very concept of «system», an internal immanent of its definition.

The author, in solidarity with G.P. Melnikov, in terms of rehabilitation of the concept of substance, as an integral conceptual component of the definition of the concept of a system, considers such an extension to be insufficient and suggests supplementing it with another conceptual component.

This refers to the following.

---

[49] Preskill Dzh. Kvantovaya informaciya i kvantovye vychisleniya. T. 1. Moscow-Izhevsk, 2008. 464 s.; Valiev K.A. Kvantovaya informatika. Komp'yutery, svyaz' i kriptografiya.*Vestnik RAN*. 2000. 70, No. 8. S. 688—695; Chivilihin S.I. Kvantovaya informatika: Ucheb. posobie. SPb., 2007; PiotrowskiEric S. Toward a Quantum Linguistics: Possibilities for Change in the Delta Zone, 2000. Stapp Henry P. Quantum Nonlocality and the Description of Nature. In:James T. Cushing and Ernan McMullin, eds., *Philosophical Consequences of Quantum Theory.* Notre Dame Press, 1989; Strehle Susan. Fiction in the Quantum Universe. Chapel Hill: Un-ty of North Carolina Press, 1992.

[50] We also note his work: Systemic linguistics and its relation to structural. *Problemy yazykoznaniya: Dokl. i soobshch. sov. uchenyh na X Mezhdunar. kongresse lingvistov*. Moscow: Nauka, 1967. S. 98–102 [in Ukrainian].

Let's start with some simple examples that apply both to the area of the language and those outside it.

Let's consider such a relatively simple and formally well-defined object as a system of linear equations. In what sense and under what conditions can we consider it as a «system» within the meaning of the systems approach? In standard notation, the system of linear equations of the *n*-th order has the following form:

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \tag{21}$$

where $i, j = 1, 2, \ldots, n$; $a_{ij} \in Q$; $x_j \in Q$; $b_i \in Q$. The structure of this object is reduced to the structure of linear combinations and equality relations. At this first step, we are faced with the impossibility of a complete formal definition of the system without information on the «substance» of the quantities that make up these linear combinations. Indeed, in the above formula, the operations of addition and multiplication appear, which are defined differently for different algebraic objects (we can say different algebraic substances). The same applies to equality, the operation (or relationship) of comparison. Having fixed, for example, the elements from this formula as belonging to the field of rational numbers, we get a system whose properties are different from the one as if objects of a different algebraic nature appeared in this formula. Consequently, even in this simplest case, the property of the system involves the determination of the substantial properties of the system structure elements. Thus, the «substantial» initiative of G.P. Melnikov finds here a completely natural illustration.

But this, we believe, is not enough. After all, an object defined by this formula acquires the properties of a system, if and only if all its constituent elements and relations receive some rather definite interpretation. And for this, it is necessary to have an intelligible mechanism or tool capable of «recognizing» these elements and relationships, identifying them as elements of certain populations, performing some actions allowed for these populations according to certain rules (algorithms), and, finally, obtaining a totality of solutions of a «system» or the belief that they do not exist. It is this identification-interpretation-algorithmic component that acts as a kind of *subject — the «analytical» and «active» principles —* forms the last member of the triad «*structure - substance — subject*», which, in our opinion, is system-forming for determining the concept of a system.

Thus, the definition of the concept of a system can be represented as symbolic equality:

$$C = C + C + C, \tag{22}$$

where the «*C*» of the left side denotes the concept of «system», and the right side demonstrates the presence and interaction of the main constituent components of this concept, namely, «*structure*», «*substance*» and «*subject*» [51].

[51] The author is also aware of the difficulties that will be encountered in defining the concepts of structure, substance and subject. Therefore, we will not deal with this issue, to which, however, a colossal bibliography is devoted, appealing to the reader's intuition.

Let's make clarifications regarding this last member of the system triad. Firstly, we will find out whether the property of subjectivity is an internal property of the system or is it the result of the action of some reason external to the system itself.

From diagrams (14)—(15), (18)—(20) and the construction of the Kolmogorov information measure, it follows that the subject has a dualistic nature: it has a mechanism for direct («perceptual-sensory», reflecting the form of things) perception of objective reality *D*, and is also endowed with an apparatus for «intelligent»processing of the information received by interpreting its «content». Besides, it is endowed with a kind of «interface» between these two ways of information processing.

As it was shown, the content of formulas (9)—(12) and diagrams (14)—(15), (18)—(20) cannot be understood without introducing a subjective principle, which abstracts from the fundamental, immanent properties of matter. In formulas (9)—(12), this statement is initially introduced as something taken for granted. It is necessary to emphasize that although information characteristics (properties associated with the complexity of the object) manifest (actualize, explicate ...) in the process of «subject — object» interaction, this can happen only when these properties are in «collapsed» form hidden in the structures of both subject and object.

## Consistency and inflection of inflectional languages

The systemic triad «structure — substance — subject» is manifested in almost all areas of the theoretical description of the language. Let us consider, for example, such a characteristic lexical phenomenon of inflectional languages as inflection and analyze its system properties using the Ukrainian language as an example, taking into account the system triad established by us (22).

1. Structure:

Each word in the Ukrainian language (and in other inflectional languages) has the structure:

$$x = \rho\,(x) \cdot \omega\,(x), \tag{23}$$

where the symbol $\rho\,(x)$ denotes the quasi-base of the word $x$, that is, the part of the lexeme that remains unchanged during the inflection $x$ (this part is the same for all inflectional forms of the token); $\omega(x)$ is quasiflexion, that is, part of the lexeme $x$, which changes in the process of constructing a paradigm [52]. The symbol «$^\star$» indicates concatenation. We introduce additional notation: $[x]$ is the complete paradigm of the word $x$; $[x] = \rho\,(x) \cdot [\omega\,(x)]$; $[\omega\,(x)]$ is the set of quasiflexions that make up the paradigm $[x]$. The structure and substantial content $[\omega\,(x)]$ is determined by the inflectional classification. We give an example of a paradigm for the word інстинкт (instinct).

The structure of this paradigm is given in the Table 2.

[52] In this section, by the word *paradigm*, we mean the *inflectional paradigm of the lexeme*.

*Table 2.* **Structure of paradigm for the word**

| Case Number | [x] | ρ (x) | [ω(x)] |
|---|---|---|---|
| Nominative Singular | Інстинкт | Інстинкт | ∅ |
| Genitive Singular | Інсти́нкту | Інсти́нкт | у |
| Dative Singular | Інсти́нкту<br>Інсти́нктові | Інсти́нкт | у<br>ові |
| Accusative Singular | Інсти́нкт | Інсти́нкт | ∅ |
| Ablative Singular | Інсти́нктом | Інсти́нкт | ом |
| Locative Singular | Інсти́нкті | Інсти́нкт | і |
| Vocative Singular | Інсти́нкте* | Інсти́нкт | е |
| Noun Plural | Інсти́нкти | Інсти́нкт | и |
| Genitive Plural | Інсти́нктів | Інсти́нкт | ів |
| Dative Plural | Інсти́нктам | Інсти́нкт | ам |
| Accusative Plural | Інсти́нкти | Інсти́нкт | и |
| Ablative Plural | Інсти́нктами | Інсти́нкт | ами |
| Locative Plural | Інсти́нктах | Інсти́нкт | ах |
| Vocative Plural | Інсти́нкти* | Інсти́нкт | и |

2. Substance.

The substantial filling of this paradigm is given by a set of quasiflexions [ω(*x*)] ={∅; у; (у, ові); ∅; ом; і; е; и; ів; ам; и; ами; ах; и}. It represents the substance of the paradigm [інсти- нкт], defines the inflectional class *K* (*x*), which owns the token «інстинкт» and to which all paradigmatic attributions are attributed according to the rules of the Ukrainian inflection of nouns.

3. The subject.

This member of the systemic triad includes:

• algorithms for grammatical (morphological) identification;

• expansion algorithms $x = \rho\,(x) \cdot \omega\,(x)$ and paradigm construction: $[x] = \rho\,(x) \cdot [\omega\,(x)]$;

• inflectional classification algorithms, that is, the construction of correspondence: $[\omega\,(x)] \Leftrightarrow K\,(x)$;

• lemmatization algorithms, that is, rules for reconstructing grammatical meanings by the form of the corresponding text form. This problem may have an ambiguous solution even within a fixed paradigm due to the phenomenon of grammatical homonymy. For example, the form «*інстинкту*» has grammatical meanings: «Genitive; Singular» and «Dative; Singular». This homonymy can only be removed if there is a sufficiently broad context. Only then can the «subject» uniquely identify the grammatical state of the corresponding form.

Let's note that the elements of the system triad «structure — substance — subject» in Ukrainian Lingua-Information Fund are implemented in Virtual Grammar Lexicographic Laboratories (VGLL), namely: for the Ukrainian language (more than 258 thousand units in the register), Russian language (more than 180 thousand units in the register), German language (more than 60 thousand units in the register) and agglutinative Turkish language (noun; than 30 thousand units in the register). The relevant information is presented in our monograph «Computer lexicography» [53], published in Ukrainian. VGLL for Spanish, French, Polish languages are also created. The indicated systems are instrumental and can be used for grammar studies in the environment of the corresponding languages. The authors are convinced that the manifestations of systematicity in the form of the triad «structure — substance — subject» are universal and in one form or another are characteristic of the grammar of any language.

[53] Shirokov V.A. Komp'iuterna leksykohrafiia.

# THE THEORY OF LEXICOGRAPHIC SYSTEMS

## The structure and architecture of the lexicographic systems

The study of the informational properties of the language, the lexicographic effect, and the detailing of the properties of systems allow us to develop a certain formalized apparatus, the purpose of which is to describe the properties of the language system, in its modus consonant with the dictionary (lexicographic) description of the language. In developing this apparatus, in this case, apart from purely practical purposes, we were motivated by the consideration that it seemed not quite logical that a formal object has not yet been established for the lexicographic description of the language, as the theory of formal grammars serves for a grammatical description. Moreover, formal grammars have found their rightful place among the objects of such a respectable, well-formalized area as logical-linguistic calculations. Based on the understanding of the fact that grammatical and lexicographic methods of describing a language are additional in some way, we had the confidence that a formal object for lexicography, similar to formal grammars for a grammatical description of a language, must exist. The very structure, the «construction» of dictionaries also led to this conclusion, which, even at the most superficial glance, were presented as tables of some kind containing linguistic data, arranged in quite regular, «standard» ways.

We defined *lexicographic systems* as the basic constructs of constructions of this type (hereinafter, we will use the abbreviation L-systems). The concept of the L-system, in our opinion, is fundamental, and its definition is based on the stated phenomenology of the lexicographic effect. It turned out that L-systems represent a fairly general type of formalized constructions in a set, such as data models, formal grammars, formal systems, canonical calculations infinite alphabets, etc.

Note that special cases (or implementations) of lexicographic systems have been operated on in science and technology for a very long time. They are wide variety of diverse information systems, databases, and knowledge-bases, which include all traditional dictionaries and computer dictionary systems.

If we talk about machine dictionaries, they can effectively perform their functions only when their structure sufficiently fully reflects the form and content of language units that are the lexicography objects. The tendency to reproduce this

completeness is observed only in those cases when the design of information-linguistic systems is based on a deep study of the phenomenology of the language, which itself prompts the choice of an adequate apparatus, as well as the design of the corresponding models. Even though the goal of information science is to interpret the subject area (in our case, linguistic facts) using the data model language, the types and constructions of these models themselves should flow from the subject area and take into account the specifics of linguistic phenomena as accurately as possible. Based on the stated ideas, a structural theory of lexicographic systems was formulated, based on the phenomenology of the lexicographic effect, the consistent application of which provided us with the opportunity for the necessary systemic generalizations and the development of a methodology for constructing lexicographic models.

The starting point of the analysis, the result of which was the formulation of the theory of lexicographic systems, turned out to be a study of a significant number of structures of really existing traditional dictionaries, their generalization, and the construction of corresponding models. To detail and information-lexicographic concretization, a study was carried out of the general structure-forming effects and elements of lexicographic systems, which are abstracted from traditional dictionaries, turning into elements of infological models of general lexicographic systems. This path has led to the establishment of the concept of the lexicographic system's structure.

It is clear, the structure of traditional dictionaries is not accidental, since it has concentrated the centuries-old experience of many generations of lexicographers. Therefore, as a rule, it is quite independent of the subjective tastes of information system developers. The experience of lexicography as a variety of intellectual activity (to the extent of its accumulation) from the systematization of linguistic facts (and even from the systematization of data on *lexical units*) gradually extends to the systematization of data about the world, knowledge of which, in turn, is concentrated in the natural language as a holistic information system.

The universality of the lexicographic effect phenomenon causes the tendency of lexicography any linguistic phenomenon, that we repeatedly noted, and this very fact explains the existence in the dictionary practice of examples of dictionaries in which even those units of the language that do not have direct verbal expression are lexicography. So, an attempt to lexicography syntactic structures was made, for example, in the work of G.A. Zolotova[54], where the introduction says: «Just as the physical world around us consists of elementary particles, the smallest known particles of matter, so the syntactic structure of our speech is organized by diverse, but regular combinations of elementary, or minimal, units, further indivisible at the syntactic level. In linguistics at the present stage of its development, the need has ripened to comprehend the concept of elementary syntactic units,

[54] Shirokov V.A. Komp'iuterna leksykohrafiia.

from which, as it becomes more and more obvious, all other, more complex constructions are built». And further: «The minimum, further indivisible semantic-syntactic unit, acting both as a carrier of elementary meaning and as a constructive component of more complex syntactic constructions, characterized by a certain set of syntactic functions, is called a syntaxeme». We note an obvious analogy (sometimes it is almost a textual coincidence) with our formulation of the lexicographic effect, the scope of which isincomparably wider.

Similar attempts to lexicography semantic structures[55] not only reflect the general tendency of the lexicographic description of all linguistic phenomena but also meet the needs of the practice of developing the most advanced means of linguistic support.

From the foregoing, the methodological correctness of the inclusion of units of any language level and any quality in the composition of elementary information and lexicographic units of a particular lexicographic system follows. In this way, semantic, syntactic, cognitive, and other structures are lexicography, which, as a rule, do not have a direct verbal representation in the natural language system. Works on the creation of dictionaries are close to this type of work: of ideographic, verbal control, word equivalents, and phraseological units, etc. The latter two types of dictionaries are adjoined by a whole set of possible lexicographic works, which have not yet been created but theoretically having the full right to exist[56]. In the mentioned work, proposals are presented for creating more than 50 different dictionaries in which lexicographical units (elementary information units concerning corresponding, sometimes very exotic lexicographic effects) are, for example, vocatives, etiquette phrases, euphemisms, honorifics (politeness expressions), gumilatives (rudeness expressions), stimuli and reactions (encouragings, consents, objections, refutations), etc.

Studying the various structures of existing traditional dictionaries allows us to make certain generalizations that can not only be the basis of the theoretical lexicographic scheme but also be used in the design of specific linguistic information systems and the creation of appropriate software. Since, in lexicography, a distinction has long been made between the concepts of «dictionary» and «list of words», «list», «index», «inventory», a dictionary as an abstract lexicographic system is necessarily endowed with a structure containing at least two necessary parts: registration (left) and interpretative (right), which is a manifestation of the OFS. It is the presence of the second — the carrier of the substantial OFS component — that distinguishes the dictionary from the usual list of words. But the dictionary has a deep structure, which is reflected in the structure of the registry

[55] Zolotova G.A. Sintaksicheskij slovar'. Repertuar elementarnyh edinic russkogo sintaksisa. Moscow: Nauka, 1988. 439 s.

[56] Russkij semanticheskij slovar'. Tolkovyj slovar', sistematizirovannyj po klassam slov i znachenij / Red. N.Yu. Shvedova. Moscow: Azbukovnik, 1998. **1**. 832 s.

and interpretation parts of the dictionary as a whole and its dictionary entries, as well as in the structure of inter-article and inter-dictionary reflections. Therefore, a dictionary is a special type of text in which a description of the vocabulary of a given language (or combination of languages) is presented in a systematic and structured way. However, it is natural to consider it as a specific object of technology, namely, an information system, where, through printing, certain linguistic effects are noted using font selections, positional placement, special labels, etc., which play the role of identifiers of the corresponding information variables — metalanguage elements of a dictionary. The complexity of the structure of the dictionary lies in the fact that not all elements of its structure are explicitly by the method indicated above. In the structure of real dictionaries, as a rule, there are a large number of implicit structure-forming elements, the identification of which is often a rather difficult task. The process of abstracting the dictionary (lexicographic) structure is a kind of decoding, reconstruction of the lexicographic effect that led to the formation of this structure, and is developed using several provisions formulated first in linguistics, but in fact, having a system-wide character.

The construction of a structural model of lexicographic (dictionary) systems, focuses on a multi-aspect representation of the significant nature of lexical units as compact and informative ones in a natural language. From the standpoint of the theory of the lexicographic effect, this means the allocation in the language system understudy of the subsystem of its elementary information units and the determination of the set of their system-structural parameters.

The next point is to take into account the dichotomous structure of each elementary information unit (and their complete set), which is reflected in the multidimensional ratio of form and content, the carrier of which is a certain class of elementary information units.

The multidimensionality of the representation of the significant nature of language units in traditional dictionaries (or elementary information units in general lexicographic systems) is provided by taking into account the linguistic and cognitive features of lexicographic objects — depending on the type of dictionary and the depth of characterization of the lexicographic effect, which is the subject of research in each particular case. In the information-lexicographic model, with these features, a certain set of data complexes and/or knowledge is compared.

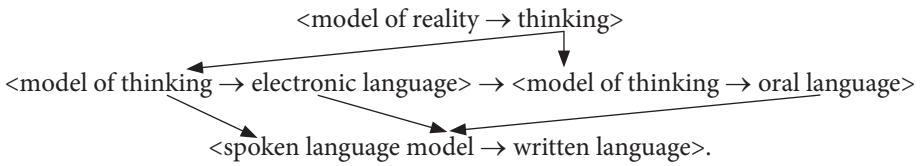
Note that in the linguistic (speech) stream, the ontological nature of the language appears indivisible into individual components that are so characteristic for conceptual representations. This implies the desire to create «integral» dictionaries, and, consequently, the need to use complex (integrated) models of linguistic phenomena. Therefore, when developing computer language processing tools, the task is to create formalized models that would be tuned to efficiently represent integration processes and at the same time take into account the specifics of linguistic objects. Thus, the criterion for the multidimensional representation of the symbolic nature of language units allows us to build complex, inte-

grated data models suitable for combining conceptual representations of language phenomena of various nature.

The dichotomy of the structure of elementary information units in the information-lexicographic model (similar to how it is done in most traditional dictionaries) is revealed in the structural organization of the lexicographic system and follows from the fundamental principles of modern linguistics, which operates with the concepts of form and content, internal and external forms of language units, whose phenomenology is deeply traced on linguistic material.

As V.M. Rusanivsky [57] noted, language has a dualistic function: on the one hand, this is the material basis on which thinking relies in the process of its functioning, and on the other hand, the material in which is fixing, becoming a fait accompli. The objects of study of the components of the thought-speech stream are both the physical (material) and the substantive (ideal) sides. Thus, the sound substance of speech can be considered its form, and information properties — it's content. Given this circumstance, the sound realization of speech can be divided into elements of varying degrees of aggregation — intonationally integral units (intonems), combinations of vowels and consonants (syllables), vowels and consonants (sounds), etc. This process is limited since the allocation and classification of the components of speech sounds depend on several factors, including the progress of acoustics, phonology, etc. The physical process of sound speech belongs to irreversible and (like many other acoustic phenomena) dissipative processes. The indicated properties of the physical substance of the speech, together with the properties of the speech apparatus, determine its external informational characteristics.

In turn, the written form of the language models its oral form, therefore the sequence is generally valid: <model of reality — thinking> → <model of thinking — oral speech> → <model of oral speech — written language>. Since the mentioned above models are physically implemented in a single system (associated with individuals, social communities, cultural systems, etc.), their interaction and mutual influence are natural and necessary. Thus, the written version of the language also acts as a function of both the model of thinking and the model of reality. Recently, in connection with the rapid development of electronic communications and software and hardware for the intellectual processing of various aspects of the natural language, this sequence should be supplemented and expanded to include linguotechnological elements:




[57] Devkin V.D. O nerodivshihsya nemeckih i russkih slovaryah. *Vopr. yazykoznaniya*. 2001. No 1. S. 8—97.

Thus, linguistic technologies are introduced into the system of thought-speech communications and even move to the central plane of the general cognitive-communicative picture.

The existence of a language in the form of speech activity, as well as writing and other ways of fixing speech acts on physical media other than natural language ones, represents the property of a language to have an external form. The latter is generally possible due to the ability of a language to be a representative of the phenomenal side of reality, and since language itself represents a certain reality, it contains potentialities for designating (representing) oneself.

A system that acts as a representative of the phenomenal side of reality must also be organized in a certain way. Since the difference between the phenomenon and the essence is relative, and between the phenomenal and the essential reality there is no clear boundary, it should not be in speech as a model of reality. This fact is implemented in the property of a word to have an inner form, which is associated with the representation of the noumenal part of its being in the language system. The external and inner forms are thus interconnected and together constitute the form of a word as opposed to its content — the sum of specific meanings.

All this gives grounds for the assertion that OFS (including the idea of the internal and external form of linguistic units) are general, represent the universal property of elementary information units, are induced in the process of decomposition of any lexicographic effect, and, being formalized in the form of corresponding data models, are capable of form a substrate for information systems models of arbitrary nature and origin, and for culturally-oriented data models, they are generally mandatory. These concepts, in our opinion, have a constructive potential, since any content exists only in a certain formal shell, which allows us to use a unified approach to the construction of their representatives in scientific theory.

Consider some object *D*, the conceptual description of which we intend to present in the form of a kind of «*lexicographic system*». Since we were previously interested in linguistic facts, the object *D* will be some natural language, or a combination of natural languages, or a complex of natural language phenomena highlighted in a certain way.

According to said above, in object *D*, considered as a kind of information system, a complex hierarchy of lexicographic effects develops. So, for the natural language system, one can cite examples of several lexicographic effects, the result of which is the distinguishing from the speech flow an individual phonemes, syllables, morphemes, words (word forms), lexems, phrases, sentences, etc. All of these structural units act as components of the corresponding classes of elementary information units concerning certain types of natural language lexicographic effects.

In the future, as a lexicographic system, we will consider special information (semiotic and semantic) environment in which a certain lexicographic effect (or a specific set of lexicographic effects) develops (implements).

To build a practically useful modeling scheme for these phenomena, it is necessary to determine a set of information constructs that specify the structural elements of L-systems, allowing to develop of specific applications. In turn, this entails the need to construct a constructive theory of L-systems — it is based on a lexicographic data model developed in works[58], the results and notation of which we will use in the following presentation.

So, according to the informational interpretation of the processes of perception, we determine the result of the reception by the subject *S* of the class of elementary information units (EIE) $I^Q(D)$ in the form of some set $V(I^Q(D))$ — the set of descriptions of units belonging to the class $I^Q(D)$. This set is the result of a process:

$$S : I^Q(D) \to V(I^Q(D)), \tag{24}$$

therefore, for each element $x \in I^Q(D)$, its description of $V(x)$ as an element of the set $V(I^Q(D)$ is uniquely defined: $V(x) \in V(I^Q(D))$; $Sx = V(x)$. Therefore, it is logical to assume that $V(I^Q(D))$ has the form of a union:

$$V(I^Q(D)) = \bigcup_{x \in I^Q(D)} V(x). \tag{25}$$

According to the information concept of representing the description of the EIE system, each *V(x)* is represented as a word (text) in some finite alphabet $\boldsymbol{A} = \{a_1, a_2, ..., a_n\}$, i.e. a finite sequence of characters (finite chains) from $\boldsymbol{A}$; for theoretical constructions, the binary alphabet is usually used — as in the paragraph on Kolmogorov complexity and information measure. For linguistic tasks, it is appropriate to introduce an alphabet whose composition correlates with the composition of the sign system of a given language, or several languages that are the subject of lexicographic modeling. In what follows, words (finite chains) in the alphabet $\boldsymbol{A}$ will be called $\boldsymbol{A}$-words. For example, for most explanatory dictionaries, alphabet $\boldsymbol{A}$ consists of the following elements:

ordinary phonetic alphabet of the language (capital and small letters);
punctuation;
Arabic numerals;
Roman numerals;
space and paragraph characters;
special characters (//, Δ, ▲, ◊, ♦, ...);
font types etc.

Thus, the description of *V(x)* of any elementary information unit *x*, $x \in I^Q(D)$, is represented by an $\boldsymbol{A}$-word of the following form:

$$V(x) = v_1(x)\, v_2(x) \ldots v_{k(x)}(x),\ v_i(x) \in A,\ i = 1, 2, ..., k(x),\ k(x) \geq 1, \tag{26}$$

where each $\boldsymbol{A}$-letter $v_i(x)$ is taken from the alphabet $\boldsymbol{A}$. Note that the length *k (x)* of the $\boldsymbol{A}$-word *V(x)* depends on the choice of the element $x \in I^Q(D)$. Formula (26),

[58] Shirokov V.A.: Informacijna teoriya leksikografichnih sistem; Fenomenologiya leksikografichnih sistem.

by definition, is a fairly complete, in a certain sense, an exhaustive description of the elementary information unit $x$ in a given lexicographic system. A definite isomorphism is established between the class EIE $I^Q(D)$ and the set of descriptions $V(I^Q(D))$ using the reflection $S$. In other words, the set of descriptions of $V(I^Q(D))$ is a definite proper subset of the set $W(\boldsymbol{A})$: $V(I^Q(D)) \subset W(\boldsymbol{A})$, and $\boldsymbol{W(A)}$ is the set of all words of a finite length over $\boldsymbol{A}$, that is, sequences of the form $v_1 v_2 ... v_q$, $q < \infty$, $v_i \in \boldsymbol{A}$, $i = 1, 2, ..., q$. We consider that the word of zero length — 0 also belongs to $W(\boldsymbol{A})$: $\forall\, a \in W(\boldsymbol{A})\ \exists\, 0 \in W(\boldsymbol{A})$, such that $a * 0 = 0 * a = a$, where $*$ is the concatenation. Closedness with respect to the concatenation operation, i.e. the requirement: $a, b \in \boldsymbol{W(A)} \Rightarrow \exists\, c \in \boldsymbol{W(A)}$, $c = a * b$, as well as associativity in relation to it: $\forall$ a, b, c $\in \boldsymbol{W}(\boldsymbol{A}) \Rightarrow a * b * c = (a * b) * c = a * (b * c)$ turns $\boldsymbol{W}(\boldsymbol{A})$ into a semigroup with the semigroup operation «$*$» and the identity element 0.

The choice of the alphabet $\boldsymbol{A}$, through which $W(\boldsymbol{A})$ and $V(I^Q(D))$ are depicted, according to the algebraic tradition, here we do not substantiate and do not specify, however, we note that its generation is also a consequence of some lexicographic effect that develops in the speech system (acoustic) and its information-graphic interpretation. If we consider ordinary dictionaries, then it is natural to interpret the $\boldsymbol{A}$-word $V(x)$ as the text of a dictionary entry with a registry unit $x$.

In general, the structure of a semigroup is rather poor, and the construction $W(A)$ is too wide for its effective reception of the characteristic properties of language systems. To achieve this goal, it is necessary to introduce additional assumptions and restrictions with the help of which substructures characteristic of a natural language are distinguished in the structure of $W(A)$. This is achieved as follows.

Since each $V(x)$, as noted earlier, is an adequate and unambiguous representative (description) of the corresponding element $x$ from the $I^Q(D)$ system, the properties of this element should be adequately reflected in its structure. Given the linear nature of the object $V(x)$, which is represented by a linear sequence of characters from $\boldsymbol{A}$, we conclude that the only natural source of its structure can be only a certain set of its $\boldsymbol{A}$-subwords and certain relations between its elements. $\boldsymbol{A}$-subwords in the description of $V(x)$ are defined as $\boldsymbol{A}$-words made up of characters from the alphabet $\boldsymbol{A}$ contained in the description $V(x)$ in question and located in the $\boldsymbol{A}$-subword in the order induced by the order of letters in the description itself. The set of all $\boldsymbol{A}$-subwords of an $\boldsymbol{A}$-word of length $n$ (i.e., an $\boldsymbol{A}$-word composed of $n$ $\boldsymbol{A}$-letters) contains $2^n$ elements. We denote the set of all $\boldsymbol{A}$-subwords of the $\boldsymbol{A}$-word $V(x)$ as $B[V(x)]$.

The structure of the set of descriptions is introduced as follows. Suppose that for all descriptions of $V(x)$ there is a unique rule according to which from any $\boldsymbol{A}$-word $V(x)$ we can distinguish the set of $\boldsymbol{A}$-subwords $\beta(x) = \{\beta_i(x)\}$ with the following properties:

- the element $x$ belongs to the set $\beta(x)$;
- the whole description of $V(x)$ is an element of the set $\beta(x)$;

• the rule by which elements of the set β $(x)$ are distinguished is the same for all $V(x)$.

In the described manner, from any $V(x)$, the set β $[V(x)]$ of quantities (***A***-subwords) $\beta_i(x)$ of the following form is distinguished:

$$\beta\,[V(x)] \equiv \{\beta_i(x),\ i = 1, 2, ..., q\} \subseteq B[V(x)], \tag{27}$$

where $B\,[V(x)] = \{v_{i1}v_{i2}...\,v_{ip}\,,\ 1 \le i_1 < i_2 < ... < i_p \le k\,(x),\ p = 1, 2, ..., k\,(x)$, moreover:

$$v_{ij\hat{1}}\,\{v_{1(x)},\ v_{2(x)},\ ...,\ v_{k(x)}\,(x)\};\ x \in \beta\,[V\,(x)];\ V\,(x) \in \beta\,[V\,(x)],$$

$$\beta_{k(x)} \neq \beta_{m(x)} \text{ при } k \neq m. \tag{28}$$

Put by definition:

$$\beta\,[V\,(I^Q\,(D))] = \cup\limits_{x\,\in\,I^{Q\,(D)}} \beta\,[V\,(x)], \tag{29}$$

It's obvious that $V\,(I^Q\,(D)) \in \beta\,[V\,(I^Q\,(D))]$. Denote:

$$\beta_i = \cup\limits_{x\,\in\,I^{Q(D)}}^{q} \beta_i\,(x),\ i = 1, 2, ..., q, \text{ as well as } \beta = \cup\limits_i \beta_i. \tag{30}$$

It is clear that $\beta \equiv \beta\,[V\,(I^Q\,(D))]$. Note that some of the elements $\beta_i(x)$, $i = 1$, 2, ..., $q$ may be empty for certain $x \in I^Q\,(D)$; in this case, they are omitted in formulas (29), (30).

We introduce some structure on $\beta$ (and, accordingly, on $V\,(I^Q\,(D))$) — we denote it by the symbol $\boldsymbol{\sigma}\,[\beta]$; below we will call $\boldsymbol{\sigma}\,[\beta]$ the macrostructure $V\,(I^Q\,(D))$. The restriction of $\boldsymbol{\sigma}\,[\beta]$ to $V\,(x)$: $\boldsymbol{\sigma}\,[\beta]\ \big|_{\,V(x)} \equiv \boldsymbol{\sigma}\,(x)$ generates microstructure $V\,(x)$. The active formulation of this fact is to establish a procedure (operator, process...) $\boldsymbol{\sigma}$ that generates the structure $\boldsymbol{\sigma}\,[\beta]$ on $\beta$:

$$\boldsymbol{\sigma}: \beta \to \boldsymbol{\sigma}\,[\beta]. \tag{31}$$

On $\beta$, it is possible to generate several nonisomorphic structures $\boldsymbol{\sigma}[\beta]$, representing various aspects of the description of the corresponding language objects. Formally, the role of these structures can be any of the known data models (hierarchical, network, relational, object-relational, etc.), logical-mathematical models (in particular, logical calculi such as predicate logic), formal grammar constructions, etc. One of the possible mechanisms of structure formation may be as follows. We construct a table of structural elements β of the following form:

| $\beta_1$ | $\beta_2$ | ... | $\beta_q$ |
|---|---|---|---|
| $\beta_1(x_1)$ | $\beta_2(x_1)$ | ... | $\beta_q(x_1)$ |
| $\beta_1(x_2)$ | $\beta_2(x_2)$ | ... | $\beta_q(x_2)$ |
| . | . | | . |
| . | . | | . |
| . | . | | . |
| $\beta_1(x_M)$ | $\beta_2(x)$ | ... | $\beta_q(x_M)$ |

Some of the elements $\beta_i(x_j)$, obviously, can be empty, therefore, the lengths of the columns of the table, generally speaking, are different. The quantities $\beta_i$, $i = 1, 2, ..., q$ are interpreted as attributes (attribute names), and the sets $\mathrm{Dom}\,\beta_i \equiv \{\beta_i(x_1), \beta_i(x_2), \ldots, \beta_i(x_M)\}$, $i = 1, 2, ..., q$ — as the domains of these attributes. Then the structure $\sigma[\beta]$ can be realized in the form of the relational algebra $R[\beta]$ defined on the Cartesian product:

$$\prod_{i=1}^{q} \mathrm{Dom}\beta_i = \beta^{\otimes}. \tag{32}$$

In other words, if the structure **σ** is identified with some relational algebra R over $\beta^{\otimes}$, then the triple $\{V(I^Q(S)), \beta, R[\beta]\}$ is nothing more than a relational model, and the quinary $\{I^Q(S), D, V(I^Q(S)), \beta, R[\beta]\}$ defines some object-relational model. Moreover, $I^Q(S)$ represents a class of model objects; $\beta_i$ are interpreted as attributes (attribute names) with Dom $\beta_i$ domains and whose elements are $\beta_i(x)$, $x \in I^Q(S)$. It is clear that the set $\{x\}$ itself is separate domains (to shorten the presentation without further detail, we identify the element $x$ as belonging to the class $I^Q(S)$, with its name in $V(x)$), and $V(I^Q(S))$ as the set $\{V(x)\}$. The relational relations of the corresponding arities are determined, as always, in the form of certain subsets of the set $\beta^{\otimes}$. Their tuples are elements of the form:

$$(\beta_{i1}(x_{ji1}), \beta_{i2}(x_{ji2}), ..., \beta_{ir}(x_{jir})),\ i_1 < i_2 < ... < i_r;$$
$$x_{jim} \in I^Q(S),\ m = 1, 2, \ldots, r \tag{33}$$

Relational calculus in this model is defined in the usual way (see, for example, the work of J. Ullman[59]).

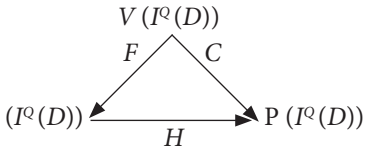
Due to the possibility of interpreting $I^Q(D)$ as classes whose elements are objects of any origin, an object-oriented interpretation of the model is natural. The relations between the elements of class $I^Q(D)$ are induced by the system of unary relations $r[\beta_1]$ on $\beta_1 = \{x\}$, and the reflection:

$$\mathrm{D}: V(I^Q(D)) \to I^Q(D);\ D\ r[\beta_1]. \tag{34}$$

Thus, the complex $I^Q(D)$ serves as a representative of the ontological nature of reality to be modeled, while $V(I^Q(D))$, $\beta$, **σ** $[\beta]$ represent its conceptual side.

Further consideration will be concentrated around the general physical structure, the carrier of which is the $I^Q(D)$ complex and which are developed and implemented in the medium $\{I^Q(D), S, V(I^Q(D)), \beta,$ **σ** $[\beta]\}$. Moreover, the set of properties and qualities of the $I^Q(D)$ complex, as described above, is divided into two not very clear and not very clearly separable parts. Note that in the conceptual scheme implemented in the description of $V(I^Q(D))$, these parts must be separated; in other words, a necessary condition for the correctness of its construction is the existence of a procedure that carries out such separation. All this can be reflected in the commutative diagram:

[59] Ul'man Dzh. Osnovy sistem baz dannyh. Moscow: Finansy i statistika, 1983. 334 s.




(35)

$\boldsymbol{F}V\ (I^Q\,(D)) = \Lambda\ (I^Q\,(D))$; $\boldsymbol{C}V\ (I^Q\,(D)) = P\ (I^Q\,(D))$; $\Lambda\ (I^Q\,(D)) \cap P\ (I^Q\,(D)) = \varnothing$, moreover, $\boldsymbol{H} \circ \boldsymbol{F} = \boldsymbol{C}$, where the symbol «∘» denotes a composition of reflection;

$$\Lambda\ (I^Q\,(D)) = \cup_{x \in I^Q\,(D)} \Lambda\ (x);\ P\ (I^Q\,(D)) = \cup_{x \in I^Q\,(D)} P\ (x). \tag{36}$$

On $\Lambda\ (I^Q\,(D))$ and $P\ (I^Q\,(D))$ macrostructures are induced:

$$\boldsymbol{F\sigma}\ [\beta] = \lambda\ [\beta]\ \text{i}\ \boldsymbol{C\sigma}\ [\beta]) = \rho\ [\beta] \tag{37}$$

and corresponding microstructures:

$$\lambda\ [\beta]\big|_{V(x)} \equiv \lambda\ (x);\ \rho\ [\beta]\big|_{V(x)} \equiv \rho(x) \tag{38}$$

as the restriction of $\lambda\ [\beta]$ and $\rho\ [\beta]$ to $V\ (x)$.

Note that diagram (35), i.e. objects $V\ (I^Q\,(D))$, $\Lambda\ (I^Q\,(D))$, $P\ (I^Q\,(D))$ and reflections $\boldsymbol{F}$, $\boldsymbol{C}$, $\boldsymbol{H}$ are common, the non-optional element of the lexicographic model, but can be interpreted as a mandatory element of the structure $\boldsymbol{\sigma}\ [\beta]$. The object $\Lambda\ (I^Q\,(D))$ corresponds to that part of the description $V\ (I^Q\,(D))$, which, in a certain sense, represents the form $I^Q\,(D)$, and $P\ (I^Q\,(D))$, respectively, is compared to that part of the description $V\ (I^Q\,(D))$, which is responsible for the content of $I^Q\,(D)$. The given specification practically confirms the opinion that OFS is universal, characteristic for objects of any origin since they are involved in the process of interaction with the subject who perceives or studies them.

**Definition 1.** Eight objects:

$$\{I^Q\,(D),\ S,\ V(I^Q\,(D)),\ \beta,\ \boldsymbol{\sigma}\ [\beta],\ \boldsymbol{F, C, H}\}$$

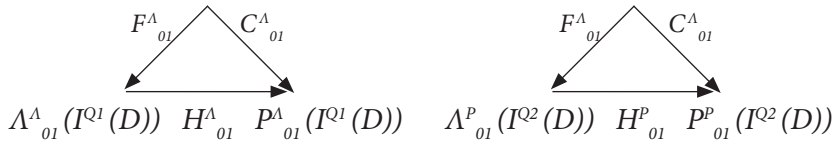
defines *an elementary lexicographic data model*, and its *concrete implementation defines an elementary lexicographic system*. Sometimes, for brevity, if there is no discrepancy, we denote the entire elementary lexicographic system by $V\ (I^Q\,(D))$.

Note that any element $\beta_i$ (or any combination of elements) that belongs to the structures $\beta$, $\boldsymbol{\sigma}[\beta]$, $\lambda\ [\beta]$, $\rho\ [\beta]$ can be interpreted as an elementary L-system. This allows us to distinguish some information-linguistic substructures in the structure of the initial elementary L-system, which we interpret as separate L-systems. Redefining the structure of the original L-system in this way, we get the L-model and the L-system in general position (not elementary); it is represented as the union of many elementary A-systems with possible reflections and connections between them. Thus, the general position L-system has the form of a graph:

$G = \{V = \{V_i\};\ R = =\{R_{kl}\}\}$, where $V = \{V_i\}$ is the set of vertices that are the elementary L-systems $\text{V}_i$ in G, and $R = \{R_{kl}\}$ is the set of edges of the graph$G$, $R_{kl}$, joining $V_k$ and $V_l$.

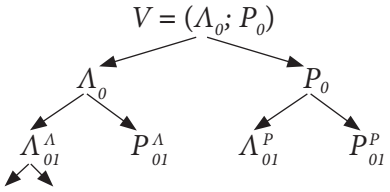
In particular, nothing prevents us from considering $\Lambda\ (I^Q(D))$ and $P\ (I^Q(D))$ as separate, autonomous elementary A-systems, and this makes this construction possible:

$$V\ (I^Q(D)) = (\Lambda\ (I^Q(D)) \equiv \Lambda_0(I^Q(D)) \xrightarrow{H_0} P_0(I^Q(D)) \equiv P\ (I^Q(D))$$




(39)

take note of the change in the type of lexicographic effect at the second level: instead of $Q$, we now have $Q_1$ and $Q_2$, respectively. Thus, we come to the complex of objects $I^{Q1}(D)$ and $I^{Q2}(D)$. Continuing this process, we obtain the recursive development of the lexicographic system $V\ (I^Q(D))$:




(40)

We call this process *a recursive reduction of the lexicographic system*. It resembles a kind of informational «microscope», which reveals increasingly finer details of the structure of the lexicographic system and induces a structure resembling a fractal one.

In what follows, we denote the process of recursive reduction of the L-system $V\ (I^Q(D))$ as $RR^-\ [V\ (I^Q(D))]$. The definition of this process includesthe characterization of all operators ***F***, ***C***, ***H*** at all available levels of recursive reduction, together with the results of their action, as well as all macro- and microstructures **σ**, λ, ρ.

The above construction makes up the content of the lexicographic model of data in general position:

$$\{I^Q(D),\ S,\ V\ (I^Q(D)),\ \beta,\ \boldsymbol{\sigma}\ [\beta],\ RR^-\ [V\ (I^Q(D))]\} \tag{41}$$

and lexicographical system:

$$\{I^Q(D),\ S,\ V\ (I^Q(D)),\ \beta,\ \sigma\ [\beta],\ RR^-\ [V\ (I^Q(D))],\ \boldsymbol{\Sigma}\}, \tag{42}$$

where the symbol **Σ** denotes its *architecture* as an information system.

The definition of individual elements of the model in formulas (41), (42) is represented by formulas (27)—(38). The **Σ** architecture is usually chosen in three levels, corresponding to ANSI/X3/SPARK (or simply ANSI/SPARK)[60]. In slightly

[60] ANSI/X3/SPARK/DBMS study group interim report.*FDT-Bull. ACM SIGMOD*. 1975. Vol. 7, No. 2. 140 p.

different wording, a similar architecture was presented in the report[61], which also introduced three levels of data abstraction: conceptual — physical — representations. With certain reservations, it is fair to say that the above three levels correspond to the conceptual, inner, and external levels of the ANSI/SPARK architecture. Mention should also be made of the work of Cycries and Klug[62], which is an informal introduction to the revised version of the report. Three levels of abstraction are represented in a large number of existing databases. The main components of the ANSI/SPARK architecture will be used in this interpretation:

$$ARCH\_LS = \{CM, EXM, INM; \Phi, \Psi, \Xi\}, \tag{43}$$

where the symbol CM denotes the conceptual model of the lexicographic system *LS*. The symbol $EXM = \{exM\}$ denotes the set of its external models that correspond to this conceptual model *CM*, and $INM = \{in\ M\}$ is the corresponding set of its inner models. The symbol $\Phi = \{\varphi\}$ denotes the set of reflections from *CM* in *EXM*:

$$\varphi : CM \rightarrow exM, \tag{44}$$

where $exM \in EXM$; accordingly, $\Psi = \{\psi\}$ is the set of reflections of *CM* in *INM*:

$$\psi : CM \rightarrow in\ M, \tag{45}$$

where $in\ M \in$ INM; $\Xi = \{\xi\}$ is the set of reflections of INM in EXM:

$$\xi\ (in\ M) = exM. \tag{46}$$

At the same time, we dwell on such an interpretation of the elements of architecture.

The conceptual model has the following properties:

1. *Semiotics*. The conceptual model is implemented in the environment of a certain sign system. Its construction is based on a joint examination of the subject area, the field of thinking, and the sign area.

2. *Semantics*. In the categories of the model objects are displayed and the relationships between them, essential from the point of view of an adequate description of knowledge about the subject area. The model should not depend on the logical, physical and external representation of the data.

3. *Unambiguity*. The conceptual model describes the subject area according to the principle of unique naming, according to which each sign of the model has one meaning, one sense. Homonymy and polysemy in the definition of model elements are removed. The use of qualifications and contextual mechanisms is not allowed.

4. *Consistency*. For each state of the conceptual model, the boundaries between its categories are absolute: When classifying or describing objects of a sub-

[61] CODASYL DBTG 1971. CODASYL Data Base Task Group April 71 Report. ACM New York, 1971.

[62] Tsichritzis D. and Klug A. (eds.).The ANSI/X3/SPARK Framework? AFIPS Press, Nontvale N. J., 1978.

ject area, each of them corresponds to a uniquely defined set of states that do not intersect in pairs.

5. *Integratedness*. The conceptual model combines the ideas of various specialists about the subject area. The inconsistency of these representations is fixed using integrity restrictions and is eliminated through special procedures.

6. *Specialized*. All conceptually defined properties of model objects must be interpreted on certain types of data.

7. *Algorithmization*. All objects of the conceptual model, as well as relations, reflections, and operations on them should have a finite description, i.e. have an interpretation in the form of algorithms of finite complexity over constructive objects.

Thus, the conceptual model (conceptual level of presentation) of the subject area is a sign, semantic model in which the ideas of various specialists about the subject area are integrated into an unambiguous, final, and consistent form.

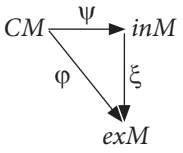
In the internal model (internal level of presentation), the types, structures, and formats of presentation, storage and manipulation of data, the algorithmic base, and the operating-software environment into which the conceptual model is «immersed» when it is implemented, are determined.

The external model (external level of presentation) reflects the views of end users and, therefore, applied programmers on the information system. It implements a set of procedures that allow the user to carry out authorized contacts and manipulate data presented at the inner level.

Several inner and external models can correspond to one conceptual model; therefore, the reflections $\Phi$: $CM \to EXM$ и $\Psi$: $CM \to INM$, generally speaking, are not injective. But the reflections $\varphi$: $CM \to exM$, и $\psi$: $CM \to inM$ are constructed by injective (but usually not bijective). Define the set $\Xi$ of reflections INM in EXM: for $\forall\ inM \in INM$ and $\forall exM \in EXM \exists\ \xi \in \Xi$ such that:

$$\xi\ (in\ M) = exM. \tag{47}$$

Moreover, the reflections $\varphi$, $\psi$, $\xi$ have constructed in such a way that the diagram:




(48)

is commutative: $\xi \circ \psi = \varphi$. The commutability requirement for this diagram is essential because it guarantees consistency between all levels of the system architecture. In turn:

$$\begin{aligned} CM = \{&Ob\ (LS);\ RelOb\ (LS);\ Mor\ (Ob\ (LS),\ RelOb\ (LS));\\ &Res\ (Ob\ (LS),\ RelOb\ (LS);\ Mor\ (Ob\ (LS),\ RelOb\ (LS)))\}, \end{aligned} \tag{49}$$

where *Ob(LS)* is a set of objects and categories of*LS; RelOb (LS)* is a set of connections (relations) between objects and categories LS *LS; Mor (Ob(LS), RelOb (LS))* is

a set of possible operations (processes) on *Ob(LS)* and *RelOb(LS); Res(Ob, RelOb, Mor(Ob,RelOb))(LS)* is a set of integrity constraints. All these elements have an unambiguous interpretation in terms of the lexicographic data model $\{I^Q(S), d, V(I^Q(S)), \beta, \sigma[\beta], R\, R\downarrow [V(I^Q(S))]\}$.

The inner model of the lexicographic system includes the following elements:

$$INM = \{D(t, s, f), ALG(D(t, s, f)), OS, PL\}, \quad (50)$$

where *D(t, s, f)* is the set of data specified by types, structures, and formats *t, s, f*, respectively, with the help of which SM elements are represented at the inner level; *ALG(D (t, s, f))*is a set of algorithms (processes) of data processing and manipulation; OS is a set of operating platforms and *PL* is a set of programming languages (including procedural languages such as DBMS), which implement *D(t, s, f)* and *ALG(D (t, s, f))*.

The external model of the lexicographic system is presented in the form:

$$EXM = \{IF, SC, FUNC, PROC, APR\}, \quad (51)$$

where *IF* is the system interface; *SC* is a set of scenarios; *FUNC* is a set of features; *PROS* is the set of valid processes, *APR* is the set of applications.

The specific information and linguistic implementation of the lexicographic system in a specific computer environment will be called the lexicographic database (LBD).

## Lexicographic systems and dictionaries

When applied to traditional dictionaries and vocabulary complexes, lexicographic systems acquire a simple and transparent interpretation. Indeed, the alphabet of the L-system is identified with the sign system of the dictionary (including special characters), the class EIE $I^Q(D)$ — with a set registry units (objects of lexicography — $\{x\}$), a set of descriptions $V(I^Q(D)) = \{V(x)\}$ — with a set of dictionary entries, where the uppercase units run over the set $\{x\}$, and $\Lambda(x)$ and $P(x)$ — with the left and right parts of the corresponding dictionary entries, etc.

The use of L-system constructs provides means for far-reaching generalizations of the traditional problems of lexicography, and many of its unsolved problems acquire their natural solution in the L-system paradigm.

Let us give an example of the problem of organizing dictionary entries by various criteria, which is a traditional dictionary does not have a natural solution. In the L-system, this problem reduces to defining a classification system on the set $I^Q(D)$, which serves as the basis for constructing the corresponding search apparatus and is implemented by the appropriate means of the internal and external models of the L-system. The simplest of these classifications — an alphabetical one — generates the so-called lexicographical ordering of the set $I^Q(D)$ and, therefore, the entire dictionary. Morphemic classification, which consists of distin-

guishing classes of words with the same bases, leads to the nesting principle, which is characteristic of word-formation type dictionaries and is also used in bilingual translation dictionaries. Grammar classification can give rise to a whole range of dictionaries. This list can be continued.The traditional paper dictionary is naturally ordered only according to one of the classification principles (although there are attempts to combine different classification principles in one dictionary). Quite different possibilities open up for general lexicographic systems and their implementations — computer dictionaries, the search apparatus of which can simultaneously include all of the above, as well as many other classifications.

In the structure σ [β] of the L-system $V(I^Q(D))$, we select a subset **A** of a special type whose elements $A$ ($A \in$ A) will be called automorphisms of the L-system $V(I^Q(D))$. The meaning of these elements is that they provide internal displays $V(I^Q(D))$, that is, a display:

$$A\colon V(I^Q(D)) \to V(I^Q(D)) \tag{52}$$

of a special character, more precisely — a display between separate dictionary entries $A$: $V(x) \to V(y)$ for different $x$ and $y$. In specific lexicographic works, automorphism $A$ can, in particular, ascertain the presence of reference dictionary entries such as, for example, ***x*** *see* ***y***. The indicated automorphism defines such a display of dictionary entries: $V(x) \to V(y)$. Its identifier is, as a rule, some reference pseudoword (in the given example — *see* ***y***), which matches the dictionary entry $V(x)$ to its correspondence $V(y)$. But the structure of the automorphism ***A*** can be more complicated than in this example.

First, the length of the references set can be greater than «one», that is, have a chain of recursive nature of deployment:

$$V(x) \to \{V(x')\} \to \ldots \to \{V(x'')\} \to \ldots$$

Also, display $V$ (x) $\to V$ (y) can represent a whole bunch of references. This is implemented, for example, when the dictionary entry $V(x)$ has the following structure: **x**, **x′** , **x″** , ... *см.* **y**, **y′**, **y″**, .... .

Elements of the automorphisms set $A$ do not have to be specified explicitly. Moreover, the establishment of vocabulary automorphisms, as a rule, is not formalized, since it is a rather difficult task related to the disclosure of internal laws and the hidden structure (symmetry) of the L-system. One of these hidden symmetries, namely, the hidden symmetry of the verb subsystem of the L-system of the academic explanatory Dictionary of the Ukrainian language, will be investigated below.

The sets of displaying ***H*** and ***A*** generate the macrostructure of the L-system. The macrostructural elements of the L-system also include ***Fσ*** [β] = λ [β] and ***Cσ*** [β]) = ρ [β] defined by formula (37). Local displaying (they are constructed as a restriction of the corresponding macrostructures on $V(x)$): $H|_{V(x)}$, and also $\lambda[\beta]|_{V(x)} \equiv \equiv \lambda(x)$; $\rho\ [\beta]|_{V(x)} \equiv \rho\ (x)$ — the last two are defined by formula (38)) determine the microstructure, which implicitly reflects the semantics of the subject domain, which is the object of this particular L-system.

The establishment and determination of the microstructures of lexicographic systems allow us to formalize, and in many cases automate the process of constructing the structures of the corresponding vocabulary databases, which provides the structural approach with significant advantages in designing the elements of linguistic support for information systems. The second part describes in detail a specific example of the application of the theory and methodology of lexicographic systems to the construction of a lexicographic database of the academic explanatory dictionary of the Ukrainian language based on its text.

Below we introduce some concepts that allow us to formally determine structures that are induced on lexicographic systems by special form displaying. These displaying are not only induce natural structures on specific lexicographic systems but also provide certain tools for generating structural classification of lexicographic systems and a classification methodology for dictionaries. Using these displaying, it is possible to formulate the concept of proximity on a set of words (in general, units of any level) — we use the term *pseudotopology* — and even the distance (pseudo-distances) between words.

Let's consider some elementary L-system $V\,(I^W(L)) = \{V\,(x)\}$. Any its dictionary entries *Vi* can be represented as:

$$V_i = x_0^{\,i}\,\pi_1\,\xi_1^{\,i}\,\pi_2\,\xi_2^{\,i}\,\pi_3 \ldots \pi_n\,\xi_n^{\,i}, \tag{53}$$

where

$$\cup_i \cup_j \xi_i^j = I^W(L), \tag{54}$$

where $I^W(L)$ is the set of words of the language $L$ contained in all dictionary entries of the L-system $V\,(I^W(L))$; $\pi_i$ — separators between words (punctuation marks, function marks, special characters, abbreviations, etc., they are summarized under the name mark-up symbols). So, an arbitrary dictionary entry $V_i$ is presented in the form of a union:

$$V_i = \partial V_i \cup M_i, \tag{55}$$

where the following notations are accepted: $\partial V_i \equiv x_0^{\,i}$ — the boundary element of the dictionary entry, its headword; $M_i$ — the inner part of the dictionary entry:

$$M_i \equiv \pi_1\,\xi_1^{\,i}\,\pi_2\,\xi_2^{\,i}\,\pi_3 \ldots \pi_n\,\xi_n^{\,i}. \tag{56}$$

Thus, any dictionary entry is a union of the inner part and the border, and the entire dictionary is represented as a union of many inner parts with a set of borders. Denote:

$\partial V = \cup\, \partial V_i$ is the boundary of the elementary L-system $V\,(I^W(L))$;

$M = \cup\, M_i$ is the inner part of the elementary L-system $V\,(I^W(L))$.

Thus:

$$V = \partial V \cup M. \tag{57}$$

**Definition 2.** The lexicographic system $V(I^W(L))$ is called closed if for $\forall\, \xi \in I^W(L)$ $\exists\, V\,(x_0\,\alpha) \equiv V\,\alpha \in V$, such that $x_0\alpha \equiv x_\xi$, where $x_\xi$ is the original form of the word $\xi$.

This definition means that all the words in the dictionary entries of the L-system are also the headwords of some other dictionary entries.

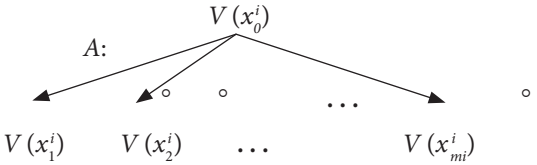
We define an automorphism $A$ of a special type as follows. Let's will name a set of words:

$$<V_i> \equiv (x_1^i, x_2^i, ..., x_{mi}^i); m_i \leq n_i \quad (58)$$

as a tuple of the dictionary entry $V_i$, if it is a set of canonical (dictionary) forms for the words $\xi_1^i, \xi_2^i, ..., \xi_{mi}^i$, respectively. All words with $M_i$ are necessarily included in the tuple. When considering specific examples, in particular, when studying the structure of sensible lexicographic systems, lexicographic expediency leads to the restriction of the tuple to lexemes that are included only in interpretation formulas (dictionary definitions), and except for the so-called stop words (those that do not give a significant contribution to semantics). We define an automorphism $A$ using the formula:

$$A : V(x_0^i) \equiv V_i \rightarrow \{V(x_k^i), k = 1, 2, ..., m_i\}. \quad (59)$$

The graphic representation of the formula (58) has the following form:




(60)

Formulas (59), (60) mean that a dictionary entry $V(x_0^i)$ with a headword (border) $x_0^i$ is associated with a dictionary entry $V(x_k^i)$, k = 1, 2, ..., $m_i$ with a headword $V(x_k^i)$, $k = 1, 2, ..., m_i$, respectively, if, of course, they do exist in $V(I^W(L))$. We recursively define the action of the operator $A$ on $V(x_1^i)$, $V(x_2^i)$, ..., $V(x_{mi}^i)$, and then on the results of its application to $V(x_1^i)$, $V(x_2^i)$, ..., $V(x_{mi}^i)$, etc., until the objects $V(x_j^i)$, $i, j = 1, 2, ...$ begin to repeat. Let's denote the set of dictionary entries $\{V(x_0^i), V(x_1^i), V(x_2^i), ..., V(x_{mi}^i), ...\}$, obtained as a result of the action of operator $A$ defined above, as $V^A[x_0^i]$ ($V^A[x_0^i] \subseteq V$).

**Definition 3.** The set $V^A[x_0^i]$ is called $A|[x_0^i]$-sub-dictionary of the dictionary $V(I^W(L))$, if $A\ V^A[x_0^i] = V^A[x_0^i]$.

Thus, the $A$-sub-dictionary $V^A[x_0^i]$ is an invariant set in $V(I^W(L))$ concerning the action $A$.

**Definition 4.** Elements $V(x_0^i)$, $V(x_1^i)$, $V(x_2^i)$, ..., $V(x_{mi}^i)$, ..., $A$-sub-dictionary $V^A[x_0^i]$ will be called $A$-equivalent elements.

The last definition becomes clear if we take into account that the set of $A$ displays, that generates the $A$-sub-dictionary $V^A[x_0^i]$, induces an equivalence relation on the set of its dictionary entries, we will denote it by the symbol $EV^A[x_0^i]$. Let us denote as:

$$W = V \setminus EV^A[x_0^i] \quad (61)$$

is a factor set in $V$to $EV^A[x_0^i]$. Therefore: $V = W \cup V'$. Accordingly with the definition of $V'$:

$$AV' = V'AM\ W = V' \text{ for some } M \geq 0. \tag{62}$$

Then we say that $V$ is a half-direct sum of the dictionaries $W$ and $V'$:

$$V = W \succ V'. \tag{63}$$

A dictionary $V$ with such a structure will be called $A$-indecomposable.

**Definition 5.** Dictionary $V\ (I^W(L))$ is called $A$-irreducible (completely irreducible) if it does not have its own $A$-sub-dictionary.

From the last definition, it follows that if $V$ is an $A$-irreducible dictionary, then for arbitrary $x, y \in I^W(L)\ \exists\ N \geq 0$, such that $V(y) \subseteq AN\ V(x)$.

**Definition 6.** Dictionary $V$ is called completely $A$-reducible if it can be represented as:

$$V = \cup\ V^i, \text{ moreover, } V^i \cap V^j = \varnothing \text{ when } i \neq j, \tag{64}$$

where $V^i$ is an $A$-irreducible dictionary. In this case, we say that $V$ decomposes into a direct sum of $A$-dictionaries $V^i$:

$$V = \sum_i \oplus\ V^i. \tag{65}$$

Consider some $A$-irreducible dictionary $V$:

$$V = \cup_{x_0^i \in S_0} V(x_0^i), \tag{66}$$

where $V(x_0^i)$ is a dictionary entry with the headword $x_0^i$. The automorphism $A$ induces the map$S_0 \to S_0$, i.e. it defines some displays of headwords set into *itself*. The specified display is defined as follows: we assume that if:

$$A : V(x_0^i) \equiv V_i \to \{V(x_k^i), k = 1, 2, ..., m_i\},$$

TO

$$A : x_0^i \to \{x_k^i, k = 1, 2, ..., m_i\}. \tag{67}$$

From the foregoing, it follows

**Theorem 1.** For the $A$-irreducible dictionary $V$, there are no $A$-invariant subsets in $S_0$. Therefore, for any $x, y \in S_0\ \exists\ N(x, y) \geq 0$ such that $A^N x = y$.

In other words, the path $A^N x$ for sufficiently large $N$ passes through all points of the set $S_0$. Therefore,

**Theorem 2 holds.** The complete $A$-path on the graph $G^A(S_0)$ is always closed.

Denoteinf $N(x, y) = \rho(x, y)$.

**Definition 7.** The number $\rho(x, y)$ is called the $A$-pseudo-distance from the word $x$ to the word $y$.

The number $\rho(x, y)$ shows the minimum number of steps that can be taken from the word $x$ to the word $y$ using algorithm $A$.

Let the set of words:

$$(x_1^i, x_2^i, ..., x_{mi}^i; mi \leq ni) = \tau^i \equiv \tau(x_0^i) \tag{68}$$

is a tuple of the dictionary entry $V_i$, that is, it is a set of source forms for the words $\xi_1^{\,i}, \xi_2^{\,i}, ..., \xi_{mi}^{\,i}$, respectively. It is clear, that $x_0^{\,i}$ also belongs to $\tau(x_0^{\,i})$.

**Definition 8.** We call $\tau(x_0^{\,i})$ a closed area of the point $x_0^{\,i}$.

**Definition 9.** A set

$$\{\varnothing, \tau(x_0^{\,i}), i = 1, 2, ..., \text{Card } S_0\} \tag{69}$$

is called the $(V, A)$-pseudotopology of an elementary L-system.

The concept of pseudotopology can play an important role in the theory of lexicographic systems since this concept it becomes possible to formalize the concept of the proximity of elementary information units (words). Namely, the intersection of the areas $\tau(x_0^{\,i}) \cap \tau(x_0^{\,j})$ determines the degree of proximity of the lexemes $x_0^{\,i}$ and $x_0^{\,j}$, which follows from the L-system.

## Lexicographic environment

In reality, language objects function in their integrity, not divided into separate components of the conceptual representation. In the lexicographic system, this is manifested in the fact that when modeling language objects using the theory of L-systems, the task of integrating various types of lexicographic effects, as well as combining and reconciling heterogeneous (heterogeneous) lexicographic structures, organically arises. In turn, this requires coordination between all elements of the architecture of L-systems subject to the integration process.

Numerous specific computer implementations of integrated linguistic objects experiments have led to the conclusion that it is necessary to make a special cultural information environment, which from the very beginning would be adapted to the integration processes of various lexicographic systems and contain the necessary tools and constructs for implementing these processes, as well as fixing its results in the form of integrated lexicographic systems having excellent lexicographic structures. As a result, the concept of a lexicographic environment was suggested [63].

**Definition 10.** We consider that the lexicographic medium (L-medium) ***ML*** is given if:

1. A class of elements *Ob**ML*** has been defined, each of which is a diagram of the form (48) and represents a certain L-system (not necessarily elementary one). Elements *Ob**ML*** are called objects of the L-medium ***ML***, and we will denote them by capital Latin letters: A, B, C, ...

2. For each pair of objects *A, B* from ***ML***, a set *HomML (A, B)* is given, which is called the set of morphisms A in B, instead of $f \in Hom_{ML}(A, B)$ it also can be written:

[63] Shyrokov V.A., Rabulets O.H. Formalizatsiia u haluzi linhvistyky. *Aktualni problemy ukr. linhvistyky: teoriia i praktyka: Zb. nauk. prats*. Kyiv, 2002. **5**. S. 3—27; Rabulets O.H. Intehrovani leksykohrafichni systemy: Avtoref. dys. … kand. tekhn. nauk. 05.13.06. Nats. bibl. Ukrainy im. V.I. Vernadskoho. Kyiv, 2002. 18 s.

$$f\colon A \xrightarrow{f} B \text{ or } A \to B.$$

At the same time:

$$\begin{gathered} f\colon CM_A \to CM_B; f\colon INM_A \to INM_B; f\colon EXM_A \to EXM_B; \\ f(\varphi_A) = \varphi_B; f(\psi_A) = \psi_B; f(\xi_A) = \xi_B \text{ и } f(\xi_A) \circ f(\psi_A) = f(\varphi_A). \end{gathered} \tag{70}$$

3. For each triple of objects *A, B, C*) from ***ML***, the reflection is given:

$$\mu\colon Hom_{ML}(A, B) \times Hom_{ML}(B, C) \to Hom_{ML}(A, C) \tag{71}$$

(image μ (*f, g*) of a pair *(f, g)*, where $f \in Hom_{ML}(A, B)$, $g \in Hom_{ML}(B, C)$, will be denoted by $f \circ g$ or $f g$, and will be called the composition of morphisms *f* and *g*).

4. The sets $Hom_{ML}(A, B)$ and the composition of morphisms satisfy the following axioms:

(a) Associativity: for each of the three morphisms *f, g, h*:

$$A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{h} Df(g\,h) = (f\,g)\,h. \tag{72}$$

(b) The existence of a unit: for each $A \in Ob\ \mathbf{ML}$there is a morphism $1_A\colon A \to \to A$, ($1_A \in Hom_{ML}(A, A)$), such that $1_A f = f$ і $g\,1_A = g$ for arbitrary morphisms $f \in Hom_{ML}(B, A)$ and $g \in Hom_{ML}(A, B)$.

(c) If the pairs $(A, B)$ and $(A', B')$ are different, the intersection of $Hom_{ML}(A, B)$ and$Hom_{ML}(A', B')$ is empty.

Let two lexicographic environments $\mathbf{ML}_1$ and $\mathbf{ML}_2$ be given. The covariant functor *F* from $\mathbf{ML}_1$ to $\mathbf{ML}_2$ consists of:

(a) the reflection $A \to F(A)$, which compare each object $A \in Ob\ \mathbf{ML}_1$ to the object $F(A) \in Ob\ \mathbf{ML}_2$;

(b) reflections $F(A, B) : Hom_{ML1}(A, B) \to Hom_{ML2}(F(A), F(B))$ for covariant, and $F(A, B) : Hom_{ML1}(A, B) \to Hom_{ML2}(F(B), F(A))$ for the contravariant functors defined for all pairs $(A, B)$ of objects from $\mathbf{ML}_1$ and such that (if instead of $F(A, B)(u)$ we write $F(u)$) $F(1_A) = 1_{F(A)}$ і $F(vu) = F(v)\,F(u)$ ($F(vu) = F(u)\,F(v)$ respectively).

The structure of the lexicographic environment includes a set of elements necessary for representing the effects of grammatical and lexical semantics, as well as many other natural language phenomena. This also applies to the representation of individual relations and linguistic phenomena, which are abstracted from the linguistic continuum — inflection, word formation, orthoepy, phraseology, synonymy, antonymy, etc. Lexicographic media are especially useful in modeling integrated and heterogeneous L-systems, which will be discussed in the next two subsections. In particular, etymological L-systems, being very heterogeneous in the sense described below, receive a natural and, in our opinion, extremely convenient cultural and informational environment, since the described construction is easily adapted to model each of these phenomena and relationships separately with subsequent integration of them into a single lexicographic complex.

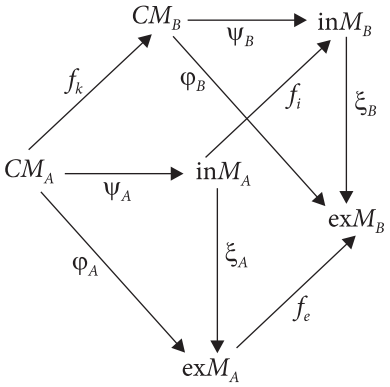
## Integrated L-systems and methods for their construction

The structure of the L-medium introduced by the method described above is a convenient formal object for the formation of complex lexicographic constructions, which combine into a whole set of separate heterogeneous L-systems. The heterogeneity of L-systems to be integrated is a multi-faceted concept. We consider L-systems that are heterogeneous at all levels of architecture: conceptual, internal, and external. Such an approach involves the development of methods for integrating conceptual models, methods for presenting data and operational-software platforms, as well as coordinating external representations of the corresponding conceptual schemes and their internal representations.

By the integration of L-systems as information systems, we mean the achievement of the possibility of simultaneous and joint use by an application of several information systems as a whole. So, for an application, an integrated set of different lexicographic databases (LBD) should look like a single LBD. To a certain extent, these representations can be transferred to traditional dictionaries, which can also be integrated L-systems, in particular, the explanatory dictionary is one of the brightest examples of such an L-system.

In this section, the maximum emphasis is placed on the integration of L-systems at a conceptual level. The basic objects to be integrated, are diagrams of the form (48), which we consider as objects of a certain L-medium.

Let there be two objects $A$ and $B$, and also $f\colon A \to B$, $f \in Hom_{ML}(A, B)$. Let's make $f$ as a special morphism in the form of a three-component vector: $f = (fc, fi, fe)$ such that the diagram:




(73)

is commutative. The last condition is equivalent to the validity of such equalities:

$$\psi_B \circ f_k = f_i \circ \psi_A;\ \varphi_B \circ f_k = f_e \circ \varphi_A;\ \xi_B \circ f_i = f_e \circ \xi_A. \tag{74}$$

Diagram (73) and equalities (74) provide necessary formal means in the language of which the processes of building an integration architecture can be formulated, which follows from the further discussion.

Morphism $f = (f_k, f_i, f_e)$, $f: A \to B$ is called regular if it satisfies the conditions:

• of full certainty, according to which any state of $A$ corresponds to one and only one state of $B$;

• of interpretability, according to which any element that belongs to the set $\{\boldsymbol{\sigma}_A [\beta_A], RR \downarrow [V(A)\}$ uniquely corresponds to an element from the set $\{\boldsymbol{\sigma}_B [\beta_B], RR \downarrow [V(B)]\}$;

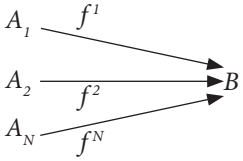
• of reproducibility, according to which a change in any state of $A$, which is carried out by a certain operator from $\{\boldsymbol{\sigma}_A [\beta_A], RR \downarrow [V(A)]\}$, corresponds to an adequate change in the corresponding state, which is performed by a certain operator from $\{\boldsymbol{\sigma}_B [\beta_B], RR \downarrow [V(B)]\}$.

Along with the ability to integrate conceptually heterogeneous lexicographic systems representing various linguistic phenomena, an equally important aspect of the integration architecture is the ability to achieve a high degree of independence of applications from a database management system (DBMS) and ensure their mobilityto various types of DBMS. The problem of program mobility is generally formulated as providing the ability to run a particular program on different computer platforms without changing it. The issues of program mobility concerning the DBMS are determined similarly.

Theoretically, there are the following possibilities: 1) the creation of a universal programming language and the requirement of its universal application; 2) providing each computer with compilers for all programming languages with proper standardization; 3) the introduction of a platform-independent intermediate language with its instrumental implementation at the virtual machine level with built-in interfaces in any platform; 4) the use of emulation methods; 5) the use of computer networks under the assumption that at least one of the network computers is provided with the necessary compiler.

An analysis of these suggestions leads to the conclusion that the third one is rational. In different variations, they are trying to implement it, for example, in systems such as JAVA, standardization of machine-independent languages SQL, SQL2, etc. Note: These suggestions are outwardly similar to the construction of a common conceptual model, which once again confirms the centrality of the conceptual model in the architecture of the information system. Therefore, we will continue to consider the development of integration processes of L-systems at a conceptual level (or the integration of conceptual representations of L-systems).

The meaning of the integration procedure is as follows. Assume that instead of a single object $A$, which is present inthe diagram (73), we have a set of objects $A_1$, $A_2$, ..., $A_N$. We construct a fan-shaped reflection with morphisms $f = (f^1, f^2, ..., f^N)$:




(75)

$$f^p : A_p \to B,\ p = 1, 2, \ldots, N,$$

where each $f^p$ is given by the three-component vector $f^p = (f_c^p, f_i^p, f_e^p)$ with properties (74). In the process of applying these morphisms, an integration procedure is used, which has the following steps.

First, an auxiliary L-system $B$ is constructed from the following elements: sign systems $\boldsymbol{A}(A_p)$; structures $\beta_p$, $\sigma_p$ $[\beta_p]$ и and processes $RR_p \downarrow [V(I^Q(S))]$ of all $A_p$. At this stage, the reflection $f^p : A_p \to B$, $p = 1, 2, ..., N$ is interpreted as an embedding reflection. So we get some non-elementary $L$-system B with independent components $A_p$, $p = 1, 2, ..., N$. The sign system for the L-system $B$ is built as a union: A $\boldsymbol{A}(B) = \cup \boldsymbol{A}(A_p)$.

Next, we introduce a special semantic procedure SEM, which ensures the identification of structural elements that coincide for at least two systems from the set $A_p$, $p = 1, 2, ..., N$. We assume that this procedure identifies not only the names of semantically identical attributes that exist in different L-systems $A_p$, $p = 1, 2, ..., N$, but also the corresponding domains of their meanings (domains). There can be various causes of bonds of structures belonging to different $A_p$, $p = 1, 2, ..., N$. Let's consider them in more detail.

We denote by $\varepsilon_{p1p2\,...\,pk}[\beta]$ the set of structural elements (with their domains) that belong to each of $A_{p1}$ $A_{p1}$, $A_{p2}$, ..., $A_{pk}$, $1 \le p_1 < p_2 < ... < p_k \le N$. We apply the semantic operation to it:

$$SEM\,(\varepsilon_{p1p2\,...\,pk}[\beta]) = \varepsilon^{SEM}_{p1p2\,...\,pk}[\beta] \tag{76}$$

and thus we obtain semantically identical general elements of the structure of L-systems with numbers $p_1, p_2, ..., p_k$. The presence of elements (74) makes it possible to construct an L-system with the following structure:

$$\boldsymbol{\sigma}^{SEM}_{p1\,p2\,...\,pk}[\beta] =$$

$$= (\varepsilon^{SEM}_{p1p2\,...\,pk}[\beta], [b^{SEM}_{p1}], [\beta^{SEM}_{p2}], ..., [\beta^{SEM}_{pk}]; R^{SEM}), \tag{77}$$

where $[\beta^{SEM}_i]$, $i = p_1, p_2, ..., p_k$ — structural elements of L-systems $A_i$, in which the identification of common elements is performed; $R^{SEM}$ is a combination of operations that operate in each L-system. The remaining structural elements of the L-systems $A_i$, $i = p_1, p_2, ..., p_k$, we denote as $S_i$, $i = p_1, p_2, ..., p_k$, , remain without changes.

In this way, the L-system $\underline{B}$ with the structure:

$$\{\sigma^{SEM}_{p1p2\,...\,pk}[\beta]; S_i, i = p_1, p_2, ..., p_k\} \tag{78}$$

integrates L-systems $A_i$, $i = p_1, p_2, ..., p_k$.

The set of structural elements (together with their domains that belong to all $A_p$, $p$= 1, 2, ..., $N$) we denote as $\varepsilon[\beta]$:

$$\varepsilon[\beta] = \overset{N}{\cap}[\beta]\,A_p. \tag{79}$$

As a result, we get an L-system $\underline{B}$ with the sign system $\boldsymbol{A}(\underline{B}) = \cup \boldsymbol{A}(A_p)$ and the structure $\beta(\underline{B})$, $\sigma[\beta](\underline{B})$, $RR \downarrow [V(\underline{B})]$, obtained as described above. Particu-

larly interesting is the case when a certain element or certain structural elements are present in all L-systems to be integrated, i.e. when ε [β] ≠ ∅. Then it is advisable to conduct continuous indexing of all $A_p$ according to the meanings of the corresponding domains belonging to the elements ε [β]. These structural elements acquire the status of included in each $A_p$. As experience shows, the lexical arrays (sets of word forms) of all languages participating in $A_p$ are elements of the structure of the marked type for natural language systems. Therefore, the problem arises of constructing a procedure of natural language indexing.

The L-system $\underline{B}$ constructed in the described way is called the integration of L-systems $A_p$, $p = 1, 2, ..., N$.

## Inhomogeneous L-systems and their indices

The presented theory of lexicographic mediums and the integration technique of L-systems provides effective tools for modeling complex language phenomena. Such phenomena are, as a rule, characteristic of large lexicons, such as, for example, multivolume explanatory dictionaries of national languages, large etymological dictionaries. Such examples include the Integrated Lexicographic System «Dictionaries of Ukraine», an open online version of which is available on the website of the Ukrainian Linguistic Portal at http://lcorp.ulif.org.ua/dictua/.

These and many other L-systems contain in their structure certain linguo-informational elements, often related to different, sometimes very loosely connected, sections of the information-cultural environment, which (sometimes for a methodical purpose, and sometimes due to the inability to establish systemic links between these sections) should be considered as distinct and independent. This introduces additional complexity into the structure of the L-system, giving it signs of heterogeneity. This is how, for example, phenomena of this type are displayed in etymological lexicography.

As the famous Russian etymologist Zh. Zh. Varbot notes: «Etymology (the science of the origin of words) is one of the oldest branches of linguistics. Its founders were ancient Greek philosophers, in whose writings the term ετυμολογία had appeared, formed from έτιμον «truth» and λογος «word, the doctrine», and originally it meant «science of truth», or the true meaning of words. During the existence of etymology as a science, understanding of its goals has repeatedly changed; the methodology of etymological studies has undergone even greater changes. In modern linguistics, the ultimate goal of the etymological analysis is to determine where (in which language, dialect), when, how (from which components — morphemes, according to what word-formation model), with what meaning the word arose and what changes in its form and meaning had created sound form and meaning of the word, known to the researcher» [64].

[64] Varbot Zh.Zh. Etimologiya. *Russkij yazyk: Enciklopediya*. Moscow: Drofa, 1997. S. 643—647.

So, the fundamental fact, which is the basis for the existence of etymology as a science, is the multiplicity of human languages that undergo significant changes during their evolution. This quote identifies the aspects that are required when conducting an etymological analysis. In particular, aspects are highlighted: linguo-geographical (the aspect of linguistic space), glottochronological (the aspect of linguistic time), morphonematic and derivational (structural-grammatical aspect), semantic, pragmatic, and other factors that have led to the emergence of modern forms and meanings of words. To clarify the role and essence of these aspects, it is necessary to carry out several complex comparative historical studies in a wide range of languages (including «dead» languages), dialects, dialects, dialects, etc. When performing this kind of research, one has to study various processes of interlanguage interaction (interference, contamination, divergence, convergence, assimilation, absorption, borrowing, ...), and to do this at different levels of different, sometimes insufficiently studied, language systems.

The fact of multiple languages is quite indicative and, in our opinion, is a consequence of the general law of evolution of complex systems, namely, the law of increasing the complexity of a system during its evolution. It is natural to assume that in the process of the evolution of a language as the main communicative system of mankind, its complexity at some stage (or stages) has already ceased to satisfy the needs of this communication, which has led to several bifurcations of this system, resulting in the emergence of a large number of languages. This was a one-time process because linguistic bifurcations arise in our time in different historical and geographical situations. Without analyzing the specific historical causes of linguistic bifurcations, we note their evolutionary aspect, which is accompanied by an increase in the complexity of the integrated cultural and communication information system of mankind. However, complexity, as an objective property of things, is directly related to the property of heterogeneity (which directly follows from the circle of ideas associated with the Kolmogorov definition of complexity) — therefore, bifurcation processes in a cultural and communication system must also cause an increase in its heterogeneity. In our opinion, it is the heterogeneity of etymological processes that is their distinguishing feature, which distinguishes them from many other language processes.

What are the determinants of this heterogeneity and how are they reflected in the system of lexicographic description of etymological processes?

The most significant of the factors of heterogeneity of the etymological system is the integral parameter, which «absorbs» the continuum of various linguistic and extralinguistic factors (phonological, grammatical, semantic, cognitive, historical, cultural, ...) that determine whether a certain language unit belongs to a particular language, and, finally, acquires an identifier in the form of the name of a particular language. In the lexicographic representation, this parameter, except, of course, the language marker, is identified (although not always complete-

ly) by its sign system. Thus, the etymological L-system (EL-system), the fundamental parameter of which is the type of lexicographic effect, determines parameter «Language» (or «Dialect», etc.) as the main parameter of heterogeneity. So, the EL-system initially acquires the features and designs of an integrated L-system with the lexicographic parameter «Language», which simultaneously represents the type of lexicographic effect with the same identifier *Li*. The lexicographic model, compared with a single language, takes on the form:

$$\{I^{Li}(S), A^{Li}, V(I^{Li}(S)), \beta, \sigma[\beta], RR\downarrow[V(I^{Li}(S))]\}, \tag{80}$$

where *Li* is the lexicographic effect that is identified with the language*Li*, $A^{Li}$is the corresponding sign system; other parameters are determined, as always. Integration over the parameter *Li* into a single heterogeneous lexicographic environment is nontrivial and is based on the etymological relation ε, which is decisive for a particular EL system and structurally is a certain specification of the semantic relation (76). If the base language $L^0$ is defined for the EL-system, then we assume that the units $x^0$ and $x^i$ (respectively: the word in the base language and the language *Li*) are in the etymological relation ε (related by the etymological relation ε):

$$<x^0 \mid \varepsilon \mid x^i>, \tag{81}$$

provided that their lexicographic development occurs in one dictionary entry with the head element $x^0$. The relation ε is naturally factorized into the direct sum of the individual relationships for each language:

$$\varepsilon = \sum_{i=1}^{N}{}^{\oplus}\,\varepsilon^i, \tag{82}$$

where *N* is the number of languages in the EL-system and can be represented by a diagonal matrix with elements $\varepsilon^i$, $i = 1, 2, ..., N$, on the diagonal.

In this form, the relation ε provides the basis for integrating the EL-medium into an inhomogeneous EL-system according to the parameter «Language». Given that the number of different languages in representative etymological dictionaries can reach several tens or even hundreds, it becomes clear that this parameter of heterogeneity is non-trivial both in linguistic and technological terms.

Lexicographically ordered set:

$$\begin{array}{c} \cup^{[\mathrm{LORD}\ xi]} x^i \\ x \in < x^0 \mid \varepsilon^i \mid\ x^i > \end{array} \tag{83}$$

we will call the index of the EL-system in the language *i* and denote it by the symbol IND[*i*]. The full index of the EL-system is a combination of indices for all languages available in the EL-system.

A generalization of the concept of heterogeneity of an L-system (L-medium) is obvious and is carried out as follows.

First, the class *Ob* ***ML*** is defined, whose objects are defined by diagrams of the form (48). For each object *A*, *A* ∈ *Ob* ***ML***, among the elements of the lexicographic structure:

$$Q^A; \beta^A; \boldsymbol{\sigma}^A[\beta^A], \tag{84}$$

where $Q^A$ is the set of lexicographic effects that appear in the definition of the L-system *A* (not necessarily elementary) and $\beta^A$; $\boldsymbol{\sigma}^A[\beta^A]$ are the elements of its structure; a subset of the elements that qualify as elements of heterogeneity is determined:

$$Q^{AI}; \beta^{AI}; \boldsymbol{\sigma}^{AI}[\beta^{AI}]. \tag{85}$$

Similarly, for each pair of objects *A, B* (*A, B* ∈ *Ob* ***ML***)**,** and the corresponding set of morphisms $Hom_{\boldsymbol{ML}}(A, B)$, a subset is defined:

$$I\_Hom_{\boldsymbol{ML}}(A, B), \tag{86}$$

which is non-trivial only on the elements of heterogeneity:

$$Q^{AI}; \beta^{AI}; \boldsymbol{\sigma}^{AI}[\beta^{AI}]\ ; Q^{BI}; \beta^{BI}; \boldsymbol{\sigma}^{BI}[\beta^{BI}]. \tag{87}$$

A lexicographic medium with distinguished elements of heterogeneity $Q^{AI}$; $\beta^{AI}$; $\boldsymbol{\sigma}^{AI}[\beta^{AI}]$, where *A* runs through the class of objects *Ob* ***ML***, and a subset of morphisms $I\_Hom_{\boldsymbol{ML}}(A, B)$ for each $Q^{AI}$; $\beta^{AI}$; $\boldsymbol{\sigma}^{AI}[\beta^{AI}]$, where *A, B* ∈ *Ob* ***ML***, will be called an inhomogeneous A-medium. Determining the parameters of heterogeneity is a non-trivial linguistic problem, which in the case of two languages reduces to the construction of the so-called intermediate language (tertium comparationis).

# THE THEORY OF LINGUISTIC STATES

## A. Kolmogorov's formalization of the grammatical concept of the case, the concepts of states of linguistic units, and grammatical observables. Development of these ideas by A. Zaliznyak [65]

Based on our belief that neither language units immediately nor grammatical or semantic categories should be the objects of conceptual representation in linguistics but entities «intermediate» in respect of language with psychophysical conditions and processes that have a place in the language-thinking apparatus of human as their phenomenological correlates we are to find out their role in the modeling language and recall in this regard the initiative of the repeatedly cited in this book great Russian mathematician A.N. Kolmogorov, which, as far as we know, was the first to use the concept of word states when attempting to build a strict definition of the case in Russian.

A.N. Kolmogorov himself did not publish his works on linguistics, so in our presentation of his ideas, we follow V.A. Uspensky's paper [66], where they were originally presented. Let us turn to the Kolmogorov — Uspensky approach concerning the definition of the case that we present below with our comments. We give V. Uspensky's text in italics while our comments are numbered and presented in Roman type.

Thus, A.N. Kolmogorov proposed a definition of the case as follows.

### § 1

*Objects can be in different states. Thus, the object named «молоко» («milk») in Russian may be in the following states: it can boil, it can be absent, it can be drunk by a cat, it can be drunk by a dog, etc. The states are expressed in language through sen-*

[65] This section uses the results of the works: Shirokov V.A. Grammatika kak fenomenologicheskaya problema; Shirokov V.A., Shevchenko I.V. Gramatika u fenomenologichnomu vymiri; Shirokov V., Shevchenko I. On the phenomenological approach to grammar.

[66] Uspenskij V.A. K opredeleniyu padezha po A.N. Kolmogorovu. *Byull. Ob'edineniya po problemam mashinnogo perevoda*. 5. Moscow: I MGPIIYA, 1957. S. 11—18.

*tences that involve a noun that is the name of this object. The above states of milk are expressed in Russian sentences such as «молоко кипит» («milk is boiling»), «молока нет» («there is no milk»), «кошка пьёт молоко» («cat drinks milk»), «собака пьёт молоко» («dog drinks milk»). When expressing states in which there is an appropriate object with the help of sentences, its name is used in one form or another (in the examples above —«молоко», «молока»).*

**Comment 1.** As you can see, there Kolmogorov speaks about the state of the object. The states of the object are given in some contexts, which may contain either the name of the object in the appropriate form or some of its abstraction, which allows the substitution in its place of this name (or some other) of the object in the appropriate form (forms). Contexts of the said kind serve as determinants of states and therefore may in some way be identified with them. That is why below we discuss only the forms of nouns that serve as names of the relevant items, the contexts in which they can operate, and some relationships between these contexts. We, therefore, still keeping the concept of «state of the object» without the interpretation, will continue to try to operate only with states of these forms, suggesting that they are a result of a perceptual-sensory and language-thinking processes, outlined below in «Linguistic world view and its conceptual representation», which has led to the formation of the psychophysical states of the language-thinking apparatus so that the forms themselves are observable components of the state, and the complete state of the word is a certain conceptual abstraction of a sum of all the reasonable contexts.This implies (so far implicitly) the existence of a correspondence between a word and its state: $s: x \to s(x)$ where $x$ — a certain word; $s$ — correspondence between$x$and$s(x)$ — an object representing the state of the word$x$, the determinants of which are elements of the material means for expressing semantics (both grammatical and lexical). Thus, $s(x)$ expresses the full state of the word$x$ in some context. This means that this state includes its grammatical part, including the one responsible for displaying partial grammatical categories of the word $x$, and its case among them.Thus, the value of $s(x)$ is a function of class $X$ of the corresponding language words. By the word $x \in X$ hereinafter (unless otherwise stated) we mean a set of all conceivable forms of the given word in the language. Thus, the word$x$ is, in fact, the set of all of its forms in all the grammatical meanings: $x = \{x^1, x^2, \ldots\}$, i.e. partial forms of the word $x$ are denoted with superscripts. There are also states of more a general form that meets a certain subclass of $X$. When substituting words from this division into a generalized state, we get logically, a state of the chosen word.

*Two states are called equivalent in respect of the appropriate object if the terms of these states of the object in the language in both cases are used inthe same form. For example, two states, the first of which is that the given object is boiling, and the second is that the cat is drinking this object are equivalent in respect of the object of «молоко» («milk»). These states are not equivalent on the object of «вода» («water»), since «вода кипит» («water is boiling») but «кошка пьёт воду» («cat is*

*drinking water»). We say that two states are completely equivalent if they are equivalent in respect of any object that may be in these states. For example, two states, the first of which is that the cat likes the given object, and the second is that the dog is drinking this object are equivalent in respect of any object that may be in these states, and therefore, are completely equivalent. The set of all classes are divided into subclasses that do not overlap so that any two states of the same class are completely equivalent, and any two states from different classes are not completely equivalent.*

**Comment 2.** A.N. Kolmogorov offered to call these classes as cases. It is a natural desire of A.N. Kolmogorov, as a mathematician, to establish some equivalence on the set of objects under consideration (in this specification — on the set of states of appropriate words), since the basic property of the equivalence relation is a partition of the set the relation operates on into an association of mutually disjoint subsets (equivalence classes), their elements being homogeneous in respect of certain relations what simplifies the picture considerably. Having certain linguistic skills (says the author) the indicated equivalence can be arranged so neatly that only states of nouns in a particular case — the traditional linguistic sense of the term «case» — prove to be the elements of each class. Then each equivalence class can be assigned to a certain name, which, in turn, can be identified with the name of the case, and, thus, we get a desired formal definition of this grammatical category.

*However, (notes V.A. Uspensky further) unfortunately, this definition is not entirely correct. The fact is that the same state for the same object can be expressed through a variety of sentences, and the name of the object can be in various forms. For example, Russian «мальчик идёт по берегу» («boy is walking on the shore») and «мальчик идёт берегом» («boy is walking shore»), «рабочий строит дом» («a worker is building a house») and «дом строится рабочим» («a house is being built by a worker»). Consequently, the definition of equivalence in respect of the subject ceases to be clear. (One of the possible ways to resolve this ambiguity is to regard two states with different linguistic expressions, as different states, since different sentences differ always — at least slightly — in content).*

## § 2

*Exchange of views on the definition of the case that took place in the workshop and its sidelines (meaning, in particular, the ideas expressed by R.L. Dobrushin and I.A. Melchuk), suggests the possibility of such a path.*

*The final orderly line which has either a word or three dots in every place, three dots occurring only once, is to be called a «set of words with space» (shortly — merely a «set»).*

*For example:*

*1. ..... кипить*

*2. кішка п'є .....*

*3. кішці п'є .....*

*4. кішка любить .....*

*5. ..... п'є молоко*

*are five different sets of words with space. When substituting a word (in some form) into a set of words with space instead of the three dots you can get a correct sentence. For example, when substituted in the second and fourth of the above, the word «вода» in the form of «воду» correct sentences «кошка пьёт воду» and «кошка любит воду» appear, with the same substitution in the third set we get«кошке пьёт воду», which is not a correct sentence. The word that is substituted in some of its forms in a matching set turns this set into a correct sentence is called admissible for that set. Set for which there is at least one admissible word, is also called admissible. Note that we do not deal with this question of what is «correct sentence» or whether it is, for example, the expression of some actual circumstance or merely a set of words combined according to some fixed grammatical rules (depending on the choice of a particular point of view the result of the substitution of the form «воду» in the fifth example is or is not a true sentence).*

*We call two sets equivalent in respect of the corresponding noun admissible for each of them if the substitution of the same form of the noun turns both sets into correct sentences. For example, the first and the second sets of the prescription above are equivalent in respect of the noun «молоко», because, to transform these sets into a correct sentence we should substitute the considered noun in each of them in the same form of «молоко», and at the same time, these same sets are not equivalent in respect of the noun «вода», since the form «вода» is to be substituted into the first of them, and the second and the form «воду» into the second. Sets «мальчик идёт .....» and «мальчик идёт по .....» are not equivalent in respect of the noun «берег» and sets «..... строит дом» and «дом строится .....» are not equivalent in respect of the noun «рабочий». We call, then, two sets directly equivalent if for them there exists at least one common admissible word and they are equivalent in respect of any word admissible for each of them. For example, set «..... бежала» and «..... бежит» are directly equivalent sets. Finally, we call two sets P and Q are completely equivalent if there exists a chain of sets $X_1$, $X_2$, ..., $X_n$, that:*

*1) for each i sets $X_i$ and $X_{i+1}$ are directly equivalent;*

*2) $X_1 = P$;*

*3) $X_n = Q$.*

*For example, «..... бежала» and «..... бежал» are completely equivalent sets as the corresponding chain consists of the sets «..... бежала», «..... бежит», «..... бежал».*

*The set of all admissible sets of words with space is partitioned into classes that do not overlap so that any two sets of the same class are completely equivalent, and any two sets of different classes are not completely equivalent. You can offer these classes to be called cases.*

**Comment 3.** The author notes that there may be situations (even with identical states of words) when their forms are not the same It means that the fact of

$s_1(x) = s_2(x)$ still does not imply the forms concurrence in states *1* and *2*. Therefore, it is necessary to first clarify Kolmogorov's equivalence. In particular, the author does not directly use here the term «state of the object», but makes a fundamental step to consider a sum of contexts with one free word — namely, to review «sets with space», which we henceforth also call «contexts with one (marked) variable word» or «context with one variable», or «context». Here a future relationship between the sum of contexts of a word and its grammatical state is noticed not quite obvious yet butenough as we see it. It is, in fact, the recognition that some states are not sufficient to determine the grammatical features, such as case. Therefore, some additional structures, such as sets of strings with spaces, are required. In this case, we note the properties of reflexivity, symmetry, and transitivity of certain sets of strings as useful for further consideration, since they allow us to introduce a relationship of complete equivalence. However, this step does not remove all the difficulties that necessitate further clarification.

As V.A. Uspensky notes *the sets «я вижу синий…» and «синий ... стоит» are directly equivalent and therefore completely equivalent, although we are dealing here with different cases. To avoid such incidents, we suggest (as if not the best, but the easiest way out) to prohibit the use of sets of adjectives, ordinal numbers, etc. (this prohibition extends to the next paragraph).*

*There remain, however, difficulties of the same nature as in the previous paragraph. The same set can be converted to a correct sentence by using the substitutions of different forms of the same noun. Forexample, «не читал газеты»and«не читал газету»; «дал кошке»and«дал кошку». Therefore, the definition of equivalence in respect of the appropriate noun turns to be unclear. (To resolve this ambiguity we should possibly consider only rather extensive sentences. Another way out is to consider only those sets that can be converted into a sentence by substitution of not more than one form of the same noun.)*

## § 3

*The combination of the views expressed in § 1 with those from § 2 brings us to the following approach to defining the notion of case.*

*Consider a certain state A, in which objects can be, and any set of words with a space B. We call a set B is coherent to the state of A, if, for any object that can be in the state A, the following statement is valid: to express in the language the fact that the subject is in state A it is enough to substitute some admissible form of the name of the object instead of three dots in the set B. If, for example, state A is that the appropriate object is building a house, then this condition is coherent both to the set of «..... строит дом», and then set «дом строится .....».*

*If state A means that someone did not read the appropriate object the set coherent with it is «не читал .....»; this example shows that a coherent set can be turned into a sentence that expresses the underlying state of the appropriate object also*

*when substituted of more than one form of the name of the subject: «не читал газету» and «не читал газеты».*

*The set «дал .....» is coherent both to the state meaning that someone gave something to this subject and with the state meaning that someone gave the subject to someone else. A pair (A, B), where A is a state and B is a set of words with a space coherent to this state is called a coherent pair. Let us call objects that can be in state A, as admissible for the pair (A, B).*

*Two coherent pairs ($A_1$, $B_1$), ($A_2$, $B_2$) are called equivalent in respect of the appropriate object if for any form of the name of the subject the following two assertions are valid:*

*1) If the substitution of this form in the set $B_1$ turns it into a sentence that says that the object under consideration is in state $A_1$, then the substitution of the same form in the set $B_2$ turns it into a sentence that says that the object under consideration is in the state $A_2$;*

*2) if the substitution of this form in the set $B_2$ turns it into a sentence that says that the object under consideration is in the state $A_2$, then the substitution of the same form in the set $B_1$ turns it into a sentence that says that the object under consideration is in the state $A_1$.*

*We call two coherent pairs are directly equivalent if there is at least one object admissible to them both and if they are equivalent in respect of any admissible object to both of them. Call finally two coherent pairs (P, Q) and (U, V) completely equivalent if there exists a chain of coherent pairs ($X_1$, $Y_1$), ($X_2$, $Y_2$), ..., ($X_n$, $Y_n$) such that*

*1) for each pair ($Xi$, $Yi_{)}$ and ($Xi_{+1,}$ $Yi_{+1)}$ are directly equivalent;*

*2) ($X_1$, $Y_1$) = (P, Q);*

*3) ($X_n$, $Y_n$) = (U, V).*

*The set of all the coherent pairs is partitioned into sub-classes that do not overlap, while any two pairs of the same class are completely equivalent, and any two pairs from different classes are not completely equivalent. It is these classes that are offered to call cases. Fully aware of the inconclusiveness of the newly formulated definition of the case, the author nevertheless considers it appropriate to bring it here, at least as a material for discussion.*

**Comment 4.** Thus, in the above snippet, the author proposes to eliminate the difficulties in turning to the consideration of the concept of «state of the object» in conjunction with the construction of «a set with space», i.e., from the context with one (marked) variable word. The concept of coherent pairs is introduced as <«state of the object», «set with space»>. Thus, in the first phase, the «state» allows to carry out a certain selection of contexts and form a set of «coherent pairs». The introduction of relations of equivalence and direct equivalence for the coherent pairs lets the author define the property of symmetry, and the introduction of absolute equivalence — also the property of transitivity. Thus, the relation of complete equivalence turns to be reflexive, symmetric, and transitive — that is an

equivalence relation in the usual sense of the set theory. As such, it divides the set of all the coherent pairs into classes that are mutually disjoint. The author suggested to identify them as cases.

True, there is a certain reticence. If, for example, the author notes, state A is that the **appropriate object is building a house**, both the set $B_1$ = «..... building a house» and set $B_2$ = «house built .....» are coherent with this state, and dots here can be replaced by a noun in a suitable form such as «worker is building a house» and «house is being built by the worker». It is good that states 1 and 2 deny the equivalence of pairs. But is not entirely clear what is the status of the expression **«appropriate object is building the house»**, which is used as a definition of the state — in fact, due to the uncertain definition of **«appropriate object»**, this expression is virtually indistinguishable from the set of «..... building a house». Nor is it entirely clear whether the different states of «relevant objects» emerge when substituting a specific noun in this formula. Because states of nouns in first place in the contexts of «a worker is building a house», «a foreman is building a house», «the factory is building a house», «the State is building a house», are hardly the same. Also, still, the identity of the states **«the appropriate object builds the house»** and **«the house is being builtby the appropriate object»** is not very clear either And in general, it is not clear which state of a worker, for example, is defined with the expression «a worker is building a house» (the worker is laying bricks in masonry, mixing cement, just having a smoke or like Shura from the famous film by Leonid Gaidai running away from a bully delivered for re-education). For comparison, consider the similar structure of the state **«the object likes vodka»** and **«vodka is liked by the object»**, **«THE OBJECT LOVES A COUNTRYWOMAN»** and **«a countrywoman is loved by the object»**, **«the object makes the revolution»** and **«revolution is made by the object»**. Substituting the word «worker» in matching sets in the due form informs a lot of details to participants of the speech process to never clarify the point at all.

Here, however, we cannot fail to note a certain congeniality of approaches developed by Isaac Newton and A.N. Kolmogorov who both tried to bring the properties of the language of the «nature of things». At the same time, it appears that by introducing the concept of «state of the object» Kolmogorov had in mind, not so states of real objects but rather a certain semantic lingual construct, the use of which in the procedure Kolmogorov developed is instrumental in selecting forms of words, allowing to «cut off» unacceptable, too grammatically irrelevant situations in respect of that one given by the context under consideration. Of course, Andrei Nikolaevich himself was aware of a certain insufficiency of this definition, noting that «... *Fully aware of the inconclusiveness of the definition of the case just formulated, the author nevertheless considers it appropriate to bring it here, at least as a material for further discussion*». However, in the next paragraph, he gives quite a meaningful algorithmic scheme of the definition of the noun case, which is as follows.

## § 4

*To determine in which case the appropriate noun is in a particular sentence, act as follows:*

*1. Determine the state in which there is a subject denoted by this noun;*

*2. Replace this noun with three dots and get the same set of words with space;*

*3. We note that the resulting pair (set, state) is coherent, and determine which class, i.e. the case it belongs to.*

*To extend our construction to nouns in the plural it is enough to admit that every such noun denotes a particular subject (other than the subject designated by the same noun, but in the singular). Thus, the noun «glass» refers to the subject of «glass», and the noun «glasses» refers to an object that consists of some set of glasses. (Note that because of our agreement «профессора» and «профессоры» («professors») are merely different names for the same object).*

## § 5

*The answer to the question, how many cases are in a particular language, should be given through a particular linguistic analysis of the language. If we proceed with the proposed definition in § 3, we find that in the Russian language in addition to the traditional six cases there are still the following cases:*

*1. Localcase: «в лесу», «в году»etc.*

*2. Quantitative-ablativecase«выпить чаю», «прибавить ходу», «дать воды» etc.*

*If both sentences «не читал газету» and «не читал газеты» are correct and express the same state of the object «newspaper», it indicates that there is a special case («derivative»), which is used after verbs of denial and has two forms (one of which is identical with the accusative, and the other is a form of the genitive case). If «не читал газету» is right, and «не читал газеты» is wrong in the first of these sentences, we deal with the accusative. If «не читал газету» is wrong, but «не читал газеты» is right, in the second of these sentences we are dealing with the genitive case. If, finally, both sentences are correct, but they represent different states of the subject «newspaper» then in the first sentence there is accusative, and the second there is genitive.*

*There may be other cases. It would be interesting to enumerate all the cases of the Russian language.*

**Comment 5.** If we ignore some details related to the need to attract many useless formalization and undetectable aspects, the Kolmogorov-Uspensky scheme can be reduced to the following.

We introduce the set $K(X)$ of contexts of a single variable $x$, where $x$ ranges over the class $X$ of nouns of a certain language in which the category of case is defined.

The notion of state $s(x)$, $x \in X$, which, however, does not have a clear definition, but operates on the set $K(X)$ in a well-defined way described above. The re-

sult of this action is a certain selection of contexts into coherent pairs <state-context> and the establishment of an equivalence relation which divides *K(X)* into mutually disjoint subsets, i.e. equivalence classes:

$$K(X)=\bigcup_{i=1}^{N} K_i(X);\ K_i(X)\cap K_j(X)=\varnothing,\ \text{when } i\neq j, \tag{88}$$

where *N* — number of cases in that language, so contexts $k_i^j(x)$ and $k_i^r(y)$, which belong to the class $K_i(X)$ are equivalent in the above sense. Thus, the name of a subset of $K_i(X)$, $i = 1, 2, ..., N$, can be identified with the name of a relevant case, and all labeled $x \in X$, belonging to the class $K_i(X)$ are available in it only in the forms of *i*-th case. It follows from this that the partial grammatical class (a part of the overall state of *s (x)* related to the case properties) is an abstraction of certain elements of $K_i(X)$.

In this sense, the notion of Kolmogorov — Uspensky state considered in a single complex process of constructing equivalence classes of the case appears as an operator (call it the operator of the case or the Kolmogorov — Uspensky operator and denote by the symbol [KU]); it operates on a set of word contexts with *X* and breaks the corresponding set into many mutually disjoint subsets. Moreover, each of the subsets contains contexts in which words from the *X* are found in only one case characteristic of this subset. As noted above, each such subset can be marked by the symbol (name) of the case and used as the case's definition.

However, a slightly different look is also possible. As the design and the procedure described above can be applied to any new context (since *K (X)* is never final) in order to determine the case of a word ξ from *k (ξ)*, it is logical to assume that in such a way the «case» part of its full linguistic state *s (ξ)* is defined, and the value of the word case acquires the status of an eigenvalue of the operator [KU]. Thus, we could argue that in the considered context, the word ξ is in a linguistic state with a certain value of case, and the mentioned linguistic state plays a role of its own state of the Kolmogorov — Uspensky operator. There are, of course, contexts of uncertain value (the classic example is «день преодолевает ночь»); to describe them one uses the procedure of state superposition which is discussed in the next section.

According to the logic of our presentation, the value of case received by the Kolmogorov — Uspensky procedure («eigenvalues» of the operator [KU]), play a role as observed values of the theory, and the «eigenfunctions» of the operator can be regarded as partial grammatical states that identify the case status of the corresponding words. As a result, in this case, we can formalize the problem as an eigenvalue problem (typical of mathematical physics). The fact that we are dealing with mathematically undefined objects should not embarrass us, since this situation is typical of theories with formally (topologically and algebraically) uncertain baseline parameterization of its objects — and such are mainly linguistic theories.

It should be noted that the formal definition of the observed values is rather a complicated procedure for any theory that works with real objects, and that Kolmogorov did manage almost 60 years ago to advance so far in the formal definition of a very difficult grammatical concept, inspires deep respect. We note occasionally that Kolmogorov may have not set out as a special goal to define the concept of the state of a linguistic unit — the object is subsidiary for him and, so to speak, of a technical nature. However, this very proposal (i.e., putting per the linguistic state to a language unit) is, in our opinion, the most fundamental and valuable in his work. Indeed, the concept of the case (as well as other categories of language) will still be refined and specified continuously when further exploring the properties of the linguistic substance, while the proposals concerning a fundamentally new base for the conceptual description of the phenomenology of language (and that is the approach of linguistic states) appear much more seldom. Here, as it seems to us, it is just the case.

Unfortunately, these ideas remained practically unnoticed and unused by the linguistic community. In particular, we do not know any works that would attempt to extend the above methodology to determine other grammatical categories and other parts of speech (including the determination of the parts of speech themselves). Only A.A. Zaliznyak perceived the Kolmogorov — Uspensky ideology and used it in his excellent book «The Russian nominal inflection» (hereinafter abbreviated use *RNI*), published in 1967. The reprint of this book in 2002[67] also contains several other works by A.A. Zalizniak, including the article «On the understanding of the term «case» in linguistic descriptions», first published in the book «Problems of grammatical modeling»[68], where the author develops the concept of the case system, indicating, in particular, the place the Kolmogorov — Uspensky approach occupies in it. Here are some snippets of this article. Here are some fragments of this article[69].

A.A. Zalizniak notes that *two different understanding of the term «case» that can be (somewhat arbitrarily) designated as «semantic» and «formal». The main difference between them is whether there is allowed the existence of two different cases without a difference between them that can be expressed outwardly. Under the semantic understanding of the case, the answer is positive, under the formal one it is negative. A particular case in semantic understanding is a certain element of meaning, namely, a certain semantic relation, for example, «to be the subject of action», «be an instrument of action», «belong to someone (something)», «be inside of*

[67] Zaliznyak A.A. «Russkoe imennoe slovoizmenenie» s prilozheniem izbrannyh rabot po sovremennomu russkomu yazyku i obshchemu yazykoznaniyu. Moscow: Yazyki slavyanskoj kul'tury. 2002. I—VIII. 752 s. (Studia phililogica).

[68] Problemy grammaticheskogo modelirovaniya. Red. A.A. Zaliznyak. Moscow: Nauka, 1973. 262 s.

[69] The reader will easily separate A.A. Zaliznyak's text in italics from our comments on it.

*something».Cases in the semantic sense can somehow be expressed in any language, so if considered without regard to the mode of expression, they act as an element of a universal system units of meaning. We can describe as «narrowly formal» such understanding of the case, in which whole word forms alone are allowed as case forms and two cases are regarded as different only if at least a part of the conjugated words they correspond to have different external forms.*

*A specific case in the narrow formal sense (for nouns) can be described in the first approximation as follows: it is a set of word forms (or, to say, something that all the word forms of the set have in common), each of which is capable of expressing one or more of cases in the semantic sense in addition to its core, objective meaning.*

*For example, in Russian ablative meets some word forms, including, among many others, the word-forms «пером», «человеком», «стрелой», «лесом», «мыслью», «перьями», «детьми» etc. Each of them can express a series of cases in the semantic sense (the meaning of an instrument, doer, reference standard, traffic areas, etc., cf. above). Thus, the cases in the narrowly formal sense are one of the possible means (along with prepositions, postpositions, word order, etc.) to express the cases in the semantic sense.*

Note: here A.A. Zaliznyak believes that the semantic understanding of the case provides finer gradations within the semantic structure inside its proposed narrowly formal interpretation and, thus, the semantic understanding «splits» the narrowly formal level (the actual «case») into some «sublevels». This interpretation is supported with the following remark in *footnote 2,* p. 616 of the Russian edition of RNI, which states: «*the relation between cases in narrowly formal and semantic understanding isdemonstrated in the titles traditional for the Latin grammars: genetivus possessivus, genetivus partitivus, genetivus subjectivus, genetivus objectivus, dativus commodi, dativus possessivus, dativus finalis, etc.: here the first part of the name corresponds to the case in the narrowly formal sense, and the second — in the case in the semantic sense*». So, within the «narrow formal», for example, *genetivus* as many as four «semantic» cases are contained: *genetivus possessivus, genetivus partitivus, genetivus subjectivus, genetivus objectivus.*

*For the recognition of cases (in the narrow formal sense) in a particular language, there must be at least two of them, otherwise, it is said in that language there are no cases at all.*A.A. Zaliznyak notes that *the cases in the formal (and in particular, the narrowly formal) sense are not universal, and always make an element of the grammatical structure of a particular language: one language, for example, has 6 cases (in the formal sense), in another, there are 15, in the third, they do not exist.* And further: *in modern linguistic literature, especially in specific descriptions the narrow formal understanding of case certainly prevails, and that is what will be of interest to us below.* Accordingly, in the cited papers A.A. Zalizniak considers only such descriptions of case systems that are based explicitly or implicitly on a formal description of the narrow case. Later under the «case» (without qualification), the author means a case in the narrowly formal sense. On the same page in

*footnote 4* A.A. Zalizniak notes thatthe *case in the semantic sense corresponds to the notion of «semantic role» he introduces below.*And we note that statement because subsequently, we have to understand the relationship of such concepts as the *case in the semantic sense, the (semantic) state, and the semantic role*.

Regarding this last remark, we will do one more clarification. A.A. Zalizniak further on in his article again emphasizes that *he uses a vague notion of a«semantic» role that matches in its content to what was the above-named case in the semantic sense. This concept is related to the usage specified below. Let the phrase «Он послал родителям книгу своего друга». («He sent his friend's book to his parents»). We say that the segment of «родителям» expresses in this phrase a semantic role of «recipient of action» for the nominateme «родители»; segment «книгу» in this phrase expresses the semantic role of «object of action» for the nominateme «книга», the segment «друга» may express in this sentence the semantic role of «owner of something» or the semantic role «author of something» for the nominateme «друг». Semantic roles can be expressed by segments not only in phrases but also collocations. For example, in the collocation «послать родителям книгу своего друга» the segments «родителям», «книгу», «друга» of the same semantic roles (for appropriate nominatemes) as in the phrase analyzed above.* In *footnote 14* on page 623, A.A. Zalizniak notes that the term «context» corresponds to the term «set of words with space» in the V.A. Uspensky's paper from 1957, the term «semantic role» to the «state», the term «situation» to the term «coherent pair» (we also refer to the RNI, § 2.3). Thus, the above three concepts, namely *the case in the semantic understanding, the (semantic) condition, and the semantic role* in the author's interpretations appear identical (at least within the problem under consideration).

Next, A.A. Zalizniak develops his theory of the case, which, in our opinion, is an updated and detailed version of the Kolmogorov — Uspensky approach. We will not dwell here on the details of the presentation, because we intend to focus on the *case in terms of semantic understanding, the (semantic) condition, and the semantic role*. In A.A. Zalizniak's paper, as already mentioned, they are regarded as identical, although, generally speaking, are different. Thus, by the very sense of the notion of «case», it looks like the number of cases in the semantic sense should be small and (be still better), they should be named (listed), at least in the very rough approximation. In Latin, the author gives the list — *(genetivus)*: *possessivus, partitivus, subjectivus, objectivus, (dativus): commodi, possessivus, finalis*, ..., but without specifying whether it is complete and versatile. It is not clear either in what accordance are semantic roles to the list of cases in the semantic sense. And even more uncertain seems to match these two concepts to the concept of the state according to Kolmogorov. At the same time, judging from the A.A. Zalizniak's description of the procedure to establish the case system this procedure, except for some details, follows the above procedure by Kolmogorov — Uspensky, so this common, unified procedure could easily be named as Kolmogorov — Uspensky — Zalizniak, setting the line not only between the procedures applicable in both concepts but also be-

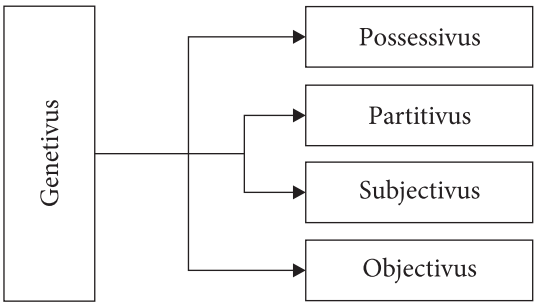



***Fig. 3.*** «Fine semantic structure» of the Latin case (on the example of *genetivus*)

tween the relevant parts of the algorithm. At least for today in terms of function the use of all the three above basic concepts is quite identical, despite their possible conceptual differences and uncertainties in their definitions.

The above uncertainties, we think, maybe allowed under a slightly different conceptual paradigm with states of linguistic units, corresponding grammatical observed values, as well as procedures that bind together the linguistic states and linguistic observables to make the paradigm's core. Namely, if we follow the general concept of linguistic states, the categories of the case should match a certain operator acting in the space of grammatical states so that its eigenvalues are the values of specific cases. That is the Kolmogorov — Uspensky — Zalizniak procedure that acts as a grammatical operator of the case. That the state by Kolmogorov is involved in its definition should not be confusing, since the notion of space of linguistic states is still, unfortunately, far from a mathematical formalization and we have to be content with its purely symbolic representation. Therefore, components of the linguistic stateguessed can be nevertheless quite legitimately used in the process of determining the eigenvalues of the case operator, considering the approach to be specific and characteristic of the method to be formulated.

In this picture, it is necessary to bear in mind some other effects that occur in theories that work with observables formalized as the eigenvalues of the corresponding operators in the relevant state space.

Firstly, it is the invariance of the subspace of states corresponding to a particular eigenvalue. Here it meets the equivalence class of case identified by the Kolmogorov — Uspensky — Zalizniak procedure on a set of coherent pairs (situations).

Secondly, it is the so-called effect of the fine structure[70] when the level corresponding to a particular eigenvalue being split into some «sublevels» with the inclusion of an additional «semantic interaction» or, equivalently, considering the semantic factors not taken into account earlier. So, for example, the value of Latin cases given by A.A. Zalizniak <genetivus>; <dativus>; ...are interpreted as «eigenvalues» of the operator of «the narrowly formal case», while <genetivus possessivus>, <genetivus partitivus>, <genetivus subjectivus>, <genetivus objectivus>.... are the eigenvalues of the operator of «the semantic case», which includes some additional data in comparison with «narrowly formal» aspects of the grammatical meaning. The «narrowly formal case» presents a «degenerate» level, and this

[70] Fine structure in quantum mechanics.

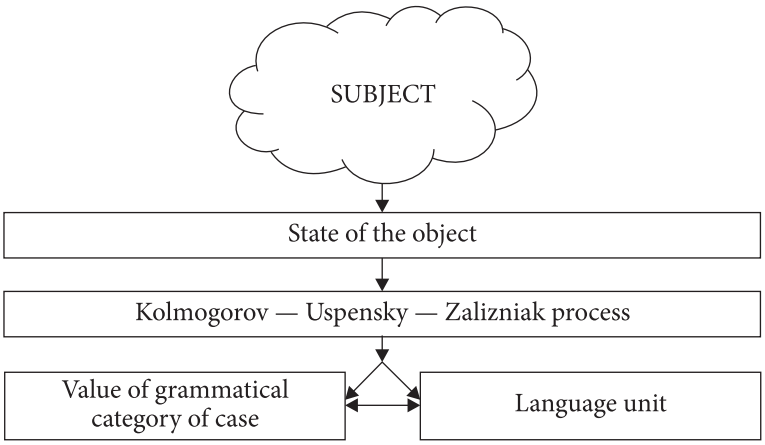



***Fig. 4.*** States in the picture of Kolmogorov —Uspensky —Zaliznyak

«degeneracy» is removed when taking into account more a subtle semantic factor defined by a «semantic case» and the original case state is split, showinga «subtle semantic structure» that on the above A.A. Zalizniak's example can be schematically shown in Latin as follows (Fig. 3).

Similar comments can be made on the concept of semantic role, which, given generally, should serve (and it does in the above case) rather as a part of the procedure (operator) of the case system than as its state. In this sense, the meaning of the grammatical category of case both in its narrowly formal and semantic understanding plays the role of eigenvalues of the corresponding grammatical operators that operate on the set of grammatical states of words. It should therefore be considered that the note expressed by V.A. Plungian in his very detailed book[71] saying that the «semantic role» is analogous to a «more impressionistic» concept of the «state» does not match the actual situation, namely, the «semantic role» is in no way a «state», but plays (as a part of the Zalizniak procedure) the role of an operator with the states as its functions, and specific (and manifested in contexts) values of the case as its eigenvalues. And the fact that the semantic role is involved in the definition of the case looks almost trivial since the said role corresponding to the case in the semantic sense, already is, so to say, «inside» of the fine structure of the case and detected in the context of the case in the semantic sense, as it follows from the scheme shown in Fig. 1, it brings us very close to the definition of case class of equivalence on the set of coherent pairs (situations), and therefore to the definition of the case in the narrowly formal sense.

Thus, in Kolmogorov's theory (and therefore in the Zalizniak procedure) the states should occupy a place as indicated in the following diagram (Fig. 4).

[71] Plungyan V.A. Vvedenie v grammaticheskuyu semantiku: grammaticheskie znacheniya i grammaticheskie sistemy yazykov mira. Moscow: Ros. gos. gumanitarnyj un-t, 2011. S. 183 (*69).

## The general concept of linguistic states and the concept of linguistic observables associated with it

The employment of the concept of the state of language unit was further developed in our works and the works of our fellows[72], according to them any word (to say generally — any language unit) in the context of the speech flow is in a certain semantic state. For units of the lexical level, the state is in some way a sum of signs of grammatical and lexical semantics somehow arranged and provides a way to summarize the concepts of grammatical and lexical meaning, which is to be the topic of our further narrative.

The process of language understanding can be represented as a reduction in the a priori distribution of lexemes according to the total characteristics of the grammatical and lexical semantics inherent in the subjective vocabulary of the recipient, to a specific grammatical and lexical meaning, characteristic of that particular context that is in the field of attention of the recipient and is currently subject to the process of its language processing. Therefore, the above diagram can be generalized and detailed as follows (Fig. 5).

Even though a formal apparatus for determination of all possible states of any lexemes (more generally — of any unit of language) has not yet been created it is natural to assume that a set of grammatical and lexical meanings as presented, for example, in a large explanatory dictionary of a language can serve as a sufficiently adequate model of such a system. In general, we intend to use the concept of the system state (we consider every unit of language as a system and at the same time a part of a higher-level system), within the paradigm of modern science and technology[73]. Thus, when considering the formal aspects of semantics, we shall proceed with the existence of a correspondence between the language unit and its state:

$$s : X \to s\,(X), \tag{89}$$

where $X$ — is a certain language unit; $s$ — correspondence between $X$ and $s(X)$ — a formal object that represents the state of unit $X$ with elements of the material means

[72] Shirokov V.A. Leksykohrafichne predstavlennia semantychnykh staniv.*Matematychni mashyny i systemy*. 1999. No. 3. C. 21—32; Shirokov V.A. Fenomenolohiia leksykohrafichnykh system; Shirokov V.A. Semantychni stanymovnykh odynyts ta yikh zastosuvannia v kohnityvnii leksykohrafii. *Movoznavstvo*. 2005; Shirokov K.V. Imenna slovozmina u suchasnii turetskii movi. Kyiv: Dovira, 2009. 318 s.; Potapova E.V. Model lynhvystycheskoi ontolohyy predmetnoi oblasty c nechetkymy semantycheskymy sostoianyiamy termynov. *Byonyka intellekta*. 2012. No. 2(79). S. 9—102; Shirokov V.A. Systemna semantyka tlumachnykh slovnykiv. *Zb., prysviach. 75-littiu V.H. Skliarenka*. Kyiv: Nauk. dumka. 2012.

[73] Boum A. Kvantovaya mekhanika: osnovy i prilozheniya. Moscow: Mir, 1990. 720 s. (Gl. IV); Dirak P. Principy kvantovoj mekhaniki. Izd. 2. Moscow: Nauka, 1979. 480 s.; Landau L.D., Lifshic E.M. Kvantovaya mekhanika (nerelyativistskaya teoriya).Izd. 4. Moskow: Nauka, 1989. 768 s. (Teoreticheskaya fizika. T. III).

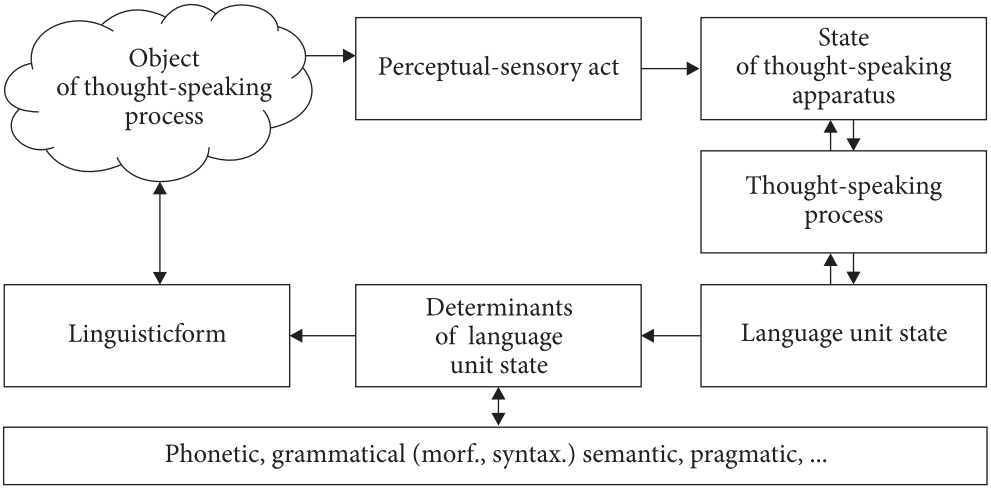



***Fig. 5.*** Extended scheme of states of linguistic units

for expressing the semantics as its determinants as indicated in the diagram above. For every unit *X*, its states form a certain set (for simplicity we consider it to be finite but unlimited, although this is not essential) — usually denoted by the symbol as {*s (X)*}. Class of units of a certain type in the language *L* is denoted by the symbol *W (L)* or simply *W* if only one specific language is meant, belonging of *X* to the class *W* is denoted by $X \in W$; set of all states for all $X \in W$ is denoted by $S \equiv \{s\ (X); X \in W\}$.

Suppose that there exists an operator *F*, the effect of which is defined on the set of semantic classes *S* and which we interpret as a value operator of a certain semantic category (they may be also categories of grammatical semantics as in the example analyzed in the previous paragraph. If, for example, *F* is the operator of the language, then its values are $f_1$ — «noun», $f_2$ — «verb», $f_3$ — «adjective», etc.). This means that *F* is a kind of intelligent mechanism that analyzing the state of *X*, namely *s (X)*, identifies the value of a particular semantic category that precisely matches this very state. The definition of operators of type *F* must be based on some semantic theory formalized properly.

Using the mathematical language the action of the operator *F* can be expressed as follows:

$$F\, s_i\,(X) = f_i\, s_i\,(X), \tag{90}$$

where $f_i$ — a specific value of semantic categories, functions $s_i\,(X)$, $i = 1, 2, ...$, represents the states of the unit *X* that mark its affiliation with the values of $f_i$ of the category *F*. The values of $f_i$, $i = 1, 2, ...$, are called the eigenvalues of the operator *F*, that correspond to states $s_i\,(X)$, $i = 1, 2, ....$ The set of all states that correspond to the eigenvalues $f_i$, are called the set of partial semantic classes and marked by the symbol $S\,(f_i)$:

$$S\,(f_i) := \{s: Fs = f_i\, s_i\}. \tag{91}$$

By definition of sets $S(f_i)$ consists only of those semantic classes characterized by a definite value of the semantic category $F$, — namely, the value of $f_i$.

Equation (3) in cases where the set ranging over the index $i$ consists of more than one element is a formal expression of the phenomenon of semantic ambiguity. For units of the lexical level, this is polysemy and homonymy (both lexical and grammatical). In particular, in the case of homonymy, this reflects a situation where the word form $X$ in one context may be, for instance, in the state of a noun, andanother — in the state, for example, of a verb. This type is usually called inter-part-of-speech homonymy.

Situations are possible when in the equation (90) there is not one semantic condition $s_i(X)$ but several that corresponds to a value of $f_i$: $s_i^j(X)$, $j = 1, 2, ....$ Such conditions will be called degenerate; the presence of degeneracy is signaled by the superscript character in the semantic state $s_i^j(X)$. For example, in the Ukrainian language the word form «МАТИ» when getting the value of «noun» of the category «part of speech» has two grammatical states g (МАТИ):

$$g^{\underline{feminine}\ \underline{\mathbf{singular}},\ nominative}_{noun},$$

where the word «МАТИ» has a lexical meaning «woman concerning the child that she gave birth» and:

$$g^{\underline{noun}\ \underline{\mathbf{masculine}}\ \underline{plural},}_{nominative},$$

where the word «МАТИ» has a lexical meaning «sports mattresses». This example illustrates the phenomenon of *homonymy inside the same part of speech*.

The number of intrinsic semantic states of a unit $X$ corresponding to a certain value of the category $F$ will be called the multiplicity of the degeneracy of the semantic state. Thus, in the above example, the state $g$ (МАТИ) has a multiplicity of degeneracy that equals 2.

Semantic states in which the operator $F$ has one value only is called *net*. However, a priori it is impossible to forbid the existence of semantic states for which the operator $F$ acquires more than one particular value, but can be characterized, for example, by two. Formally, this situation can be represented using the relation:

$$F(X) = f_1\alpha_1(s_1)\, s_1(X) + f_2\alpha_2(s_2)\, s_2(X), \qquad (92)$$

where s is a semantic state of the linguistic unit $X$ when exposed to it by the operator $F$ is split into two, namely: $s_1(X)$ and $s_2(X)$, where $s_1(X)$ corresponds to values of semantic categories $f_1$, and $s_2(X)$ to values of semantic categories $f_2$; the linguistic sense of the functions $\alpha_1(s_1)$ and $\alpha_2(s_2)$ is explained below.

States for which the operator $F$, acting on the function of the semantic state is a combination of a certain number of net semantic states corresponding to different eigenvalues of this operator are called *semantic mixed states*.

Thus, equation (92) defines the semantic state in which the unit $X$ is and which is a peculiar phenomenon of *superposition* («mixture») of net semantic classes $s_1(X)$ and $s_2(X)$ corresponding to the eigenvalues $f_1$ and $f_2$, respectively. The

linguistic interpretation of equation (92) is that the linguistic unit *X* has the semantic features of both $f_1$ and $f_2$. Corresponding figures, the identifiers of the specified belonging are located in the expressions for semantic states $s_1(X)$ and $s_2(X)$.

This situation is quite typical for the language. For example, Ukrainian and Russian participles contain the properties of the verb and the adjective. Consider, for example, the Russian participle lexemes «*ведущий*» (leading) and «*ведомый*» (driven). They decline according to the adjective inflectional paradigm (six cases in the masculine, feminine and neutral gender of singular and in plural), having besides in its structure a verbal morphological feature of active or passive voice — it is materially expressed by the suffix *-ущ* and *-ом*, respectively. The noted morphological characters do not look strong in terms of belonging to the verb, since in the verbal paradigm it is not inflectional. Also, it is typical for the full inflectional paradigm, not only for its members — this provides a basis for distinguishing a class of words with such properties as an independent part of speech —«participle». The Ukrainian title for the term «дієприкметник» (participle) reflects both features — those of the verb and the adjective [74].

## Linguistic states in the formalism of fuzzy sets

We emphasize the fundamental difference between the above situation and the usual homonymy. The phenomenon of homonymy can be described by the equation of the type (92):

$$Fs(X) = f_1\alpha_1(s_1)\, s_1(X) + f_2\alpha_2(s_2)\, s_2(X) + \ldots,$$

where different members of the right-hand side correspond to different homonymic states of the word *X*— if it is exposed to a particular kind of ambiguity. But in the process of language processing, where there is disambiguation, the right side of this equation is reduced to one member that represents the net grammatical status of the given word *X* in a particular context. The situation is completely different when *X* is in a mixed state — then no linguistic processing can reduce the number of members of equation (92), which is the final result. This situation indicates the existence of language units that operate in a context in multimodal semantic states.

The conclusion about the existence of units for which the language admits only those contexts to function in mixed semantic states looks like rather an interesting evidence of the impossibility to formalize the language system in full, rather,evidence of the limits of its formalization. This particularly means that even the most suitable for formalization and, indeed, the most formal sector of

[74] An example of superposition of grammatical classes in Turkish language is given in the book byShirokov K.V.: Nominal inflection in modern Turkish language [in Ukraine]. It demonstrates that some inflectional paradigm of the Turkish noun has certain categorical and material features of the verb.

the theory of language, namely grammar, has features of fuzziness that is in disharmony with the tradition of grammatical determinism and encourages the development of a special language to describe ambiguous grammatical situations. The conceptual framework of this language, we believe, is given in the fuzzy sets theory by L. Zadeh. We demonstrate below how this is achieved.

Let us turn to the definition of the set of semantic classes $S$ and sets of partial semantic classes $S(f_i) = \{s: Fs = f_i s_i\}$. If in the language processing there were in principle a possibility of reduction of any state to the semantic net, the set $S$ could have been represented as a union of subsets of $S(f_i)$ that did not overlap, i.e. the following formula would have been valid:

$$S = \cup_f S(f_i);\ S(f_i) \cap S(f_j) = \varnothing, \text{ where } i \neq j. \tag{93}$$

The situation is quite different if we consider the possibility of the described items that are both characterized by, for example, two (or more) values of a certain semantic category. Semantic condition s ($X$) of such a unit $X$ will no longer belong to only one of the subsets $S(f_j)$, but also two, and possibly more.

A formal mechanism for describing such a phenomenon operates as follows. Define on the set $S = \cup_f S(f_i)$ the structure of a fuzzy set in the sense of Zadeh. For this purpose, for each of the subsets $S(f_i)$ we define the membership function $\alpha_i(s)$, which for each $s \in S(f_i)$ takes a certain numerical value in the interval [0,1]:

$$\alpha_i(s) \in [0,1]. \tag{94}$$

Moreover, we assume that if $\alpha_i(s) = 1$, then s is a net state. If $\alpha_i(s) < 1$, it corresponds to the s, which is a component of the mixed state, while its second component s′ (for two-component states) belongs to a subset of $S(f_j)$, $i \neq j$ with the membership function value $\alpha_j(s')$, also less than 1, but such that the condition:

$$\alpha_i(s) + \alpha_j(s') = 1. \tag{95}$$

Then each of the subsets $S(f_i)$ is transformed into a fuzzy set with the membership function $\alpha_i(s)$:

$$S(f_i) \rightarrow \{S(f_i), \alpha_i(s), s \in S(f_i)\}. \tag{96}$$

Pairs (s ($f_i$), ι($s$)) serve as elements of the fuzzy set. At the same time, the structure of the fuzzy set is induced on the entire set $S$ as the union of fuzzy sets (96) under condition (8). Membership function $\alpha_i(s)$ in this case is interpreted as a measure of acquisition of value properties $f_i$ of the category $F$ by the semantic state $s$. The function reaches the maximum value equal to 1 on net states, describing words with clearly defined indicators of belonging to the corresponding semantic category. For words with mixed semantic states of the form:

$$s(X) = \alpha_1(s_1)\, s_1(X) + \alpha_2(s_2)\, s_2(X) \tag{97}$$

the value of $\alpha_1(s_1)$ demonstrates the degree to which the linguistic object $X$ shows the property of the value $f_1$, while $\alpha_2(s_2)$, respectively, indicate the degree to which

$X$ shows the property of the value $f_2$. Condition (95), namely: $\alpha_1 (s_1) + \alpha_2 (s_2) = 1$, ensures the completeness of the semantic properties of the proposed facility and the insularity of its semantic description.

## Synthesis of the linguistic states of lexical units

Along with the theoretical provisions on the nature of grammatical and lexical meanings and their reflexive definitions, the operational aspect of these concepts, which requires formalization of actions related to the process of abstracting grammatical and lexical meanings, is no less important for lexicography. Namely, the question arises: how (in what various ways) it is possible to establish different meanings of a word, by what means they can be distinguished, how relations are established between them, in particular, relations of proximity, or, say, opposites; how their semantic determinants are determined, what these determinants can be, etc.

The meaning of the word is revealed only in context. Therefore, theoretically, to explicate the meanings of a particular word, it is necessary to collect all — in a certain sense — its contexts where it functions, to distribute them among groups that are homogeneous in a certain (semantic) sense, each of which is a representative of a certain lexical meaning. Further, studying these groups of contexts, the lexicographer derives from each group a separate lexical meaning of the lexeme under consideration and qualifies the corresponding grammatical meanings.

Such a scheme is an idealization, because the sum of contexts is infinite, and in practice, the lexicographer always deals with a limited (often very limited) set of contexts from which, based on their knowledge of the language system, linguistic experience, and intuition, builds a system of lexical meanings.At the same time, it is difficult to get rid of the subjectivity inherent in any, even very objective, researcher.

There are a few more questions. Does any context $k$ ($x$) of the word $x$ have a specific lexical meaning? Could it be a situation when it is necessary to ascribe two or more lexical meanings to the same context, or even ascribe meanings that, due to desemantization or any other semantic processes, cannot be interpreted as lexical? How to perceive «wrong» contexts in which the speaker demonstrates «negative»language competence, or «language play» when the use of the word is deliberately deviant? The answers to these questions, typical of cognitive situations characteristic of the formalism of semantic states, can be obtained only thorough analysis of fairly large, representative collections of contexts.

As such collections, which serve as a factual basis for the definition of lexical semantics, lexicography traditionally used lexical card indexes in which certain works, entirely or selectively, were recorded on lexical cards: fiction, journalistic, scientific, official, folklore, etc. They were arranged in the alphabetical order — linguistic centers of relevant contexts and were studied by lexicographers when they established and described lexical semantics. Currently, tasks for the collection and explication of contexts are assigned to linguistic corpuses — significant

in volume electronic text libraries that are labeled (marked) with grammatical parameters of all the words in these texts that make up a special search index, the use of which makes it possible to almost instantly obtain a complete set of contexts of any selected word from all sources presented in the corpus. Nowadays large linguistic corps have volumes of hundreds of millions of word usages — computer analogs of traditional lexical cards.

Based on the study of one way or the other received sets of contexts, complexes of lexical meanings of words that are objects of lexicography are formed.

Let's summarize the above in the form of a specific scheme or sequence of steps.

At the first stage, a set of $K(x)$ contexts of a language unit (lexeme [75]) $x$ is formed, where $x$ runs through the class of words of a certain language. Thus, a «set of sets» of contexts is obtained.

At the second stage, the set of $K(x)$ for each lexeme $x$ is distributed over sets of disjoint subsets:

$$K(x) \supseteq \bigcup_{i=1}^{N(x)} K_i(x);\ K_i(x) \cap K_j(x) = \varnothing, \text{ when } i \neq j, \tag{98}$$

where $N(x)$ is an integer that is identified with the polysemy multiplicity of the lexeme $x$. Partitioning $K(x)$ into separate subsets of $K_i(x)$, $i = 1, 2, ..., N(x)$ is an informal intellectual process since it is in it that a set of $K_i(x)$ correlate with a certain lexical meaning. In other words, we assume that there is a certain one-to-one correspondence between the lexical meaning $l_i(x)$ and the set of contexts $K_i(x)$:

$$l_i(x) \leftrightarrow K_i(x), \tag{99}$$

which lexicographer sets and determines. Moreover, for simplicity of presentation, we assume that the sets $K_i(x)$ are formed in such a way that in each context $k_i^r(x) \in K_i(x)$ is the lexeme $x$ could occur in one and only one lexical meaning, although more than once.

We'll make some remarks since in reality there is a more complex process. In practice, not all contexts of the lexical unit $x$ are taken into consideration when determining lexical meanings. Only those contexts are considered and analyzed in which the value of the corresponding unit is «socialized». That is, there are a lot of such contexts and the application of the unit in question is socially accepted in them, the one that can be introduced and fixed in the conventional mode. Too individualized, «occasional» contexts, as a rule, qualify as exponents of «meanings». Thus, the extralinguistic content of the lexical unit takes on a two-part character, namely, «sense» and «meaning» [76].

[75] Here, a lexeme (in contrast to the definition adopted by the Moscow Semantic School) is a word with homonymy removed together with the totality of all its inflectional forms, which have a certain common set of lexical meanings.

[76] Compare, for example, what was said with the definition of meaning and sense, which is given by I.M. Kobozeva (Kobozeva I.M. Linguistic semantics. Moskow: KomKniga,

Thus, it should be noted that in formula (98) the subsets $K_i(x)$, $i = 1, 2, ..., N(x)$, generally speaking, do not completely cover the whole set of contexts $K(x)$. Together with and near to $K_i(x)$, certain «exceptional» contexts can exist and function in speech, which manifest meanings that are not reduced to «standard» lexical meanings attributed to $K_i(x)$ contexts. Strictly speaking, formula (98) should be rewritten as follows:

$$K(x) = \bigcup_{i=1}^{N(x)} K_i(x) + Q(x), \tag{100}$$

where $Q(x)$ is the set of «non-standard» contexts in which meanings, that are not reduced to the conventional lexical meanings attributed to the contexts $K_i(x)$, are manifested. The dividing of $K(x)$ into two groups, presented in formula (100), but rather the existence (even if potential) of the set of «meanings» $Q(x)$ is the evidence of the semantic openness of the language system and the practically unlimited expressive capabilities of the language. As a consequence of the above, we state that the contexts $Q(x)$ are «non-dictionary» and could not be lexicographic standardly.

Note that the formulation of the lexical meaning $l_i(x)$ (or the construction of its vocabulary definition) occurs already at the next third stage. It is where, based on the analysis of the members of the set $K_i(x)$ the elements of dictionary definition are categorized and formalized, and the definition is arranged in a form that corresponds to the concept of a specific dictionary and lexicographic tradition.

The complexity of these stages makes the lexicography process.

As it was said, the first stage consists in the formation of the set of contexts $K(x)$ of the lexeme $x$, when $x$ runs through the class of words of a certain, in our case, Ukrainian, language. We denote this class by the symbol $W^U$, i.e. $x \in W^U$. We believe that the class $W^U$ contains all Ukrainian words. We consider that their quantity is finite but unlimited. The finiteness of $W^U$ is a completely understandable property because otherwise, we would have to state the presence among the members of $W^U$ of such elements that have infinite length, which is impossible. At the same time, $W^U$ should be considered an unlimited complex, because at any time it can be expanded by adding some new element.

To form the set $K(x)$, it is necessary to have a certain set of texts (oral and/or written) that would be indexed by elements $x \in W^U$. This indexing provides direct access to all contexts of the selected lexeme $x$. Each specific context $k(x)$ that relates to $K(x)$: $k(x) \in K(x)$ is a text of a certain length containing the word $x$. The length $k(x)$ must be sufficient to determine the meaning of the word $x$. As a rule, this length is limited to one sentence, but it happens that one sentence is not

2007. 352 p. [in Russian]): «The meaning of X is an information associated with X-th conventionally, i.e. According to the generally accepted rules for using X-a as a means of transmitting information. The meaning of X for Y in T is the information associated with X-th in Y's mind during the time period T, when Y produces or perceives X as a means of transmitting information».

enough. In lexical card indexes, determining the length of the context was a very difficult problem. In modern linguistic corpuses, the length of contexts is not constant and can be changed at the request of the user.

From formula (99), in particular, it follows that all contexts from the set $K_i$ (x) are equivalent: $k_i^r(x) \sim k_i^p(x)$, $k_i^r(x) \in K_i$(x); $k_i^p(x) \in K_i$(x), belonging to $l_i(x)$. However, in different contexts $k_i^r(x)$, $r = 1, 2, ...$, the lexeme $x$ can take on different grammatical meanings. Therefore, using the formalism of the theory of semantic states, we can represent the differential semantic state $s_i^r(x)$ of the lexeme $x$ in the context $k_i^r(x)$ in the following form:

$$s_i^r(x) = g_i^r(x)\, l_i(x), \tag{101}$$

where the symbol $g_i^r(x)$ denotes the grammatical state (formal representation of the grammatical meaning) of the lexeme $x$ in the context $k_i^r(x)$, and $l_i(x)$ is the lexical state (formal representation of the lexical meaning) of the lexeme $x$ in the same context. The change in the index $r$, i.e. the variation of the grammatical meaning of the lexeme $x$, does not affect its lexical meaning. This means that the semantic state of the lexeme $x$ could be determined in the form of a weighted sum:

$$s_i(x) = [\sum_r \alpha_i^r(x)\, g_i^r(x)]\, l_i(x), \tag{102}$$

where the index $r$ marks different grammatical meanings of the lexeme $x$, and the weights $\alpha_i^r(x)$ are chosen so that $\Sigma\alpha_i^r(x) = 1$; $\alpha_i^r(x) \geq 0$.

The factor in $\sum_r \alpha_i^r(x)\, g_i^r(x)$ formula (102) is responsible for the contribution of individual grammatical meanings to the semantic state of the lexeme $x$. Note that today we are not aware of work where a differential study of the contribution of individual grammatical meanings to the general semantics of lexemes would be carried out.

Formula (101) can be generalized to the case when it is necessary to take into account the interaction of grammatical and lexical semantics:

$$s_i^r(x) = g_i^r(x)\, I\,(g;\, l;\, x)\, l_i(x), \tag{103}$$

where the member $I\,(g;\, l;\, x)$ is responsible for the interaction between the grammatical and lexical semantics of the lexeme $x$ in the state $s_i^r(x)$.

Let us consider separately the case when different lexemes are characterized by a single complex of lexical meanings. Moreover, we are talking about such lexemes that are not absolute synonyms but vary according to a certain grammatical attribute, which cannot be qualified as inflectional. Examples of this type are quite common, in particular, in the system of the Ukrainian verb, which will be studied in detail in the section «Semantics of Explanatory Dictionaries». Here, grammatical semantics motivates the affiliation of morphological variants of verbs to various lexemes, although they are characterized by a common system of lexical meanings. Such verbs are characterized by one complete semantic state:

$$\mathrm{s}\,(X) = \sum_i \beta_i s_i\,(X);\ \sum_i \beta_i = 1,\ \beta_i \geq 0, \tag{104}$$

where $s_i(X)$ is a semantic state defined by formula (102), in which, however, some clarification should be made regarding the concept of «lexeme». Namely, this concept should be generalized to the complex of all variants of the forms of the word *x*, which have a common complex of lexical meanings. Thus, the concept of a semantic state is more general than the concept of grammatical and lexical meaning. Thus, grammatically-based variability can be typified and formally taken into account in the member $\alpha_i^r(x)\, g_i^r(x)$ of formula (102) by introducing a special parameter *t*, by which the marking of various types of lexeme *x* variability is carried out, which preserve complex of its lexical meanings:

$$\sum_r \sum_t \alpha_i^{rt}(x^t)\, g_i^{rt}(x^t). \tag{105}$$

As a result, the formula for representing the semantic state of the lexeme that taking into account the described variability, which does not lead to a change in the full lexical meaning, takes the form:

$$s(x) = \sum_i \beta_i \left[\sum_r \sum_t \alpha_i^{rt}(x^t)\, g_i^{rt}(x^t)\, I(g; l; x^t)\right] l_i(x), \tag{106}$$

where the coefficients $\beta_i$ can be interpreted as relative weights (middle frequencies) with which various lexical meanings $l_i(x)$, $i = 1, 2, \ldots$ are found in the set K (x).

According to the previous subsection, the formalized description of lexical meanings will be mainly carried out according to Yu. D. Apresyan [77], who, using the semiotic approach of C. Morris [78] and his linguistic interpretation of I.A. Melchuk, suggested to take into account semantic, syntactic and pragmatic aspects when determining lexical meanings. In our formalism using the apparatus of states, this suggestion can be represented as follows:

$$l_i(x) = \sigma_i(x)\, Z_i(x) + \tau_i(x)\, U_i(x) + \pi_i(x)\, V_i(x). \tag{107}$$

In the latter formula, we consider that $Z_i(x)$ represents the semantic parameters of unit *x* in the meaning $l_i(x)$, $U_i(x)$ are the parameters of syntactic, $V_i(x)$ are the parameters of pragmatics; $\sigma_i(x)$, $\tau_i(x)$ and $\pi_i(x)$ are numerical coefficients that represent the contribution (weight, amplitude) of semantic, syntactic and pragmatic aspects, respectively, to the lexical meaning $l_i(x)$, and:

$$\sigma_i(x) + \tau_i(x) + \pi_i(x) = 1;\ \sigma_i(x) \geq 0;\ \tau_i(x) \geq 0;\ \pi_i(x) \geq 0. \tag{108}$$

Thus, the full semantic state of the lexeme *x* takes the following form:

$$\begin{aligned} s(x) = \sum_i \beta_i \left[\sum_r \sum_t \alpha_i^{rt}(x^t)\, g_i^{rt}(x^t)\, I(g; l; x^t)\right] \times \\ \times \left[\sigma_i(x)\, Z_i(x) + \tau_i(x)\, U_i(x) + \pi_i(x)\, V_i(x)\right], \end{aligned} \tag{109}$$

where individual parameters are defined by formulas (99)—(108). Fixating the meanings of these parameters, we obtain various differential semantic states of the lexeme *x*.

[77] Apresyan Yu.D. Izbr. trudy. T. 2: Integral'noe opisanie yazyka i sistemnaya leksikografiya. Moscow: Shkola «Yazyki russkoj kul'tury», 1995. 767 s.

[78] Morris Ch. Signs, language and behaviour. NY: Moss, R.N., 1947.

It should be noted that the operations of multiplication and addition of structural elements of semantic states in formulas (99)—(109) are symbolic. Since as of now we do not have the means to compare specific elements of algebraic structures to these elements, at this stage these formulas give us an intuitive idea of the construction of semantic states rather than their «strict» definition. But, despite this, the analysis of formula (109) allows us to compare the semantic state of the lexeme with the corresponding dictionary entry in the explanatory dictionary, and it presents rather unambiguous interpretations of the composition and structure of this dictionary entry, as well as the dictionary as a whole. And, conversely, from the construction of formula (109), we can deduce the constructions of dictionary entries for semantic dictionaries of different types. The multi-parameter nature of this formula and the substantiality of its components demonstrate its flexibility and great possibilities of explication and detailing a wide range of lexical-grammatical and lexical-semantic phenomena using constructs of semantic states. Thus, the stated formalism has significant potentials for generalizations. From the examples we have examined, we have drawn the conviction that in terms of semantic states, almost any lexico-grammatical and lexico-semantic effect can be formulated. Moreover, the above scheme has not only significant formalization potential but can also be used to construct a meaningful theory of the lexicographic description of the language system. It contains the answer to the question of universal mechanisms operating in the information environment that induce a linguistic substance to acquire a dictionary form.

## Synonymy and linguistic state. The analogy to the definition of translation equivalents

In this section, we consider some formal aspects of the relationship of lexical synonymy, since this relationship plays an important role in the language, in particular, in the processes of interlanguage communication.

We assume that the lexemes $x$ and $y$ are in synonymy relation to $S$, i.e. $xSy$ if there are semantic states (states of lexical semantics) with $c\,(x)$ and $c\,(y)$, which are close: $c\,(x) \approx c\,(y)$.It means that:

$$\left| c\,(x) - c\,(y) \right| < \varepsilon, \tag{110}$$

where ε is some «sufficiently small» quantity.

It is clear that to estimate the quantity $\left| c\,(x) - c\,(y) \right|$, it is necessary to have such formal definitions for c *(x)* and c *(y)* that allow us to calculate the difference c *(x)* — *c (y)* and estimate its absolute value. This is a rather complicated task, which is usually solved by an expert assessment carried out by a linguist who, analyzing the meanings (semantic states) c *(x)* and c *(y)*, and relying on their language and linguistic competence, establish their proximity by themself.

Without dwelling on purely linguistic aspects, we note some formal properties of the relationship of synonymy.

It is reflective and symmetrical, that is $xSx$ and $xSy \Rightarrow ySx$. Moreover, the synonymy relation is not necessarily transitive, i.e., from the facts $xSy$ and $ySz$, generally speaking, $xSz$ does not follow. This means that the differences $|c(x) - c(y)|$ and $|c(y) - c(z)|$ can accumulate so that $|c(x) - c(z)| > \varepsilon$. We consider only such synonymic series (synsets) $SX$ that if $X = \{x_1, x_2, \ldots, x_n\}$ is the set of lexemes forming some synset $SX$, then $x_i \, S \, x_j$ for all $i, j = 1, 2, \ldots, n$ ... This makes it possible to designate the set of meanings of the syxet $SX$ as $C(SX)$. So, for any $c(x) \in C(SX)$ and $c(y) \in C(SX)$, $|c(x) - c(y)| < \varepsilon$.

Note the fundamental difference between the description of lexical semantics in explanatory and synonymous dictionaries. Namely, the explanatory dictionary seeks to present, so to speak, an absolute semantics, i.e. for each lexeme $x$, to determine and describe as fully as possible the set of its semantic states:

$$C(x) = \{c_i(x) \mid i = 1, 2, \ldots, k \ldots\}, \qquad (111)$$

moreover, do it in such a way, that for any context $M(x)$ that contains the word form $x$, it is possible to find an element $c_M(x) \in C(x)$ representing the meaning of the lexeme $x$ in this particular context. The set $C(x)$ contains both pure and mixed semantic states.

At the same time, the semantics of synonymy is relative, and in the formalism of semantic states, it seems to be just the difference $|c(x) - c(y)|$ of semantic states $c(x)$ and $c(y)$ satisfying the condition of the proximity of meaning described above. In practice, the semantics of explanatory dictionaries, as a rule, and unfortunately, are not consistent with the semantics of synonyms, i.e. the definition of semantic states under synonymy, fixed in formula (110), is not always consistent with the definitions of the states of lexical semantics presented in formula (111). The noted fact leads to significant problems in the development of semantic analyzers, bilingual translation dictionaries, and lexicographic systems of machine translation.

Indeed, there is a deep analogy between the definition of synonymy and the finding of translation equivalents. It is based on the fact that if finding a synonym is reduced to searching for close meanings of lexemes of one language, then finding a translation equivalent is the same search for the similar meaning of lexemes, but in different languages. We introduce the superscript for the language unit and its semantic state to indicate the language in which this unit (or state) is represented. Then the synonymy is reflected by the relation $|c^1(x) - c^1(y)| < \varepsilon$.

At the same time, the ratio of «translation equivalent»: $x^1 T x^2$ is represented by the relation $|c^1(x^1) - c^2(x^2)| < \varepsilon$. It is clear that for the formal definition of the latter procedure, it is necessary to bring the meanings $c^1(x^1)$ and $c^2(x^2)$ to some kind of unified representation that, for example, the intermediary language can provide, we denote it by the index 0. Thus, first, we have the reflections $c^1(x^1) \rightarrow c^0(x)$ and $c^2(x^2) \rightarrow c^0(y)$. Then we obtain the equivalence:

$$xSy \Leftarrow |c^0(x) - c^0(y)| < \varepsilon \sim |c^1(x^1) - c^2(x^2)| < \varepsilon \Rightarrow x^1 T x^2. \qquad (112)$$

From the latter formula, it follows that the search for a translation equivalent from one language to another is equivalent to selecting a synonym in some abstract intermediary language. Although such a language remains an unknown quantity of conceptual modeling (in particular, its universality remains non-obvious), formula (112) gives a key to understanding the role of synonymy in translation processes: the construction of the quantities $c^0(x)$ and $c^0(y)$ allows the presence of the sets $C^1(SX^1)$ *и* $C^2(SX^2)$ and the establishment of certain correspondence between them — even if we do not have a formal method for determining such a correspondence.

Note the crucial role of context in translation processes. Consider an example. The title of Alekseyev's novel «Хлеб — имя существительное» («Bread is a noun») can be translated into Ukrainian language, so to speak, in «point blank», as «Хліб — іменник» because «*имя существительное*» in Ukrainian is «*іменник*». However, in Russian, the phrase *X* = «*noun*» is in a mixed semantic state:

$$c\,(X) = \alpha_1\,(c_1)\,c_1\,(X) + \alpha_2\,(c_2)\,c_2\,(X), \tag{113}$$

where in the latter formula the full semantic state $c(X)$ is a combination of two elementary semantic states $c_1(X)$ $and c_2(X)$ with the following interpretation: $c_1(X)$ represents the usual lexical meaning for *X* as part of speech, and the element $c_2(X)$ represents a new lexical meaning, which is a reflection of a completely different cognitive situation that has nothing to do with grammatical terminology. Namely, in this second meaning, «*имя существительное*» acquires the meaning of «something that ensures the possibility of human existence» (that is, «хлеб» («bread») is the name of a substance that ensures the existence of people and to some extent acts as a symbol of human existence). You can consider this meaning as figurative, thinking that in some initial meaning of the phrase «имя существительное», when its components were integrated into a separate term, their partial desemantization had occurred. The formally described situation can be interpreted by the following diagram:

«Хлеб — имя существительное» → «Хліб — іменник»

$$\begin{array}{ccc} \text{«Имя существительное»} & \xrightarrow{C^R} & c^R(X) = \alpha_1(c_1)c_1(X) + \alpha_2(c_2)c_2(X) \\ \Big\downarrow t^{RU} & & \Big\downarrow C^{RU} \\ \text{«Іменник»} & \xrightarrow{C^U} & c_1(X) \end{array} \tag{114}$$

The requirement of adequate translation leads to the need for the inverse operators $(t^{RU})^{-1} \equiv t^{UR}$ and $(C^{RU})^{-1} \equiv C^{UR}$. This is impossible since the $C_{RU}$ operator is a projection and the inverse one does not exist, it is clear that knowledge of only the quantity $c_1(X)$ is not enough to restore three more quantities: $\alpha_1(c_1)$, $\alpha_2(c_2)$ *и* $c_2(X)$.

# DEMONSTRATION OF SYSTEMIC EFFECTS IN THE SEMANTICS OF EXPLANATORY DICTIONARIES

## Introductory remarks

The systematic nature noted above has a very non-trivial embodiment in the structure of large explanatory dictionaries. It is so peculiar that it allowed us to interpret it as a separate phenomenon, which we qualify as the *semantics of explanatory dictionaries*. To clarify what has been said, it is necessary to say a few words about the role that explanatory dictionaries play in the scientific description of a language. Let us begin with some cultural-historical distraction.

Callimachus (ancient Greek. Καλλίμαχος; about 310, Cyrene — about 240 BC, Alexandria), one of the most prominent representatives of Alexandrian poetry, a critic, bibliographer, compiler of the Alexandria Library multivolume catalog, probably impressed by the titanic work he did, spoke as follows: «*Μέγα βιβλίον μέγα κακόν*» («*The big book is a big evil*»). Compilers of large explanatory dictionaries, at least of the Oxford English Dictionary, from where we borrowed this quote, experienced (and continue to experience) similar feelings. To summarize, so to speak, the theoretical justification for this worldview, which (justification) we see in the system properties of the lexicographic system, which claims to be the most complete (systematic) description of the language system.

We believe that the systemic effects of the language are most fully represented in monolingual dictionaries, which, in our opinion, in addition to the reference function, play a much more fundamental role in the system of language scientific description. Here we are talking about large multivolume lexicons that contain the bulk of the national vocabulary and phraseology and are characterized by a detailed description of the lexico-grammatical and lexico-semantic system of the language. Due to the large volume, complexity, and, at the same time, the regularity of the structure, as well as the completeness of the lexicographic coverage of the language material, such dictionaries are carriers of a huge number of implicitly defined linguistic, cognitive, logical, and other connections and relationships. That marked connections and relations, unfortunately, are mostly uncontrollable, which turns these large lexicographic systems into a kind of «things in themselves». Indeed, the user of a large vocabulary «glides» along its surface only, without having effective means to penetrate the depth of its content. And, even worse, such an op-

portunity is not available even for compilers of the dictionary, who are forced to take into account the systemic relationships of the language, relying only on their own experience and lexicographic intuition. Aforesaid, in our opinion, is the basis for the mentioned above statement that «the big book is a big evil».

We emphasize the role of large dictionaries in the modern era as semantic analyzers, taggers, elements of conceptual and orthographic systems, resident language support elements for various types of applied language information systems and processes (text examination, legislation, and law enforcement, knowledge engineering, etc.). Thus, the importance of dictionaries in modern information technologies is constantly growing, which makes it necessary to convert them into digital formats. But this same factor puts forward new requirements for the quality of dictionaries and the means and methods of their inner semantic and logical structure explication. Therefore, arises the problem of identifying and analyzing systemic relationships concentrated in explanatory dictionaries. Some of them are well known and understood. However, here we will focus on the analysis of relationships that are not usually considered in standard lexicography.

Our study is based on the material of the Dictionary of the Ukrainian language in 11 volumes (SUM-11)[79] and the Dictionary of the Ukrainian language in 20 volumes (SUM-20)[80], created in the Ukrainian Lingua-Information Fund of the National Academy of Sciences of Ukraine with the participation of Potebnia Institute of Linguistics of the National Academy of Sciences of Ukraine. Note that we are talking about electronic versions of these dictionaries, because, as it will be shown in the course of the presentation, it is almost impossible to identify the patterns described by us in the so-called «manual mode». To create SUM-20, the Ukrainian Lingua-Information Fund of the National Academy of Sciences of Ukraine developed a special computer lexicography technology embodied in the Virtual Lexicographic Laboratory «Dictionary of the Ukrainian Language» (VLL_SUM), which not only ensures the joint, coordinated efforts of a large, geographically distributed lexicographers team in the implementation of a large vocabulary project but also provides the necessary tools for conducting studies of the system properties of the language, implicitly represented in the dictionary. The first step in creating VLL_SUM was the parsing of SUM-11, that is, the digitalization of the text of this dictionary, which was published in the 70s of the last century in traditional paper form. To create this digital-analog, the formal structure of this dictionary was developed (in our terminology, the structure of the lexicographic system of SUM-11), which turned out to be rather rich. Formal features

[79] Slovnyk ukrainskoi movy: V 11 t. / Hol. red. I.K. Bilodid. Kyiv: Nauk. dumka, 1970—1980.

[80] Slovnyk ukrainskoi movy: T. 1 / Hol. red. V.M. Rusanivskyi. Kyiv: Nauk. dumka, 2010. T. 2 / Hol. red. V.M. Rusanivskyi. Kyiv: Nauk. dumka., 2011. T. 3 / Hol. red. V.M. Rusanivskyi. Kyiv: Viakhyr, 2012. T. 4 / Hol. red. V.M. Rusanivskyi. Kyiv: Ukr. movno-inform. fond, 2013. T. 5—7 / Hol. red. V.A. Shirokov. Kyiv: Ukr. movno-inform. fond, 2014—2016.

playing the role of markers of elements of this structure were identified. Then, the digital text was marked up using found and identified structure markers, a lexicographic database structure corresponding to the formal structure of the SUM-11 lexicographic system was developed, and a program procedure for automatically transporting the marked SUM-11 text to the corresponding database was implemented, which, were the contents of the dictionary parsing procedure. This approach allowed us to avoid many errors that would be inevitable in the «manual» mode of forming the lexicographic database, not to mention the fact that automatic parsing saved a lot of working time. Subsequently, an appropriate interface was developed, which gave this system the properties of instrumentalism and virtuality. Details of this process and the system as a whole are described in our monographs [81]. Thus, the VLL_SUM platform was created, on which some system effects were investigated, and we will pass on to presenting them.

## The first systemic relation: lexicographic completeness and fullness. Automorphisms of L-systems, hyperchains, and hypercycles

We start with the systemic relationship induced by the so-called hyperchains on lexicographic structures, which are introduced as follows.

Let a lexicographic system (hereinafter: L-system) $V(U_0)$ be given, where $U_0$ is the set of headwords of the corresponding dictionary. In the structure σ [β] of the L-system $V(U_0$, we highlight a subset **A** whose elements $A$ ($A \in$ **A**)are called *automorphisms* of the $V(U_0)$. The meaning of these elements is that they determine the inner reflection of the L-system:

$$A\colon V(U_0) \to V(U_0) \tag{115}$$

of a special type, namely, reflection between dictionary entries:

$$A\colon V(x) \to V(y)$$

for various headwords $x$ and $y$ belonging to the set $U_0$. The link chain length can be longer than 1, so there can be chains of a recursive type:

$$V(x) \to \{V(x')\} \to \ldots \to \{V(x'')\} \to \ldots .$$

Thus, automorphisms from the set **A** (we shall call them A-automorphisms) induce the reflection of the set of headwords into itself:

$$A\colon U_0 \to U_0,\ A \in \mathbf{A}. \tag{116}$$

Some general properties of automorphisms of L-systems were considered in the section «Lexicographic Systems and Dictionaries».

[81] Shirokov V.A., Bilonozhenko V.M., Buhakov O.V. tain. Linhvistychni ta tekhnolohichni osnovy tlumachnoi leksykohrafii / Vidp. red. V.A. Shirokov. Kyiv: Dovira, 2010. 295 s.; Shirokov V.A. Komp'iuterna leksykohrafiia. S. 163—222.

Consider one of the algorithms of such reflection, which has a transparent semantic interpretation.

As it is known, the obligatory, non-optional element of the dictionary entry of the explanatory dictionary is the lexical meaning, presented in the lexicographic structure of the dictionary definition. We postulate two basic systemic principles that we consider mandatory when compiling large explanatory dictionaries.

The first of them is ***lexicographical closeness***, requiring that all words included in the dictionary definitions are present in the dictionary registry $U_0$ as headwords for the corresponding dictionary entries.

The second principle is ***lexicographic completeness***, which requires that all the meanings of the words presented in the definitions, be developed in the corresponding dictionary entries, where these words are head ones.

We study the lexicographic closeness and completeness with the help of A-automorphisms as follows. Consider a chain of registry words extracted from dictionary definitions using A-automorphisms:

$$\begin{array}{ccccccccc} & A_1 & & A_2 & & A_3 \ \ldots \ A_{n-1} & & A_n & \\ x_0 & \rightarrow x_1(x_0) & \rightarrow & x_2(x_1) & \rightarrow & \ldots & \rightarrow & x_n(x_{n-1}) & \rightarrow \ x_n(x_{n-1}) \end{array} \qquad (117)$$

The process of building a chain is as follows. In the L-system associated with the dictionary $V$, we select a specific dictionary entry with the registry unit $x_0$. Ithas a structure, $V_0 = \partial V_0 \cup M_0$, where $\partial V_0 \equiv x_0$, and the inner part of $M_0$ has a tuple($x_1^0$, $x_2^0$, ..., $x_{r0}^0$), which is formed as a sequence of words from s definition presented in initial form. A specific element $x_1(x_0)$ could be selected from this tuple, which, in turn, is the headword of the dictionary entry $V_1$ through the procedure $A_1$. The following steps of the algorithm are similar. Thus, we obtain a chain of words (let's call it a hyperchain on a lexicographic structure):

$$x_2(x_1), \ldots, x_{n-1}(x_{n-2}), x_n(x_{n-1}), \qquad (118)$$

which is obtained by the consistent application of automorphisms $A_2, A_3, \ldots, A_n$.

Consider an example of constructing a hyperchain using the SUM-20 dictionary entry with the registry unit $x_0$ = АБАЖУР. Its lexical meaning is given by definition: «*Частина світильника, звичайно у вигляді ковпака, призначена для зосередження і відбиття світла та захисту очей від його впливу*». We choose the word «*світильник*» as the element $x_1(x_0)$, which is in the relationship «part — whole» with the word «абажур». One of the meanings of the selected word is: «*Прилад для освітлення — електричний або у вигляді лампи, плошки, куди наливають олію і вставляють ґніт*». Applying the automorphism $A$, we obtain: $A_2[x_1(x_0)] = x_2(x_1)$, where $x_2(x_1) = прилад$. The word «прилад» is in the relationship «genus—aspect» with the word «світильник». Continuing this process, we obtain the following hyperchain:

$Z_1 = абажур \rightarrow світильник \rightarrow прилад \rightarrow пристрій \rightarrow обладнання \rightarrow прилад.$

We see that at the fifth step the chain is closing, and turns into a cycle. Such varieties of hyperchains are called hypercycles.

This phenomenon was very successfully used by Stanislav Lem in his «Star Diaries of John the Quiet» for a comic effect:

«...I went to the Tarantoga to read about the sepulcas. And had found the following brief information:

«SEPULKI — a prominent element of the civilization of Ardrites from the planet of Enteropia (see). See SEPULKARIA».

I followed this advice and read:

«SEPULKARIA — establishments used for sepuling (see)».

I searched for «Sepuling»; it said:

«Sepuling — an activity of Ardrites from the planet of Enteropia (see). See SEPULKI».

The circle closed, there was nowhere else to look».

In this passage, the lexicographical style is played out and parodied — the style of compilers of dictionaries, reference books, encyclopedias. However, we recognize that the requirement of completeness and closeness leaves them no other opportunity. And, if you look at it, this phenomenon is quite didactic, and very non-trivial consequences can be deduced from it. It turns out that the described structures have another interesting and unexpected logical-linguistic effect related to Gödel's incompleteness theorem[82]. The fact that a hyperchain of sufficient length is closed demonstrates a certain logical defect of definition (tautology)since, in the end, it turns out that some word *x* is interpreted as some *y*, and *y* is interpreted as the same *x*. Considering formally dictionary definitions as axioms of some theory, and applying Gödel's incompleteness theorem to such a «formal» system, we conclude: in this theory, there must necessarily exist statements that can neither be proved nor disproved by the means of the theory itself. The closeness of hyperchains and their reduction to hypercycles is a manifestation of the incompleteness of the lexicographic system in the sense of Gödel.

Consider another example of a hypercycle:

$$Z_2 = \textit{абажур} \rightarrow \textit{ковпак} \rightarrow \textit{накривка} \rightarrow \textit{накривання} \rightarrow \textit{накривати} \rightarrow$$
$$\rightarrow \textit{закривати} \rightarrow \textit{загороджувати} \rightarrow \textit{загорода} \rightarrow \textit{загоро́дження} \rightarrow$$
$$\rightarrow \textbf{\textit{перешкода}} \rightarrow \textit{перепона} \rightarrow \textbf{\textit{перешкода}}.$$

In this hypercycle, there are not only nouns but also verbs. As it could be seen, hyperchains and hypercycles represent some semantic characteristics of the initial words. They have a linguistic meaning since any hyperchain word appears in a certain («hyperonymic») relation to its predecessor. The complex of relations embodied in hyperchains is much richer because they include the relations «complex-element», «cause-effect», as well as still not satisfactorily specified associa-

[82] Uspenskij V.A. Teorema Gyodelya o nepolnote. Moscow: Nauka, 1982. 110 s.

tion relations and many others. At least, it is immediately obvious that hyperchains generalize the relations of hypohyperonymy.

On the other hand, if one succeeds in constructing, in an explicit form, a system of explanatory dictionary hyperchains and meaningfully marking the relations existing in these hyperchains, then the L-system obtained in this way will turn into a powerful lexico-semantic analyzer that can be used in the contours of text mining systems.

Finally, the hyperchain method is a powerful tool for ensuring the quality of the lexicographic description of a language. If a lexicographer, while developing dictionary entries, also examines hyperchains on the corresponding lexicographic structures, then they will be protected from many mistakes and inaccuracies. Unfortunately, this mode of operation is possible only in the environment of a digital instrumental lexicographic system, while in manual mode it is practically impossible due to its enormous complexity.

## The second systemic relation: requisites sets structure

One of the principles of building the Dictionary of the Ukrainian language is the provision that in the registry area of a dictionary entry all lexemes are presentedand that they are characterized by a single, common complex of lexical meanings. And here we are not talking about lexemes, which are absolute synonyms, but about certain cognate words, varying according to some morphological characteristics, which cannot qualify as inflectional. Examples of this kind are quite common, in particular, in the system of the Ukrainian verb. Consider for illustration a dictionary entry from the Dictionary of the Ukrainian language with the registry unit «НАЛАШТОВУВАТИ»:

**НАЛАШТО́ВУВАТИ,** ую, уєш, *недок.*, **НАЛАШТУВА́ТИ,** у́ю, у́єш, *док.*, *що.* 1. Готувати що-небудь або приводити до стану готовності. *Сакен почав налаштовувати парус з таким розрахунком, щоб поманити ще турків* (С. Добровольський); *— Не знаєте, чи хазяїн уже устали? — спитався він. — Казали, сьогодні кудись їхати, так щоб віз налаштувати* (Панас Мирний); *Уляна висипала з пелени квасолю і вже хотіла йти до хати, щоб налаштувати їжу кабанцеві* (Григорій Тютюнник);. *кого.* Настроювати кого-небудь на певний лад, настрій і т. ін. *Галина ж, ідучи сюди, налаштувала себе на звичайність, навіть мимохідність свого завітання* (М. Олійник).

2. *що, чого.* Робити що-небудь придатним для роботи, користування і т. ін. *— Під час вечері Мишуня налаштовував радіоприймача* (Ю. Яновський); *— Братчики, до гармат! — закричав Гулик. З ним було до десятка старих гармашів, які одразу ж кинулись до ворожих єдинорогів і почали налаштовувати їх до стрільби* (С. Добровольський).

The left side of this dictionary entry:

**НАЛАШТО́ВУВАТИ,** ую, уєш, *недок.,* **НАЛАШТУВА́ТИ,** у́ю, у́єш, *док.* is represented by the aspect pair «**налаштовувати**» (imperfective verb) and «**налаштувати**» (perfective verb). The indicated verbs belong to different verb inflectional classes and different inflectional paradigms. Thus, grammatical semantics motivates the belonging of these verbs to various «lexemes». Despite this, they are characterized by a common system of lexical meanings, as can be seen from the lexicographic development of the given dictionary entry.

Let us call sequences of lexemes with the defined common property of a complex of lexical meanings *equisemantic series*. As analysis shows, in the Ukrainian language, the above variability can be carried out through various morphological means. Here are some of them.

**Prefixal variation:**

а) вбік~убік, вбілений~убілений, вбогість~убогість, вболівати~уболівати, (phonetic prefixal variation; provided by the phonetic variation of the prefix: «в- ~ у- »);

б)зглянутися~ізглянутися~зоглянутися;згарячу~ізгарячу;ззаду~іззаду, зобганий~зібганий; відбивати~одбивати (structural-phonetic prefixal variation; provided by complex phonetic and structural variation of the prefix «з- ~ із- ~ зо- ~ зі-»).

**Root variation:**

кмітити~кметити; кмітувати~кметувати;

**Suffix variation:**

а) УСТРІНУТИ~УСТРІТИ (does not lead to a change of the inflectional class and verb type);

б) БУЛЬКОТАТИ (очу́, о́чеш) ~ БУЛЬКОТІ́ТИ (очу́, оти́ш) — (leads to a change in the inflectional class, but does not lead to a change in the verb type);

в) ВИГОВО́РЮВАТИ (*недок.*) ~ ВИ́ГОВОРИТИ (*док.*) — (leads to a change in theverb type).

A combination of different types of variation is also possible. Here are few examples of equisemantic series of various multiplicity from the Dictionary of the Ukrainian language (here are left (registry) side of the corresponding dictionary entries; in parentheses in front of the registry unit is the multiplicity of the registry series, i.e. the number of words in this series):

**1. (1) АБЕТКУВА́ТИ**, у́ю, у́єш, *недок., перех.*

**2. (2)ІТИ́ (ЙТИ),** іду́ (йду), іде́ш (йдеш); *мин. ч.* ішо́в (йшов), ішла́ (йшла), ішло́ (йшло), ішли́ (йшли); *наказ. сп.* іди́ (йди); *недок.*

**3. (3) ЗГЛЯ́НУТИСЯ** *і рідше* **ІЗГЛЯ́НУТИСЯ** *і* **ЗОГЛЯ́НУТИСЯ**, нуся, нешся, *док.*

**4. (4) УСТРІ́НУТИ** (**ВСТРІ́НУТИ**) *і* **УСТРІ́ТИ** (**ВСТРІ́ТИ**), і́ну, і́неш; *мин. ч.* устрі́нув, нула, ло *і* устрі́в, рі́ла, ло; *док., діал.*

**5. (5) СПИНА́ТИСЯ**, а́юся, а́єшся, *недок.,* **СП'ЯСТИ́СЯ** *і* **ЗІП'ЯСТИ́СЯ, ЗІПНУ́ТИСЯ** *і* **ЗІП'Я́ТИСЯ**, зіпну́ся, зіпне́шся, *док.*

**6. (6) ЗЛА́ЗИТИ**[1], зла́жу, зла́зиш *і рідко* **ІЗЛА́ЗИТИ**, ла́жу, ла́зиш *і* **ЗЛІЗА́ТИ** *і рідко* **ІЗЛІЗА́ТИ**, а́ю, а́єш, *недок.,* **ЗЛІ́ЗТИ** *і рідко* **ІЗЛІ́ЗТИ**, зу, зеш; *мин. ч.* зліз *і рідко* ізлі́з, ла, ло; *док.*

**7. (8) ВПИХА́ТИ** (**УПИХА́ТИ**), а́ю, а́єш, *недок.,* **ВПХНУ́ТИ** (**УВІПХНУ́ТИ, УПХНУ́ТИ**), ну́, не́ш *і* **ВПХА́ТИ** (**УВІПХА́ТИ, УПХА́ТИ**), а́ю, а́єш, *док., перех.*

As can be seen, the number of members of the equisemantic series in the SUM varies from 1 to 8, and no verb series with a multiplicity of 7 were found in the Dictionary. The question arises: Is there a formal law governing the filling rule and, therefore, the principle of constructing equisemantic series? The answer to this seemingly complex question is positive, at least for the Ukrainian verbal equisemantic series, the structure of which we managed to establish in the first half of the 90s of the last century.

Namely: if we consider the texts of the left parts $\Lambda(X)$ of dictionary entries, with headwords $X$, where $X$ runs through many verbs of the Ukrainian language, abstracting from their content, i.e. as linear chains of symbols, it can be noted that certain «subchains» are distinguished in the structure of these chains so that the above-mentioned selections have the nature of enclosures. Moreover, these enclosures are distinguished in a certain invariant way, such that the various subchains do not intersect with each other. The noted patterns of the structure of the chains $\Lambda(X)$ are so regular that it can be formulated even in axiomatic form, which we give here in the form of three statements.

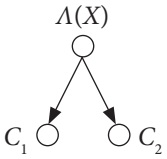
The first statement is quite obvious:

(1) Each Ukrainian verb is presented in the language in a form of a lexeme (with defined and fixed lexical semantics), which has one or two meanings of a category of an aspect.

The lexicographic representation of the lexeme belonging to a certain meaning of the «ASPECT» attribute (as presented in the SUM) is called an aspect complex or simply a complex and denoted by the symbols $C_1$ and $C_2$. In the presence of one aspect complex, we have an enclosure of the first level $C_1 \subset \Lambda\,(X)$, that have the following graphic representation:


(119)

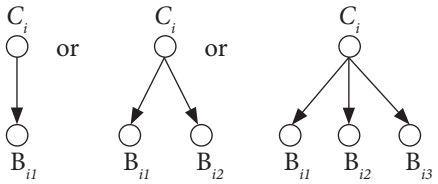
where the attachment ratio «⊂» is represented by the arrow. Similarly, the two-complex $\Lambda(X)$ is graphically represented as follows:




(120)

The second statement:

(2) Each verb lexeme with a specific and fixed lexical semantics and a specific meaning of the aspect can be represented by verbs belonging to no more than three different inflectional (paradigmatic) classes. Identifiers of inflectional classes in the SUM are paradigmatic indicators, which, as a rule, are sets of quasiflexions of the first and second person singular of the present tense; if they are not enough to determine the inflection system, then some other forms are given — one way or another, the set of paradigmatic indicators is chosen to be sufficient to identify the inflectional paradigm of a particular lexeme. To lexicographically represent these facts, we introduce the term paradigmatic block, or simply a block, denoting it with the symbol $B_{ij}$, where the index «*i*» corresponds to the number of the aspect complex to which this block belongs and takes meanings 1, 2, and the index «*j*» numbers the blocks inside aspect complex and takes on meanings from 1 to 3. Thus, each aspect complex has one, two, or three inflectional blocks. Graphically, this is represented as follows:




(121)

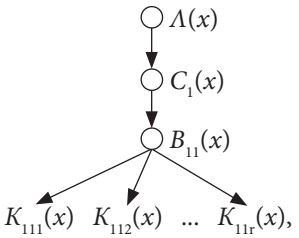
Finally, the third statement:

(3) Each verb lexeme with a specific meaning of the form, defined by belonging to a particular inflectional class, and defined and fixed by lexical semantics, can be represented by verbs that have no more than four phonetic variations (as a rule, this is a prefixal and root variation, but also suffixed).

For the lexicographic presentation of this fact, we introduce the term «component». Components are indicated by $K_{ijr}$ symbols, where the «*i*» index corresponds to the number of the aspect complex to which the block with «*j*» belongs, and the «*r*» index runs through meanings from 1 to 4 depending on how many components the block in question has. We introduce the denomination. By the symbol $I^1$ we denote the set of numbers 1, 2, 3, 4:

$$i \in I^1,\ i = 1, 2, 3, 4, \tag{122}$$

by the elements of which we enumerate verb components defined by our 3rd axiom, namely phonetic versions of verbs, characterized by specific meanings of inflectional parameters. Then, a setting of the $r$ ($r \in I^1$) signals the presence of a verb, characterized by *one* meaning of the «ASPECT» attribute with *one* set of meanings of paradigmatic indicators inherent in it and the number of phonetic variants equal to *r*. Denoting the attachment ratio by the arrow, i.e.: $A \supseteq B \Rightarrow A \rightarrow \rightarrow B$, this fact can be represented by the graph:




(123)

where the symbol *Λ(x)* denotes the left side of the dictionary entry with the registry unit *x*; $C_1(x)$ is its only complex; $B_{11}(x)$ is the only block in this complex, and $K_{111}(x)$, $K_{112}(x)$, ..., $K_{11r}(x)$ are, respectively, the 1st, 2nd, ..., *r*-th components of this block, where*r* can run through meanings from 1 to 4.

The Cartesian product $I^1 \times I^1 \equiv I^2$ parameterizes the structures of two-block complexes, and $I^1 \times I^1 \times I^1 \equiv I^3$ — structures of three-block complexes. Thus, the structure of any aspect complex is represented by one of three signatures:

$$(i),\ (i\ j),\ (i\ j\ k), \tag{124}$$

in each of which the numbers *i, j, k* are run independently through the set $I^1 = \{1, 2, 3, 4\}$. Signature (*i*) corresponds to a graph of type (123); the signature (*i j*) represents a complex having two paradigmatic blocks, in the first of which there are *i* components, and in the second — *j* components; a signature (*i j k*) represents a complex having three paradigmatic blocks; in the first of them there are *i* components, in the second — *j* components, and in the third — *k* components.There can be only 4 one-block signatures, 16 two-block signatures, and 64 three-block signatures. Thus, the aspect complex can be realized with a maximum of 84 different patterns: (4 + 16 + 64 = 84). In terms of signatures, we get the Table 3.

A dictionary entry, containing both aspect complexes, is represented by a pair (α . β), where α and β are run independently through the set of signatures (*i*), (*i j*), (*i j k*).Signatures of type (α. β) can have a maximum of 7056 = (84 × 84).

So, the total number of structures satisfying axioms (1)—(3) can theoretically be 7140 = (84 + 7056). Denote them by: (α) are single-complex signatures and (α. β) are two-complex signatures. The graphical interpretation of the signatures (α) and (α. β) is obvious.

It can be determined what signatures are used to implement verbal dictionary entries containing a given number of lexemes in the registry. The maximum possible number of registry entries for a single-complex Λ(*X*) is 12 (a maximum of three blocks and a maximum of four components in each block). The general distribution of signatures by the number of elements in the registry series is given in the Table 4.

Thus, the latent symmetry of the verbal Λ(*X*) leads to the fact that the unlimited (at first glance) number of structures of dictionary entries is reduced to a finite, and precisely to a defined integer. The constructed formalism, the structure of signatures, and their distribution are direct consequences of axioms 1—3. We,

in the early 90s, when the computer database of the Dictionary was not yet made, formulated these axioms as a result of observations on the texts of verb entries in the 11-volume Dictionary of the Ukrainian language. Then we raised the question: Do all verbal dictionary entries of the Dictionary satisfy these axioms and are there no exceptions to the «1-2-3-4» rule?

*Table 3.* **Signatures of aspect complexes**

| 1-block ($I^1$) | 2-block ($I^2$) | | | |
|---|---|---|---|---|
| 1<br>2<br>3<br>4 | 11<br>21<br>31<br>41 | 12<br>22<br>32<br>42 | 13<br>23<br>33<br>43 | 14<br>24<br>34<br>44 |

| 3-block ($I^3$) | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 111 | 112 | 113 | 114 | 211 | 212 | 213 | 214 | 311 | 312 | 313 | 314 | 411 | 412 | 413 | 414 |
| 121 | 122 | 123 | 124 | 221 | 222 | 223 | 224 | 321 | 322 | 323 | 324 | 421 | 422 | 423 | 424 |
| 131 | 132 | 133 | 134 | 231 | 232 | 233 | 234 | 331 | 332 | 333 | 334 | 431 | 432 | 433 | 434 |
| 141 | 142 | 143 | 144 | 241 | 242 | 243 | 244 | 341 | 342 | 343 | 344 | 441 | 442 | 443 | 444 |

*Table 4.* **The general distribution of signatures**

| Number of words in the equisemantic series of the aspect complex | Signatures by which vocable entries with a certain number of members of the equisemantic series of the aspect complex are represented | | | | | | | | | | | | | | Number of signatures |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | |
| 1 | 1 | | | | | | | | | | | | | | 1 |
| 2 | 2 | 11 | | | | | | | | | | | | | 2 |
| 3 | 3 | 21 | 12 | 111 | | | | | | | | | | | 4 |
| 4 | 4 | 31 | 22 | 13 | 121 | 112 | 211 | | | | | | | | 7 |
| 5 | 41 | 32 | 23 | 14 | 131 | 122 | 113 | 221 | 212 | 311 | | | | | 10 |
| 6 | 42 | 33 | 24 | 141 | 132 | 123 | 114 | 231 | 222 | 213 | 321 | 312 | 411 | | 13 |
| 7 | 43 | 34 | 142 | 133 | 124 | 241 | 232 | 223 | 214 | 331 | 322 | 313 | 421 | 412 | 14 |
| 8 | 44 | 143 | 134 | 242 | 233 | 224 | 341 | 332 | 323 | 314 | 431 | 422 | 413 | | 13 |
| 9 | 144 | 243 | 234 | 342 | 333 | 324 | 441 | 432 | 423 | 414 | | | | | 10 |
| 10 | 244 | 343 | 334 | 442 | 433 | 424 | | | | | | | | | 6 |
| 11 | 344 | 443 | 434 | | | | | | | | | | | | 3 |
| 12 | 444 | | | | | | | | | | | | | | 1 |
| | | | | | | | | | | | | | | | 84 |

The answer to this question at that time was very difficult since it was simply unrealistic to view the structures of more than 41 thousand verbal dictionary entries developed in the Dictionary in the «manual» mode. We solved this problem only in 2001 after the Dictionary database was built, which allowed us to create an instrumental complex for its modernization and a whole series of studies. Among them, the structures of the verb $\Lambda(X)$ were studied for their compliance with axioms 1—3 and the rule «1-2-3-4». As a result of a computational experiment on the Dictionary database, out of the potentially possible 7140 classes of verbal $\Lambda(X)$, only 52 classes were identified. Their list is given in Appendix 1.

We make a few comments.

Firstly, the resulting model determines a certain classification on the set of Ukrainian verbs. Indeed, we denote the obtained verb classes, marked with the signatures presented in Table 4, by $\lambda_1, \lambda_2, ..., \lambda_{52}$, and the sets of verbs belonging to the corresponding class — by $q(\lambda_1), q(\lambda_2), ..., q(\lambda_{52})$. Then it is obvious that:

$$q(\lambda_i) \cap q(\lambda_j) = \varnothing, \text{ when } i \neq j, \tag{125}$$

that is, any of the verbs can belong (by modulo homonymy) to one and only one of the classes given in the table[83]. This means that this classification is correct.

Besides it is quite accurate, because we, through automatic analysis of the lexicographic structures of the Dictionary, conducted a computational experiment to determine exactly those signatures $(\alpha)$ and $(\alpha.\ \beta)$ that were implemented on its entire array. As a result, out of the potentially possible 7140 classes of verbal $\Lambda(X)$, it was identified precisely those 52 classes that are presented in the Appendix. But most importantly, no exceptions were found from the «1-2-3-4» rule. Moreover, the «1-2-3-4» rule and the following classification does not give errors on the array of the new 20-volume version of the Dictionary, created in the Ukrainian Lingua-Information Fund, and numbering more than 48 thousand verbs.

The theory of semantic states provides a natural interpretation of the derived regularity, and this is done as follows. Let's suppose that there is an intelligent mechanism that, analyzing the structure of the verb lexeme $X$, identifies its semantic state $\psi(X)$, i.e., catalogs its grammatical and semantic characteristics. Formally, such mechanism can be represented by the formula:

$$S\,\psi(X) = \sigma\,\psi(X), \tag{126}$$

where the symbol S denotes the operator, which we interpret as the symmetry operator of the equisemantic series, and $\sigma$ is the eigenvalue of this operator: $\sigma \in \{(\delta); (\alpha.\ \beta)\}$. For example, we will consider the equation (18) for $X$ = «ЗСО́ХНУТИ»**:**

$$S\,\psi(\text{ЗСО́ХНУТИ}) = (\alpha.\ \beta)\,\psi(\text{ЗСО́ХНУТИ}), \tag{127}$$

where the signatures $\alpha$ and $\beta$, respectively, are equal to: $\alpha = (1, 2)$; $\beta = (1, 4)$.

[83] The classifier for verbs, of course, is the complete set of 7140 classes, parameterized by the signatures $(\alpha)$ and $(\alpha.\ \beta)$.

This fact is a consequence of the structure of the dictionary entry:

ЗСИХА́ТИ *і рідко* ІЗСИХА́ТИ, а́ю, а́єш, *недок.,* ЗСО́ХНУТИ *і рідко* ІЗСО́ХНУТИ, ЗСО́ХТИ *і рідко* ІЗСО́ХТИ, хну, хнеш; *мин. ч.* зсох *і* зсо́хнув, ла, ло; *док.,*

Which itself defines this structure:

$$(X, Y)\ par\ (X, Y)\ недок.,\ (Z.Q.P.T)\ par\ (Z.Q.P.T)\ док. \qquad (128)$$

Text correlates of elements of this structure have the following representation:

Complex: $C_1$ = ЗСИХА́ТИ *і рідко* ІЗСИХА́ТИ, а́ю, а́єш,

Unit: $B_{11}$ = ЗСИХА́ТИ *і рідко* ІЗСИХА́ТИ

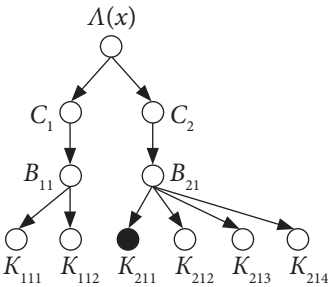
Components: $K_{111}$ = ЗСИХА́ТИ; $K_{112}$ = ІЗСИХА́ТИ;

Complex: $C_2$ = ЗСО́ХНУТИ *і рідко* ІЗСО́ХНУТИ, ЗСО́ХТИ *і рідко* ІЗСО́ХТИ, хну, хнеш; *мин. ч.* зсох *і* зсо́хнув, ла, ло

Unit: $B_{21}$ =ЗСО́ХНУТИ *і рідко* ІЗСО́ХНУТИ, ЗСО́ХТИ *і рідко* ІЗСО́ХТИ

Components: $K_{211}$ = ЗСО́ХНУТИ; $K_{212}$ = ІЗСО́ХНУТИ; $K_{213}$ = ЗСО́ХТИ; $K_{214}$ = ІЗСО́ХТИ.

A graphic representation of the structure (α. β) on the state ψ (ЗСО́НУНИ), where α = (1, 2); β = (1, 4), is represented by the graph:



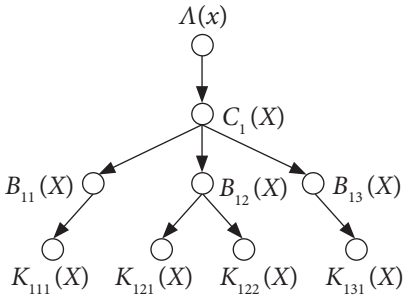

(129)

where the element $K_{211}$, corresponding to the element «ЗСО́ХНУТИ» of the equisemantic series, is darkened.

A few words about the «empty» classification structures, those for which no suitable verbs were found in the Dictionary. This fact does not bother us. Let us recall, for example, the history of the periodic classification of chemical elements — the number of cells in it is theoretically not limited at all, although during the time of D.I. Mendeleev knew no more than 63 elements, and many of them were not very well-identified. Table created D.I. Mendeleev's had some gaps, but the system — a periodic law — made it possible to predict the properties of «missing» elements. Note that even in our time only 118 elements are discovered, the last of which are superheavy, unstable, have a very short lifetime, do not occur in nature, and can only be obtained artificially by accelerators.

Likewise, in our case, the existence of certain verbs related to «missed» classes is not ruled out. The presence of vacant classes signals the hitherto unused word-building potentials of the Ukrainian verb. Perhaps more in-depth studies

and the development of the language system will make it possible to identify such classes, especially since the system itself «predicts» their morphological properties. For example, among the single-complex signatures identified in the Dictionary, there is a signature (11), and it corresponds, for example, Λ(*X*) = БУЛЬКОТА́ТИ, очу́, о́чеш *і* БУЛЬКОТІ́ТИ, очу́, оти́ш, *недок.*, and signature (12) corresponds to, for example, Λ(*X*) = РИБА́ЛИТИ, лю, лиш, РИБА́ЛЧИТИ *і рідко* РИБА́ЧИТИ, чу, чиш, *недок*, but signature (121) is missing. It follows from our classification that signature (121) defines an aspect complex with three paradigmatic blocks, in the first of which there is one component, in the second — two, and the third — one component. So, for the elements of this class, the graphic representation is valid:




(130)

and, by analogy to the signature option (12), the signature (121) can correspond to the model *Λ*(X):

$$R + \text{А́ЛИТИ, лю, лиш, } R + \text{А́ЛЧИТИ } і \text{ рідко } R + \text{А́ЧИТИ, чу, чиш,}$$
$$R + \text{А́ЧУВАТИ, ую, уєш, } недок, \qquad (131)$$

where the symbol*R* denotes the root of a hypothetical lexeme. Thus, there is a theoretical possibility of the development of the lexical system in the direction of the formation of structural classes, which until now were absent in the Ukrainian language, but which are allowed by its morphosemantic system and consistent with the law of hidden symmetry «1-2-3-4».

On the other hand, it is intuitively clear that the implementation of «superheavy» classes represented by signatures, for example (344.444), (444.344), (444.444), is very unlikely, the morphological system of the Ukrainian language simply will not «withstand» them. Consequently, a logical question arises about establishing the boundaries of this system, the answer to which can be obtained on the way of conducting comprehensive lexico-grammatical and lexical-semantic studies.

So, using the example of the Ukrainian verb, we showed that the language system induces a certain structure of the verbal equisemantic series. In this case, the substantial features of this structure were determined, the regularity and the formal algorithm for its construction were established, i.e., all system-forming components were involved, namely: «Structure» — «substance» — «subject». Significantly, the accuracy of this pattern turned out to be very high, no exceptions

were found for it on an array of more than 48 thousand Ukrainian verbs. It can be confidently stated that the discovered effect not only demonstrates the rule for constructing dictionary entries in the Dictionary of the Ukrainian language but represents a certain, still unknown systemic morphosemantic law.

Note that in the Ukrainian Lingua-Information Fund similar works are also performed for the Ukrainian noun.

And finally, it seems to us that similar studies can be carried out for other Slavic and, possibly, some other languages that have a developed system of inflection and word formation.

## The third systemic relation: quasi semantics formula and lexical system core

Among the many types of interpretations in the Dictionary of the Ukrainian language, a group of definitions is distinguished, which uses a method of a link to a dictionary entry being in a certain derivational relation with the signified one. Such definitions are characterized by a certain standard set of interpretation formulas, peculiar semantic cliches that mark the type of relationship between the signified word and its corresponding definition. So, in the SUM-20 database at the time of writing of this work, there were 1952 dictionary entries with the definition of «Abstr. noun to*» (an asterisk denotes a reference word that acts as a source of lexical semantics of the signified). For example:

**ВІКОВІ#ЧНІСТЬ,** ності, *ж.* Абстр. ім. до**вікові́чний.** *Яскравим доказом віковічності патріотизму може бути епос народів, їхня усна творчість* (П. Тичина); *Вірили трипільці у віковічність смерті-воскресіння природи, а отже, і безперестанність вмирання та народження усього світу* (з наук.-попул. літ.); *Метою зустрічі було допомогти подолати хибні погляди, віковічність яких зумовлюється відсутністю контактів між людьми, а також ранами історії* (з газ.).

**ВІЗАНТІ#ЙСТВО,** а, *с.* Абстр. ім. до **візанті́йський** 2. *А панство буде колихать, Храми, палати муровать* [мурувать]. *Любить царя свого п'яного Та візантійство прославлять* (Т. Шевченко).

**ВЕЛЕМО#ВСТВО,** а, *с.* Абстр. ім. до **велемо#вний.** *А як молитеся, не проказуйте зайвого, як ті погани, бо думають, ніби вони будуть вислухані за своє велемовство* (Біблія. Пер. І. Огієнка).

**ДАВНОЛІ#ТТЯ,** я, *с., розм.* Абстр. ім. до **давноліт́ній.** *Вийде* [баба], *бувало, у чисте поле, — Гей, давноліття ти молоде* (В. Бичко).

**МОКРОТА#,** и, *ж., розм.* Абстр. ім. до **мо́крий** 1, 2. *Зі сходом сонця випогодилось. Запахло цвітом, мокротою*(В. Бабляк).

As you can see, this interpretation formula is associated with only four affix word-formation models, and the actual vocabulary material indicates the abso-

lute dominance (approximately 99%) of the «*X* + ість» model, where *X* is the basic part of the derivative.

Dictionary definitions of this type are characterized by certain features that distinguish them from ordinary interpretations with the obligatory archiseme (semantic dominant) and a certain set of differential semes. As noted, they are standard and fully justify the name of the interpretation formulas or lexicographic cliches. Secondly, their structure does not include the necessary semantic components, which were mentioned above (archi- and differential semes). Finally, the interpretation of lexical semantics through grammatical is often used here (іменник до..., прикметник до..., прислівник до... (noun to..., adjective to..., adverb to...)). The foregoing provides a basis for highlighting the definitions of the marked type in a separate class, the elements of which we call quasisemantics formulas. The statistics on the use of these formulas in SUM-20 are summarized in the Table 5.

The table shows that 27 models — quasisemantics formulas ensure the existence of more than 45 thousand semantically derived lexical units. Without touching the content of the indicated formulas, their interpretation, as well as the com-

*Table 5.* **The statistics on the use of these formulas in SUM-20**

| Quasisemantics formula | Qty in SUM-20 | Example |
|---|---|---|
| Абстр. ім. до | 1952 | БАРВИСТІСТЬ, тості, *ж.* Абстр. ім. до барви́стий |
| Вищ. ст. до | 55 | ВУЖЧИЙ, а, е. Вищ. ст. до ву#зький |
| Властивість за знач. | 2197 | ЧУДОТВОРСТВО, а, *с.* Властивість за знач. чудотворний |
| Властивість і стан за знач. | 214 | БЕЗЖУРНІСТЬ, ності, *ж.* Властивість і стан за знач. безжу́рний |
| Властивість і якість за знач. | 32 | ЗАДУШЕВНІСТЬ, ності, *ж.* Властивість і якість за знач. задуше́вний |
| Дієпр. акт. до | 862 | ВІДМЕРЗЛИЙ, а, е. Дієпр. акт. до відмерзнути, відме#рзти |
| Дієпр. пас. До | 8419 | ЗАХИЩЕНИЙ, а, е. Дієпр. пас. до захисти́ти |
| Дія за знач. | 9661 | ВИШІПТУВАННЯ, я, *с., заст.* Дія за знач. вишіптувати 2 |
| Дія і стан за знач. | 439 | ВИЖИВАННЯ, я, *с.* Дія і стан за знач. виживáти 1, 2, 4 |
| Док. до | 976 | ВИБАНИТИ, ню, ниш, *док., кого, що, діал.* Док. до ба#нити |

*End of Table 5*

| Quasisemantics formula | Qty in SUM-20 | Example |
|---|---|---|
| Жін. до | 2746 | ДУБЛЕРКА, *ж.* Жін. до дублер<br>ДРІБОТУХА, и, *ж., розм.* Жін. до дріботун<br>ДРЕСИРУВАЛЬНИЦЯ, і, *ж.* Жін. до дресирува́льник<br>МОНАХИНЯ, і, *ж.* Жін. до мона́х 1, 2 |
| Збільш. до | 276 | БАБЕГА, и, *ж., зневажл.* Збільш. до ба́ба 1 2, 3 |
| Збірн. До | 241 | БАДИЛИННЯ, я, *с.* Збірн. до бадили́на |
| Зменш. до | 1402 | БАЗАРЧИК, а, *ч.* Зменш. до базар 1 |
| Зменш.-пестл. до | 2023 | БДЖІЛКА, и, *ж.* Зменш.-пестл. до бджола́<br>БДЖІЛОНЬКА, и, *ж.* Зменш.-пестл. до бджі́лка<br>БДЖІЛОЧКА, и, *ж.* Зменш.-пестл. до бджі́лка |
| Найвищ. ст. до | 74 | НАЙТРУДНІШИЙ, а, е. Найвищ. ст. до трудний |
| Однокр. до | 796 | АГАКНУТИ, ну, неш, *док., розм.* Однокр. до ага́кати |
| Пас. до | 3346 | БІЛИТИСЯ, білю́ся, бі́лишся, *недок.* 1. Пас. до біли́ти 1 |
| Пестл. До | 92 | ВУЗЕСЕНЬКИЙ, а, е, *пестл.* Дуже вузький |
| Підсил. до | 317 | ГИДЮЧИЙ, а, е, *розм.* Підсил. до гидки́й |
| Прикм. до | 8242 | ВЕДМЕЖИЙ, а, е. 1. Прикм. до ведмі́дь 1 |
| Стан за знач. | 589 | БЕЗШЛЮБНІСТЬ, ності, *ж.* Стан за знач. безшлю́бний |
| Уживається як пред. за знач. | 12 | СМИКЗ, *розм.* Уживається як пред. за знач. смикнути і смикнутися |
| Уживається як присудок за знач. | 160 | ЛЯСЬ, *виг.* Уживається як присудок за знач. ля́скати |
| Числівник порядковий, відповідний до кількісного числівника | 11 | СТОМІЛЬЙОННИЙ, а, е. 1. Числівник порядковий, відповідний до кількісного числівника сто мільйонів |
| Якість за знач. | 158 | ВИСОКОТОЧНІСТЬ, ності, *ж.* Якість за знач. високоточний |
| Якість і властивість за знач. | 25 | ПРУЖНІСТЬ, ності, *ж.* Якість і властивість за знач. пружни́й |
| ∑ 45 317 | | |

pleteness of this set, possible connections between individual formulas, etc., we will focus on systemic issues related to the definition of semantically and grammatically primary and derivative vocabulary in a language.

It is clear that such a definition, in addition to the examples given in the table, should also cover other linguistic mechanisms. After all, these such models as «Prefix + *X*» were almost not taken into account, where *X* are nouns, verbs, adjectives, to which prefixes *а-, ви-, от-, від-, до-, із, з-, зо-, по-, на-, над-, по-, про-, роз-, с-, в-, у-*, etc can be added. There are more than 30 thousand such lexemes in SUM. Also, it is necessary, as structural derivatives, to take into account the words forming the equisemantic sets, which were mentioned in the previous paragraph — there are more than 30 thousand units for the verb only.

So, even a cursory analysis allows us to raise the question of determining the core of the lexical system of the Ukrainian language and the qualifications and codification of generative mechanisms operating in this system. In particular, it is advisable to study what affix (substantial) models ensure the functioning of quasisemantics formulas, and vice versa: what affix word-building models are involved in the formation of vocabulary related to a particular quasisemantic type. The result of this work should be a generative grammar, which should be introduced into the formal structure of the lexicographic system of the Dictionary of the Ukrainian language in the form of a set of operators that, acting on a specific lexical core, deploy a complete class of Ukrainian lexemes in the noted sense.

## The fourth system relation: words and collocations

Recently, in applied linguistics, there has been a significant increase in the flow of scientific papers devoted to the linguotechnological description of various kinds of phrases. The immediate source of such interest is the need for automatic speech processing. Given that the main algorithmic solutions in this area are still concentrated mainly within the sentence, the importance of the task of developing formal models and mechanisms in the triad «word — phrase — sentence» becomes clear. In this case, the class of phrases during automatic processing gradually moves to the center of attention of applied linguistics, which is motivated by their prevalence in texts. In the work[84], the statistics of the distribution of phrases by the number of words (phrase length) is investigated. Assuming formally separate words with phrases of length 1, the authors of this article provide the following data on their distribution in texts (Table 6).

[84] Belonogov G.G., Zelenkov Yu.G., Kuznecov B.A., Novoselov A.P., Horoshilov A.A. Sistemy frazeologicheskogo mashinnogo perevoda politematicheskih tekstov s russkogo yazyka na anglijskij i s anglijskogo na russkij (sistemy RETRANS i ERTRANS). *MFID*. 1995. **20**, No. 2. S. 20—24.

The table shows that the phrases in the text occupy about 77% of its volume, while individual words that are not related by the corresponding grammatical and semantic relations, typical for phrases only, a little more than 23%. Such importance of the class of word combinations, their role in the language cannot but prompt attempts to build formal models of units of this class as deep as possible. It is also clear that the first steps along this path should have a grammatical orientation. No less significant are the tasks of phrases lexicography. According to tradition, in explanatory dictionaries it is customary to submit certain types of phrases; as a rule, these are units of the phraseological level (phraseological mergers (idioms), phraseological unity, phraseological compounds, and phraseological expressions). In explanatory dictionaries, a zone for such units is usually marked with a rhombus symbol (◊), which is why it has come to be called the «after-rhombic zone», extending this «corporate» lexicographic term even to those types of dictionaries that contain other types of phrases. In particular, in SUM-20 the «after-rhombic zone» is represented by four types of phrases:

established collocation (e.g. **Бити / вдарити в долоні**);

terminological phrases (marked with the symbol «Δ»; e.g.

Δ **Топографічна анатомія**;

phraseological units (e.g. ◊ **Байдики (байди, *діал. гандри)* бити**);

and the so-called equivalents of the word (marked with the symbol «**О**». E.g. **О До безмежжя**).

Thus, it can be argued that SUM-20 is an integrated lexicographic system consisting of two lexicographic subsystems — words and phrases. Symbolically, this fact is represented by the formula:

$$V[\text{СУМ}] = V^{\text{word}}[\text{СУМ}] \oplus V^{\text{collocat}}[\text{СУМ}]. \qquad (132)$$

In this regard, several questions arise. Firstly, the typification of phrases, which in linguistic theory is still far from complete. Since in the future we will consider all types of phrases as arbitrary syntactically organized groups of words interconnected by any type of syntactic connection, we use the common name collocation for them. Secondly, the development of a theoretical apparatus for the interpretation, explication, and lexicographic representation of collocations in dictionary systems. We will dwell on this issue in more detail.

Considering the systemic role of collocations and the similarity of many of their linguistic properties with the corresponding properties of words, it is logical that the dictionary was built from the very beginning as an integrated (uni-

*Table 6.* **The data on distribution in text**

| Collocation length (word count) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| The share in the text, % | 23.4 | 48.3 | 16.0 | 6.6 | 3.0 | 1.5 | 0.7 | 0.3 | 0.2 |

fied) system of words and collocations. However, this approach involves the application of a unified theory of these units. So, a correctly constructed integrated L-system of SUM should be built based on a unified theory of words and collocations.

Let us consider the possible contours of a unified formal grammatical calculus of words and collocations. For this, we need the concept of the grammatical state of a linguistic unit (as a special case of a semantic state) and its parameterization. The starting point is the provision that in a person's thought-speech apparatus linguistic units are (and in a speech flow they are functioning in) in certain states, which we call semantic states. So, the object $\psi(X)$ is formally introduced — the state of unit $X$, which is given by the correspondence $\psi: X \to \psi(X)$. Moreover, we consider that factorization of the full state $\psi(X)$ into two components is proper:

$$\psi(X) = G(X)\, S(X), \tag{133}$$

where $G(X)$ is the grammatical state of unit $X$, $S(X)$ is the semantic state.

The class of main objects $W$ in this theory is made up of collocations of length $n$, $n = 1, 2, \ldots$, moreover, for $n = 1$ these are ordinary words (denoted as $W_1$), for $n = 2$ they are two-word ones, and for $n = 3$ they are three-word ones combinations, etc.: $W_2$, $W_3$, $W_4$, ..., — collocations of length 2, 3, 4, ... . We introduce the concept of elementary grammatical states ($G$-states), which are analogs of markers of grammatical meanings): $G^1 = \{g_1, g_2, \ldots, g_k\}$. For example, the elementary $G$-states of Ukrainian nouns are g = (ім.; відмінок; число). (g = (noun; case; number).) Thus, the following qualification is typical for the theory of semantic states: «The linguistic unit $X$ is in the grammatical state of a noun with a certain meaning of case and number». The phenomenon of superposition, characteristic of the theory of semantic states, opens up wide possibilities for considering, qualifying, and interpreting (including lexicographic) the numerous grammatical ambiguities that function in the language system.

We formalize two-conditional collocations as a Cartesian product of one-word: $W_2 \subset W_1 \times W_1$. x grammatical states belong to the Cartesian product of elementary $G$-states :: $G_2 \subset G_1 \times G_1$. This means that on $G_1 \times G_1$a relation $c_2$ is defined such:

$$c_2 : G_1 \times G_1 \to G_2, \tag{134}$$

where the relation $c_2$ is a qualifier of the type of connection between the words $w_1$ и $w_2$ in the collocation $w_1 w_2$.

A generalization of this approach to the case of collocations of arbitrary length is obvious:

$$W_3 \subset W_1 \times W_1 \times W_1;\ W_4 \subset W_1 \times W_1 \times W_1 \times W_1; \ldots\ldots \tag{135}$$
$$c_3 : G_1 \times G_1 \times G_1 \to G_3$$
$$c_4 : G_1 \times G_1 \times G_1 \times G_1 \to G_4$$

etc.

$W_2$

| | |
|---|---|
| $G_2$: AN | Економічна самостійність, вагітна жінка |
| $G_2$:NNрод | Конституція України |

$W_3$

| | |
|---|---|
| $G_3$:AAN | Тимчасова слідча комісія,<br>Вищий спеціалізований суд |
| $G_3$:ANNрод | Верховна Рада України |
| $G_3$:NAдавNдав | Недовіра генеральному прокурору |

$W_4$

| | |
|---|---|
| $G_4$:Np2AродNрод | Половина від конституційного складу |
| $G_4$: Np5AорNор | Розподіл між територіальними громадами |

***Fig. 7.*** Collocation models

More than 170 grammatical collocation models up to nine in length were identified in the Ukrainian Lingua-Information Fund, NAS of Ukraine based on an analysis of a significant amount of textual material. Some examples of collocation models of lengths 2, 3, and 4 are given below (Fig. 7).

Currently, the development of an algorithmic-software system capable of automatically searching and identifying collocations in the text is ongoing. The indicated work is carried out using the method of lexicographic media and agents, and the software agent, which in this case belongs to the recognizing ones, actually carries out the marking of the text according to the collocations present in it, assigning the latter their grammatical parameters according to the stated general theoretical scheme.

## Concluding remarks to the section on system semantics in explanatory dictionaries

Given the above four systemic relations, of course, do not cover the entire set of systemic relations of the lexicographic system of the explanatory Dictionary of the Ukrainian language, and any other, because it is open — just like the open system is the language itself.

However, the role of systemic relations in the fundamental explanatory dictionary is already obvious. After all, the Dictionary is not a mechanical sum of elements — dictionary entries; it is the carrier of a very large number of implicit system connections of the system that form its semantic frame. For the correct

presentation of semantics, the lexicographer (and the user) must «go through» these system connections, supporting the semantic structure of both the dictionary and the language as a whole with an invisible framework. Thus, the task of developing effective calculus of systemic relations and creating effective tools for working with them seems to be relevant.Such a perspective can be implemented only in electronic lexicographic systems, so we are forced to state that the fundamental paper lexicography is perhaps drawing to an end. As a result of this state of affairs, the role and importance of linguistic, in particular, lexicographic technology, is growing. Orientation to computer and communication technologies, undertaken in the Ukrainian Lingua-Information Fund in the 1990s, stimulated a significant revision of the theory and practice of lexicography. It was for this purpose that we developed the theory of lexicographic systems, following which dictionaries of a new generation began to be created in the form of «lexicographic systems», and not just as «electronic dictionaries». Based on this theory, a new — computer-communication — technology was developed for creating dictionaries in the form of virtual lexicographic laboratories (VLL).

The system engineering of virtual lexicographic laboratories is based on the concept of a virtual lexicographic system, which we introduced in the monograph [85], and is defined as follows: VLL is an integrated lexicographic system distributed in the physical space, the components of which are represented by dots in the network address space. In this case, the coordination of the corresponding data models at the conceptual and internal levels is not even required. Only consistency is necessary at the level of external models, in particular, at the level of network protocols, which ensures the minimum integrity of such an integrated lexicographic system and the possibility of its identification as an integrated object.

The main purpose of virtual lexicographic laboratories was to ensure effective professional interaction of geographically distributed teams of lexicographers performing large vocabulary projects. The first one was the project of creating a fundamental academic explanatory Dictionary of the Ukrainian Language in 20 volumes (Ukrainian Language Dictionary in 20 volumes — SUM-20) [86]. It was put

[85] Shirokov V.A. Informatsiina teoriia leksykohrafichnykh system. S. 138—139.

[86] Shirokov V.A. Humanitarna tradytsiia i tekhnolohichnyi status movy. *Movoznavstvo*. 2001. No. № 3. S. 128—132; Shirokov V.A., Manako V.V. Tekhnolohichnyi kompleks zi stvorennia fundamentalnoi akademichnoi leksykohrafichnoi systemy «Slovnyk ukrainskoi movy». *VII Mizhnar. nauk.-prakt. konf. «Informatsiini resursy nauk.-tekhn. informatsii: problemy stvorennia i vykorystannia*. Kyiv: UkrINTEI, 2000. S. 121—124; Rusanivskyi V.M., Shirokov V.A. Informatsiino-linhvistychni osnovy suchasnoi tlumachnoi leksykohrafii. *Movoznavstvo*. 2002. No. 6. S. 7—48; Shirokov V.A., Rabulets O.H., Shevchenko I.V., Kostyshyn O.M., Yakymenko K.M. Tekhnolohichni osnovy suchasnoi tlumachnoi leksykohrafii. *Movoznavstvo*. 2002. No. 6. S. 49—86; Shirokov V.A. Fenomenolohiia leksykohrafichnykh system. S. 293—311.

into production in 2002 as the Virtual Lexicographic Laboratory «SUM» (VLL_SUM). Thanks to it, in 2007 the formation of the first edition of the lexicographic database of this twenty-volume book has been almost completed. Unfortunately, financial problems did not allow completing the paper version as quickly as we would like, so there are only seven volumes of SUM-20 have been published; their electronic versions are available on the website of the Ukrainian Linguistic Portal http://lcorp.ulif.org.ua.

The structure and functions of the VLL «SUM-20» are described in the monograph [87].

The method of virtual lexicographic laboratories is very fruitful. To date, about fifty different VLLs have been created in the Ukrainian Lingua-Information Fund; among them should be mentioned such fundamental works as «The Etymological Dictionary of the Ukrainian Language» [88], explanatory «Dictionary of the Russian Language», explanatory «Dictionary of the Turkish Language» [89], Grammar Dictionary of the Ukrainian Language, Grammar Dictionary of the Russian Language [90], Grammar Dictionary of the Turkish Language [91]; as part of the Mondilex project, 30 bilingual VLLs for six languages were developed: Bul-

[87] Shirokov V.A., Bilonozhenko V.M., Buhakov O.V. tain.

[88] Ostapova I.V. Digital Etymology (illustrated by the example of the Etymological Dictionary of Ukrainian language). *Organization and Development of Digital Lexical Resources. Proceedings of Mondilex Second Open Workshop*. Kyiv, 2009. pp. 66—72; Ostapova I.V. Leksikograficheskaya struktura etimologicheskogo slovarya i ego predstavlenie v cifrovoj srede. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii. Materialy ezhegodnoj Mezhdunar. konf. «Dialog 2009»* (27—31 maya 2009, Bekasovo). 8(15).Moscow: RGGU, 2009. XII. 620 s. (S. 359—365); Ostapova I.V., Shirokov V.A. Virtual'naya leksikograficheskaya laboratoriya dlya tolkovyh slovarej. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii. Materialy ezhegodnoj Mezhdunar. konf. «Dialog»* (26—30 maya 2010, Bekasovo). 9(16). Moscow: RGGU, 2010. 920 s. (S. 363—367).

[89] Shirokov K.V. Imenna slovozmina u suchasnii turetskii movi. S. 119—135.

[90] Lyubchenko T.P. Morfologichna model' slovozmini flektivnoï movi ta elektronnij gramatichnij slovnik. *Bionika intellekta*. 2006. No. 1 (64). S. 72—77; Gryaznuhina T.A., Lyubchenko T.P., Rabulec A.G. Elektronnaya versiya Grammaticheskogo slovarya russkogo yazyka A. A. Zaliznyaka kak instrument avtomaticheskogo morfologicheskogo analiza russkogo teksta. *Korpusnaya lingvistika i lingvisticheskie bazy dannyh: Dokl. nauch. konf.* SPb.: Izd-vo S.-Peterb. un-ta, 2002. S. 63—70; Gryaznuhina T.A., Lyubchenko T.P., Shevchenko I.V. Novaya versiya elektronnogo grammaticheskogo slovarya russkogo yazyka s uchetom akcentuacii.*Slovo i slovar' = Vocabulumetvocabularium : Sb. nauch. tr. po leksikografii*. Grodno: GrGU, 2005. S. 188—193; Lyubchenko T. Modelling of the Digital Grammar Dictionary of Russian. *Organization and Development of Digital Lexical Resources. MONDILEX 2nd Open Workshop*. Kyiv, Ukraine, 2—4 February, 2009. P. 73—84.

[91] Shirokov K.V. Imenna slovozmina u suchasnii turetskii movi. S. 136—145.

garian, Polish, Russian, Slovak, Slovenian, Ukrainian [92]; several terminological VLLs have also been developed [93].

An idea of virtual lexicographic laboratories created in the Ukrainian Lingua-Information Fund can be obtained at http://lcorp.ulif.org.ua/LSlist/ «Resources on the website of the Ukrainian Lingua-Information Fund» (some screenshots of this website are posted in Appendix 2).

It is characteristic that all the marked VLLs are full-scale instrumental systems and give a fairly complete lexicographic representation of the corresponding linguistic facts and phenomena in their representative volume. For example, grammar VLL operates with a register of more than 570 thousand units, which implements the paradigmatic inflectional classification of the Ukrainian language, which includes (accentuation including) about 3,000 inflectional classes.

[92] Shirokov V.A. Proekt universalnoi systemy slov'ianskoi leksykohrafii «Leksyka-Slavika». *Slov'ianski obrii*. 2. XIV Mizhnar. z'izd slavistiv (10—16 veresnia 2008, Okhryd, Respublika Makedoniia). Kyiv, 2008. S. 801—829; Dimitrova L., Koseska-Toszewa V., Garabik R., Erjavec T., Iomdin L., Shirokov V. Mondilex — towards the research infrastructure for digital resources in Slavic lexicography. *Cognitive Studies. Etudes Cognitives*, 10. Warsaw: SOW Publishing House, 2010. P. 155—170; Shirokov V. 3. Software enviromens for digital lexicography; 4. Technological platform for research infrastructure for digital resources in Slavic lexicography. In: Dimitrova L., Koseska-Toszewa V., Garabik R., Erjavec T., Iomdin L., Shyrokov V. Conceptual Scheme for a Research Infrastructure Supporting Digital Resources in Slavic Lexicography. Sofia, 2010. 130 p.; Shirokov V.A. Komp'iuterna leksykohrafiia. S. 283—294.

[93] Shirokov V., Ostapova I., Nadutenko M., Verbynenko Yu. Ontologized lexicographical systems in modern terminography. *Cognitive Studies. Études Cognitives*. Warsaw, 2016. **16**. P. 90—99.

# ON THE WAY TO «ABSOLUTE» DICTIONARY

The ideas of the integrated presentation of diverse aspects of dictionary information in one publication represent the most secret desires of lexicographers. «The ideal dictionary of the theory of lexicography remains a universal dictionary that could fully and accurately reflect the language element, all ranks of units of the lexicon with all their inherent attributes: for the theory of lexicography, the concept of a maximum lexical system and the concept of a universal dictionary, which captures a maximum lexical system, are necessary. Using such a dictionary, one can communicate in a specific language in all situations, talk on any topic with any native speaker of that language. It is easy to understand that this is a conscious idealization, the theoretical world of lexicography» [94].

In 1976, when these lines were written, the idea of a universal, «absolute» dictionary was nothing more than idealization, a dream of a lexicographic ideal. But now, in the realities of computer lexicography, the question of an «absolute» dictionary can and should be completely objectively discussed constructively.

Nowadays, two scientific and technological revolutions continuing — digital and communication. They are not isolated from each other, but interdependent; between them, there is approximately the same ratio as between the concepts of information resource and information infrastructure. The first of them is characterized by the rapid implementation of digitalization technologies for the entire information and cultural heritage of mankind and the results of its current intellectual activity — the creation of digital copies or analogs of all objects that are carriers of any information and knowledge. The second one aims at the creation and implementation of communication tools that allow a person from anywhere to instantly contact any other person, institution, or source of information. The quantitative and qualitative characteristics of the respective systems and technologies (both resource and infrastructure) are being improved very quickly, providing their users with more and more wide possibilities. In a certain, not too distant future, we can expect that the communication-digital world will turn into

[94] Denisov P.N. Sistemnost'i svyazannost' v leksike i sistema slovarej. *Problematika opredelenij terminov v slovaryah raznyh tipov* / Red. kol. S.G. Barhudarov i dr. Leningrad: Nauka, 1976. S. 63—73.

a complete correlate of the evolutionary paradigm of modern civilization, such a prospect seems even more realistic given the mechanism of information-energy transformations, which, in our opinion, form the natural-scientific basis of the information society and society of knowledge [95].

Meanwhile, a specific «restriction of the integrity» of the world information system associated with multilingualism can become an obstacle to such futurology, after all, the division of the information medium into separate language segments (and, accordingly, into separate language pictures of the world) is a fait accompli, while modern technologies of interlanguage communication still far from perfect and cannot overcome the «Babylonian syndrome», caused by the simultaneous active functioning of tens and hundreds of languages, very different in their designs and informational status. Thus, the question of interlanguage adaptation somewhat unexpectedly turns into a different angle in the context of the civilizational shifts that are currently undergoing, and, as a result, the linguistic problems, more precisely, the development of linguistic technology, are confidently moving from the periphery to the central areas of scientific and technological development.

From the aforesaid, the need for purposeful fundamental research of the language system to obtain results adapted for implementation in highly effective intellectual language technologies follows; after all, as noted in the Preface, artificial intelligence is such a form of systems existence when they individualize themselves precisely through their language status. Given the applied, technological aspect of modern linguistic researches, it is not necessary to further argue the priority of the lexicographic description of the language system, because the quantitative and qualitative parameters of the lexicography of all units, relationships, and aspects of the language ultimately depend on the effectiveness of linguistic technologies built on this basis. And taking into account the importance of solving the problem of multilingualism in the global information environment, the logical task is to integrate the lexicographic description of all languages, i.e., create a kind of dictionary of the entire human race — a kind of Summa lexicographical.

It is clear that at the moment such a task may not seems quite realistic, because a complete lexicographic description has not been created even for any single language, and it is not very clear based on which theoretical and linguistic scheme such work can be performed. However, this does not mean at all that such tasks should not be set. We are convinced that general statements of problems also encourage the development of common theoretical approaches, methods, and concepts.

Note that for the author, these problems have already risen in practical terms when implemented in 2008—2010 an international linguistic project within the framework of the seventh EU Framework Program for Scientific Research FP7, the aim of which had been to create a unified lexicographic system for many

[95] Shirokov V.A.: Informatsiino-enerhetychni transformatsii ta informatsiine suspilstvo; Informatsiina teoriia leksykohrafichnykh system.

Slavic languages (Bulgarian, Polish, Slovak, Slovenian, Russian, and Ukrainian)[96]. We believe that at the next stage all Slavic languages should be involved in the project, which will mark the beginning of an all-Slavic linguistic dialogue. Thus, the task of creating an integrated lexicographic system of Slavic languages seems to us quite real, and the experience of computer etymological lexicography confirms its relevance for almost all languages of the world.

What approaches should be applied to solve this problem?

To cover the latter, we focus on the problem of the integral lexicographic description of the language system. Theoretically, this problem, in our opinion, is most deeply developed in the writings of Yu.D. Apresyan and other representatives of the Moscow Semantic School[97]. Its main features are the installation of the reconstruction of the linguistic picture of the world, on the principle of integrality itself (we will call it hereinafter the principle of linguistic integrity, the concept of the lexeme and its integral lexicographic representation, as well as the concept of the lexicographic type[98].

From Apresyan's definition of integral lexicographic description, the lexicographic representation of various ethnospecific language pictures of the world, as well as various implementations of the principle of linguistic integrality, characteristic of each language, follows. The picture of the integral lexicographic description of the language system is, in fact, even more complicated, which becomes clear from the analysis of the interaction of grammatical and lexicographic descriptions of the language system. In linguistic science, the idea has long been strengthened that the basis for describing any language is vocabulary and

[96] Project acronym: MONDILEX. Project full title: Conceptual Modelling of Networking of Centres for High-Quality Research in Slavic Lexicography and Their Digital Resources. Grant agreement No. 211938. Participants:Institute of Mathematics and Informatics, Bulgarian Academy of Sciences; Institute of Slavic Studies, Polish Academy of Sciences; Ľ. Štúr Institute of Linguistics; Jožef Stefan Institute; Institute for Information Transmission Problems, Russian Academy of Sciences; Ukrainian Lingua-Information Fund of the National Academy of Sciences of Ukraine.

[97] Apresyan Yu.D. Izbr. trudy. T. 1: Leksicheskaya semantika: sinonimicheskie sredstva yazyka. 2 izd., ispr. i dop. Moscow: Shkola «Yazyki russkoj kul'tury», 1995. 472 s.; Apresyan Yu.D. Izbr. trudy. T. 2: Integral'noe opisanie yazyka i sistemnaya leksikografiya. Moscow: Shkola «Yazyki russkoj kul'tury», 1995. 767 s.; Apresyan Yu.D. Integral'noe opisanie yazyka i tolkovyj slovar'. *Vopr. yazykoznaniya*. 1986. No. 2. S. 57—70.

[98] According to the principle of integrability according to Yu.D. Apresyan, formulating a certain rule of the language, grammarian should work on the whole set of lexemes and take into account all the lexemes obeying this rule, if this form of their behavior is not recorded directly in the dictionary entry; in some cases, this requires entering information about specific lexemes directly into the rules. On the other hand, when describing another lexeme, the lexicographer must work on the whole set of rules of the language and attribute to this lexeme all the properties that are mentioned in rules; in some cases, this requires entering information about the rules into the dictionary entries of the dictionary.

grammar, although they function in different ways and consider various objects: the vocabulary traditionally describes the units of the language, and the grammar formulates the laws and rules for their change and combination, that is, morphology and syntax in the general understanding of these concepts. Vocabulary and grammar in the aforementioned sense constitute a certain «opposition», and at the same time complement each other, creating together a complete picture of the linguistic description. The effectiveness of this description depends on the consistency between its lexicographic and grammatical components. This means that, on the one hand, lexicographic units in the vocabulary should explicitly assign all those properties that are required by the grammar rules, and on the other hand, all (if possible) types of behavior of language units that are not considered in the vocabulary should be described in the grammar. Otherwise, the vocabulary and grammar will not be able to effectively interact with each other and give a correct, self-consistent description of the language system.

Thus, an ideal linguistic description that satisfies these conditions should be systemically integrated, combining both methods of describing a language system — lexicographic and grammatical, although the very requirement of integration, i.e., a comprehensive, multilevel and diverse description of a language, despite all its naturalness, is theoretical and has no simple implementation in practice.

At the same time, it seems that this requirement is a consequence of some more general and fundamental principle or has a deeper phenomenological nature. This, in particular, is signaled by the fact that in the definition of concepts and constructions of formal grammars two whole lexicographic objects appear — the so-called non-terminal and terminal dictionaries. Consequently, formal grammar (which, of course, is a grammar system by definition) must contain at least two lexicographic elements and cannot function without them in principle. On the other hand, the theory of lexicographic systems, which we use as the formal basis of the lexicographic description of the language, contains (and also necessarily) elements and means, the natural interpretation of which is precisely the grammatical interpretation, and therefore, in this case, we need to talk about grammar as an organic, non-optional element of the structure of the lexicographic system. It can be argued that both types of description of the language system — grammatical and lexicographic — are complementary in the sense of N. Bohr and, therefore, there must be a theoretical scheme that unites, «integrates» both of the mentioned types in a single conceptual object. We believe that the theory of semantic states can take on the role of such a conceptual scheme.

However, both the grammatical and lexicographic description of the language — each of them — is already integrated and should represent an appropriate, specific for each of the integrated language and information environment in which various levels, units, language relations, etc. are represented. Indeed, the integral vocabulary and integral grammar should ideally give an interpretation of not only individual, specially identified and prepared elements of the language

system (which, in particular, are its phonetic, morphemic, lexical, syntactic, and other levels), but must contain adequate instruments for the processing of whole objects language and speech (text). A universal, in our opinion, conceptual tool for representing such a situation provides a lexicographic effect in information systems that explains the mechanism for generating complexes of discrete units in each relatively stable, system-defined level of perception of language substance. The integration schemes of various grammatical and lexicographic systems can be quite different. We give one of them, possibly the simplest. Denote by

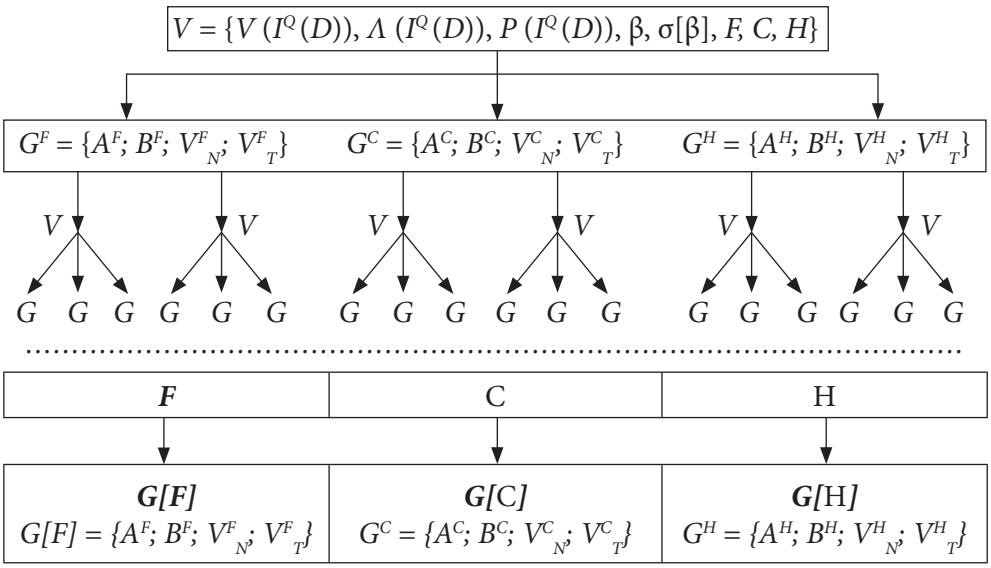
$$G = \{A;\ B;\ V_N;\ V_T\} \tag{136}$$

some formal grammar representing a formal grammatical description of a particular element of the language system. In this formula, as always, the symbol $A$ denotes the class of initial elements (axioms) of the grammar, $B$ is the set of base rules, and $V_N$ and $V_T$ are non-terminal and terminal dictionaries, respectively. The symbol

$$V = \{V\,(I^Q(D)),\ \Lambda\,(I^Q(D)),\ P\,(I^Q(D)),\ \beta,\ \sigma\,[\beta],\ F,\ C,\ H\} \tag{137}$$

denote some elementary L-system, where $I^Q(D)$ is the class of elementary information units to the lexicographic effect $Q$ over $D$; $\Lambda\,(I^Q(D) = FV\,(I^Q(D))$ and $P\,(I^Q(D)) = CV\,(I^Q(D))$ are the elements of the formal content of the lexicographic description of $V\,(I^Q(D))$, respectively; $\beta$ and $\sigma[\beta]$ represent the microstructure of the L-system; the operator $H$ makes the connection between $\Lambda\,(I^Q(D)$ and $P\,(I^Q(D)$.

The scheme for constructing an integrated grammar and lexicographic description consists of a sequential interpretation of the operators $F, C, H$ as some grammars (G-systems), and the lexicographic elements $V_N$ and $V_T$ as some L-systems:




(138)

We can hope that at a certain stage of this process we will get an integrated grammar and lexicographic description for a particular language. A certain technologically verified experience in building integrated lexicographic systems is available in the Ukrainian Lingua-Information Fund, where the Integrated Lexicographic System «Dictionaries of Ukraine» has been created produced for several years, it combines in a single language-information object, on a fairly large array of Ukrainian vocabulary (more than 250 thousand units), relations of inflection, phonemic transcription, phraseology, synonymy, and antonymy. This experience makes it possible to assert, with restraint optimism, that the theory of lexicographic systems and the theory of semantic states have sufficient structural potential for constructing integrated grammar and lexicographic objects (we will call them GL-systems in the future) of almost unlimited volume and complexity.

At a conceptual level, the formalism of lexicographic environments and the methodology of integrated lexicographic systems provide theoretically unlimited possibilities for combining in one object any number of lexicographic effects and their dictionary representations. However, in practice, building such a dictionary is far from an easy task.

At the stage of conceptual modeling, a whole series of rather complicated tasks arise.

The first one is the volume of language material that the lexicographic system is capable of supporting while remaining within the limits of integrity signs for the user's perception. This issue has both an objective and a subjective component. Moreover, it is the subjective component that has been least studied and obvious to developers and designers of L-systems. However, without solving this issue, the creation of super-large L-systems, that is, effective lexicographic implementation of the concept of «Big Data», is impossible. The next question is what kind of a linguistic theory it should be to be capable of providing a conceptual apparatus for designing super-large L-systems? The existing courses and monographs on various sections of grammar, semantics, syntax, etc. are hardly suitable for use in designing super-large L-systems, primarily due to the insufficient level of formalization of the corresponding linguistic material. On the other hand, too high formalization leads to such abstract theoretical schemes that can hardly be meaningfully projected onto real linguistic phenomenology. Thus, the implementation of super-large L-systems is impossible without an adequate linguistic and lexicographic theory. The author realizes that theories of lexicographic systems and semantic states in their modern form are too general for a real solution to the problem of creating an «absolute» dictionary. At the same time, their conceptual potential seems quite adequate for this purpose.

## Integration strategies

In practice, a slightly different path is possible, which consists of the gradual expansion of the structure, content, functions, implementations, etc. of some al-

ready created and working lexicographic system. At the same time, the general direction is to integrate specially developed L-systems intended for integration purposes in this «initial» L-system. In this case, the mechanisms of recursive reduction and construction of lexicographic environments can be involved. It is clear that for this it is necessary to have a certain maximally wide and constantly expanding certain field of worked out and technically tested L-systems. In this regard, there is a task of developing a fairly universal and at the same time simple parsing technology, with which it would be possible in a permissible short time to «accumulate» a sufficient number of meaningful, voluminous, structurally meaningful L-systems, and conduct experiments on their integration. In this case, it is desirable to bring the integration procedures to the most automated level. This will allow experimentally working out various modes of integration of L-systems, determining the potentials forthe integration of various L-systems depending on the types of lexicographic effects, branching of the vocabulary structure, functional implementations, etc.

## The parameter of time (evolution)

An important aspect is related to ensuring the continuous evolution of lexicographic systems. After all, the language is constantly changing, and practice requires lexicography to track the changes that occur in the usage as quickly as possible. In this regard, it is necessary to introduce a time parameter into the lexicographic system. Moreover, this parameter should be not only explicit but also conceptually meaningful. Indeed, in this case, it is not just about physical time (although this is not so simple), but about the specific time of a language, the pace of which is not constant, but can vary, responding to changes in a particular language-social environment.

The introduction of the time parameter will allow not only to respond quickly to linguistic dynamics but also to study the patterns of language structure evolution. This approach allows us to raise the question of creating such L-systems that are constantly evolving, mutating, responding to changes in the «external» world. For this, of course, the L-system must have means of «connecting» to this world, be able to respond to changes in it, analyze these changes, and make adequate changes in its structural and substantive part. In this case, the role of the «subject» is activated, that is, the subjective component of the triad «structure-substance-subject» objectified in an integrated super-large L-system. It is clear that to solve this problem, an adequate model of the evolution of the world is also necessary, which seems to be a very difficult, but practically necessary, problem.

Summing up the experience of lexicography, we present a set of linguistic objects and relationships, which, in our opinion, can quite naturally be integrated into the general («absolute») lexicographic system of a particular language.

**OBJECTS (LANGUAGE UNITS):**

Vocabulary: (general, dialect, terminological, abbreviations, obscene, pro-national, ...)

Collocations: (free phrases, phraseological units, terminology, word equivalents, etiquette phrases, discourse phrases, occasional, obscene, etc.).

Onomasticon: (toponymicon, cosmonimicon, anthroponymicon, mythologicon, organimicon, ...).

Morphemes, affixes: (repertoire and classifications; models and invariants of inflection and derivation).

Syntax: (syntaxeme, syntactic models of phrases, syntactic models of sentences, ...).

«Non-dictionary» language units.

Protologisms.

**RELATIONS:**

Grammatical semantics (grammatical meanings, grammatical states, lexical and grammatical classes, lexical and grammatical parameterization, grammatical classifications, grammaticalization...)

Lexical semantics (lexical meanings, semantic states, lexical functions, semantic classifications, synonymic relations).

Hyperchains and hypercycles.

Dictionary definitions and their structures.

Semantic dynamics (neologisms, metaphorization, termination, determinologization, ...).

Linguopragmatics, speech acts, speech tonality.

Etymology (glottochronology, linguogeography, glottosemantics, ...).

**CONCEPTOGRAPHY AND ONTOLOGY**

Language pictures of the world and linguistic ontologies, ontologies of subject areas, cognitive-communicative, system-functional and text-discursive conceptography.

The previous presentation concerned monolingual lexicography. For many of the objects and relationships mentioned above, the corresponding lexicographic structures are already well developed. For the rest, such work should still be done. To support multilingualism, we consider promising the environment of the MONDILEX lexicographic platform, in which multilingualism is represented through many bilingual systems. Effective and correct operation in such a system is ensured by incorporating intermediary languages (tertium comparationis). A specific example of the development of such a language, which also covers theoretical issues of constructing tertium comparationis, is the work of V. Koseska-Tosheva. An important question is the functional characteristics of L-systems and their reflection in external models (interfaces).

# QUANTUM LINGUISTICS

We decided to devote the last section of the book to a brief outline of the most exotic trend in linguistics, which is interpreted as «quantum linguistics». We believe that the reader is already prepared in some way by the preceding text for such a spread of exposition, in which a lot was said about quantum phenomenology and methodology in connection with the general problems of cognition. It should be noted that quantum presentations and, so to speak, a quantum view of the world in recent years are becoming increasingly popular in more and more unexpected areas. Such disciplines as quantum chemistry and biochemistry[99], quantum biology[100], even quantum medicine, and quantum pharmacology[101] are already quite traditional. In this set, quantum psychology[102] and quantum economics[103] are distinguished by their extravagance (the last book, however, was written by a major scientist and stands out for its high scientific level).

What about quantum linguistics? It should be noted that at present it is even difficult to outline the phenomenological boundaries of what we can attribute to this area, althoughthere are already several publications on this topic[104]. On the website http://www.neurochromatics.com, in the section designated «QUANTUM-LINGUISTICS», it is noted: «*Quantum Linguistics is a language system that takes into account the Quantum nature of the human nervous system: it sees the*

---

[99] Ejring G., Uolter Dzh., Kimball Dzh. Kvantovaya himiya / Per. s angl. Moscow: GIIL, 1948. 528 s.; Pyul'man B., Pyul'man A. Kvantovaya biohimiya / Per. s angl. Moscow: Mir, 1965. 654 s. Gel'man G.G. Kvantovaya himiya. Moscow: ONTI, 1937. 546 s.; Cirel'son V.G. Kvantovaya himiya. Molekuly, molekulyarnye sistemy i tverdye tela: Ucheb. posobie. Moscow: BINOM. Laboratoriya znanij, 2010. 495 s.

[100] Davydov A.S. Biologiya i kvantovaya mekhanika. Kyiv: Nauk, dumka, 1979. 296 s.; Ball Ph. Physics of life: The dawn of quantum biology. *Nature*. 2011. 474. P. 272—274; Bordonaro M., Ogryzko V.V. Quantum biology at the cellular level — Elements of program. *Biosystems*. 2013. 112(1). P.11—30.

[101] Quantum Medicine. Quantum Therapy. URL: http://www.qrs.com/quantum_medicine.php; Chekman I.S. Quantum pharmacology: new direction in materia medica. *Nauka ta innovacii*. 2010. **6**, No. 2, P. 29—35. https://doi.org/10.15407/scin6.02.029

[102] Panov A.D. Tekhnologicheskaya singulyarnost'...; Uilson R. Kvantovaya psihologiya. Moscow: Sofiya, 2016. 224 s.

[103] Maslov V.P. Kvantovaya ekonomika. Moscow: Nauka, 2006. 96 s.

[104] Shirokov V.A. Ocherk osnovnyh principov kvantovoj lingvistiki. *Bionika intellekta*. 2007. No. 1(66). S. 25—32.

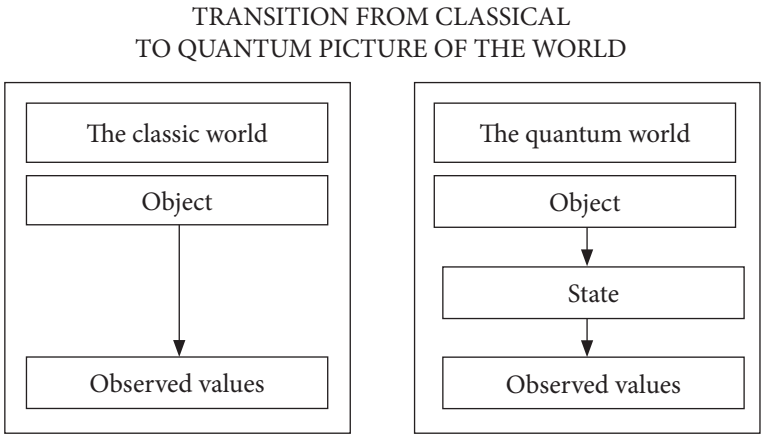



***Fig. 8.*** Scheme

*human nervous system as a holographic unit, a quantum mechanical device powered by the mind — powered by the pictures and self-talk we generate.* It is difficult to say what is more in this statement: literal meaning or metaphor. We are more impressed by the approach of V.P. Maslov, who, considering the Planck formula well-known in physics as a quantum continuation of the Stefan-Boltzmann formula, in a similar sense implements the «quantum continuation» of the Zipf-Mandelbrot and Pareto laws for the distribution of capitals, as well as for dependence of sales on stock market prices. It can be argued that, in a certain sense, V.P. Maslov «quantized» the rank distributions, which are the laws of Zipf-Mandelbrot and Pareto.

We, declaring an informational approach to the language, believe that the quantum interpretation of language systems also primarily involves a quantum interpretation of the most profound phenomena of the language, which, in our opinion, are informational phenomena. Therefore, the quantum approach to language dictates the transition to a quantum interpretation of the informational components of language systems. The first step towards a quantum interpretation of linguistic phenomena, we believe, should be done by replacing the classical picture of the world with a quantum one. Schematically, this can be represented on Fig. 8.

So, the first step in the transition from the classical to the quantum picture of the world is that a new object of conceptual description, state, is introduced into the theory. Thus, now the theory begins to operate not with directly observable quantities (grammatical and semantic parameters), but with some intermediate objects — states that acquire the status of the main object of the theory. We justified this approach in the section «Phenomenology of the Language and the Linguistic Picture of the World», where it was determined: «*Own objects of the language are certain psychophysical states and processes that occur in the human mental-speech apparatus, and its oral and written forms serve as elements of the infrastructure of the speech process*». In the same section, the concept of observability was analyzed. Fur-

ther development of these ideas was made by us in the theory of semantic states, the beginning of which was laid by the initiative of A.N. Kolmogorov.

A remarkable (specifically «quantum»!) property of states is the *principle of superposition*; we defined it by formula (92) in the paragraph «General concept of linguistic states and the concept of linguistic observables associated with it». The principle of superposition provides an excellent (and practically the only) theoretical tool for the formal representation of linguistic ambiguities. Namely, it follows from the principle of superposition that if $c_1(X)$ and $c_2(X)$ are different states of the unit $X$, then there is a state $c(X)$, such that:

$$c\,(X) = \alpha_1\,(c_1)\,c_1\,(X) + \alpha_2\,(c_2)\,c_2\,(X),$$

where

$$\alpha_1\,(c_1) + \alpha_2\,(c_2) = 1;\ 0 \le \alpha_i\,(c_i) \le 1;\ i = 1, 2. \qquad (139)$$

We give examples of various linguistic ambiguities and their interpretation in terms of the principle of superposition.

The first example is from the Constitution of Ukraine. The preamble to this document states: «Офіційне тлумачення до Конституції див. у Рішеннях Конституційного Суду» («For an official interpretation of the Constitution, see Constitutional Court Decisions»). In this context, the word «Офіційне» («Official») functions simultaneously in two lexical meanings, that is, it is in a superposition of two semantic states:

$$C\,(X) = a\,C_1\,(X) + b\,C_2\,(X), \qquad (140)$$

which, accordingly, have the following dictionary interpretations:

$C_1\,(X)$ = *Який запроваджується, регулюється державою, державною установою або службовою особою; державний, урядовий, службовий.*

$C_2\,(X)$ = *Повідомлений, оголошений, опублікований державною установою; який надходить від держави, державної установи або службової особи.*

At the same time, no linguistic processing removes this unambiguity, since, according to Ukrainian legislation, the law acquires official status only after it is *approved by the Verkhovna Rada of Ukraine* (Ukrainian parliament), *signed* by the President, and it is *published* in the official state-owned newspaper «Voice of Ukraine».In the stated approach, this semantic ambiguity is formalized through a superposition of states.

The effect of a language game is often based on semantic ambiguities. Consider the example from the «Twelve Chairs» by I. Ilf and E. Petrov, namely, a fragment when the journalist Persitsky parodies the style of the hack poet Nikifor Lyapis-Trubetskoy:

Гаврила стул купил на рынке.
Был у Гаврилы стул плохой.

Here, in the first sentence, the word «стул» («chair») functions in the pure semantic state $C_1$ *(стул)*, namely, the literal meaning of the word as a type of furniture. But in the second sentence, the semantic state of the same word is transformed and takes on the form of a superposition of two states: $C_1$ *(стул)* + $C_2$ *(стул)*, where $C_1$ *(стул)* is the same as in the first sentence, and $C_2$ *(стул)* is determined by the semantics of the terminology «*плохой стул*» and has a «gastrointestinal» connotation. The ambiguity of the semantic structure is the reason for the comic effect inherent in this language situation. Despite this, a discerning reader with a sense of humor gets the impression of its unambiguous understanding, namely, its unambiguous understanding of its ambiguity.

Similar examples can be cited in hundreds. We are convinced that almost any polysemic word can be synthesized in a context in which it functions in a superposition of semantic states. This device is practically indispensable in many applications, such as the building of linguistic ambiguity generators, developing intelligent tools for studying linguistic ambiguity removal algorithms, creating software for automatic marking, interpretation, intellectual analysis, and examination of natural language texts, developing natural-language intellectual dialogue systems and many others.

The most consistent advancement into the field of quantum linguistics, we believe, consists in «quantizing» the information itself. It consists of replacing classical information with quantum information. As noted in the paragraph «Structure, substance and subject in the definition of the concept of a system»: «*...quantum information has properties that are fundamentally different from classical information (we are talking about quantum superposition and quantum entanglement), so computer systems created based on such a «substance» will also have system properties that are fundamentally different from classical computers*». A huge flow of recent works has been devoted to the problems of quantum information.

We will not present here a consistent presentation of the essence of quantum information; we will explain only two of its fundamental differences from the traditional one noted above. The first of them is the possibility to implement a superposition of states. This, in turn, provides an opportunity for colossal parallelization of information-computational processes, one that makes it possible to solve problems that even belong to the class of exponential computational complexity. The second one is the so-called quantum entanglement, which paves the way for long-range information, that is, ensuring information interaction with a speed that is hundreds of thousands of times faster than the speed of light.

Linguistic information, based on quantum technologies, and the technological implementation of these possibilities opens the way to a completely new mechanism of evolution, and provides the means for what is now hidden behind the metaphor of «technological singularity».

# FINAL REMARKS AND CONCLUSIONS

One of the main tasks of our monograph is to expand the phenomenological and conceptual base of linguistics, in a certain sense, a change in the research paradigm. This epistemology (a kind of epistemological turn) means the achievement of the maximum correspondence of the research tools of linguistics to the nature of the language (as we currently imagine it) and to the tasks that have already been set by the cognitive-communicative stage of evolution.

This state of affairs in the field of linguistic science and technology is fully consistent with the forecast of Jean Piaget, made back in 1970: «After the stage of interdisciplinary research, we should expect a higher stage — a transdisciplinary one, which will not be limited to interdisciplinary relations, but will place them inside the global system, without strict boundaries between disciplines». This attitude is fully consistent with the definition of transdisciplinarity as a way of expanding the scientific worldview, which consists of considering a particular phenomenon outside the framework of any one scientific discipline. It is in this meaning that transdisciplinarity is used as the principle of the organization of scientific knowledge, which opens up wide opportunities for the interaction of many disciplines and cognitive paradigms, as well as the mobilization of diverse research resources to solve complex and critical problems.

We believe that such an approach in linguistic science and especially in linguistic technology forms the main perspective for the development of the cognitive-communicative, network-centric world. The evolution of information processing systems already gives reason to assert that their «perceptual-sensory» capabilities are developing in the direction of convergence, «convergence» with the human perceptual-sensory apparatus. At the same time, the perceptual-sensory capabilities of intellectual artifacts can even surpass the human ones, this is the main content of the hypothetical «technological singularity» that we mentioned in the Preface.

said above can be summarized in the following theses:

1. The information society is a necessary stage in the development of human civilization, the first stage of a post-industrial society[105].

[105] Shirokov V.A. Informatsiino-enerhetychni transformatsii ta informatsiine suspilstvo; Shirokov V.A. Informatsiina teoriia leksykohrafichnykh system; Bell D. TheSocialFrameworkofInformationSociety. Leningrad, 1981. P. 4; Brzesinski Z. Between Two Ages: America's Role in Technotronic Era. NY, 1970. P. 52; Dizard W. P. The coming information age. An owerview of Technology, Economics and Politics. Longman, 1982. P. 152; Masuda T. The Information Society as a Postindustrial Society. Tokyo, 1983. P. 28, 47, 61, 74.

2. In the era of the formation of the information society, the foundations of the noosphere are laid[106].

3. The second phase of the information society is the so-called knowledge society. Its characteristic feature is the widespread adoption of intelligent information systems.

4. The digital and communication scientific and technological revolutions that accompany the formation of a knowledge society determine the role of a language and determine the noospheric effects of this stage of human civilization development.

5. The amount of information produced by one or another biosocial community can be considered a formal correlate and a measure of evolutionary processes.

6. At each stage of development, evolution had created increasingly sophisticated and powerful information systems, which culminated in the creation of the human brain. The possibilities of biological evolution were exhausted on this, which led to the emergence of the technotronic era, and after it to the information society and knowledge society, representing threshold, limit states regarding the ability to produce information, and, therefore, threshold, limit states for the evolution of such type as a whole.

7. Some futurological forecast can be made: if ways to combine the information-technological evolution of human society and the biological evolution of matter will be found, that is, if these two lines of evolution will intersect at some stage (and the modern development of genetic engineering, micro- and nanoelectronics, nanotechnology, neuro- and cognitive science gives reasons for such a development scenario), the emergence of a new form of intelligent life that integrates the biological and technotronic substrate in a single cognitive organism seems very likely.

*The author is deeply grateful to the colleagues who co-authored many of our works.*

[106] Shirokov V.A. Noosfernye izmereniya informacii i znaniya.

# APPENDIXES

*APPENDIX 1*

**Structural classes of equisetites series of verbs that are represented in the dictionary of the ukrainian language**

| Structure type Λ (x) (classc) | | Number of dictionary entries per class | Number of verbs in class | Example (class representative) |
|---|---|---|---|---|
| Signature | Number of words in the registered range | | | |
| 1 | 1 | 16 340 | 16 340 | АБЕТКУВА́ТИ, у́ю, у́єш, *недок., перех.* |
| 2 | 2 | 509 | 1018 | ІТИ́ (ЙТИ), іду́ (йду), іде́ш (йдеш); *мин. ч.* ішо́в (йшов), ішла́ (йшла), ішло́ (йшло), ішли́ (йшли); *наказ. сп.* іди́ (йди); *недок.* |
| 3 | 3 | 10 | 30 | ЗГЛЯ́НУТИСЯ *і рідше* ІЗГЛЯ́НУТИСЯ *і* ЗОГЛЯ́НУТИСЯ, нуся, нешся, *док.* |
| 4 | 4 | 10 | 40 | УСТРІ́НУТИ (ВСТРІ́НУТИ) *і* УСТРІ́ТИ (ВСТРІ́ТИ), і́ну, і́неш; *мин. ч.* устрі́нув, нула, ло *і* устрі́в, рі́ла, ло; *док., діал.* |
| 11 | 2 | 684 | 1368 | БУЛЬКОТА́ТИ, очу́, о́чеш *і* БУЛЬКОТІ́ТИ, очу́, оти́ш, *недок.* |
| 12 | 3 | 5 | 15 | РИБА́ЛИТИ, лю, лиш, РИБА́ЛЧИТИ *і рідко* РИБА́ЧИТИ, чу, чиш, *недок.* |
| 21 | 3 | 5 | 15 | ПЛИСТИ́ *і* ПЛИВТИ́, пливу́, пливе́ш *і* ПЛИ́НУТИ, ну, неш, *недок.* |
| 22 | 4 | 5 | 20 | ЗАВОРКОТА́ТИ *і* ЗАВУРКОТА́ТИ, кочу́, ко́чеш *і* ЗАВОРКОТІ́ТИ *і* ЗАВУРКОТІ́ТИ, очу́, оти́ш, *док.* |
| 111 | 3 | 15 | 45 | ПОЗАКУПО́ВУВАТИ, ую, уєш *і розм.* ПОЗАКУПЛЯ́ТИ, я́ю, я́єш *і* ПОЗАКУ́ПЛЮВАТИ, юю, юєш, *док., перех.* |
| 112 | 4 | 1 | 4 | КМІТИ́ТИ, кмічу́, кмітиш, *рідко* КМЕТИ́ТИ, кмечу, кметиш *і* КМІТУВА́ТИ, *рідко* КМЕТУВА́ТИ, у́ю, у́єш, *недок.* |

*Continuation of APPENDIX 1*

| Structure type Λ (x) (classc) | | Number of dictionary entries per class | Number of verbs in class | Example (class representative) |
|---|---|---|---|---|
| Signature | Number of words in the registered range | | | |
| 1.1 | 2 | 7564 | 15128 | ВИГОВО́РЮВАТИ, юю, юєш, *недок.,* ВИ́ГОВОРИТИ, рю, риш, *док.* |
| 1.2 | 3 | 218 | 654 | ДОСТЯГА́ТИ, а́ю, а́єш, *недок.,* ДОСТЯГТИ́ *і* ДОСТЯГНУ́ТИ, стягну́, стя́гнеш; *мин. ч.* достя́г, ла́, ло́ *і* достягну́в, ла, ло; *док., заст.* |
| 1.3 | 4 | 9 | 36 | ПРИЙМА́ТИ, а́ю, а́єш, *недок.,* ПРИЙНЯ́ТИ, *рідко* ПРИНЯ́ТИ *і діал.* ПРИЙМИ́ТИ, прийму́, прийме́ш, *док., перех.* |
| 1.4 | 5 | 1 | 5 | СПИНА́ТИСЯ, а́юся, а́єшся, *недок.,* СП'ЯСТИ́СЯ *і* ЗІП'ЯСТИ́СЯ, ЗІПНУ́ТИСЯ *і* ЗІП'Я́ТИСЯ, зіпну́ся, зіпне́шся, *док.* |
| 1.11 | 3 | 196 | 588 | ЗВЕЛИ́ЧУВАТИ, ую, уєш, *недок.,* ЗВЕЛИ́ЧИТИ, чу, чиш *і* ЗВЕЛИЧА́ТИ, а́ю, а́єш, *док., перех.* |
| 1.12 | 4 | 9 | 36 | ОГЛЯДА́ТИ, а́ю, а́єш, *недок.,* ОГЛЯ́НУТИ, ну, неш *і розм.* ОГЛЕ́ДІТИ, *рідко* ОГЛЯ́ДІТИ, джу, диш, *док., перех.* |
| 1.21 | 4 | 9 | 36 | ЗАСИХА́ТИ, а́ю, а́єш, *недок.,* ЗАСО́ХНУТИ *і* ЗАСО́ХТИ, хну, хнеш і *діал.* ЗАСХНУ́ТИ, ну́, не́ш, *док.* |
| 1.111 | 4 | 3 | 12 | СПЛИВА́ТИ, а́ю, а́єш, *недок.,* СПЛИСТИ́, спливу́, спливе́ш; *мин. ч.* сплив, сплила́, ло́; *і* СПЛИВТИ́, ву́, ве́ш; *мин. ч.* сплив, спливла, ло́; *і рідко* СПЛИ́НУТИ, ну, неш, *док.* |
| 11.1 | 3 | 800 | 2400 | БЛАГОСЛОВЛЯ́ТИ, я́ю, я́єш *і рідко* БЛАГОСЛОВИ́ТИ, влю́, ви́ш; *мн.* благословля́ть; *недок.,* БЛАГОСЛОВИ́ТИ, влю́, ви́ш; *мн.* благословля́ть; *док., перех.* |
| 11.2 | 4 | 40 | 160 | ПІДТЯГА́ТИ, а́ю, а́єш *і* ПІДТЯ́ГУВАТИ, ую, уєш, *недок.,* ПІДТЯГТИ́ *і* ПІДТЯГНУ́ТИ, тягну́, тя́гнеш, *док., перех. і без додатка.* |
| 12.1 | 4 | 2 | 8 | ПРИВИ́ДЖУВАТИСЯ, уюся, уєшся, ПРИВИДЖА́ТИСЯ *і рідко* ПРИВИЖА́ТИСЯ, а́юся, а́єшся, *недок.,* ПРИВИ́ДІТИСЯ, джуся, дишся, *док.* |

*Continuation of APPENDIX 1*

| Structure type Λ (x) (classc) | | Number of dictionary entries per class | Number of verbs in class | Example (class representative) |
|---|---|---|---|---|
| Signature | Number of words in the registered range | | | |
| 12.2 | 5 | 1 | 5 | РОЗКОРІ́НЮВАТИСЯ, ю́ється *і* РОЗКОРЕНЯ́ТИСЯ, РОЗКОРІНЯ́ТИСЯ, я́ється, *недок.,* РОЗКОРЕНИ́ТИСЯ *і* РОЗКОРІНИ́ТИСЯ, и́ться, *док.* |
| 11.11 | 4 | 63 | 252 | СПОНУКА́ТИ, а́ю, а́єш *і* СПОНУ́КУВАТИ, ую, уєш, *недок.,* СПОНУ́КАТИ, аю, аєш *і рідко* СПОНУКНУ́ТИ, ну́, не́ш, *док., перех., до чого, на що, з інфін. і без додатка.* |
| 11.12 | 5 | 1 | 5 | ВІДДИХА́ТИ, а́ю, а́єш, *рідко* ВІДДИ́ХУВАТИ, ую, уєш, *недок.,* ВІДДИ́ХАТИ, віддиха́ю, віддиха́єш *і* віддиши́ю, віддиши́ш *і* ВІДДИХНУ́ТИ, ВІДІТХНУ́ТИ, ну́, не́ш, *док.* |
| 11.21 | 5 | 3 | 15 | СТРІЧА́ТИСЯ, а́юся, а́єшся *і рідко* СТРІВА́ТИСЯ, а́юся, а́єшся, *недок.,* СТРІ́НУТИСЯ *і* СТРІ́ТИСЯ, стрі́нуся, стрі́нешся, *діал.* СТРІ́ТИТИСЯ, стрі́чуся, стрі́тишся, *док.* |
| 12.11 | 5 | 3 | 15 | ПРОУ́ЧУВАТИ, ую, уєш *і* ПРОВЧА́ТИ *і рідко* ПРОУЧА́ТИ, а́ю, а́єш, *недок.,* ПРОВЧИ́ТИ, чу́, чи́ш *і рідко* ПРОУЧИ́ТИ, учу́, у́чиш, *док., перех.* |
| 11.111 | 5 | 2 | 10 | НАКУПО́ВУВАТИ, ую, уєш *і* НАКУПА́ТИ, а́ю, а́єш, *недок.,* НАКУПИ́ТИ, куплю́, ку́пиш; *мн.* наку́плять; НАКУПУВА́ТИ, у́ю, уєш, *розм.,* НАКУПЛЯ́ТИ, я́ю, я́єш, *док., перех.* |
| 111.1 | 4 | 12 | 48 | ОБ'ЇЖДЖА́ТИ [2], а́ю, а́єш *і* ОБ'ЇЗДИ́ТИ, їжджу́, їзди́ш *і рідко* ОБ'ЇЖДЖУВАТИ, ую, уєш, *недок.,* ОБ'ЇЗДИТИ, їжджу, їздиш, *док., перех.* |
| 112.2 | 6 | 1 | 6 | ЗЛА́ЗИТИ[1], зла́жу, зла́зиш *і рідко* ІЗЛА́ЗИТИ, ла́жу, ла́зиш *і* ЗЛІЗА́ТИ *і рідко* ІЗЛІЗА́ТИ, а́ю, аєш, *недок.,* ЗЛІ́ЗТИ *і рідко* ІЗЛІ́ЗТИ, зу, зеш; *мин. ч.* зліз *і рідко* ізліз, ла, ло; *док.* |
| 111.11 | 5 | 5 | 25 | ЗАСТИЛА́ТИ, а́ю, а́єш *і* ЗАСТЕЛЯ́ТИ, я́ю, я́єш *і рідко* ЗАСТЕ́ЛЮВАТИ, юю, юєш, *недок.,* ЗАСЛА́ТИ, стелю́, сте́леш *і* ЗАСТЕЛИ́ТИ, стелю́, сте́лиш, *док., перех.* |

*Continuation of APPENDIX 1*

| Structure type Λ (x) (classc) | | Number of dictionary entries per class | Number of verbs in class | Example (class representative) |
|---|---|---|---|---|
| Signature | Number of words in the registered range | | | |
| 2.1 | 3 | 54 | 162 | ПОМО́ВЧУВАТИ *і рідко* ПОМО́ВКУВАТИ, ую, уєш, *недок., розм.,* ПОМО́ВЧАТИ, чу, чиш, *док.* |
| 2.2 | 4 | 512 | 2048 | ВБАЧА́ТИ (УБАЧА́ТИ), а́ю, а́єш, *недок.,* ВБА́ЧИТИ (УБА́ЧИТИ), чу, чиш, *док.* |
| 2.3 | 5 | 4 | 20 | УРИВА́ТИ [1] (ВРИВА́ТИ), а́ю, а́єш, *недок.,* УВІРВА́ТИ (ВВІРВА́ТИ) *і* УРВА́ТИ, ву́, ве́ш, *док., перех.* |
| 2.4 | 6 | 26 | 156 | ЗСИХА́ТИ *і рідко* ІЗСИХА́ТИ, а́ю, а́єш, *недок.,* ЗСО́ХНУТИ *і рідко* ІЗСО́ХНУТИ, ЗСО́ХТИ *і рідко* ІЗСО́ХТИ, хну, хнеш; *мин. ч.* зсох *і* зсо́хнув, ла, ло; *док.* |
| 2.11 | 4 | 31 | 124 | СПЛЕ́СКУВАТИ *і* СПЛІ́СКУВАТИ, ую, уєш, *недок.,* СПЛЕСНУ́ТИ, ну́, не́ш *і рідко* СПЛЕСКА́ТИ, сплещу́, спле́щеш, *док.* |
| 2.12 | 5 | 1 | 5 | ЗГНИВА́ТИ *і рідше* ЗОГНИВА́ТИ, а́ю, а́єш, *недок.,* ЗГНИ́ТИ, згнию́, згние́ш *і рідше*ЗІГНИ́ТИ *і* ЗОГНИ́ТИ, ию́, ие́ш, *док.* |
| 2.21 | 5 | 1 | 5 | ЗШТО́ВХУВАТИ *і* ЗІШТО́ВХУВАТИ, ую, уєш, *недок.,* ЗШТОВХНУ́ТИ *і* ЗІШТОВХНУ́ТИ, ну́, не́ш *і розм.* ЗШТОВХА́ТИ, а́ю, а́єш, *док., перех.* |
| 2.22 | 6 | 8 | 48 | ВГЛЯДА́ТИСЯ (УГЛЯДА́ТИСЯ), а́юся, а́єшся, *недок.,* ВГЛЕ́ДІТИСЯ (УГЛЕ́ДІТИСЯ), джуся, дишся *і* ВГЛЯДІ́ТИСЯ (УГЛЯДІ́ТИСЯ), джу́ся, ди́шся, *док.* |
| 2.33 | 8 | 2 | 16 | ВПИХА́ТИ (УПИХА́ТИ), а́ю, а́єш, *недок.,* ВПХНУ́ТИ (УВІПХНУ́ТИ, УПХНУ́ТИ), ну́, не́ш *і* ВПХА́ТИ (УВІПХА́ТИ, УПХА́ТИ), а́ю, а́єш, *док., перех.* |
| 2.42 | 8 | 2 | 16 | УПЛИВА́ТИ [1] (ВПЛИВА́ТИ), а́ю, а́єш, *недок.,* УПЛИВТИ́ (ВПЛИВТИ́) *і* УПЛИСТИ́ (ВПЛИСТИ́), иву́, ве́ш; *мин. ч.* упли́в, упливла́, ло́ *і* уплив, уплила́, ло́; *і рідко* УПЛИ́НУТИ (ВПЛИ́НУТИ), ну, неш, *док.* |
| 2.112 | 6 | 2 | 12 | ЗВИВА́ТИ *і рідко* ІЗВИВА́ТИ, а́ю, а́єш, *недок.,* ЗВИ́ТИ, зів'ю́, зів'є́ш *і рідко* ІЗВИ́ТИ, ізів'ю́, ізів'є́ш *і діал.*ЗВИ́НУТИ *і рідко* ІЗВИ́НУТИ, ну, неш, *док., перех.* |

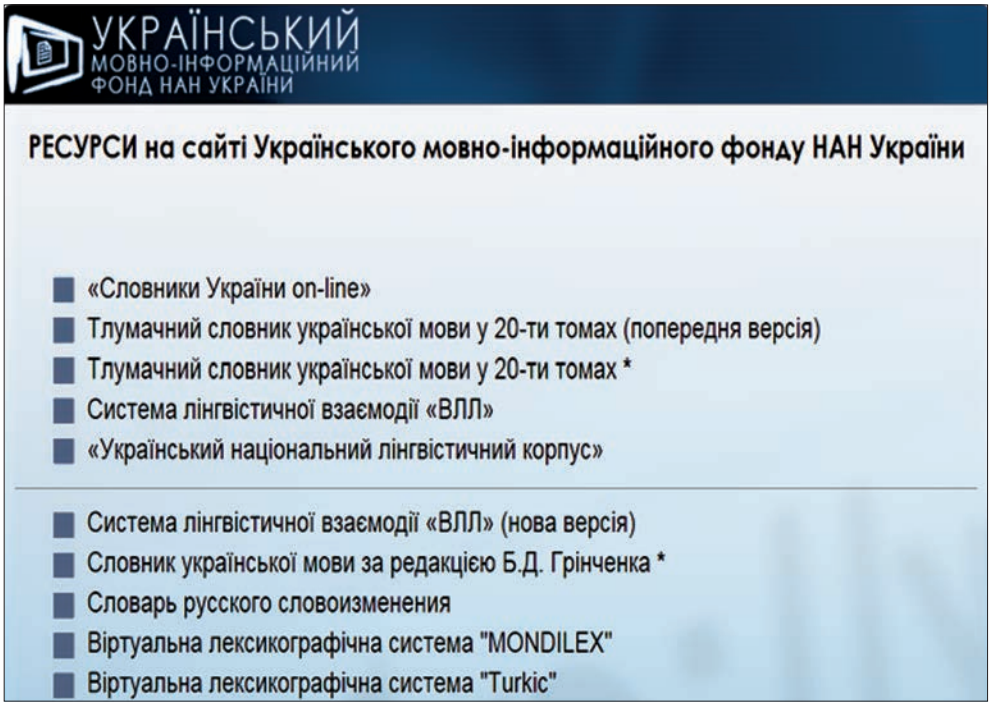
*Continuation of APPENDIX 1*

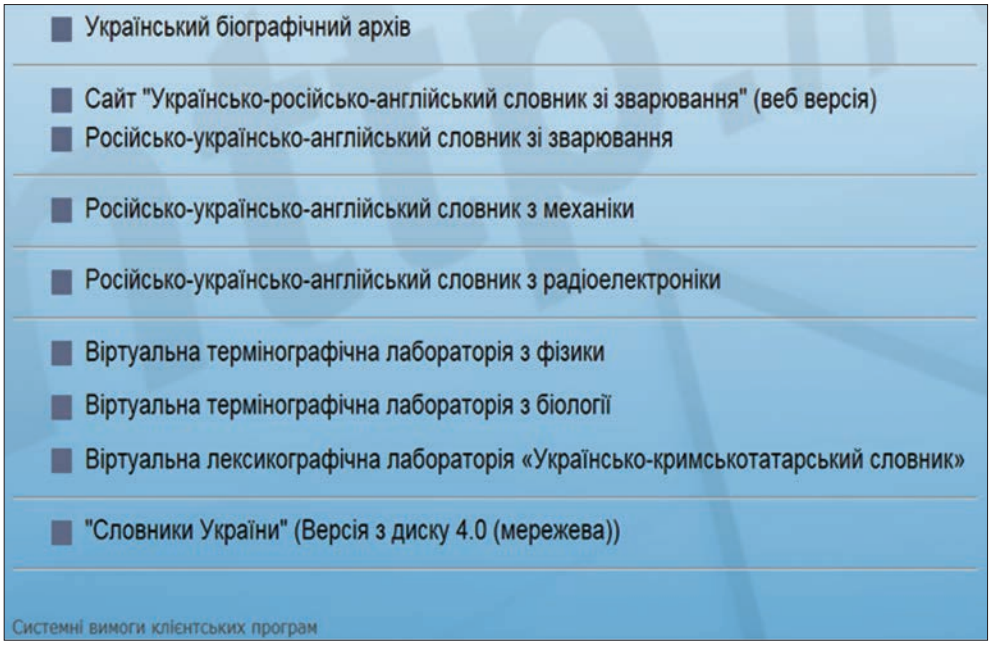
| Structure type Λ (x) (classc) | | Number of dictionary entries per class | Number of verbs in class | Example (class representative) |
|---|---|---|---|---|
| Signature | Number of words in the registered range | | | |
| 21.1 | 4 | 3 | 12 | ПІДСЛУ́ХУВАТИ *і* ПІДСЛУХО́ВУВАТИ, ую, уєш *і* ПІДСЛУХА́ТИ, а́ю, а́єш, *недок.,* ПІДСЛУ́ХАТИ, аю, аєш, *док., перех. і без додатка.* |
| 21.2 | 5 | 3 | 15 | ВРІ́ЗУВАТИ[1] (УРІ́ЗУВАТИ), ую, уєш *і* ВРІЗА́ТИ, а́ю, а́єш, *недок.,* ВРІ́ЗАТИ (УРІ́ЗАТИ), врі́жу, врі́жеш, *док.* |
| 22.2 | 6 | 55 | 330 | ВГЛИ́БЛЮВАТИ (УГЛИ́БЛЮВАТИ), юю, юєш *і* ВГЛИБЛЯ́ТИ (УГЛИБЛЯ́ТИ), я́ю, я́єш, *недок.,* ВГЛИ́БИ́ТИ (УГЛИ́БИ́ТИ), вгли́блю́, вгли́би́ш; *мн.* вгли́бля́ть; *док., перех., рідко.* |
| 22.4 | 8 | 4 | 32 | ВГАНЯ́ТИСЯ[1] (УГАНЯ́ТИСЯ), я́юся, я́єшся *і рідко* ВГО́НИТИСЯ (УГО́НИТИСЯ), нюся, нишся, *недок.,* ВВІГНА́ТИСЯ (УВІГНА́ТИСЯ) *і рідше* ВГНА́ТИСЯ (УГНА́ТИСЯ), вжену́ся, вжене́шся, *док.* |
| 222.2 | 8 | 1 | 8 | ВСЛУХА́ТИСЯ (УСЛУХА́ТИСЯ), а́юся, а́єшся, ВСЛУ́ХУВАТИСЯ (УСЛУ́ХУВАТИСЯ), уюся, уєшся *і рідко* ВСЛУХО́ВУВАТИСЯ (УСЛУХО́ВУВАТИСЯ), уюся, уєшся, *недок.,* ВСЛУ́ХАТИСЯ (УСЛУ́ХАТИСЯ), аюся, аєшся, *док., у що і без додатка.* |
| 212.111 | 8 | 1 | 8 | ПРОВО́ДИТИ, ПРОВА́ДИТИ, джу, диш, ПРОВОДЖА́ТИ, а́ю, а́єш, *рідко* ПРОВО́ДЖУВАТИ *і* ПРОВА́ДЖУВАТИ, ую, уєш, *недок.,* ПРОВЕСТИ́, еду́, еде́ш; *мин. ч.* прові́в, вела́, ло́, *рідко* ПРОВОДИ́ТИ, воджу́, во́диш *і діал.* ПРОВА́ДИТИ, джу, диш, *док.* |
| 3.3 | 6 | 1 | 6 | ЗШКРІБА́ТИ, ЗІШКРІБА́ТИ *і* ЗСКРІБА́ТИ, а́ю, а́єш, *недок.,* ЗШКРЕБТИ́, ЗІШКРЕБТИ́ *і* ЗСКРЕБТИ́, бу́, бе́ш, *док., перех., розм.* |
| 3.11 | 5 | 2 | 10 | ПЕРЕПЛІ́СКУВАТИ, ПЕРЕПЛЕ́СКУВАТИ *і* ПЕРЕПЛЮ́СКУВАТИ, ую, уєш. *недок.,* ПЕРЕПЛЕСНУ́ТИ, ну́, не́ш *і* ПЕРЕПЛЮ́СНУТИ, ну, неш, *док., перех., через що.* |

*The End of APPENDIX 1*

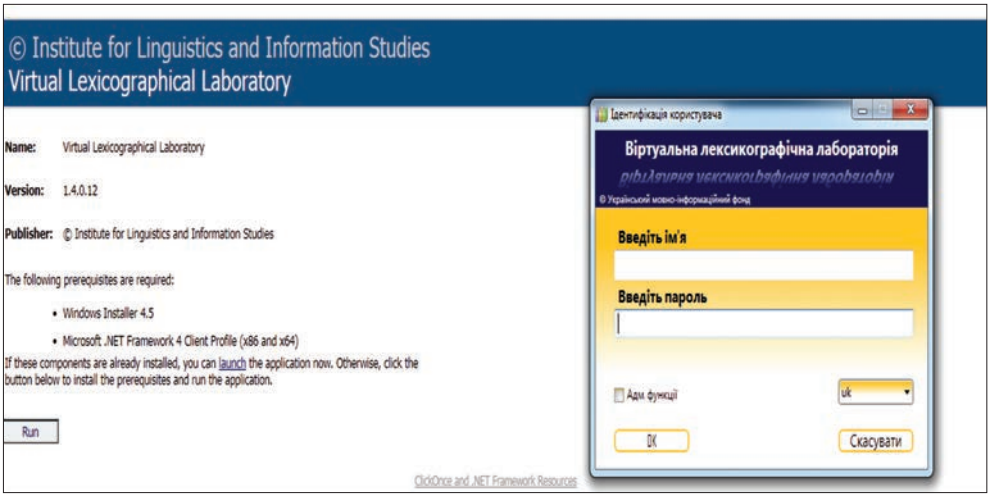
| Structure type Λ (x) (classc) | | Number of dictionary entries per class | Number of verbs in class | Example (class representative) |
|---|---|---|---|---|
| Signature | Number of words in the registered range | | | |
| 4.2 | 6 | 1 | 6 | ВХО́ДИТИ і ВВІХО́ДИТИ (УХО́ДИТИ, УВІХО́ДИТИ), джу́, дииш, *недок.,* ВВІЙТИ́ (УВІЙТИ́), увійду́, увійдеш, *док.* |
| 4.4 | 8 | 1 | 8 | ВТОВКМА́ЧУВАТИ (УТОВКМА́ЧУВАТИ) і ВТОКМА́ЧУВАТИ (УТОКМА́ЧУВАТИ), ую, уєш, *недок.,* ВТОВКМА́ЧИТИ (УТОВКМА́ЧИТИ) і ВТОКМА́ЧИТИ (УТОКМА́ЧИТИ), чу, чиш, *док., перех., розм.* |
| 4.22 | 8 | 2 | 16 | УДОВОЛЬНЯ́ТИ (ВДОВОЛЬНЯ́ТИ) *і заст.* УДОВОЛЯ́ТИ (ВДОВОЛЯ́ТИ), я́ю, я́єш, *недок.,* УДОВОЛЬНИ́ТИ (ВДОВОЛЬНИ́ТИ), ню́, ни́ш *і заст.*УДОВОЛИ́ТИ (ВДОВОЛИ́ТИ), лю́, ли́ш, *док., перех.* |
| **To t a l :** | | **27 243** | **4 1402** | |

**Screenshots «resources on the website of the ukrainian lingua-information fund» (http://lcorp.ulif.org.ua/lslist/)**

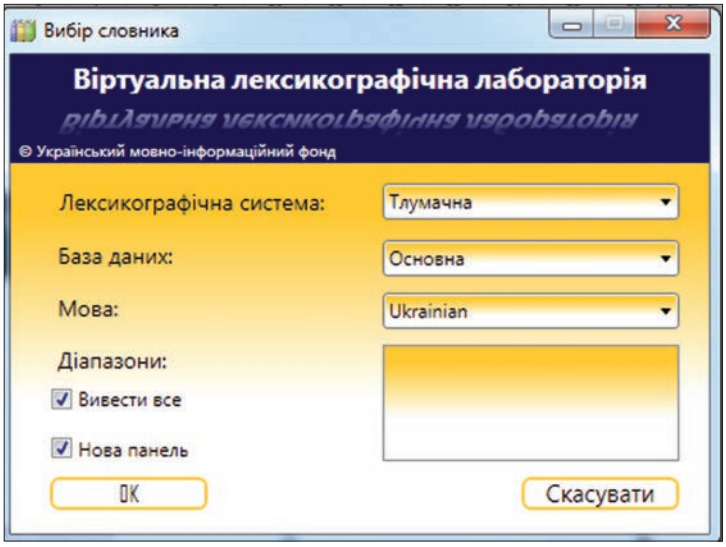


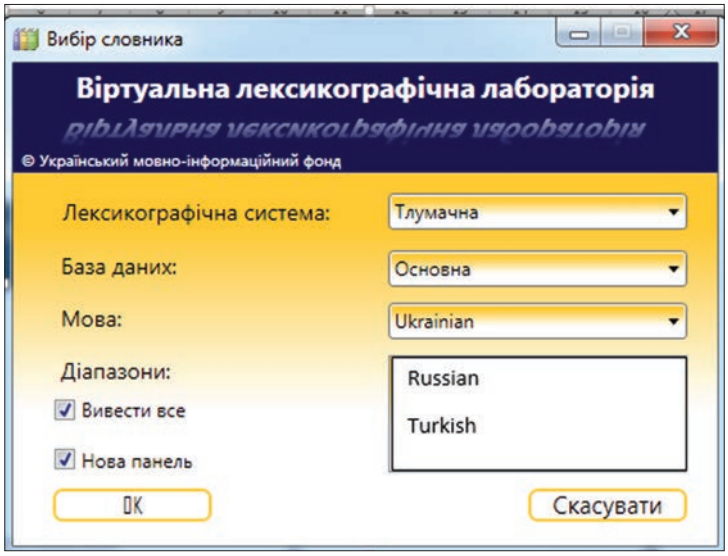
At the moment, most of the resources of the Ukrainian Lingua-Information Fund (and all instrumental ones) operate in the corporate network mode with regulated access through an authentication system, the elements of which are presented below.



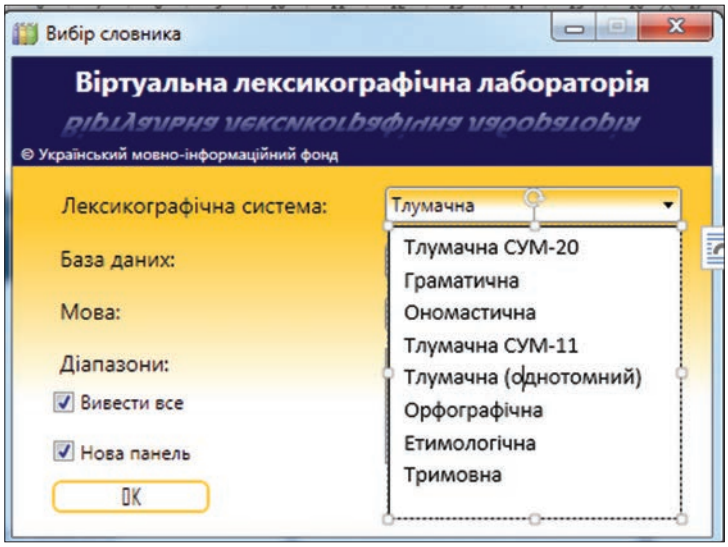

The bottom slide shows the actual entrance to the Virtual Lexicographic Laboratory.




It consists of several separate VLL. For example, lexicographic VLs of an explanatory type are represented by three VLLs: for Ukrainian, Russian and Turkish languages.




Also, for the Ukrainian language, the following VLLs are functioning in the system: grammatical, onomastic, spelling, etymological, and others, as demonstrated on the slide.



# REFERENCES


ANSI/X3/SPARK/DBMS study group interim report.*FDT-Bull. ACM SIGMOD*. 1975. **7**, No. 2. 140 p.

Apresyan Yu.D. Integral'noe opisanie yazyka i tolkovyj slovar'.*Vopr. yazykoznaniya*. 1986. No. 2. S. 57—70.

Apresyan Yu.D. Izbr. trudy. T. 1: Leksicheskaya semantika: sinonimicheskie sredstva yazyka. 2 izd., ispr. i dop. Moscow: Shkola «Yazyki russkoj kul'tury», 1995. 472 s.

Apresyan Yu.D. Izbr. trudy. T. 2: Integral'noe opisanie yazyka i sistemnaya leksikografiya. Moscow: Shkola «Yazyki russkoj kul'tury», 1995. 767 s.

Ball Philip.Physics of life: The dawn of quantum biology. *Nature*. 2011. **474**, P. 272—274.

Bekhterev V.M. Ob'ektivnaya psihologiya. Moscow: Nauka, 1991. 480 s.

Bell D. The Social Framework of Information Society. Leningrad, 1981. P. 4.

Belonogov G.G., Zelenkov Yu.G., Kuznecov B.A., Novoselov A.P., Horoshilov A.A. Sistemy frazeologicheskogo mashinnogo perevoda politematicheskih tekstov s russkogo yazyka na anglijskij i s anglijskogo na russkij (sistemy RETRANS i ERTRANS). *MFID*. 1995. **20**. No. 2. S. 20—24.

Bordonaro M., Ogryzko V.V. Quantum biology at the cellular level — Elements of the research program. *Biosystems*. 2013. 112(1), pp. 11—30.

Boum A. Kvantovaya mekhanika: osnovy i prilozheniya. Moscow: Mir, 1990. 720 s.

Brzesinski Z. Between Two Ages: America's Role in Technotronic Era. NY, 1970. 52 p.

Chekman I.S. Quantum pharmacology: new direction in materia medica. *Nauka ta innovacii*. 2010. **6**, No. 2, P. 29—35. https://doi.org/10.15407/scin6.02.029

Chivilihin S.I. Kvantovaya informatika: Ucheb. posobie. SPb., 2007.

Cirel'son V.G. Kvantovaya himiya. Molekuly, molekulyarnye sistemy i tverdye tela: Ucheb. posobie. Moscow: BINOM. Laboratoriya znanij, 2010. 495 s.

CODASYL DBTG 1971. CODASYL Data Base Task Group April 71 Report. ACM New York, 1971.

Davydov A.S. Biologiya i kvantovaya mekhanika. Kyiv: Nauk. dumka, 1979. 296 s.

Denisov P.N. Sistemnost' i svyazannost' v leksike i sistema slovarej. *Problematika opredelenij terminov v slovaryah raznyh tipov* / Red. kol. S.G. Barhudarov i dr. Leningrad: Nauka, 1976. S. 63—73.

Devkin V.D. O nerodivshihsya nemeckih i russkih slovaryah. *Vopr. yazykoznaniya*. 2001. No. 1. S. 85—97.

Dimitrova L., Koseska-Toszewa V., Garabik R., Erjavec T., Iomdin L., Shirokov V. Mondilex — towards the research infrastructure for digital resources in Slavic lexicography. *Cognitive Studies* / *Etudes Cognitives*, 10. Warsaw: SOW Publishing House, 2010, pp. 155—170.

Dimitrova L., Koseska-Toszewa V., Garabik R., Erjavec T., Iomdin L., Shirokov V. Conceptual Scheme for a Research Infrastructure Supporting Digital Resources in Slavic Lexicography. Sofia, 2010. 130 p.

Dirak P. Principy kvantovoj mekhaniki. Izd. 2. Moscow: Nauka, 1979. 480 s.
Dizard W.P. The coming information age. An overview of Technology, Economics and Politics. Longman, 1982. 152 p.
Eco Umberto. Opera aperta. Forma e indeterminazione nelle poetiche contemporanee. Milano: Bompiani, 1967. 284 p.
Ejring G., Uolter Dzh., Kimball Dzh. Kvantovaya himiya / Per. s angl. Moscow: 1948. 528 s.
Eko Umberto. Imya rozy / Per. s ital. E. Kostyukovich. SPb.: Simpozium, 2005. 440 s.
Eric C., S. Piotrowski. Toward a Quantum Linguistics: Possibilities for Change in the Delta Zone, 2000.
Gejzenberg V. Fizika i filosofiya. Chast' i celoe / Per. s nem. Moscow: Nauka, 1989. S. 191—196.
Gel'man G.G. Kvantovaya himiya. Moscow: ONTI, 1937. 546 s.
Golovko G.G., Shirokov V.A. Teorema Levengejma — Skolema kak formal'nyj korrelyat leksikograficheskogo effekta v informacionnyh sistemah. MegaLing'2006. *Horyzonty prykladnoi linhvistyky ta linhvistychnykh tekhnolohii. Dop. Druhoi mizhnar. nauk. konf.* (20—27 veresnia, 2006, Ukraina, Krym, Partenit). S. 201—202.
Gryaznuhina T.A., Lyubchenko T.P., Rabulec A.G. Elektronnaya versiya Grammaticheskogo slovarya russkogo yazyka A.A. Zaliznyaka kak instrument avtomaticheskogo morfologicheskogo analiza russkogo teksta. *Korpusnaya lingvistika i lingvisticheskie bazy dannyh: Dokl. nauch. konf.* SPb.: Izd-vo S.-Peterb. un-ta, 2002. S. 63—70.
Gryaznuhina T.A., Lyubchenko T.P., Shevchenko I.V. Novaya versiya elektronnogo grammaticheskogo slovarya russkogo yazyka s uchetom akcentuacii. *Slovo i slovar' = Vocabulum et vocabularium: Sb. nauch. tr. po leksikografii* / Otv. red. L.V. Rychkova. Grodno: GrGU, 2005. S. 188—193.
Gumboldt V. Ob otlichiyah v stroenii chelovecheskih yazykov i ih vliyanii na duhovnoe razvitie chelovechestva. SPb., 1859. 366 s.
Gusserl E. Sobr. sochinenij. T. III (1). Logicheskie issledovaniya. Issledovaniya po fenomenologii i teorii poznaniya / Per. s nem. V.I. Molchanova. Moscow: Dom intellektual'noj knigi, 2001.
Hrennikov A.Yu. Vvedenie v kvantovuyu teoriyu informacii. Moscow: FIZMATLIT, 2008. 284 s.
Kobozeva I.M. Lingvisticheskaya semantika. Moscow: KomKniga, 2007. 352 s.
Kolmogorov A.N. Teoriya peredachi informacii. *Sessiya AN SSSR po nauch. problemam avtomatizacii proizvodstva* (15—20 okt. 1956: Plenar. zasedaniya). Moscow: Izd-vo AN SSSR, 1957. S. 66—99.
Kolmogorov A.N. Tri podhoda k opredeleniyu ponyatiya «kolichestvo informacii». *Teoriya informacii i teoriya algoritmov*. Moscow: Nauka, 1987. S. 220.
Kurzweil Ray. The singularity is near: when humans transcend biology. Viking, 2006. 652 p.
Landau L.D., Lifshic E.M. Kvantovaya mekhanika (nerelyativistskaya teoriya). Izd. 4. Moscow: Nauka, 1989. 768 s. (Teoreticheskaya fizika). T. III.
Linhvistychni ta tekhnolohichni osnovy tlumachnoi leksykohafii / V.A. Shyrokov, V.M. Bilonozhenko, O.V. Buhakov ta in.; Vidp. red. V.A. Shyrokov. Kyiv: Dovira, 2010. 295 s.
Liubchenko T.P. Morfolohichna model slovozminy flektyvnoi movy ta elektronnyi hramatychnyi slovnyk. *Byonyka intellekta*. 2006. No. 1 (64). S. 72—77.
Lyubchenko T. Modelling of the Digital Grammar Dictionary of Russian. *Organization and Development of Digital Lexical Resources. MONDILEX 2nd Open Workshop*. Kyiv, Ukraine, 2—4 February, 2009. P. 73—84.
Maslov V.P. Kvantovaya ekonomika. Moscow: Nauka, 2006. 96 s.
Masuda T. The Information Society as a Postindustrial Society. Tokyo, 1983. P. 28, 47, 61, 74.

Medvedev D.A. Sverhtekhnologii i obshchestvo v XXI veke. Kurs lekcij [Videolekciya 10. Tekhnologicheskaya singulyarnost']. 2008. URL: http://univertv.ru/video/sociologiya/ (access from 21.12.2015).

Mel'nikov G.P. Sistemnaya lingvistika i ee otnoshenie k strukturnoj. *Problemy yazykoznaniya: Dokl. i soobshch. sov. uchenyh na X Mezhdunar. kongresse lingvistov*. Moscow: Nauka, 1967. S. 98—102.

Mel'nikov G.P. Sistemnyj podhod v lingvistike. *Sistemnye issledovaniya. Ezhegodnik*. 1972. Moscow: Nauka, 1973. S. 183—204.

Morris Ch. Signs, language and behaviour. NY: Moss, R.N., 1947. https://doi.org/10.1080/00043079.1949.11407849

Muzyakov S. Informacionnaya sreda i usloviya eksponencial'nogo rosta ob'ema znanij v sovremennom obshchestve. *Vlast'*. 2012. No. 4. S. 42—46.

Ostapova I.V. Digital Etymology (illustrated by the example of the Etymological Dictionary of Ukrainian language). In: I.V. Ostapova. Organization and Development of Digital Lexical Resources. Proceedings of Mondilex Second Open Workshop. Kyiv, 2009, pp. 66—72.

Ostapova I.V. Leksikograficheskaya struktura etimologicheskogo slovarya i ego predstavlenie v cifrovoj srede. In: I.V. Ostapova. Komp'yuternaya lingvistika i intellektual'nye tekhnologii. *Materialy ezhegodnoj Mezhdunar. konf. «Dialog»* (27—31 maya 2009, Bekasovo). 8(15). Moscow: RGGU, 2009. XII. 620 s. S. 359—365.

Ostapova I.V. Virtual'naya leksikograficheskaya laboratoriya dlya tolkovyh slovarej. In: I.V. Ostapova, V.A. Shyrokov. Komp'yuternaya lingvistika i intellektual'nye tekhnologii. *Materialy ezhegodnoj Mezhdunar. konf. «Dialog»* (26—30 maya 2010, Bekasovo). 9(16). Moscow: RGGU, 2010. 920 p.

Palagin A.V., Shirokov V.A. Principles of cognitive lexicography. *Informational theories & application*. 2000. Vol. 9, No. 2, pp. 43—51.

Panov A.D. Tekhnologicheskaya singulyarnost', teorema Pernouza ob iskusstvennom intellekte i kvantovaya priroda soznaniya URL: http://www.intelros.ru/pdf/metafizika/2013_3/7.pdf

Piotrowski Eric S. Toward a Quantum Linguistics: Possibilities for Change in the Delta Zone. 2000.

Plungyan V.A. Vvedenie v grammaticheskuyu semantiku: grammaticheskie znacheniya i grammaticheskie sistemy yazykov mira. Moscow: Izd-vo RGGU, 2011. S. 183.

Potapova E.V. Model' lingvisticheskoj ontologii predmetnoj oblasti c nechetkimi semanticheskimi sostoyaniyami terminov. *Bionika intellekta*. 2012. No. 2(79). S. 95—102.

Preskill Dzh. Kvantovaya informaciya i kvantovye vychisleniya. T. 1. Moscow-Izhevsk: Regulyarnaya i haoticheskaya dinamika, 2008. 464 s.

Pribram K. Yazyki mozga. Eksperimental'nye paradoksy i principy nejropsihologii / Per. s angl. Ya.N. Danilovoj i E.D. Homskoj; Red. i predislovie akad. A.R. Luriya. Moscow: Progress. Red. lit-ry po filosofii, 1975.

Problemy grammaticheskogo modelirovaniya / Red. A.A. Zaliznyak. Moscow: Nauka, 1973. 262 s.

Puaza Bruno. Kurs teorii modelej. Almaty, 2001. 460 s.URL: https://people.maths.ox.ac.uk/zilber/poizat/poizatkoi8.html (access from 20.08.2020).

Pyul'man B., Pyul'man A. Kvantovaya biohimiya / Per. s angl. Moscow: Mir, 1965. 654 s.

Quantum Medicine. Quantum Therapy. URL: http://www.qrs.com/quantum_medicine.php

Rabulets O.H. Intehrovani leksykohrafichni systemy: Avtoref. dys. ... kand. tekhn. nauk. 05.13.06. Nats. bibl. Ukrainy im. V.I. Vernadskogo. Kyiv, 2002. 18 s.

Rabulets O.H., Sukharyna N.M., Shirokov V.A., Yakymenko K.M. Diieslovo v leksykohrafichnii systemi. Kyiv: Dovira, 2004. 260 s.

Rusanivskyi V.M. Struktura leksychnoi i hramatychnoi semantyky. Kyiv: Nauk. dumka, 1988. 237 s.

Rusanivskyi V.M., Shirokov V.A. Informatsiino-linhvistychni osnovy suchasnoi tlumachnoi leksykohrafii. *Movoznavstvo*. 2002. No. 6. S. 7—48.
Russkij semanticheskij slovar'. Tolkovyj slovar', sistematizirovannyj po klassam slov i znachenij / Red. N.Yu. Shvedova. Moscow: Azbukovnik, 1998. **1**. 832 s.
Selfridge O.G. Pandemonium: a paradigm for learning. *Mechanisation of thought processes*. London: HMSO, 1959, pp. 511—531.
Shilejko A.V., Kochnev V.F., Himushin F.F. Vvedenie v informacionnuyu teoriyu sistem. Moscow: Radio i svyaz', 1985. 278 s.
Shirokov K.V. Imenna slovozmina u suchasnii turetskii movi. Kyiv: Dovira, 2009. 318 s.
Shirokov V. System relations in explanatory dictionaries. *Proceedings of the VI Int. Sci. and Tech. Conf*. CSIT, 2011. Lviv, 2011. P. 260—264.
Shirokov V., Ostapova I., Nadutenko M., Verbynenko Yu. Ontologized lexicographical systems in modern terminography. *Cognitive Studies.Études Cognitives*. Warsaw, 2016. **16**, pp. 90—99.
Shirokov V., ShevchenkoI. On the phenomenological approach to grammar. *Cognitive studies. Etudes cognitives*. Warsaw: SOW Publishing House, 2015. **15**, P. 3—34.
Shirokov V.A. Elementy leksykohrafii. Kyiv: Dovira, 2005. 304 s.
Shirokov V.A. Fenomenolohiia leksykohrafichnykh system. Kyiv: Nauk. dumka, 2004. 326 s.
Shirokov V.A. Grammatika kak fenomenologicheskaya problema. *Bionika intellekta*. 2013. 1 (180). S. 3—14.
Shirokov V.A. Humanitarna tradytsiia i tekhnolohichnyi status movy. *Movoznavstvo*. 2001. No. 3. S. 128—132.
Shirokov V.A. Informatsiina teoriia leksykohrafichnykh system. Kyiv: Dovira, 1998. 331 s.
Shirokov V.A. Informatsiino-enerhetychni transformatsii ta informatsiine suspilstvo. *Ukr.-polsk. nauk.-prakt. zhurn. «Nauka, innovatsiia, informatsiia»*. Kyiv, 1996. No. 1. S. 48—66.
Shirokov V.A. Komp'iuterna leksykohrafiia. Kyiv: Nauk. dumka, 2011. 356 s.
Shirokov V.A. Leksykohrafichne predstavlennia semantychnykh staniv. *Matematychni mashyny i systemy*. 1999. No. 3. C. 21—32.
Shirokov V.A. Noosfernye izmereniya informacii i znaniya.*Noosferologiya: nauka, obrazovanie, praktika* / Red. O.A. Gabrielyan. Simferopol: TNU im. V.I. Vernadskogo, 2008. 463 s.
Shirokov V.A. Ocherk osnovnyh principov kvantovoj lingvistiki. *Bionika intellekta*. 2007. No. 1 (66). S. 25—32.
Shirokov V.A. Proekt universalnoi systemy slov'ianskoi leksykohrafii «Leksyka-Slavika». *Slov'ianski obrii*. 2. XIV Mizhnar. z'izd slavistiv (10—16 veresnia 2008, Okhryd, Respublika Makedoniia). Kyiv, 2008. S. 801—829.
Shirokov V.A. Semantychni stany movnykh odynyts ta yikh zastosuvannia v kohnityvnii leksykohrafii. *Movoznavstvo*. 2005.
Shirokov V.A. Sistemna semantika tlumachnih slovnikiv. *Zb., prisvyach. 75-littyu V.G. Sklyarenka*. Kyiv: Nauk. dumka, 2012.
Shirokov V.A., Bilonozhenko V.M., Buhakov O.V. tain. Linhvistychni ta tekhnolohichni osnovy tlumachnoi leksykohrafii / Vidp. red. V.A. Shirokov. Kyiv: Dovira, 2010. 295 s.
Shirokov V.A., Buhakov O.V., Hriaznukhina T.O., Kostyshyn O.M., Kryhin M.Yu. Korpusna linhvistyka. Kyiv: Dovira, 2005. 472 s.
Shirokov V.A., Manako V.V. Tekhnolohichny i kompleks zi stvorennia fundamentalnoi akademichnoi leksykohrafichnoi systemy «Slovnyk ukrainskoi movy». *Informatsiini resursy nauk.-tekh. informatsii: problemy stvorennia i vykorystannia: VII Mizhnar. nauk.-prakt. konf*. Kyiv: UkrINTEI, 2000. S. 121—124.
Shirokov V.A., Rabulets O.H. Formalizatsiia u haluzi linhvistyky. *Aktualni problemy ukrainskoi linhvistyky: teoriia i praktyka: Zb. nauk. prats*. Kyiv, 2002. **5**. S. 3—27.

## REFERENCES


Shirokov V.A., Rabulets O.H., Shevchenko I.V., Kostyshyn O.M., Yakymenko K.M. Tekhnolohichni osnovy suchasnoi tlumachnoi leksykohrafii.*Movoznavstvo*. 2002. No. 6. S. 49—86.

Shirokov V.A., Shevchenko I.V. Hramatyka u fenomenolohichnomu vymiri. *Movoznavstvo*. 2014. No. 4. S. 3—27.

Slovnyk ukrainskoi movy u 20 tomakh: T. 1/ Hol. red. V.M. Rusanivskyi. Kyiv: Nauk. dumka, 2010; T. 2 / Hol. red. V.M. Rusanivskyi. Kyiv: Nauk. dumka, 2011; T. 3 / Hol. red. V.M. Rusanivskyi. Kyiv: Viakhyr, 2012; T. 4 / Hol. red. V.M. Rusanivskyi. Kyiv: Ukr. movno-inform. fond, 2013; T. 5—7 / Hol. red. V.A. Shyrokov. Kyiv: Ukr. movno-inform. fond, 2014—2016.

Slovnyk ukrainskoi movy: V 11 t. / Hol. red. I.K. Bilodid. Kyiv: Nauk. dumka, 1970—1980.

Solomonik A. Semiotika i lingvistika. Moscow: Molodaya gvardiya, 1995. 352 s.: il.

Stapp Henry P. Quantum Nonlocality and the Description of Nature. *Philosophical Consequences of Quantum Theory* / James T. Cushing and Ernan McMullin (eds.). Notre Dame Press, 1989.

Stratonovich R.L. Teoriya informacii. Moscow: Sov. radio, 1975. 423 s.

Strehle Susan. Fiction in the Quantum Universe. Chapel Hill: Un-ty of North Carolina Press, 1992.

Tsichritzis D. and Klug A. (eds.). The ANSI/X3/SPARK Framework. AFIPS Press, Nontvale N. J., 1978.

Uilson R. Kvantovaya psihologiya. Moscow: Sofiya, 2016. 224 s.

Ukrainskyi Linhvistychnyi Portal. URL: http://lcorp.ulif.org.ua/dictua/

Ukrainskyi movno-informatsiinyi fond NAN Ukrainy.URL: http://lcorp.ulif.org.ua/LSlist

Ul'man Dzh. Osnovy sistem baz dannyh. Moscow: Finansy i statistika, 1983. 334 s.

Uspenskij V.A. K opredeleniyu padezha po A.N. Kolmogorovu. *Byull. Ob'edineniya po problemam mashinnogo perevoda*. 1957. No. 5. S. 11—18.

Uspenskij V.A. Teorema Gyodelya o nepolnote. Moscow: Nauka, 1982. 110 s.

Valiev K.A. Kvantovaya informatika. Komp'yutery, svyaz' i kriptografiya. *Vestnik RAN*. 2000. **70**, No. 8. S. 688—695.

Varbot Zh.Zh. Etimologiya. *Russkij yazyk: Enciklopediya*. Moscow: Drofa, 1993. S. 643—647.

Vinge Vernor. The Coming Technological Singularity: How to survive in the post-human era. *Vision-21. Interdisciplinary Science and Engineering in the Era of Cyberspace. Proceedings of a symposium consposored by the NASA Lewis Research Centre and the Ohio Aerospace Institute and held in Westlake.* (March 30—31, 1993, Ohio).URL: https://edoras.sdsu.edu/~vinge/misc/singularity.html

Vol'kenshtejn M.V. Teoriya informacii i evolyuciya. *Kibernetika zhivogo: biologiya i informaciya*. Moscow: Nauka, 1984. S. 45—53.

Vol'kenshtejn M.V. Entropiya i informaciya. Moscow: Nauka, 1986. 190 s.

Vygotskij L.S. Myshlenie i rech. Izd. 5, ispr. Moscow: Labirint, 1999. 352 s.

Zaliznyak A.A. «Russkoe imennoe slovoizmenenie» s prilozheniem izbrannyh rabot po sovremennomu russkomu yazyku i obshchemu yazykoznaniyu. Moscow: Yazyki slavyanskoj kul'tury, 2002. I—VIII. 752 s. (Studia phililogica).

Zolotova G.A. Sintaksicheskij slovar'. Repertuar elementarnyh edinic russkogo sintaksisa. Moscow: Nauka, 1988. 439 s.

Монографія висвітлює фундаментальні інформаційні проблеми лінгвістики і є результатом багаторічних досліджень і спостережень автора у сфері інформаційних властивостей природної людської мови. Ці дослідження змусили автора заглибитись у феноменологію пізнавально-комунікаційних процесів, заснованих на природній мові, вивчати їхні інформаційні властивості і будувати на цій основі лінгвістичні моделі, орієнтовані на конкретне застосування.

Розвиток цих ідей у сучасному цивілізаційно-еволюційному контексті стимулює розвиток ідей квантової лінгвістики, яка нині перебуває в області наукових прогнозів.

*Наукове видання*

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
УКРАЇНСЬКИЙ МОВНО-ІНФОРМАЦІЙНИЙ ФОНД НАН УКРАЇНИ

ШИРОКОВ Володимир Анатолійович

# МОВА ІНФОРМАЦІЯ СИСТЕМА

Англійською мовою

Редактор *В.К. Рего*

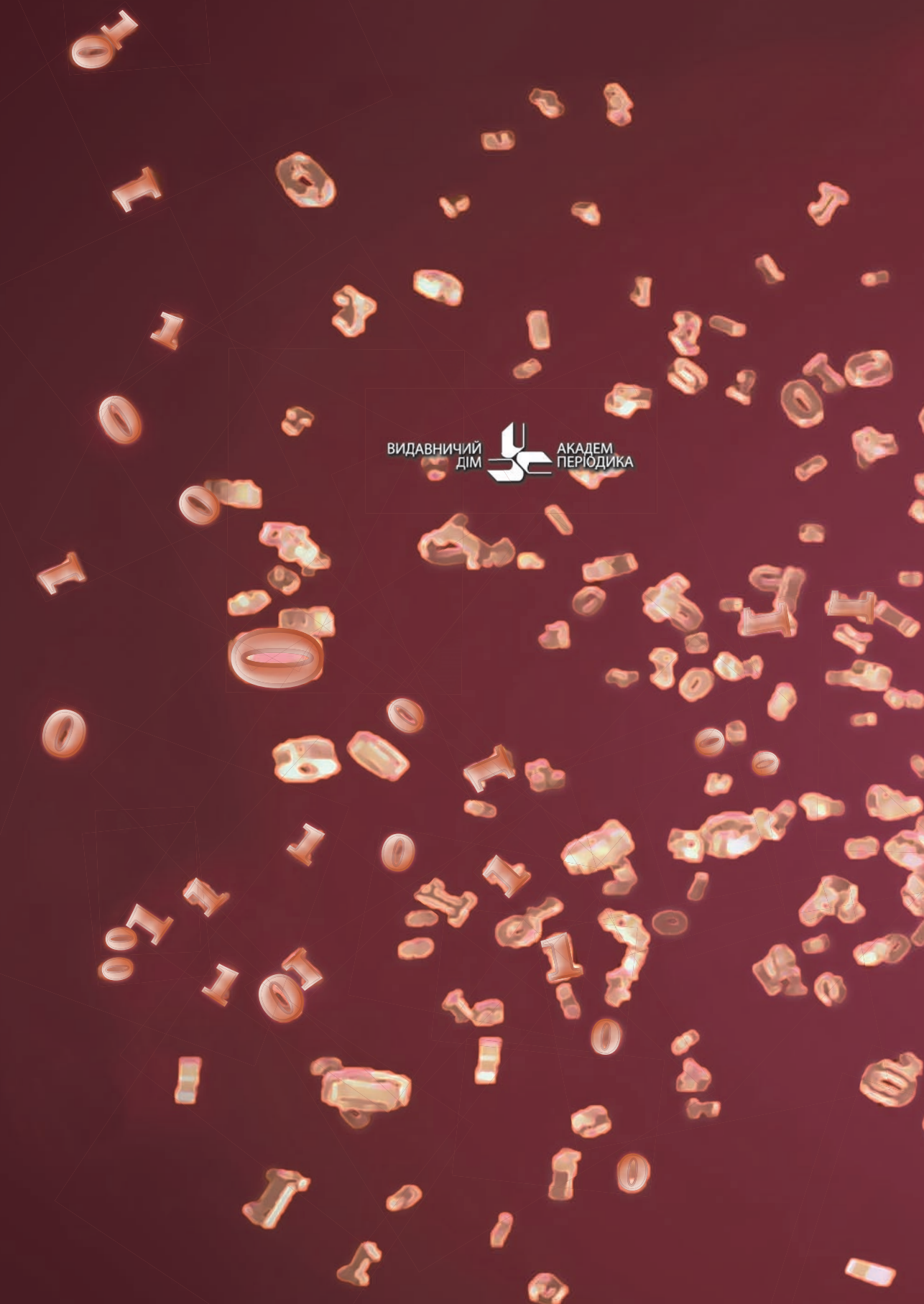
Художнє оформлення та комп'ютерна верстка
*О.А. Бурдік*

Технічне редагування
*Т.М. Шендерович*, *О.А. Бурдік*

Підп. до друку 21.12.2021. Формат 70 × 100/16. Гарн. Minion Pro.
Ум. друк. арк. 13,00. Обл.-вид. арк. 12,52. Тираж 150 прим. Зам. № 6528.

Видавець і виготовлювач Видавничий дім «Академперіодика» НАН України
01024, Київ, вул. Терещенківська, 4

Свідоцтво про внесення до Державного реєстру суб'єктів
видавничої справи серії ДК № 544 від 27.07.2001


ВИДАВНИЧИЙ ДІМ
АКАДЕМ ПЕРІОДИКА