ON THE PHENOMENOLOGICAL APPROACH TO GRAMMAR

Abstract

Phenomenological foundations for grammar description of the natural language are under considerations. The picture of language world based on the factorization of communication and cognitivity is proposed. The problem of observability–unobservability in linguistics is discussed. The concept of the states of language units is formulated both with its interpretation in the formal definition of the noun’s case. System analysis of the complexity in cognitive processes is considered.

Keywords: phenomenological approach; grammar description of the language; observability in linguistics; semantic states of the language units

Logical and phenomenological foundations for grammar description of language

Grammatical problems have lately been in the range of interests of not only theoretical linguists that have to deal with them, so to speak, on duty, but more — applied scientists who create systems of computer processing of language. The concept of this study includes a wide range of problems related to information retrieval, translation from one language to another, knowledge engineering, speech recognition and synthesis of text and images, the interpretation of the content of texts, various linguistic expertise, language training, etc. Grammatical description for the selected range of tasks is interested mainly in the sense that it provides an interpretation of linguistic objects by their form which is crucial in formalizing herein since formalization (as deeply as possible) is a necessary condition for the creation of efficient artifacts endowed with intellectual properties that carry out the linguistic processing.

The experience of building any theory says that the success of the formalization description of a subject area is related directly to the choice of adequate facilities for conceptual representation of phenomena. The above fully applies to the grammar. It should, however, be admitted that when creating grammatical patterns their
authors often tend to narrow the field of linguistic phenomenology, so that their constructions are sometimes of quite a limited scope. Of course, this restriction greatly simplifies the task of researchers, but the pay for this simplification is lack of adequacy of the description, its correspondence to the phenomenology of language rather incomplete and insufficient for applications and, ultimately, ineffectiveness of the final product, namely linguistic technologies.

It is clear that the scope of the field of linguistic phenomenology is not fixed and is subject to confirmation in connection with the tasks that are to be addressed in the research process. Moreover, there may be very different approaches to determining the very principles on which the said specification is to be implemented. In this regard, we want to quote the words of Isaac Newton from his first scientific study written in the age of eighteen, called «The Universal Language» (Newton, 1957) and devoted to fundamental grammatical problems, «The Dialects of each Language being soe divers & arbitrary A generall Language cannot bee so fitly deduced from them as from ye natures of things themselves wch is ye same to all Nations & by wch all Languag was at ye first composed. Now ye frame of Beings is made up of several substances (bodys & spirits) each of woh are in this or that state case or Condition And ye use of Language is yt one man may signify to another in wt state any substance is, hath beene, shall bee, may bee, should bee, is wished to bee, is commaunded to bee &c These & such like considerations may bee an Induction to ye following rules».

From this brief snippet it follows quite unequivocally that Isaac Newton, in formulating the original principles of the theoretical and linguistic concepts, suggested the lack of what nowadays is called «linguistic material», and expanded the boundaries of the phenomenology of language until the «nature of things themselves, which is the same for all people and on which the whole language was originally created» (Newton, 1957). In the 21st century one quite understands the whole enormous complexity of Isaac Newton’s attempts to derive the system language from «nature of things themselves», but today we cannot fail to admire the intellectual athleticism of the young genius who made so a bold attempt more than 350 years ago.

On the other hand, this attempt can be described as the apotheosis of the phenomenological approach to the study of language. Because the nature of things «appears» to a subject or is «given» to him through his awareness of a set of events that eventually become facts of language. The methodological setting on the output of the language from the nature of things seems to be especially valuable, as it introduces the language into the general range of phenomena and entities as a manifestation of the general nature of things.

Of course, in determining the boundaries of linguistic phenomenology some moderation must be followed so that theoretical constructs can possibly end not only in the formulation of general principles, but also results useful for practical applications. And despite the fact that the self-esteem of many linguists may be flattered with the expression of the Apostle John «In the beginning was the Word,
and the Word was with God, and the Word was God»,\(^2\) that introduces original language tools into the arsenal of the Creator the reasonable beginnings of the nature of things in relation to language learning should, perhaps, be attributed to the later stages of the evolution of the world.

However, speaking about the evolution of language, we cannot forget about the creationist view of its nature. Here is how, somewhat ironically, describes the act of creation of language Umberto Eco by mouth of one of the main characters of the novel «The Name of the Rose» Honorable William of Baskerville:

«God had told Adam not to eat of the tree of good and evil, and that was divine law; but then He had authorized, or, rather, encouraged, Adam to give things names, and on that score He had allowed His terrestrial subject free rein. In fact, though some in our times say that nomina sunt consequentia rerum,\(^3\) the book of Genesis is actually quite explicit on this point: God brought all the animals unto Adam to see what he would call them: and whatsoever Adam called every living creature, that was the name thereof. And though surely the first man had been clever enough to call, in his Adamic language, every thing and animal according to its nature, nevertheless he was exercising a kind of sovereign right in imagining the name that in his opinion best corresponded to that nature. Because, in fact, it is now known that men impose different names to designate concepts, though only the concepts, signs of things, are the same for all. So that surely the word «nomen» comes from «nomos», that is to say «law», since nomina are given by men ad placitum, in other words by free and collective accord.» (Eco, 1983, p. 93).

Thus, we can assume that Isaac Newton attempted to do what had actually been done by Adam. In fact, according to William of Baskerville, Adam “in calling every thing and every animal in his Eden language, was guided by the nature of the called.” Although it is not entirely clear where the First Man knew this nature from (perhaps he was advised by the Creator), the methodological setting to be guided by the nature of things when naming them is quite clear. Similarly, it becomes clear why Newton, not having a sufficient knowledge of the nature of things by his age of eighteen, but still wanting to implement this knowledge in his theory of language, devoted most of his academic life to just knowing the nature of things.

Speaking of evolutionism and creationism in relation to language, we cannot fail to mention opinions on the subject of such authority as Wilhelm von Humboldt, who said: «Whatever natural may seem to us to assume a gradual formation of the language, it could appear just at once. <...> In order that people could understand even one single word, it all well and in all its relationships should be embedded in them» (Humboldt, 1999). Our own observations on the language system relations that occur in large glossaries (Shyrokov, 2011, pp. 260–264), confirm the Humboldt’s opinions, but today we are not prepared to give preference in the subject either to evolutionism or creationism.

\(^2\)The Gospel of St. John, 1:1

\(^3\)The names of the original essence of things (Latin).
In connection with the phenomenology of language we turn to the views of the brightest representative of the phenomenological trend in the 20th century philosophy, namely Edmund Husserl. In the third volume of his «Logical studies», 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 general and particularly a priori grammar is (because we discover the laws that determine the possible forms of meanings) an indubitable foundation and at least some definitely highlighted scope of its significance (Gültigkeit). To what extent can be detected even other areas of the proper grammatical Apriori, is beyond the scope of our interests. Inside the pure logic there is a sphere of laws abstracted from any objectivity. These laws, in contrast to the usual laws of logic in their strict and usual sense, one could justifiably call {pure logico-grammatical (reinlogisch-grammatische)}. {And it would be even better oppose the pure doctrine of significance (reine Geltungslehre) to the pure doctrine of form of meanings, preceded by the former.}

At first glance it seems that in this case phenomenologist Husserl opts in favor of apriorism. However, the past century, with its great discoveries in the field of logic (associated with the names of F. Frege, G. Cantor, B. Russell, A. N. Whitehead, D. Hilbert, K. Gödel, L. Wittgenstein, A. Tarski, L. Brouwer, L. Zadeh and others), as well as quantum-relativistic physics (A. Einstein, H. Poincaré, M. Planck, N. Bohr, L. de Broglie, W. Heisenberg, E. Schrödinger, P. Dirac, M. Bourne, J. von Neumann, R. Feynman and others) so expanded the concept of logic while bringing it closer to the subject and the substance that the combination of logic and grammar by Husserl already looks not the way it was seen, for example, by rationalist R. Descartes in the 17th century. So Husserl’s remarks, we think, may be interpreted in a phenomenological spirit, and «pure logico-grammatical (reinlogisch-grammatische) laws» by Husserl in its conceptual epistemology is quite comparable to the «nature of things» by Isaac Newton.

What are the common features of the «nature of things» that are reflected in linguistic forms? For an answer we turn to what we call the picture of the lingual world. L. Wittgenstein put it aphoristically: «The limits of my language define the limits of my world». Earlier W. von Humboldt in a letter to F. Wolf wrote in 1804: «I was able to open — and I am more and more concerned about it — that through language one can look around the highest and deepest areas and the diversity of the world.»

**Picture of the language world and its conceptual representation**

The picture of the lingual world can be clearly represented as follows in Fig. 1. In this scheme the external impressions of the world through perceptual-sensory apparatus of the human (the subject of the language-thinking process), are transformed and fall into his/her thought-speech apparatus to be there processed and acquire a linguistic form. The exchange of information expressed verbally with another subject of the thought-speech process is being carried out through the communication system, i.e. the infrastructure of language.

Even from such a lapidary scheme fairly certain conclusions about the structure of language can be derived. First, it is the conclusion on a relative stability of forms the linguistic structures acquire in evolution. Communicants (participants of the
communication process) must exchange information in the agreed data «formats» and «protocols», since without them the communication is impossible at all. It is clear that these tools of data exchange «standardization» certainly should not be modified in the communication process (or can change only slightly), what provides the necessary stability of forms of the linguistic structures. The second point concerns specificity and diversity of forms, in which the linguistic data elements are implemented in the circuits and paths of thought-speech and communication processes. As noted by Pitts and McCulloch: 4 “The language by which information is transmitted (in the brain) . . . cannot and should not meet the language that people use to communicate with each other.” This is understandable, since different tasks are being solved, in fact, in the language-thinking and the communication processes to require different tools for their solution. While the communication system’s function is only to transfer voice data without or only with a minimal distortion, the language-thinking apparatus solves much more diverse tasks: encoding, decoding, analysis, synthesis, transformation, interpretation, conceptualization, comparison, memorization, short- and long-term storage, etc. These processes occur in different environments and in different, so to speak, hardware and software implementations. We believe that the properties of information and peculiarities of information processes, in general, depend on the physical substrate in which they are implemented. The last statement finds its justification in quantum information, its research having been particularly active lately. 5

The first question that arises in the formulation of principles of modeling the language substance is the issue of objects to be modelled, namely: What are the objects of language and what, in fact, are we going to model? In this paper we take as a starting position that the language’s own objects are certain psychophysical states and processes in the language-thinking apparatus of the human, and both

---

4See: Prybram, 1975.
oral and written forms of language are the elements of the infrastructure of the lingual process. This statement is slightly different from the point of view usual for the traditional linguistics, according to which the substance of language is a sound substance that forms a lingual substratum to be studied and simulated. We proceed from the obvious fact that the language does not occur in the vocal cords — its origin is much deeper and the sound form already belongs to the infrastructure of the lingual, or more precisely, the language-thinking process.

Obviously, the language-thinking process itself is integrated and contains both linguistic and mental components. In the language-thinking apparatus it is realized as a dynamic system of interrelated reflexes. Their content and nature are studied, for example, in the book by V. M. Bekhterev published as early as in 1909 but having not lost since then its relevance, as well as the above-mentioned monograph by K. Prybram and a number of other works. According to the views of V. M. Bekhterev (1991) the natural language is one of the so-called connective reflexes that occur in the human brain. Thus, the separation of the lingual processes from the mental ones adopted by many linguists as well as attempts of learning language «by itself» seems to us an unreasonable and methodologically incorrect simplification. The language system should be regarded as an open one, providing both a significant expansion of the phenomenological basis of language, and the corresponding modification of conceptual tools.

Oral and written forms of the language in this sense act as models of speech-thinking processes and at the same time — of their communicative environment (infrastructure). With this factorization we can argue that they are the linguistic periphery.

We make, however, a reservation regarding a possible underestimation of the infrastructure components of language that may be due to the fact that psychophysical conditions and processes of speech-thinking system are considered by us to be the «main» primary lingual objects. The fact is that the current evidence suggests that knowledge of language and its proficiency are not innate human qualities. It is only the ability of language acquisition that is innate and for this purpose there are certain areas responsible for language in the human brain from birth. A process of «installation» of language in humans necessarily requires such infrastructure elements as the so-called «external» and «egocentric» speeches that function on the early stages of phylogenetic development of language in children and end with the formation of his «inner» speech to crown the process of creating a full speech apparatus (Vygotski, 1999). Thus, the language periphery is an inseparable element of the language system. In addition, it is one of the elements that provide its information openness.

It should be noted that the psychophysical conditions and processes (and the language-thinking one among them) are usually not fully accessible to direct observation, and certainly not to objective fixation. Oral and written forms of the language, in fact, are representatives of the observed components of states of language objects and processes that occur in the language-thinking apparatus. As such, they are used as basic objects in the conceptual modeling of language. This circumstance further emphasizes the importance of the language periphery in the study of language. Thus, in the picture of lingual world we can identify observable
and directly unobservable components of the language-thinking process as shown on the Fig. 2.

We are aware of this factorization to be conventional, because some elements of the language-thinking processes nowadays are observable (but only some of them). However, in what kinds of form, shape, formats the speech data are stored, what is the model structure and the types of these data — all of this still cannot be observed directly, and it is unknown if it is possible in principle.

In this regard, we turn to the analysis of the idea or principle of observability. This principle laid as the basis for A. Einstein’s special theory of relativity and developed in some detail in the justification of quantum mechanics, has been very productive as a method in describing complex processes. As A. Einstein noted in his discussion with W. Heisenberg in 1926 when disputing about the foundations of quantum mechanics, a theory cannot be constructed with observables alone (although without them any scientific theory is impossible at all). Only the theory itself must determine which of its variables are directly unobservable, and which are not. In particular, in quantum mechanics the states, generally speaking, are not directly observable. The latter ones match operators of a certain type that act in the space of states, their eigenvalues do represent the observed values.

Thus, in the world and therefore in the knowledge there is «invisible» that cannot be observed directly. This gives us the basis for recognition of the role of faith in the cognitive and intellectual processes, which in the words of the Apostle Paul is «confidence in the unseen». It should, however, be recognized that some deeper epistemological reasons of observability–unobservability are not disclosed in the

---

6This discussion is described in detail by W. Heisenberg in his book «Part and the whole»: (Geizenberg, 1989, pp. 191–196).
7«Now faith is the assurance of things hoped for, the conviction of things not seen». The Letter to the Hebrews, 11:1.
Volodymyr Shyrokov & Ihor Shevchenko

quantum theory. In our opinion, these reasons are rooted in the phenomenology of complexity of the manifested existence. Any reality being an object of observation (in other words, entering into a relationship «subject–object») exerts its essence in some limited forms accessible for perception by the subject. Shown by the complexity of these forms (potentially infinite — as matter has no limits of divisibility, so to speak, «in depth» and does not allow control of all of its relationships and mediations) appears compensated, consistent with the «perceptual–sensory» devices subject or device (which, in fact, one and the same), which accepts these forms. The mechanism of self-compensation of complexity seems to us as so a universal «universal» of the relation «subject–object» that it should perhaps be attributed to the basic principles of system analysis. Note that here we understand the complexity as a positive scientific category for which some precise and even formal definitions can be formulated like those that occur in the theory of Kolmogorov (1987, pp. 213–223) and in numerous applications and generalizations of this theory, where the concept of complexity is manifested as directly related to the concept of information.

The first idea of linguistic states and the associated notion of grammatical observed

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 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 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 case in Russian.

A. N. Kolmogorov did not publish their papers in linguistics, so in our exposition of his ideas we follow V. A. Uspensky (1957, pp. 11–18), as in this paper they were originally presented. Let us turn to the Kolmogorov–Uspensky approach concerning the definition of case that we present below with our comments. We give the V. Uspensky’s text in italics while our comments are numbered and presented in Roman type.

Thus, Kolmogorov proposed a definition of case as follows.

§1

Objects can be in different states. Thus, the object named «молодёжь» («youth» — editor’s note) 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 sentences which involve a noun that is the name of this object. The above states

In this regard, recall the famous saying of Einstein: “God is quaint, but not evil,” which emphasized that at each stage of the knowledge the world manifests itself in finite forms of complexity, which make knowledge possible.
On the Phenomenological Approach to Grammar

of milk are expressed in Russian sentences such as «молоко кипит», «молока нет», «кошка пьёт молоко», «собака пьёт молоко». 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). In fact, that is the contexts of the said kind that 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 «The picture of the world of language 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 sum of all the reasonable contexts. This implies that (so far implicitly) the existence of a correspondence between a word and its state: \( s : s \rightarrow s(x) \) is postulated where \( x \) — a certain word; \( s \) — correspondence between \( x \) and \( s(x) \), \( s(x) \) — 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 in some context. So, 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 the class \( X \) of words of the corresponding language. The word \( x \in X \) hereafter (unless stated otherwise) is interpreted by us as a set of all conceivable forms of the 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 are denoted with superscripts. There are also states of more a general form that meet 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 in one and the 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». These states are not equivalent on the object of «вода», since «вода кипит» but «кошка пьёт воду». 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 sub-classes 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. Thus, A. N. Kolmogorov offered to call these classes as «cases». It is a natural desire of 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 one and 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 case that took place in the workshop and its sidelines (meaning, in particular, the ideas expressed by R. L. Dobrushin and I. A. Mel'chuk), 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 a space» (shortly — merely a «set»).

For example,
are five different sets of words with a space. When substituting a word (in some form) into a set of words with a 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 being 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).

Two sets are called 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, in order 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 по по .......» and «мальчик идёт по .......» are not equivalent in respect of the noun «берег» and sets «......строим 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 «......бежал» are directly equivalent sets. Finally, we say two sets P and Q are completely equivalent if there exists a chain of sets $X_1, X_2, \ldots, X_n$, 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 a 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 to call these classes as 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 form’s concurrence in the states 1 and 2. Therefore, it is necessary to first clarify the 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 a 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 obviously yet but clearly enough as we see it. It is, in fact, the recognition that some states are not sufficient to determine the grammatical features, such as case. Hence, some additional structures, such as sets of strings with spaces are required. Now 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 what necessitates further clarification.

As the V. A. Uspensky notes the sets

«я вижу синий...» and «синий ... стоят» 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. For example, «не читал газету» 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 say that 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: in order 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 $B$. For example, if state $A$ is that the appropriate object is building a house, then this condition is coherent both to the set of «... строит дом», and the set «дом строится ...». If the 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. The 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 a 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 the 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 say that 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 $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ which

1) for each pair $(X_i, Y_i)$ and $(X_{i+1}, Y_{i+1})$ 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 equiv-
alent. It is these classes that are offered to call cases. Fully aware of inconclusiveness of the newly formulated definition of 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 a 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 a space». Thus, at 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. It is those classes that the author proposes to identify 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 B1 = «......building a house» and set B2 = «house built ......» are coherent with this state. The 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 built by the appropriate object» is not very clear either. 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. It is clear that he was aware of some insufficiency of this definition, noting that «... Fully aware of inconclusiveness of the newly formulated definition of case, 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 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 a 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).

For example, 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» (Rus. «professors» — editor’s note) 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 traditional six cases there are still the following cases:
1. Local case (in Russian “местный”, as opposed to the traditional “предложный” — author’s note) «в лесу», «в году» etc.
2. Quantitative-ablative case «выпить чая», «прибавить ходу», «дать воды» and so on. 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 («deprivative»),
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.

It is possible that there are other cases. It would be interesting to enumerate all the cases of Russian.

Comment 5. If we ignore a number of 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 $s(x), x \in X$, is introduced. It does not have, however, a clear definition, but operates on the set $K(X)$ in a well-defined way described above. The result 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:

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

where $N$ — number of cases in that language, so contexts $k_i^x(x)$ and $k_i^y(y)$, 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 case, appears as an operator (call it the operator of 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 a number of mutually disjoint subsets. Thus, 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 a 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 $\xi$ from $k(\xi)$, it is logical to assume that in such a way the «case» part of its full linguistic state $s(\xi)$ 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 \(\xi\) 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», a Russian ambiguous phrase meaning «day overcomes night» and «day is being overcome by night»); 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 are able to 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 the 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 Andrei Nikolayevich may have not set out as a special goal to define the concept of 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 in accordance of the linguistic state to a language unit) is, in our opinion, the most fundamental and valuable in his work. Indeed, the concept of 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 we see it, 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). A. A. Zalizniak alone perceived the Kolmogorov–Uspensky ideology and used it in his excellent book «The Russian nominal inflection» (hereinafter abbreviated use RNI), published in 1967 (Zalizniak, 1967). The reprint of this book in 2002 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» (Zalizniak, 1973), where the author develops the concept of case system, indicating, in particular, the place the Kolmogorov–Uspensky approach occupies in it. Here are some snippets of this article.\(^9\)

\(^9\)The reader easily distinguishes the Zaliznyak's text typed in italics of our comments to it.
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. The 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 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 of meaning units. We can describe as «narrowly formal» such understanding of 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.

The 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 a number of word forms, including, among many others, the word-forms «пером», «человеком», «стрелой», «лесом», «мыслью», «перьями», «детьми» and so on. Each of them is able to express a series of cases in the semantic sense (meaning of an instrument, doer, reference standard, traffic areas, etc., cf. above). Thus, the cases in the narrowly formal sense that is one of possible means (along with prepositions, postpositions, word order, etc.) to express the cases in the semantic sense.

Note: Here A. A. Zalizniak believes that the semantic understanding of 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 is clearly demonstrated in the titles traditional for the Latin grammars: genetivus possessivus, genetivus partitivus, genetivus subjectivus, genetivus objectivus, dativus commodi, dativus possessivus, dativus finals etc.: here the first part of the name corresponds to the case in the narrowly formal sense, and the second — in case in the semantic sense.» Thus, within the «narrow formal», for example, genetivus as many as four «semantic» cases are contained: genetivus possessivus, genetivus partitivus, genetivus subjectivus, genetivus objectivus.

In 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 is
no cases at all. Obviously, says A. A. Zalizniak, that the cases in the formal (and in particular, the narrowly formal) sense are not of a universal character, 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 later, 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 that 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.

With regard to 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 «semantic» role that matches in its content to what was 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» (Rus. «He sent his friend’s book to his parents» — editor’s note). We say that the segment of «родителям» expresses in this phrase a semantic role of «recipient of action» for the nominative «родители»; segment «книгу» in this phrase expresses the semantic role of «object of action» for the nominative «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 nominative «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 nominatives) 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 a 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 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 A. N. Kolmogorov. However, 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 between 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, may be 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 case should match a certain operator acting in the space of grammatical states so that its eigenvalues are the values of specific cases. Actually, 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. So, components of the linguistic state actually guessed 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.

First, it is the invariance of 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\textsuperscript{10} when the level corresponding to a particular eigenvalue being split to some «sublevels» with the inclusion of an additional «semantic interaction» or, equivalently, considering the semantic factors not taken into account earlier. For instance, the value of Latin cases given by A. A. Zalizniak \textlt<genetivus>; \textlt<dativus>; ... are interpreted as «eigenvalues» of the operator of «the narrowly formal case», while \textlt<genetivus possessivus>, \textlt<genetivus partitivus>, \textlt<genetivus subjectivus>, \textlt<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 «degeneracy» is removed when taking into account more a subtle semantic factor defined by a «semantic case» and the original case state is split, showing a «subtle semantic structure» that on the above A. A. Zalizniak’s example can be

\textsuperscript{10}Fine structure of quantum mechanics.
schematically shown in Latin as follows (see Fig. 3):

Figure 3: «Fine semantic structure» of the Latin case (genetivus)

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 play 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 (Plungian, 2011, p. 183, footnote 69) saying that the «semantic role» is analogous to a «more impressionistic» concept of the «state» still 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 own 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 case, looks almost trivial due to the fact that the said role corresponding to case in the semantic sense, already is, so to say, «inside» of the fine structure of case and detected in the context of case in the semantic sense, as it follows from the scheme shown in Fig. 4.

It brings us very closely to the definition of case class of equivalence on the set of coherent pairs (situations), and therefore to the definition of case in the narrowly formal sense.

Thus, in the Kolmogorov’s theory (and therefore in the Zalizniak procedure) the states should occupy a place as indicated in the following diagram below (Fig. 5).

The employment of the concept of the state of language unit was further developed in our works and the works of our fellows (Potapova, 2012; Shyrokov, 1999, 2004, 2005, 2009, 2012). According to them any word (to say generally — any language unit) in the context or in 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.
Further formalization of linguistic states

The process of language understanding can be presented as a reduction of the a priori distribution of lexemes by gross signs of grammatical and lexical semantics inherent in the subjective lexicon of the recipient to a certain grammatical and lexical meaning specific to that very context, which is in the field of attention of the recipient and subject in this point to the process of his individual language processing. Thus, the above scheme can be generalized and detailed as follows (see Fig. 5):

Although 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 naturally 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 (Boum, 1990; Landau & Lifshits, 1989). Thus, when considering the formal aspects of semantics we shall proceed with the existence of a correspondence between the language unit and its state:

\[ s : X \rightarrow s(X) \]  

where \( X \) — is a certain language unit; \( s \) — correspondence between \( X \) and \( s(X) \) — a formal object that represents the state of the unit \( X \) with elements of the material means 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: \( f1 \) — «noun», \( f2 \) — «verb», \( f3 \) — «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. It is clear that the definition of operators of type \( F \) must be based on some semantic theory formalized properly.

Thus, when considering the formal aspects of the semantics we proceed with the existence of a correspondence between linguistic unit and its state. Using the mathematical language the action of the operator \( F \) can be expressed as following:

\[
Fs_i(X) = f_is_i(X)
\]  

(3)

where \( f_i \) — specific value of semantic categories, functions \( s_i(X), i = 1,2,\ldots \), represents the states of the unit \( X \) that mark its affiliation with the values of \( f_i \) of the category \( F \). The values of \( f_i, i = 1,2,\ldots \), are called the eigenvalues of the operator \( F \), that correspond to states \( s_i(X), i = 1,2,\ldots \). 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 \( S(f_i) \):

\[
S(f_i) := \{s: Fs = f_is_i\}
\]  

(4)

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, and in another — 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 (3) there is not one semantic condition $s_i(X)$ but several that correspond to a value of $f_i : s'_j(X), j = 1, 2, \ldots$. We call such states as degenerate, the presence of degeneracy is signaled by the superscript character in the semantic state $s'_j(X)$. 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 $g(X)$:

- feminine singular, nominative,
- masculine plural, nominative,

where the word «мати» has a lexical meaning «woman in relation to the child that she gave birth» and:

- feminine singular, nominative,
- masculine plural, nominative,

where the word «мати» has a lexical meaning “sport mattresses.” This example illustrates the phenomenon of homonymy inside the same part of speech.

We designate the number of own semantic states $X$ corresponding to a value of the category $F$ as multiplicity of degeneracy of the semantic state. Thus, in the above example, the state $g$ (мати) has a multiplicity of degeneracy that equals to 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 value. Formally, this situation can be represented by means of the relation:

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

where $s$ — 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 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 (5) defines the semantic state in which the unit $X$ is and which is a peculiar phenomenon of superposition («mixtures») of net semantic classes $s_1(X)$ and $s_2(X)$ corresponding to the eigenvalues $f_1$ and $f_2$, respectively. Linguistic interpretation of equation (5) 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 lexems «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 above morphological character does not look strong in terms of belonging to the verb, since in the verbal paradigm it is not inflectional. In addition, it is typical for the full inflectional paradigm, not only for its individual 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 «деприкметник» (participle) reflects both features — those of the verb and the adjective.\textsuperscript{11}

Notes on the concept of state and phenomenological complexity of the structure of the world

In this article we do not intend to consider a lot of examples that illustrate the capabilities of the formalism of states of linguistic units. Several applications of this formalism have been demonstrated in other works cited above. The purpose of this paper is to show a phenomenological character of grammar of natural language and hence the proximity of the science of language to other disciplines of science, and also those of the technological circle. We experience the urgency of this problem in connection with the linguistic load falling on the knowledge engineering that takes on features of a global industrial sector and requires ever more sophisticated intelligent linguistic tools for its development.

In connection with the above we consider it necessary to give some general scientific discourse on the concept of state. This concept is used in many natural, sociological and technical disciplines. We believe it most deeply developed theoretically and practically in quantum mechanics where it is a fundamental.

According to the canonical doctrine of quantum mechanics, any system at a given time is in a particular state. The system status is formalized as a solution of the Schrödinger equation for that system. Since the Schrödinger equation is differential equation of a certain type in partial derivatives the set of its solutions identified with the states of the system under consideration forms an infinite-dimension Hilbert space. Thus, the number of states of a quantum mechanical system is theoretically infinite. System status is in theory its most complete description and defines a probabilistic interpretation of the system’s behavior, but by itself it is, generally speaking, is not directly observable. Observed values are represented in quantum mechanics by Hermitian operators acting in a Hilbert space of states, and the possible values of the observed quantities are expressed through the matrix elements of the operators in the space of states. However in other theories the states of a system can be observable. For example, in classical mechanics the state of a material point is defined as a coordinate-momentum pair at a time \((x(t), y(t))\), and they are observable, both separately and together. In quantum mechanics there is however a fundamental restriction on the simultaneous measurement of position and momentum, which is given by the Heisenberg uncertainty correlation.

\textsuperscript{11}An example of superposition of grammatical classes in Turkish language is given in the book by (K. V. Shyrokov, 2009). It demonstrates that some inflectional paradigm of the Turkish noun has certain categorical and material features of the verb.
Thus, the concept and the status of observable values are not invariant and are defined diversely in different natural (and other) theories. This adds some spice to the use of the notion of state, for instance in the theory of calculus, which in its present form, in our opinion, ignores the phenomenon of observability–unobservability.\footnote{However, due to the quantum generalizations of formal structures, such as Petri nets (Abramsky, 2008) or neuronic nets (Lagaris, Likas, & Fotiadis, 1997) concept of observability–unobservability is likely to be introduced into the theory of objects used to describe information technology processes and systems.}

One would demand that the theory operate with observables only, but this issue is not simple. It was widely debated at the time of the formation of the quantum theory and has not lost relevance in our time. Theoretical knowledge achieved in the domain contains methodological lessons and principles so general that they may and should be used in any science that has ambitions of theoretical understanding of nature of the things it studies.

The first and foremost is probably the fact that for the characterization of states of objects both observable and directly unobservable quantities are actually used. Moreover, in the opinion of most scholars to build a theory with observables alone is impossible (recall the Einstein’s statement on the matter). But it is clear that without observables a scientific theory or science in general is basically unthinkable. Quantities observable and directly unobservable must have different logical and ontological status, but, as far as we know, the general theory of this issue is not developed yet in detail.

In light of this interpretation a relationship between the values observable and directly unobservable in a theory is suggested: they are, respectively, «formal» and «substantive» aspects of the object under observation (in the terminology of the theory of lexicographic systems (L-systems)\footnote{The most comprehensive to date lexicographical presentation of the theory of systems and its applications is given in the book “Computer lexicography” (Shyrokov, 2011).} register and interpretative parts of some hypothetical L-systems developed as a result of the manifestation of a lexicographical effect). When applied to objects such as language the interpretation can be detailed in the sense that any linguistic unit has a «two-way» nature and its state admits a decomposition to a formal component (accessible for direct perception by the subject — whether it be a sound or a picture image), while the substantive component is represented by a set of «all the contexts» in which such a linguistic unit can operate when in the proper state — this, in fact, makes the specified part of the given state unobservable. A scientific debate about the logical and psychological principles of the phenomenon of observability should mention such a philosophical setting as the Mach principle, according to which sensory impressions are arranged in human thought in a manner that provides the most economical layout of these experiences into stable complexes.

Characteristically, A. Einstein believing this principle to be too trivial for acting as a universal epistemological law noted the special role of language in ontological and psychological evolution of the process of cognition (Heisenberg, 2007). He considered the language constructs to be not only a way of fixing sensible complexes, but also reflecting the reality that exists (or just can exist) outside of these
complexes and not bound to them — recall in this connection the Newtonian «nature of things». We believe that Einstein’s observations on the role of language (note, he was extremely sensitive to the issues of the philosophy of knowledge) are not casual at all — they emphasize the idea of the universality of information and cultural processes at all levels of knowledge of reality.

Further notes concern the discussion of the criteria of simplicity of the scientific theory — it is not necessarily associated with the Mach principle. The simplicity of a scientific theory is for most researchers rather a question of aesthetics — simplicity and beauty of mathematical schemes, prompted by the nature, make them more convincing. Note that in the era when the quantum theory was formulated the notion of simplicity (and the antonymous, and hence related, concept of complexity) belonged to the common language, because then there was no formulated theory of complexity — it is known to have emerged thanks to the works by A. N. Kolmogorov only in the 1950s (see: Kolmogorov, 1968, 1987 and references in this book). Nor was then clarified the communication of complexity of the objects and their descriptions (and therefore — their simplicity!) with the concept of information. Quantitative measures to assess these variables and the relationships between them were also unknown. So, the mentioned above notion of complexity developed by Kolmogorov and other scientists, its connection with information aspects of the description of reality and the concept of information and its quantitative characteristics, are deeply connected to the criteria of simplicity and beauty of the scientific theory. Minimality of description of an object under research, which is according to A. N. Kolmogorov is an objective measure of the amount of information about the object, encourages scientists (at least as a subconscious incentive) to find a description of this very type, although it does not specify ways or gives recipes, because, to say generally, it belongs to the class of algorithmically unsolvable problems. However, the lack of ways and recipes does not deny the objective existence of a minimum description — it only proves the nonexistence of a formula or an algorithm to generate new scientific truths. And if such a description is found, it obviously has to look as simple as possible — in essence, it is really so. Thus, the criterion of simplicity (or beauty) of the scientific theory, in our opinion, is not so much a consequence of the principle of economy of thought (which A. Einstein describes as «suspiciously commercial» and that, in fact, is only very indirectly related to the matter, because it is rather about a fundamental information property of objectively existing things than the property of thinking as a subjective process), as follows from the general nature of information, and corresponds to the formal definition of measure of its quantity by A. N. Kolmogorov.

Indeed, when a description of the investigated object (process, system, etc.) that most appropriately meets its essence is obtained, this description should be minimal, since it offers only essential information about the object and does not contain descriptions of casual, minor details that «clutter» the significant with «unnecessary» items. The scientist, so to speak, instinctively seeks just such a description of an object, which is consistent with the Kolmogorov’s definition of information measure based on the minimal description. This, in our opinion, explains the psychological confidence the researcher feels when he succeeds in getting a simple (beautiful!) formula, equation, conclusion, etc.
The formalism of the theory of complexity is both clear and deep. It should be taken as an ontologically objective property of things. One of the non-trivial manifestations of the above feature is that the complexity of the composite formation, in general, is not equal to the sum of the complexities of entities that form it. To put it more precisely, the complexity is not an additive function of the system. In other words, if there is any system that includes other «smaller» sub-systems, which are its constituents, i.e. if:

$$D = \bigcup_{i} D_i$$

(6)

where the symbol $D$ indicates the proposed system, and $D_i$ — its components, then:

$$K(D) \neq \sum_{i} K(D_i),$$

(7)

where $K(D)$ is a quantitative measure of the complexity of $D$, and $K(D_i)$, respectively, quantitative measures of the complexity of its constituents $D_i$ (usually $K(D_i) < \sum K(D_i)$). These ideas, of course, apply to some $K(D_i)$, as well as their components.

In the process of formation, functioning and interaction of composite systems a phenomenon takes place that we qualify as a «self-compensation of complexity». The content of this phenomenon reduces to the following. The nature of the interaction between constituents that form a unity (a wholeness), which is identified as a composite object is such that in a «bound» state they exhibit only a portion of their total «inherent» complexity. The necessity of such behavior can be interpreted as providing a fundamental possibility of knowledge of the «manifested» being, and perhaps even its very existence. Otherwise, the complexity of any object would be actually infinite (potentially it is), while in our case the complexities of individual components look like compensating each other in the formation of the whole. So one could argue that the potential complexity of any thing is infinite, because now we do not see the boundaries of the divisibility of matter and each lower structural level has a non-zero complexity. But the kinds of component complexity do not manifest all at once but only «level by level». Therefore, the complexity in each case undergoes to «renormalization» if you take an analogy with quantum electrodynamics, where for the elimination of differences you also have to apply the procedure of «subtracting the infinities». That is the language that gives us a prime example of self-compensation of complexity. For example, we can regard the length of a particular dictionary entry, which takes into account the effects of grammatical and lexical semantics including the multiplicity of grammatical meanings, the lexical polysemy, the phraseological structure of lexemes and so on as a degree of complexity of the corresponding word. Meanwhile, the word in a sentence (a specific context) operates only in a certain sense — one meaning or a kind of «mixture» of several possible meanings for polysemical lexemes and therefore the measure of its complexity in a particular context is determined by only a part of the dictionary entry, and in some cases it may be only a few tenths or even hundredths of the full complexity of the lexeme. Thus, the complexity of
a sentence may prove to be less than the full complexity of a single word which is an integral part of the former.

The design of the being proves to be paradoxical! It turns out that complex things are actually made up of even more complex ones. In this sense, «more» is less than «less». In our opinion, the well-known effect that has both ontological and psychological dimensions is a nontrivial proof of this thesis — it relates to the complexity of scientific theories: the theory of atoms, for example, does not look simpler than the theory of molecules, theory of nuclei does not seem simpler than the atomic theory, theory of elementary particles is not simpler than the theory of nuclei, and so on. In linguistics, for instance, the theory of words («lexicology») also is not simpler than the theory of sentence («syntax»). In the light of the above the principle of reductionism, according to which complex things should be composed of simpler ones looks not only obvious, but even doubtful that leads to some revision of the standard systems analysis, taking into account the effects described by the complexity theory. At this level, the latter begins to acquire features and the status of a natural-scientific and a general scientific doctrine, rather than a purely mathematical one.

Finally, we add that our observations on the phenomenological aspects of grammar convince us that the language and the science of language have far more in common with natural phenomena and natural sciences than is assumed in conventional linguistics. We would like to express our conviction that in the future this commonality will acquire much more defined features.

**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:

\[ F_s(X) = f_1\alpha_1(s_1)s_1(X) + f_2\alpha_2(s_2)s_2(X) + \ldots, \quad (8) \]

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 a disambiguation, the right-hand side of this equation is reduced to one member that, in fact, represents the net grammatical status of the given word \( X \) in a particular context. A completely different situation takes place when \( X \) is in a mixed state — then no linguistic processing is able to reduce the number of members of equation (8), 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 language system in full, rather, an evidence of the limits of its formalization. This particularly means that even the most suitable for formalization and, indeed, the most formal sector of the theory of language, namely grammar, has features of fuzziness what 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 $S(f_i) = \{s : Fs = f_is_i\}$. 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:

$$S = \bigcup_i S(f_i) ; S(f_i) \cap S(f_j) = \emptyset \text{ when } i \neq j$$

(9)

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 $S(f_i)$, but also two, and possibly more. A formal mechanism for describing such a phenomena operates as follows. Define on the set $S = \bigcup_i S(f_i)$ 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 $[0, 1]$:

$$\alpha_i(s) \in [0, 1]$$

(10)

Thus, 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:

$$\alpha_i(s) + \alpha_j(s') = 1$$

(11)

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

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

(12)

Pairs $(s(f_i), \alpha_i(s))$ 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 (12) under condition (11). 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 as following:

$$s(X) = \alpha_1(s_1)s_1(X) + \alpha_2(s_2)s_2(X)$$

(13)

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, indicates the degree to which $X$ shows the property of the value $f_2$. Condition (11), 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.
References


Acknowledgment

This work was supported by a core funding for statutory activities from the National Academy of Sciences of Ukraine.

The authors declare that they have no competing interests.

The authors’ contribution was as follows: concept of the study: first author; data analyses: second author; the writing: first and second author.

This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 PL License (http://creativecommons.org/licenses/by/3.0/pl/), which permits redistribution, commercial and non-commercial, provided that the article is properly cited.

© The Authors 2015

Publisher: Institute of Slavic Studies, PAS, University of Silesia & The Slavic Foundation