

THE THESAURUS CONCEPTION AS A NEW VIEW ON THE EPISTEMOLOGICAL FUNCTION OF THE LANGUAGE

[Svetlana OSOKINA](#)

Associate Professor, Ph.D.

(Altai State University, Russian Federation)

osokina@filo.asu.ru, <https://orcid.org/0000-0002-9480-1364>

Abstract

The article analyzes the leading linguistic conceptions which study the epistemological function of the language and offers some perspectives of research in this direction. The main perspective is the development of the conception of the thesaurus as the language system of knowledge.

Keywords: thesaurus, conception, function, language, system

Rezumat

În articol, se analizează concepțiile lingvistice, elaborate cu referire la funcția epistemologică a limbajului și se trasează, totodată, direcții noi în abordarea acestui fenomen, de felul celor care poartă asupra tezaurului, luat drept sistem verbal de materializare a cunoștințelor.

Cuvinte-cheie: teaur, concepție, funcție, limbă, sistem

The epistemological function of the language has been under linguists' consideration for a long time. Nowadays scientists are trying to understand the rules of getting knowledge and ways of its keeping in the human head. In order to describe different aspects of the cognitive sphere several scientific terms such as "the concept", "the picture of the world", "the mental lexicon" and "the information thesaurus" have been invented. Each term refers to a certain linguistic conception. All the conceptions have contributed a lot to the advancement of linguistics; still there are some points for criticism. The aim of the paper is to analyze the given theories and to suggest a new way to research and describe the knowledge structures of the language. The paper supports the thesaurus conception as a heuristic method of thinking in this direction.

The development of the mentioned conceptions is connected with the recent intensification of the cognitive studies. The presentation of knowledge in the human head and its functioning as a system is based on the language material and defined with the help of linguistic categories. The language is almost always nothing else than the means of studying cognitive structures. This point of view underestimates the essential role of the language in forming knowledge structures. The result is that a number of very important language problems remain hidden from the eyes. To overcome this situation we need to find a new view of the epistemological nature of the language.

The system of knowledge which is formed with the help of language is called in different ways. The most recognized terms of this system are "the system of concepts", "the picture of the world", "the mental lexicon" and "the information thesaurus". The entities named by these terms are so similar that scientists have to compare them and to use each of them to explain the meaning of the others.

The development of "the system of concepts" is connected with logical studies of the language in the middle of the 20th century. In consideration of this notion most Russian linguists relied on R. I. Pavilenis works. In his conception the system of knowledge is the system of concepts, or "the system of information about the world" (Павиленис/Pavilenis, 1983, p. 101). Unlike the system of language, this system is continuous and indiscrete by nature. Its development starts on the nonverbal stage of cognition and presents a necessary condition of language acquisition.

The metaphor "the picture of the world", having a long history of its consideration, is still understood by linguists in different ways. In general, the picture of the world is compared to the system of concepts as a relatively stable and more definitely structured system because it is formed with the help of language.

There are some theories which deny the role of language in building an individual picture of the world. For instance, G. Kolshanskij writes, "The language comes out not as a demiurge of this 'picture', but as a form of expression of its conceptual... knowledge obtained by a human" (КОЛШАНСКИЙ/Kolshanskij, 1990, p. 25). Though, recently linguists have supported the idea, suggested by E. Sapir and B. Whorf (and before them by W. von Humboldt) that language structures predetermine knowledge structures.

The opposite conceptions are reconciled by the differentiation between "the conceptual picture of the world" (this is often meant by the term "the picture of the world") and "the language picture of the world". The latter provides the connection between the mental system of knowledge and the outer world because the language is the basic means of objectification of the mental picture of the world.

Paying too much attention to the signified side of language signs consistently leads to the problem of language meanings. Particularly, lots of works are devoted to the relation between the lexical and grammatical meanings. Recently linguists have made the conclusion that the lexical meaning plays the same role in forming the system of knowledge as the grammatical meaning, while traditional linguistics considered the lexical meaning as being more important in this respect. The idea of the interpenetration of the lexical and grammatical meanings is underlined in the theory of mental lexicon. A. Zalevskaya, who has developed this theory since 1980-s, defines mental lexicon as "the system of codes and code transitions that provides forming

and transmission of the sense as well as extraction of the sense from the perceived message" (Залевская/Zalevskaya, 1990, p. 66).

Nowadays we can see the development of a new conception interpreting the human system of knowledge. That is the conception of the information thesaurus. Correspondingly, the mentioned conceptions belong to logical, cognitive and psycholinguistic branches of linguistics. The conception of the information thesaurus has its roots in natural science such as information theory and informatics, or computer science. The latter pays much attention to the exploration of the information content. It researches symbols, i.e. two-sided signs. The representatives of informatics apply semiotic conceptions of the sign structure in order to give adequate evaluation of the information content. This model is based on the theory of information developed by C. Shannon and the general theory of systems.

Thesaurus is viewed as a form of information existence and storage. Yuri Shrejder introduces the notion of the semantic information and develops the conception of information thesaurus. He defines "thesaurus" as the formal model of description of the observer's presentation of the world. "A complex system has semiotic (i.e. language) nature of information relations between subsystems in opposition to systems with functional signaling" (Шрейдер и Шаров/Shrejder & Sharov, 1982, p. 16).

The idea that the system of information exists in three aspects (semantic, syntagmatic and paradigmatic) is also taken from semiotic conceptions. The semantic information transferred by the complex system is equal to the pragmatic information because regardless the quantity of information which was given, the receiving system will accept only the information which is necessary and valuable.

So, thesaurus as the semantic model of information transmission has to solve the problems of understanding information. In Shannon's theory of information the problem of understanding information by the receiver is not posed at all. It is assumed that the receiver is always properly turned up. But in reality it is not so.

Thesaurus is a semantic system which works to receive, interpret and adopt information. The same message can be taken differently by different receivers. To describe this situation it is useful to have an idea about the thesaurus of the receiver. The ability of the receiver to accept a piece of information and to interpret it depends on the state of the thesaurus.

This idea is frequently repeated in other works by other authors. For instance, N. Chursin asks a question, "Why don't academicians teach first graders? ", "Because their thesauruses are not compatible" (Чурсин/Chursin, 1982, p. 59). An academician's thesaurus is far more complicated than a schoolboy thesaurus. Out of the information transferred by the academician (out of all words said by the academician) a schoolboy can grasp only a few

things (only a few words will be familiar to him, and it does not mean that he will understand them properly in the context used by the academician).

The thesaurus organizes the system of semantic links between concepts. Each concept inside this system is defined through other concepts. In fact, the thesaurus of an individual is a verbalized set of an individual presentations of the world, including his cognition intentions. Cognition intentions define the way the given information is accepted.

Only that information is accepted which is compatible with the thesaurus of an individual and a little exceeds its present state. The exceeding information is valuable information. The exchange of information between semiotic systems happens as the exchange of valuable information. This means that the information exchange is a teleological process. If the receiving information is absolutely equal to the present state of the thesaurus then the quantity of this information is equal to zero.

In this conception, the information defined as "the quantity of information" is valuable information. This idea is absolutely different from Shannon's idea of "the quantity of information". Shannon viewed the quantity of information mathematically; in the thesaurus conception the quantity of information can be measured as the change of the thesaurus of the receiving system. "After reading a text, the thesaurus of an individual may change. The change of the thesaurus can be written algebraically: equality $b=x(a)$ means the subject with the thesaurus a , receiving text x , changes his thesaurus, turning it into b " (Шрейдер и Шаров/Shrejder & Sharov, 1982, p. 119).

The problem is that Y. Shrejder does not explain how to measure the content side of the thesaurus. He only suggests that it can be measured in some way, but he does not offer any mathematical formula for the calculation of the quantity of semantic information as Shannon's formula for the calculation of the quantity of material information.

Yuri Shrejder develops the conception of the semantic information using mainly mathematical methods. Since semantics is a branch of linguistics, the way to measure the thesaurus changes can be found with the help of linguistic methods.

In linguistics the notion of thesaurus is mostly connected with ideographic dictionaries. But Yuri Karaulov, the author of the conception of "language personality", names the cognitive level of the language personality as "the thesaurus level". It is the system of knowledge of the personality formed by the language (Караулов/Karaulov, 2007).

Karaulov's view on the thesaurus is different. The researcher gives another interpretation of the epistemological function of the language. It belongs to the branch of cognitive linguistics to a wide extend. As is the case with other cognitive studies, it is aimed at explaining mental structures with the help of the language. These conceptions try to model the mental system of

knowledge using indirect evidence, such as language facts. The main problem is to define the unit of knowledge. Language facts do not allow to set its borders properly.

We think that the modeled mental structures are epistemologically derived from language structures. Although scientists say that mental units exist before language as mental abilities of an individual precede language abilities, knowledge about mental structures comes from knowledge about the language. So, from epistemological point of view, models of mental structures are second after models of language structures. That is why to understand the essence of knowledge we must pay more attention to language epistemological structures. We must search for knowledge structures inside the language, not in the human head only.

We suppose the conception of the information thesaurus has many more advantages than other conceptions. The term "thesaurus" is not so frequently used in cognitive linguistics, so, it can find its own status in the branch of linguistics dealing with language (not mental!) structures of knowledge.

Speaking of the epistemological activity of an individual, we can say thesaurus is a system of words used by an individual not only to get and produce verbal information, but to exchange knowledge with other individuals. It orients the individual in the space of information.

The orienting function of the thesaurus is stressed in a number of sociological works. Particularly, Valeriy Lukov and Vladimir Lukov define thesaurus as "an individual configuration of orienting information" (Лыков и Лукон/Лukov & Lukov, 2005, p. 15). The thesaurus has the orienting function because an individual does not create semantic links between words, but inherits them as already formed system from previous generations. The system is formed in the process of constant verbal communication between people and presses human perceptive organs. An individual has to accept this system, there is no other choice.

Empirical forms of this system can be described as stereotype collocations of words which compose human speech. Hearing them from others, an individual repeats them. The unit of the thesaurus in this respect cannot be a word. It must be a collocation, i.e. a combination of words, because only collocations show explicit word links. A lexeme contains its links in potential but, on the one hand, not all of them are revealed in reality; on the other hand, collocations may present the links which are not set in a lexeme potential.

Valeriy Lukov and Vladimir Lukov speak about "thesaurus structures" which remind idiomatic expressions. The structure element of the thesaurus as an orienting system must be such a unit which reveals semantic links in a syntagma. Set collocations of words present such a unit. They are combinations of words used by speakers as "ready speech forms". These are such col-

locations as *Happy birthday! go shopping, watch TV, drive a car*, etc. The idea that these collocations are firmly set in the language and reproduced by speakers as speech forms becomes very obvious when we compare languages. These English collocations correspond with the Russian collocations *С днем рождения! ходить по магазинам, смотреть телевизор, вести машину*, which are made of different lexemes and have different grammar structures.

The reason why Russian and English speakers use different collocations is not connected with language grammatical peculiarities or a number of lexemes. Russian language admits such combinations of words as *Счастливого дня рождения, посещение магазинов, смотреть телевидение*, which are equivalents to the given English collocations, but these combinations are not firmly set in the Russian language. Nobody uses them because there is no reason to create new forms when it is easier to use already existing forms. Besides, creating new forms always entails the risk of misunderstanding. Stereotype collocations are not produced by speakers; they are reproduced.

The fact that set collocations are reproduced contributes to understanding because hearing one word the receiver may predict the next word. At the same time, while making a message, it is enough for a speaker to say one word and the next word will come to his mind itself. It means that the number of words which can follow after the pronounced word is limited; it consists of only those words which form set collocations with the pronounced word.

Set collocations of any language are various in their structure. It is very difficult to find any foundation for their systematization. As it is obvious from the given examples, they can be different in grammatical structures, they may have different degree of semantic connection between words, they can be stylistically different and more or less frequently used by speakers.

We cannot use semantic, syntagmatic or pragmatic criteria to form classes of all set collocations of the language. That is why in our monograph (Осокина/Osokina, 2007) we concluded that the only principle for their systematization can be the formal principle. We made groups of set collocations with one word and called them "series". For instance, the series of collocation with the word *God* consists of such collocations as *dear God! a follower of God, a crime against God, God's laws, faith in God, a house of God, thank god, my God! an Egyptian god, a male god, The Work of God, God's love*, etc.; the series of set collocations with the word *love* consist of collocations *true love, deep love, sincere love, maternal love, fraternal love, a love child, a love letter, a love story, be in love, fall in love, feel love, give love, make love, be sick of love, a token of love, love from the first sight, Love Boat*, etc.

The number of collocations in each series cannot be counted and cannot be given as a limited list. It is an open row. We can only list collocation, fixed in a certain dictionary (as we did it with collocations with the word *love* - they are taken from Cambridge International Dictionary of English) or in a

certain text (collocations with the word *God* are taken from the book by D. Brown "The Da Vinci Code").

A special work must be devoted to semantic analysis of set collocations. Right now we only can say that all these collocations are used by speakers as "ready texts", "borrowed texts". Not only such collocations as *The Work of God* or *Love Boat* which refer to certain culture phenomena (the Medieval Christian organization and a world-popular BBC film), but the rest of the collocations as well are borrowed from somebody else's speeches (e.g. from parents' speeches).

It does not mean, of course, that new word combinations are not created at all. The ability to create new collocations which are properly understood by others is a gift given to those who have an intimate feeling of the language. But it is not a good idea to make a message of mainly new word combinations – it will cause understanding problems.

Set collocations objectively exist in culture texts, most frequent of them are fixed in dictionaries. Altogether they make up the thesaurus system.

Thesaurus is the system of word links objectively given as a system of word collocations in culture texts. These texts contain knowledge. To study knowledge it is not really necessary to model mental structures which are impossible to observe. It is enough to study set collocations used in speech and fixed in texts.

We support the idea that knowledge can be studied as the change in the thesaurus system. To be able to see the changes it is efficient to analyze the thesaurus of the culture texts written years ago and modern texts.

First of all, it is necessary to find out the collocations used in a certain text (or certain texts). These must be texts of a great culture value. If, for example, we analyze texts which were very popular a hundred years ago and texts popular now, then we can see the thesaurus changes which took place during a century.

The technique of finding out the collocations was developed in our mentioned above work. The main idea of the technique consists in writing out of the text the most frequently used set collocations and in the following analysis of the collocations.

For this paper, we analyzed the text by F. Dostoevskij "The Demons" which was written more than a hundred years ago and had a great impact on Russian culture and modern fiction by V. Pelevin, S. Lukyanenko and D. Dontsova. We also analyzed modern Russian newspapers.

Each text contains a certain system of set collocations. There are collocations used only in this or that text (e.g. the collocation *иметь право голоса* is very frequent in Dostoevskij's book and is not frequent at all in the others); and there are lots of collocations which pass from one text to another. We suppose that to be able to see the thesaurus changes we must compare big series of collocations used in each text.

We found out that the analyzed texts contain such big series of collocations as with the words *дело* (*не мое дело, замаять дело, приняться за дело, общее дело, на самом деле, делать свое дело, etc.*), *время* (*тратить время, последнее время, наше время, в свое время, некоторое время, etc.*), *глаза* (*глубокие глаза, голубые глаза, раскосые глаза, зеленые глаза, карие глаза, стеклянные глаза, etc.*), *проблема* (*возникла проблема, неразрешимая проблема, есть проблемы*), *деньги* (*зарабатывать деньги, тратить деньги, нет денег, искать деньги, получить деньги*), and others (the limits of the paper do not allow to mention all the collocations from all the series).

To see clearly the changes in the thesaurus of the Russian language we made a table in which we listed the words which organize the biggest series of collocations (Table 1).

The words given in the table can be considered as the organizing units of the Russian thesaurus during 100 years. The words which form the biggest series in the most popular works literature can be called the "actual thesaurus". The collocations with these words make a language net available for and perceivable by an average speaker of Russian:

Dostoevskij	Pelevin	Lukyanenko	Dontsova	Newspapers
<i>человек</i>	<i>время</i>	<i>мир</i>	<i>дом</i>	<i>дело</i>
<i>дело</i>	<i>дело</i>	<i>дверь</i>	<i>дело</i>	<i>проблема</i>
<i>глаза</i>	<i>глаза</i>	<i>дело</i>	<i>деньги</i>	<i>деньги</i>
<i>время</i>	<i>проблемы</i>	<i>вещь</i>	<i>голова</i>	<i>время</i>
<i>идея</i>	<i>мир</i>	<i>время</i>	<i>глаза</i>	
<i>голос</i>	<i>жизнь</i>	<i>глаза</i>	<i>небо</i>	
<i>вопрос</i>	<i>вещь</i>	<i>деньги</i>	<i>жизнь</i>	
	<i>голова</i>	<i>человек</i>	<i>день</i>	
	<i>душа</i>	<i>девушка</i>	<i>время</i>	
		<i>путь</i>	<i>путь</i>	

As it is clear from the table, in all the analyzed texts we can find big series of collocations with the words *время* (*time*) and *дело* (*business/matter*). Very important is the word *глаза* (*eyes*) used in all works of literature but unused in newspapers. On the third place there is the word *деньги* (*money*) which appears in three out of five columns. We can also see the words *мир* (*world*), *жизнь* (*life*), *вещь* (*thing*), *голова* (*head*), *путь* (*way*) and *проблема* (*problem*) in two out of five columns. It gives an idea how valuable information is arranged in the actual thesaurus, i.e. which type of knowledge is considered as more important and which is less important.

The most valuable knowledge is connected with the words *время* and *дело*. Thesaurus links connected with the words *глаза* and *деньги* are very actual. The collocations with the words *проблема*, *путь*, *голова*, *вещь*, *жизнь*, and *мир* are a bit less important but they are gaining higher valuable status.

It is also noticeable that the importance of the collocations with the words *время*, *дело*, and *глаза* was enhanced by Dostoevskij's book and it made these

thesaurus units very valuable for Russian culture and language knowledge for more than a hundred years. Such thesaurus units as *деньги, проблема, путь, жизнь, мир, вещь, голова* are introduced by modern texts and present the latest values.

The role of set collocations of words in forming knowledge needs further studying. The thesaurus system made by set collocations organizes the picture of the world and the system of the mental lexicon. At the same time thesaurus is not a kind of a mental information storage. It is quite a material system which influences one's mental world from the outside.

The information thesaurus is organized by the language system of knowledge. It is the system of set collocations which compose the speech and structure the mental world.

The suggested view on the epistemological function of the language presents a new aspect of linguistics. The main idea which distinguishes the thesaurus conception from the existing views consists in recognizing the language (not mental) nature of the knowledge units and the material nature of their links. It also approves the derivative status of the mental models from language structures.

References

Залевская, А.А. (1990). *Слово в лексиконе человека. Психолингвистическое исследование*. Издательство Воронежского университета / Zalevskaya, A.A. (1990). *Slovo v leksikone cheloveka. Psiholingvisticheskoe issledovanie*. Izdatelstvo Voronezhskogo universiteta (= Залевская/Zalevskaya, 1990).

Караулов, Ю.Н. (2007). *Русский язык и языковая личность*. Издательство ЛКИ / Karaulov, Yu.N. (2007). *Russkij yazyk i yazykovaya lichnost*. Izdatelstvo LKI (= Караулов/Karaulov, 2007).

Колшанский, Г.В. (1990). *Объективная картина мира в познании и языке*. Издательство «Наука» / Kolshanskij, G.V. (1990). *Obektivnaya kartina mira v poznanii i yazyke*. Izdatelstvo «Nauka» (= Колшанский/Kolshanskij, 1990).

Луков, В.А., Луков, Вл.А. (2005). *Тезаурусный анализ мировой культуры. В: Тезаурусный анализ мировой культуры: сборник научных трудов (Выпуск 1)*. Издательство Московского гуманитарного университета / Lukov, V.A., Lukov, Vl.A. (2005). *Tezaurusnyj analiz mirovoj kultury*. In: *Tezaurusnyj analiz mirovoj kultury: sbornik nauchnyh trudov (Vypusk 1)*. Izdatelstvo Moskovskogo gumanitarnogo universiteta (= Луков и Луков/Lukov & Lukov, 2005).

Осокина, С.А. (2007). *Опыт эпистемологического анализа художественного текста*. Издательство «Графикс» / Osokina, S.A. (2007). *Opyt epistemologicheskogo analiza hudozhestvennogo teksta*. Izdatelstvo «Grafiks» (= Осокина/Osokina, 2007).

Павиленис, Р.И. (1983). *Проблема смысла: современный логико-философский анализ языка*. Издательство «Мысль» / Pavilenis, R.I. (1983). *Problema smysla: sovremennyy logiko-filosofskij analiz yazyka*. Izdatelstvo «Mysl» (= Павиленис/Pavilenis, 1983).

Чурсин, Н.Н. (1982). *Популярная информатика*. Издательство «Техніка» / Chursin, N.N. (1982). *Populyarnaya informatika*. Izdatelstvo «Tehnika» (= Чурсин/ Chursin, 1982).

Шрейдер, Ю.А., Шаров, А.А. (1982). *Системы и модели*. Издательство «Радио и связь» / Shrejder, Yu.A., Sharov, A.A. (1982). *Sistemy i modeli*. Izdatelstvo «Radio i svyaz» (= Шрейдер и Шаров/ Shrejder & Sharov, 1982).