

УДК 81.272

ББК 81.05

## Лингвистическая модель языковой системы знания: тезаурусная сеть

*С.А. Осокина*

Алтайский государственный университет (Барнаул, Россия)

## Linguistic Model of Language Knowledge System: Thesaurus Net

*S. A. Osokina*

Altai State University (Barnaul, Russia)

The article is devoted to one of the topical linguistic problems — the question about the part of the language in forming and firming the knowledge. The main purpose of the work is to present a new linguistic model of the language system of knowledge, the thesaurus net. In order to draw the features of the thesaurus model more clearly, the author considers a number of the most famous linguistic models which have dominated in works since the 20<sup>th</sup> century. There discussed such models as the language picture of the world, the conceptual system, the mental lexicon, and the information thesaurus. Critical review of linguistic works shows that the thesaurus conception has not been given sufficient consideration. However, this conception corresponds to modern science methodology and must be studied with great attention. In particular, within the thesaurus conception it is possible to restore the objective word network which serves as the main condition for human communication and storage for language knowledge; and network models are supposed to be the most perspective in studying the principles of communication and thinking nowadays. The toughest problem is to find the minimal functional unit of the thesaurus net. In contrast to the other discussed models which only show how the language expresses the knowledge, the thesaurus net is an active semiotic system which enables many different meaning systems.

**Key words:** thesaurus, network model, language picture of the world, conceptual system, mental lexicon, information thesaurus, set collocation of words.

Статья посвящена рассмотрению одной из наиболее актуальных проблем лингвистики — вопросу о роли языка в формировании и закреплении человеческого знания. Основной целью работы является представление лингвистической модели языковой системы знания, разрабатываемой в рамках тезаурусного подхода. Для достижения этой цели рассматривается ряд наиболее известных моделей соотношения языка и знания, доминировавших в лингвистике в XX — нач. XXI в., на фоне которых можно более четко обозначить отличительные черты предлагаемой концепции. Обсуждаются такие модели, как языковая картина мира, концептуальная система, ментальный лексикон и информационный тезаурус. Анализ лингвистической литературы показывает, что концепция информационного тезауруса не получила пока достаточного осмысления, однако она заслуживает пристального внимания, поскольку соответствует современной методологии науки. В частности, конструирование языкового тезауруса позволяет восстановить объективно существующую словесную сеть, являющуюся хранилищем человеческого знания и условием успешной коммуникации, а сетевые модели признаются в настоящее время наиболее перспективными при построении и изучении современной коммуникации и человеческого мышления. Наиболее важной проблемой при конструировании тезауруса является определение функциональной единицы сети. Основное отличие тезаурусной сети от других моделей заключается в том, что тезаурус не просто «отражает» или «выражает» имеющееся знание, но является активной семиотической системой, обеспечивающей возможность существования различных смысловых систем.

**Ключевые слова:** тезаурус, сетевая модель, языковая картина мира, концептуальная система, ментальный лексикон, информационный тезаурус, устойчивое сочетание слов.

One of the topical linguistic problems is the question about the part of the language in forming the knowledge and about the essence of the knowledge which belongs to the language speaking human. The fact of having the language (language capacity) determines the way of knowledge acquisition. It is absolutely obvious that the system of knowledge is formed as the result of sensual experience. However, for a human the language is not just an instrument for expressing the knowledge but natural material for its generating.

The questions discussed in the article are: how does modern linguistics model the language system of knowledge, and what are the peculiarities of every model? The ultimate purpose of the paper is to present a new linguistic model of the language system of knowledge, the thesaurus net.

The most well-known term for the language system of knowledge which one can see in most linguistic works is "language picture of the world". Meanwhile, each work provides its own understanding of the term, and the number of the methods used to explore the picture of the world is so numerous that we may conclude there is no common vision of the subject. "Language picture of the world" is just a successful metaphor which disposes to guesswork.

Besides "language picture of the world", there are three more "dominating scientific metaphors", frequently used by linguists. They are "conceptual system", "mental lexicon", and "informational thesaurus". They are used to describe so close notions (or the notion?) that the linguists have to compare them with each other.

The term "conceptual system" appeared in the 1950s–1980s due to the logic studies of the language. Russian linguists had a great influence by R. Paviolis, who presented the knowledge system as the conceptual system, or the "system of information about the world" [1, p. 101]. This system has continual, nondiscrete nature and is formed before a human starts to speak; this system creates the condition for developing of language capacity.

In general, one can say that the picture of the world differs from the conceptual system as more stable and better structured entity since it is formed with the help of the language. Though, some authors deny language factor in building the picture of the world (for instance Kolshanskiy [2, p. 25]) recently more and more linguists have supported the idea of the predetermining role of the language in knowledge generating, firstly expressed by W. von Humboldt and later developed by E. Sapir and B. Whorf.

Opposite views on the role of the language in forming the knowledge system come to agreement by distinguishing the "conceptual picture of the world" and "language picture of the world". The latter links the mental knowledge system with the objective world since only through the language mental essences may become objective.

The conception of mental lexicon is mainly developed in the works by A. Zalevskaya [3]. The conception

is raised from the popular ideas of the priority of the lexical meaning in comparison with the grammar meaning. E. S. Kubryakova says that the mental lexicon is a part of the language capacity and it can be viewed as the system of knowledge created by words and their relations. Words make meaning knots which are connected with mental concepts but not equal to them.

E. Kubryakova also points out that the system of mental lexicon can be compared with the so called "information thesaurus", the conception, discussed by a number of authors. According to Kubryakova, the information thesaurus is a kind of memory system which stores knowledge accumulated with the experience [4, p. 380].

The conception of the thesaurus as the system of knowledge about the world is being intensively developed nowadays in such fields as culturology and sociology [5]. Thesaurus is described as a system of knowledge gained during life experience and attached to words. By contrast, modern linguistics determines thesaurus as a kind of encyclopedia dictionary which stores words according to logical categories. Though the author of the conception of Russian Thesaurus, Y. Karaulov, views thesaurus as the lingua-cognitive level of the language personality [6].

The discussed conceptions demonstrate different approaches to understanding the epistemological function of the language. They create different models of the language system of knowledge but all of them state mental substance of this system and conclude that mental entities of the system can be studied through words and their relations. So, words and other language units are studied as material objects which can provide some ideas about the mental world. In other words, the language itself does not get any attention as the system of knowledge, only as the system of its expression.

At the same time, since the modeled structures appear to be secondary from the language structures (we have to underline that the language itself appears as the result of mental work but the discussed models of the knowledge are derived from language structures, thus, they are epistemologically dependent on the language), it is reasonable to consider the system of knowledge as an entirely language system. We need to understand which language (not mental!) units are supposed to be knowledge units and how they are organized in a working system.

We suggest that out of the mentioned conceptions the thesaurus model is the most perspective in this respect since it is the only conception which views words as the elements of knowledge storage, not only as the instruments of knowledge expression.

Thesaurus as the notion has recently become a subject of intensive scientific discussion, thus, it does not have an "understood-by-all" meaning in cognitive linguistics so far and there is a space to develop a new theory.

The term "thesaurus" specifies the language constituent of the knowledge because it is the "treasury of the words".

Thesaurus is a purpose dictionary, i.e. a specially organized collection of words, and a personal vocabulary at the same time. It is the system a person uses in the process of acquiring and producing verbal information, that is, in the process of knowledge exchange. The extension of the thesaurus system and the way it is structured helps to find one's own way to understand and interpret the reality in the overwhelming mass of incoming information.

To our mind, the empirical essence of the thesaurus appears to be blocks of stereotype word orders, or set expressions, which compose human speech. This mass of stereotype collocations exists as "genetically and statistically determined entity" [6, p. 53].

To make the thesaurus system work words must actively interact with each other, that's why a single word can not be a working unit of the thesaurus system. Actual units of the thesaurus are "ready-to-use" word orders, such as *a young man, have breakfast, watch TV, go shopping* etc. Such word orders match the criteria of the stability and repeatability and can be recognized as the language units.

As B.M. Gasparov points out, when people speak, they merely cite such ready blocks from their memory. He names such word orders as "communicative fragments", or the "blocks of the previous language experience" [7, p. 116]. Though it is really hard to imagine that human memory stores hundreds of diverse set expressions instead of logic language models, it is the only way to create the adequate language theory, the linguist says.

Gasparov pays much attention to understanding how communicative fragments come into speech and practically does not discuss the way they are organized in the human memory and in the language itself. Indeed, it is hard to imagine that such a number of absolutely different blocks of the language experience can be stored in the human memory and easily "dragged out" of there if they do not compose a system. B.M. Gasparov protests against any linguistic models and systematizations (at the same time, he describes some mechanisms of communicative fragments linking together which is supposed to be a manifestation of their dependence on each other, or system interaction). Denying the possibility of creating a language model, Gasparov rejects all linguistic achievements and presents language development as a chaos process. This makes impossible the mere possibility of its adequate use.

We suppose that Gasparov denies the idea of language system because none of the existing linguistic models can be used to study the type of organization that the communicative fragments have. However, this does not mean that it is impossible to create such a model.

To our mind, modern idea of the thesaurus system as the all-embracing comprehensive storage of various-structured information can help to study the organization of the communicative fragments.

There are three models which describe the structure constitution of the thesaurus — the hierarchy, the field,

the network. These models have appeared since the time of the first thesaurus dictionaries and changed each other in evolutionary progression. The reason for such evolution is connected with methodological changes in science in general, so that each succeeding model met the requirements of its time and was designed to overcome the drawbacks of the previous one.

Hierarchy model dominates in scientific studies of system objects. As a rule, scientific hierarchies are strict and inflexible comparing with the real life objects they describe. Hierarchy models usually look so logically proved that there cannot be doubts in their lack of completeness. However, hierarchy models of the word system cause questions making them open to criticism.

Particularly, comparative analysis of the thesaurus dictionaries shows that different dictionaries have different number of the basic logical categories which store the words of the language in different ways. For instance, in different editions of the famous Roget's Thesaurus one can find from six to ten initial logic categories divided into different numbers of minor rubrics. This brings up the question of possibility to reconstruct the universal logical hierarchy of notions which is supposed to be hidden in the words of the language. We can assume that the number of the logical categories and their constitution in different thesauruses is motivated either by the lexical system of the language, or by the individual preferences of the authors, and does not exist "before" language.

Analyzing defects of the hierarchy thesaurus model, linguists came to the idea that the thesaurus may have a field structure. The field approach replaces the conception of the word as a separate lexical sign by the conception of its existence inside the group of connected words, or a field. It means that the word can not be studied separately from the group it belongs to and does not have its actual meaning without being compared with other words.

In contrast to the hierarchy structure, the field structure does not have main and dependent elements. Components of the field have equal functions, and each element is connected to every next element inside the field by having at least one common feature. The elements with many common features organize the center of the field, the elements with fewer common features are at the periphery and may as well join another field.

The conception of the thesaurus allows combining the hierarchy and the field structures together. As Y.N. Karaulov points out, the thesaurus is the system of interconnected word fields but word interaction reveals the inner structure of the word shown in the definition, i.e. a hierarchy structure [8, p. 64].

Interpenetration of the hierarchy and the field structures makes the structure of the net. We think that the structure model of the thesaurus system must be viewed as a network structure.

The main work of the network conception is supposed to be C. Petri's Ph.D. dissertation, 1962 [9]. Nowadays network modes are widely used in natural science and form the basis for epistemological studies of the knowledge.

The net is composed of a multitude of positions as well as transitions, inputs and exits which lead from one position to another. The net does not have control elements that form the main characteristics of the system because there is no strict level location of the elements. The net is omnipresent.

Global information networks such as the Internet are the results of numerous studies of network structure principles. Fast-growing development of communicative networks makes us speak of network human existence and network thinking. In contrast to traditional cause-and-effect thinking, network thinking replaces the cause-and-effect principle by the interactive principle and smoothes subordination.

We think that the natural system of language thesaurus is supposed to be the prototype to all man-made communicative networks. The authors of the thesaurus dictionaries as well as other researches have marked the similarity of the thesaurus structure to the network structure many times. For instance, such opinion was expressed by L. Urdang [10] and the authors of Russian Associative Dictionary [11]. However, in linguistic studies the term "network" is rather used as a metaphor than as a conceptual system. At the same time, the language network becomes apparent in thesaurus dictionaries because they show the meaning of the word not though

the definition but through its connection with other words. The only problem is that a dictionary is always lack in space and can not show all the possible word transitions.

Thesaurus is not only a semantic network as it is described in a number of works. It is a material, real-life system of words which exists in material texts. It is a net of real word objects — ready texts, to wide extent.

We suppose that the unit of the thesaurus network is a set collocation of words because a collocation may be considered as the shortest ready to use text. Collocations are repeated in a mass of culture texts. Frequent repeatability of collocations in texts of different authors proves that they function as language unites because repeatability is a language unit quality. Having been created "by precedent", now they are used by speakers as their own expressions. Set collocations correspond with the linguistic notions that R. Barthes called "intertextual code" [12] and Val. A. Lukov together with Vl. A. Lukov — "thesaurus constructions" [5, p. 4].

Inside the thesaurus net words, being parts of set collocations, do not "express" meaning but become parts of knowledge. Meanings as well as concepts do not exist "before" words. They come with words and totally depend on words.

Comparing with hierarchy model which is completed by researches with the help of rather fragmentary evidence as separate as field components, network model of the thesaurus can be studied directly on real text material. Thesaurus network is an active semiotic system which enables many different meaning systems.

## References

1. Павиленис Р.И. Проблема смысла: современный логико-философский анализ языка. — М., 1983.
2. Колшанский Г.В. Объективная картина мира в познании и языке. — М., 1990.
3. Залевская А.А. Индивидуальное знание. Специфика и принципы функционирования. — Тверь, 1992.
4. Кубрякова Е.С. Язык и знание: на пути получения знания о языке. Части речи с когнитивной точки зрения. Роль языка в познании мира. — М., 2004.
5. Луков Вал. А., Луков Вл. А. Тезаурусный анализ мировой культуры // Тезаурусный анализ мировой культуры: сб. науч. тр. — М., 2005. — Вып. 1.
6. Караулов Ю.Н. Русский язык и языковая личность. — М., 2007.
7. Гаспаров Б.М. Язык. Память. Образ. Лингвистика языкового существования. — М., 1996.
8. Караулов Ю.Н. Общая и русская идеография. — М., 1976.
9. Peterson J. Petri Net Theory and the Modeling of Systems. Prentice-Hall, 1981.
10. Urdang L. The Oxford Thesaurus. American Edition. Oxford University Press, N. Y., 1992.
11. Русский ассоциативный словарь. Книга 1. Прямой словарь: от стимула к реакции. Ассоциативный тезаурус современного русского языка. Часть I. — М., 1994.
12. Barthes P. S/Z. Blackwell Publishing, 1990.