

ICEST 2020

International Conference on Economic and Social Trends for Sustainability of Modern Society

THESAURUS AS A RESEARCH SCHEME MODEL: "LANGUAGE OF SCIENCE" POSTGRADUATE COURSE

T. V. Tarasenko (a)*, E. I. Lotsan (b), A. A. Kurzina (c), A. K. Gorbacheva (d)

*Corresponding author

(a) Reshetnev Siberian State University of Science & Technology, 31 "Krasnoyarskiy Rabochiy" pr., Krasnoyarsk, 660037, Russia, tvt2004@mail.ru

(b) Reshetnev Siberian State University of Science & Technology, 31 "Krasnoyarskiy Rabochiy" pr., Krasnoyarsk, 660037, Russia, lotsan@mail.ru.

(c) Reshetnev Siberian State University of Science & Technology, 31 "Krasnoyarskiy Rabochiy" pr., Krasnoyarsk, 660037, Russia, Lemon_doll1991@mail.ru

(d) Reshetnev Siberian State University of Science & Technology, 31 "Krasnoyarskiy Rabochiy" pr., Krasnoyarsk, 660037, Russia, gorbunova_ak@mail.ru

Abstract

The purpose of this article is to justify the pragmatic function of postgraduate student's individual work on the glossary and thesaurus of scientific research within the study of the Language of Science course as part of the work on the formation of professional linguistic personality; to show how the modeling the semantics of a scientific text orients a young researcher to study the terminology of the studied fragment of the scientific picture of the world. As a methodological basis, the authors rely on a personal-activity approach to learning. The solution of research problems was ensured through the use of a set of interconnected methods, such as theoretical (literature, generalization of existing domestic and foreign experience), general scientific (classification, differentiation, comparison, alignment, generalization), as well as empirical (practice, content analysis of activity products – glossaries and postgraduate students' thesauri, statistical data processing). The research material is based on glossaries and thesauri as products of postgraduate students' natural written language. As a result, the sequence of design work on the thesaurus is described. Systematizing the results of the content analysis of postgraduate students' works and their profiles has demonstrated that they use personal research experience in creating a thesaurus. The obtained results help to optimize the process of preparing postgraduate students for scientific activity.

2357-1330 © 2020 Published by European Publisher.

Keywords: Thesaurus, glossary, postgraduate student, language of science.



1. Introduction

Today, terminology is an object that draws attention and interest of everyone, from the ordinary consumer, who is bombarded with different terms every day, often in a foreign language, to people who are in training (Quer et al., 2020; Nowakowski & Bernard, 2019; Tareva & Tarev, 2020). All this, on the one hand, leads to confusion, and on the other hand, provides scope for research in the field of linguistic terminology, since one of the tasks of science is to develop and harmonize appropriate terminology, as well as to identify the most important trends in terminology (Korshunova, 2011).

Since the terminological system, as well as the lexical system of the language as a whole, is open to the emergence of new attributes, it requires mandatory and constant work to streamline it.

One way to study and streamline terminology is to compile terminological dictionaries, whose principles of building are developed in terminography. The works of many scientists are dedicated to the problems of descriptive terminology and the compilation of linguistic dictionaries.

The relevance of the appeal to the thesauruses (hereinafter the TSEs) lies in the fact that the current row of terms is in varying degrees of remoteness from each other. The TS description systematizes the terminology by establishing links between the terms, which enables to concisely and adequately perceive subdiscipline under consideration.

Another reason for referring to descriptive terminology is the requirement of educational standards for the training of postgraduate students, which is history knowledge and current status of the chosen professional field, the issues of their own scientific research, knowledge of the analysis of scientific texts in the State and foreign languages and the skills of in-depth analysis of objects in the professional field, i.e. requirements for the formation of professional linguistic personality.

In terms of the linguistic personality concept given by Karaulov (2002) “the professional linguistic personality” is understood as a person who can be described by the following criteria: 1. Performing activities in a special (professional) field. This activity is formed while the scientific picture of the world is mastered as a sequential “passing” different complexity levels of this activity (socialization); therefore, it is able to carry out professional functions of varying degrees of complexity. 2. Ability to social multifunctionality, that is, to actualize several social roles that require varying degrees of world development (there is a professional role on the repertoire of social roles, which is carried out through the formation of a scientific picture of the world in the consciousness of a native speaker). 3. During training a scientific picture of the world is formed. The process of forming a scientific picture of the world is a chain of successive transitions from a naive to a scientific picture of the world, which determines the multidimensionality of the consciousness of a native speaker (Karaulov, 2002).

Kiseleva (2014) gives the following definition of a professional linguistic personality: “This is a native speaker, characterized on the basis of discourse analysis with regard to using the system tools of a given language in a discourse to reflect the vision of the surrounding reality, as well as to achieve certain goals in the world. The discourse structure is correlated with the structure of a professional linguistic personality, which, like a linguistic personality, has motivational-activity and cognitive levels, as well as a linguistic level that represents language as a text, system and ability” (Kiseleva, 2014, pp. 360-361).

2. Problem Statement

In recent times, thesauri have found their place in literally all areas of human activity. The thesauri lay the groundwork for information technology in vocabulary activities and are adequate resources for all modern lexicographic works, in particular, they set the distinguishing features of the information retrieval thesaurus as an adequate model for compiling a technical translation thesaurus.

As mentioned above, the modern requirements for the training of postgraduate students are not limited only by the knowledge in the professional scientific field, but also require to effectively master scientific communication, have the skills to analyze scientific texts in the national and foreign languages, therefore individual practice in compiling a glossary and a thesaurus of the dissertation research within the Language of Science course is one of the tools for a postgraduate student to achieve the required knowledge and skills.

3. Research Questions

The hypothesis of this work is that compiling a thesaurus as one of the types of practical and methodological work with a young researcher can be used in various fields and for different purposes. For example, in cognitive sciences, work on the TS is the work of describing the picture of the world, its components (Maksimyuk, 2014). A piece of knowledge, for example, image of a tree growing on the ground, can be represented as semantic expressive elements the tree, the ground and the semantic relationship between them to grow on. By the same token, any fragment of knowledge can be described in the TS: more complex concepts are formed from simpler ones by the aid of corresponding relations. Accordingly, such activity leads researcher to understanding that the TS has a complex structure in which some concepts and relationships, grouped together, form others, which in turn form increasingly complex concepts and relationships.

Such a methodological process allows the postgraduate student to create a model of a certain fragment of scientific knowledge, on the one hand, to look at the research problem more widely; on the other hand, to fix the “point” of research in a scientific problem in a given period; allows a young researcher to understand that the thesaurus form is a means of communication between the logical-philosophical and linguistic descriptions of the scientific language. Such an understanding, as mentioned above, is an important criterion for a professional linguistic personality.

In addition, during the work the following question was raised, whether the work on the study of the informational-semiotic nature of the scientific research terminology and its systematics allows postgraduate students to apply their knowledge when writing scientific texts of various kinds (articles, reviews, etc.), including in foreign languages, and what is more, in scientific discussions using the TS modeling method.

4. Purpose of the Study

The purpose of this article is to justify the pragmatic function of postgraduate student’s individual work on glossary and TS of research within the study of the Language of Science course during formatting the professional linguistic personality; to show how modeling semantics of the scientific text orients a young researcher to study terminology of the explored piece of the world’s scientific picture.

5. Research Methods

As a methodological principle, the authors rely on a personal-activity approach to learning. Solving the research problems was ensured through the usage of interconnected methods suite, such as theoretical (literature, generalization of existing domestic and foreign experience), general scientific (classification, differentiation, comparison, alignment, generalization), as well as empirical (practice, content analysis of activity products – glossaries and postgraduate students' thesauri, statistical data processing).

In the context of this description, a thesaurus is understood as a semantic dictionary as a way of organizing vocabulary that explicitly captures all semantic relations between its units (Serova et al., 2004).

Researchers define TS as a vocabulary of “a controlled indexing language, structured formally so that the a priori relationship between concepts is clear. This definition identifies lexical units and semantic relationships between these units as elements that compile the TS. The TS relationships (genus/species, part/whole, complex/element, cause/effect) are laid on the structure of the taxonomy, i.e., the main taxonomies of the subject area are identified” (Fedotov et al., 2015, p. 88). “The TS is defined by the International Organization for Standardization (ISO), as the vocabulary of a controlled indexing language, structured formally so that the a priori relationships between concepts are made explicit. Indexation (indexing) is the sorting (systematization) of information according to some formal principles. There are two types of indexing: classification and coordinate. The TS (paradigmatic) relationships (genus/species, part/whole, complex/element, cause/effect) are laid on the structure of the taxonomy, i.e., main taxonomies of the subject area are identified. The TS is a dictionary (ideographic or semantic dictionary) with words (lexical units, terms, concepts, descriptors) ranked according to their semantic similarity” (Fedotov et al., 2017, p. 117).

Karaulov (1981) qualifies the TS as any dictionary that explicitly captures the semantic relations between its units. This definition includes any kind of ideographic dictionary (thematic, analogous, and ideographic, actually), as well as “special thesauri: general technical and information retrieval”. Karaulov (1981) divides thesauri into general language and information retrieval, obviously, including all special TSeS in the last concept. He further emphasizes that if “one removes terminological and sectoral restrictions” at the generally accepted definition of special thesaurus, then there will be no “fundamental difference” between them and general linguistic thesauri (Karaulov, 1981, pp. 148-149).

The researcher explains:

The TS is a lexical tool of information retrieval systems. It consists of a controlled but modifiable glossary of terms between which semantic connections are indicated. Such a dictionary, comprehensively covering a certain specific field of knowledge, is a list of descriptors and nondescriptors (auxiliary terms), which is ordered in accordance with systematic and alphabetical principles and contains indications of semantic relationships between them – both hierarchical (generic) and non-hierarchical types. (Karaulov, 1981, pp. 148-149)

Karaulov (1981) writes: “The ordering of two genera (systematic and alphabetical) distinguishes this dictionary from the explanatory one and determines its main structural property – there is more than one input” (pp. 148-149).

From Morkovkin's (1970) perspective, the development of the TS is based on the logical rubricating the entire conceptual vocabulary content. He also emphasizes that "the main task for the compilers of an ideographic TS type dictionary is the identification and subsequent rational ranging of conceptual groups actually represented in the language vocabulary" (Morkovkin, 1970, p. 27).

Lukashevich (2011) relates thesauri to ontological resources. Pursuant to the researcher, information retrieval thesauri as a kind of ontological resources, including Roger's and WordNet TS, have the following distinctive features:

units of TSeS have a close relationship with the natural language, are usually supplied with variants of their expression in the natural language; thesauri usually do not have an internal structure of concepts, i.e., representation of properties and attributes in frames. Knowledge about the world and a subject area is presented in the form of relations between concepts; axioms (rules of inference) are reduced to the properties of transitivity and inheritance. (Lukashevich, 2011, pp. 103-104)

In a broad sense, the thesaurus refers to the system of knowledge about reality that a subject or group of subjects possesses. The subject is also able to receive new information, due to which the original thesaurus will change. The thesaurus contains not only information about reality, but also additional information, due to which it becomes possible to receive new information. Each word of the thesaurus is associated with a synonymous descriptor for which semantic relations are defined. There are hierarchical (generic) and associative relationships. In linguistics, the semantic relationships, that construct the thesaurus, may be antonyms, hyponyms, synonyms, paronyms, etc. If the explanatory dictionary is aimed at revealing the meaning of a word solely through definition, the thesaurus helps to identify it using the correlation of the word with other words and their groups.

Thesaurus in the general sense is a special terminology; in a more rigorous and substantive sense it is a dictionary, a collection of information, a corpus or code that fully covers the concepts, definitions and terms of a special field of knowledge or field of activity, which should contribute to the correct lexical, corporate communication (understanding in communication and the interaction of persons related to one discipline or profession). Unlike the explanatory dictionary, the thesaurus can be used to identify the meaning not only by definition, but also by correlating the word with other concepts and their groups. A thesaurus is knowledge represented in the form of semantic expressing elements/concepts and semantic relations between them, i.e. structured knowledge in a certain way.

The presentation of an interconnected conceptual framework is one of the significant abilities of the thesaurus. By fixing the semantic relations between its constituent units, the thesaurus tries to reflect the model of human thinking and ways of understanding the world by man. A person understands the world through science, so the content of his concept is determined by the results of research by each specific science. Depending on the area of application of the thesaurus, its reflection scheme differs, since, on the one hand, individual branches of scientific knowledge develop unevenly, on the other hand, the unresolved problem of synthesizing private pictures of the world prevents generalization into a single and integral system.

By presenting various inputs, the thesaurus allows you to access the information you need in different ways. In this case, the thesaurus relies on the achievements of lexicography, linguistic semantics,

ontology, cultural studies, etc. Thus, the hierarchical index primarily reflects the ontological connection between concepts, and the alphabetical index reveals the lexical-semantic connection between linguistic units.

The TS inputs also reflect ways of human understanding of the concept. But for this, the TS authors are developing special artificial methods. For example, in order to avoid complications of the semantics of a natural language, the information retrieval TS uses the information retrieval language (Babalova, 2014).

Thesaurus is defined as a means of controlling the vocabulary of a language with a view to facilitate data retrieval. Moreover, the general structure of the TS is determined on the principles of its main task, although there is no fundamental difference between the general language and information retrieval TS (Korshunova, 2011).

General linguistic Tses are based on the logical rubrication of the entire conceptual content of vocabulary, but in the terminological dictionaries such a detailed description of the descriptors is impossible, because the terms of each branch of science and technology acquire their special meaning simply within the framework of this branch. Moreover, in the last decade, thesauri have been developed for almost all sectors of activity, and many narrowly specialized Tses have been created (Faal-Hamedanchi, 2010).

According to researchers Mikhaleva and Maletina (2014) the advantages of project work on the thesaurus are as follows:

“Firstly, the TS dictionary is a projection of the fragment of the world picture, it reflects accumulated and systematized knowledge in the specific field.

Secondly, the TS represents a thematic, semantic unity, despite the fragmentation of its constituent elements.

Thirdly, the TS-type dictionary reflects a scheme of concepts and the lexical units expressing them in a hierarchical relationship: from more general to particular, more specific concepts, and vice versa. Moreover, during steps of generalization, a certain concept is localized “in a certain system of general relations, relations that are the most basic, most natural and important connections between concepts”.

Fourthly, the TS dictionary is a hierarchical structure in which the most “valuable” and common elements are the core. The lexical entry in the TS dictionary is a detailed concept in which one can discover the whole totality of its relations: if a person operates on a concept, then he operates on the whole system of concepts associated with it.

Fifthly, the TS dictionary models natural cognitive mechanisms used to organize, preserve and transmit mental experience using language tools: association mechanism, internal speech, deep prediction, mechanism for internal design of the statement, mechanism for external formulating the statement, encoding/decoding mechanism for information, output on the image of the world, taking into account the diverse derivative knowledge (encyclopedic and linguistic).

Sixthly, the TS dictionary is based on using the psychological principles of recognition, categorization, various types of supports (functionality, dynamism, integrity, code variability; the principle of reliance on signs, along with signs of signs, parallel reliance on perceptual, cognitive, emotional and evaluative (verbal and non-verbal) experience), the principle of semantic substitutions, as well as the principle of subjective equivalence.

And finally, the TS-dictionary participates in the formation of the mental TS and reflects its specifics: openness of the structure, variability of boundaries while maintaining the core, structure variability, effectiveness (influence on the behavior of the subject), etc. (Mikhaleva & Maletina, 2014).

The most characteristic feature of the TS is its competence to reflect existing relationships between its units due to its special structure. Unlike the explanatory dictionary, the TS makes it possible to reveal meaning not only by definition, but also by correlating words with other concepts and their groups. A particular advantage of the TS is that it also gives the opportunity to move from concept to word and from word to concept. Therefore, in terms of the personal-activity approach to teaching the Language of Science course, postgraduate students carry out practical work to compile a glossary and TS of a dissertation research. At the beginning of their work, postgraduate students are tested. Test objective is to identify level of knowledge in terminology and terminography; ideas about scientific picture of the world; about scientific periodicals, etc. Further, postgraduate students, together with the teacher, get acquainted with the basic concepts of terminology and terminography, the TS, models for its developing, basic dictionaries and TSes of the branch of knowledge and research.

Further, a postgraduate student carries out project work on the topic “Glossary and TS of my research”, which involves the following steps: insight into keywords on the topic “Glossary and TS”; highlighting key terms of the research; compiling a list of them in alphabetical order, since the glossary is nothing more than a dictionary of specialized terms.

After this, a postgraduate student begins to work on the compilation of glossary articles. A glossary article is a definition of a term. It consists of two parts: precise formulation of a term in the nominative case; substantial part, clarifying in greater detail the meaning of this term.

Stages of work on the TS are the following:

Firstly, formulating overall terms, concepts and definitions on the topic of the dissertation based on the literature on the subject of scientific work already studied.

Secondly, analysis of this material and its division into semantic groups.

Thirdly, analysis of the selected groups from the point of view of the hierarchical subordination of the terms included in them (more general/more private), creating their logical composition.

Fourthly, identification of potential semantic relationships between the terms:

- by their nature: causal, logical, semantic, functional, genetic;
- by the area of their occurrence: internal, external;
- by their orientation: direct, inverse.

Fifthly, constructing a thesaurus model.

At the end of the course, postgraduate students fill out questionnaires that describe possibilities for the practice results gained as part of the Language of Science course.

6. Findings

Overseeing the work within the Language of Science course for a period of five years and the content analysis of the works and questionnaires at the postgraduate level allows us to draw conclusions as follows:

the practical significance of the design work by postgraduate student on the topic “Glossary and TS of my research” is that the experience of TS modeling is an auxiliary tool for in-depth study and assimilation of not only metalanguage, but also the subject area itself during professional training of specialists; models obtained as a result of research may be used to systematize other areas of science; the lexicographic practice of TS modeling can be applicable in the compilation of terminological dictionaries as part of the material of various languages, including for the compilation of electronic dictionaries and machine translation programs (we mean linguistic support using formal methods).

The analysis of the postgraduate works showed that a PhD student strives for the type of “professional linguistic personality”, and he or she is able to independently generalize, highlight keywords and terms when compiling a glossary of the dissertation. This is indicated by both the work on compiling the glossary and the involvement of the context to show the functioning of the terms. Postgraduate students can independently draw hypothetically deductive conclusions, choose the necessary language tools when compiling the TS. Further, they demonstrate the capacity for memorizing and transmitting information by the means of language tools, which was shown by colloquiums, where students were tasked with presenting their scientific research in oral form for five minutes with the necessary detail and highlighting the facts. Students demonstrated the ability to memorize and summarize, operate with significant amounts of information, but found it difficult to show the place of their work in the scientific picture of the world, as well as to reveal the relationship and interdependence of scientific terminology.

During the training, postgraduate students conducted an analysis of the terminological dictionaries of their scientific field and concluded that mainly the compilers of dictionaries use the alphabetic or alphabetic-nesting method. The TS method still remains on the periphery of lexicographic practice. And this is understandable, since it requires close cooperation between subject matter experts and linguists.

7. Conclusion

As part of the teaching of the Language of Science course, it was shown that the personal-activity approach to teaching terminography and terminology of a future researcher is one of the tools to form professional linguistic personality. Project work on the TS as an active dictionary gives a postgraduate student information on the pronunciation of the term, its compatibility, examples of its use in a scientific text and allows to create a model of a specific fragment of scientific knowledge and find “unique” place in the studied subject. Moreover, the TS as an ideographic dictionary is able to assist a starting researcher in writing scientific articles and reports, including in a foreign language. The TS representation of knowledge fragments this knowledge, structures it in the space-time continuum so that all knowledge is divided into separate groups of concepts, interconnected by certain relationships. The compound of a TS dictionary with an explanatory one is the best way to explicate the total amount of knowledge in the topical area, which is because of definitive nature of the term (there are no definition systems outside the set of definitions either). Therefore, it is necessary to further develop the skills of an in-depth evaluation of objects in the professional field for writing and designing individual scientific research in accordance with the requirements specified for a PhD thesis.

References

- Babalova, G. G. (2014). *Programming thesaurus*. Printing Center KAN. (In Russ.)
- Faal-Hamedanchi, M. (2010). *The thesaurus representation of oil-and-gas terminology in Russian and Persian* (Thesis). Moscow. (In Russ.)
- Fedotov, A. M., Idrisova I. A., Sambetbaeva M. A., & Fedotova O. A. (2015). The use of a thesaurus in a scientific and educational information system. *Vestnik NGU. Series: Information Technology*, 13(2), 86-102.
- Fedotov, A. M., Sambetbaeva, M. A., & Fedotova, O. A. (2017). Thesauruses and Ontologies in Scientific and Educational Information Systems. In *Proceedings of the XVI All-Russian Conference DICR-2017* (pp 115-127). Novosibirsk.
- Karaulov, Yu. N. (1981). *Linguistic design and thesaurus of standard language*. Nauka.
- Karaulov, Yu. N. (2002). *Russian language and linguistic identity*. Nauka.
- Kiseleva, N. I. (2014). Necessity of mastering the conceptual and terminological apparatus science through study methods of teaching Russian Language. *Scientific notes of Orel State University*, 1(57), 360-363.
- Korshunova, S. O. (2011). Hypertext as a result of a thesaurus representation of professional cognitive activity (based on syntax terminology). In *Problems of the conceptual systematics of language, speech and speech activity: materials of the 5th All-Russian Conference* (pp 362-373). IGLU. (In Russ.)
- Lukashevich, N. V. (2011). *Thesauri in Information Retrieval Problems*. Izd-vo Moskovskogo universiteta.
- Maksimiyuk, E. V. (2014). Consciousness image reconstruction based on the results of experimental research (image of happiness for a family). *Scientific notes of Orel State University*, 1(57), 261-265.
- Mikhaleva, E. V., & Maletina, L. V. (2014). Thesaurus approach to the learner's dictionary design. *Modern problems of science and education*, 3, 248.
- Morkovkin, V. V. (1970). *Ideological Dictionaries*. Izd-vo Moskovskogo universiteta.
- Nowakowski, S., & Bernard, G. (2019). From Digital Humanities to a Renewed Approach to Digital Learning and Teaching. *Journal of Siberian Federal University. Humanities & Social Sciences*, 12(9), 1682-1693.
- Quer, G, Enrique F., & Inmaculada S. (2020). Communicative and Textual Competence as a Formative Element in the Subject of Translation in Science and Technology (Russian – Spanish). *Journal of Siberian Federal University. Humanities & Social Sciences*, 13(3), 343–353.
- Serova, T. S., Shishkina, L. P., & Kuzmina, I. P. (2004). *Ecological German-Russian lexicon-thesaurus*. TEMPUS-TACIS.
- Tareva, E. G., & Tarev, B. V. (2020). Individual Styles of Summary Writing: Approaches to Styles Description and Diagnostics. *Journal of Siberian Federal University. Humanities & Social Sciences*, 13(3), 423-432.