

Hierarchical Relations in the Synonymic System of the Russian Language

By

Liliya Abrarovna Usmanova

Ph.D. in Philology, Associate Professor, Kazan Federal University

Email: usmanova77@rambler.ru

tel.: +79178642642

Id Scopus 57045843400

ORSID: <http://orcid.org/0000-0002-6784-4035>

Irina Valerievna Erofeeva

Doctor of Philology, Professor, Kazan Federal University

Email: erofeeva89@mail.ru

Id Scopus 56260827900

ORSID: <http://orcid.org/0000-0002-6702-9129>

Abstract

The paper deals with a systematic comparison of semantic categories and taxonomic relations for the main modern lexical and semantic resources, including the RuWordNet thesaurus, the Russian National Corpus, the Russian Semantic Dictionary, and the Ideographic Dictionary of the Russian Language. It compares the peculiarities of vertical systematization of lexical sections in each of them and elicits the similarities in the principles and approaches to the problem of language modeling. A group of nouns with emotive semantics has been selected as the object of research. The research is based on the synthesis of traditional and modern methods of analysis of lexical and semantic relations of linguistic units: descriptive, onomasiological, as well as methods of cognitive and corpus linguistics. It is emphasized that in the upper tier of the classification there are coincidences in the generalization of emotive names and their attribution to the abstract sphere, to the level of spiritual culture. In the lower tiers there are some differences in the degree of abstraction of concepts, which is attributable to differences in approaches and methods used by the developers in the process of systematizing elements. The authors conclude that the peculiarities of categorization of reality in dictionaries and thesauri reflect anthropocentric tendencies in reliance on the communication requests of users and an addressee factor. A cognitive vector is associated with the ability to get an objective idea of the organization of a person's lexicon, of the body of knowledge resided in their memory.

Keywords: hierarchical system, hyperonymy-hyponymy relations, taxonomy, conceptual sphere, tree structure, ideographic dictionary, RuWordNet.

Introduction

The beliefs about the world as a cohesive whole where individual phenomena and objects are presented as parts of universals, from an ontological point of view, suggest an idea of interconnected, interdependent parts in the structure of the whole. According to M.V. Nikitin, "the "part - whole" relations permeate the universe from the bottom upwards, from micro- to macrocosm, from elementary particles to galaxies. They embrace things of all levels of complexity to organize them into diverse multilevel hierarchies of things, formed by the commonality of inherent features and patterns" [1: 387].

Published/ publié in *Res Militaris* (resmilitaris.net), vol.12, n°4, December Issue 2022

Modern linguistics with a focus on the study of language and human mentality, linguistic consciousness and thinking gives priority to the problems of categorization and conceptualization of knowledge in linguistic units, ways of representing knowledge and their objectification, etc. In this connection, the problem of hierarchical relations in the lexical system of language receives a new rethink, for these relations begin to be considered from the point of view of in-depth understanding of the objective reality and integral vision of an individual and their world.

Today, in the era of global networks and rapidly augmenting flows of information, the problem of developing information retrieval and information analytics systems, thesauri, corpora designed for automatic text processing, describing relationships between domain terms or between lexical meanings of a natural language in the form of logical semantic subordination of elements becomes urgent. As is well known, ideographic dictionaries and thesauri specialize in hierarchical conceptual relationships of vocabulary. Dictionaries of this type are high-demand in the modern context, as apart from semantic information, they also encompass pragmatic data about a word, i.e. consider the aspect of “action” of a word in different contexts.

Methods

To compare the hierarchical relations, modern lexical and semantic resources were selected, such as RuWordNet [2], National Corpus of the Russian language [3], Russian Semantic Dictionary. Explanatory Dictionary, Systematized by Classes of Words and Meanings edited by N. Y. Shvedova [4] and Ideographic Dictionary of the Russian Language compiled by O. S. Baranov [5]. The peculiarities of vertical systematization of lexical sections in each of them were compared and the similarities in the principles and approaches to the problem of language modeling were identified.

In the process of work the traditional linguistic methods were used, such as descriptive, onomasiological, and modern methods of cognitive and corpus linguistics were also applied. Also, to specify a semantic content in a number of lexical units in the context of displaying their properties of reality, a semantic-cognitive method was used.

Results and Discussion

Thesauri are a variety of dictionaries that reflect more complex relationships between lexical units of a language than dictionaries. The semantic feature of vocabulary clusterization allows for considering lexemes as part of synonymous groups, called synset. Linguistically, such groups are made up of quasi-synonyms connected not only by synonymy relations, but also paronymy, antonymy, hyponymy, hyperonymy, etc.

The first thesaurus of this type, WordNet, was developed in the United States in the 1980s. (Priston University). Today, WordNet is one of the most famous resources in the field of automatic word processing [6-8]. The creation of such super-large linguistic resources for information storage and retrieval systems means moving to a new level of semantic relations in dictionary representation.

When solving problems of natural language processing, the thesauri like WordNet are used not only as lexical resources containing knowledge about the language, but also as ontological resources containing knowledge about the real world. In such cases, synsets are interpreted as concepts, the hyperonym / hyponymic relation is interpreted as the class /

subclass ontological relation (isA), the holonym / meronym relation is interpreted as the part / whole ontological relation (partOf), etc. [9].

The digital thesaurus RuWordNet is known as a lexical resource in the field of automatic processing of texts in Russian; it was compiled on the basis of automated transformation of the RuThes thesaurus into the WordNet format by a team of Russian scientists.

In accordance with modern anthropocentric paradigm, the authors of RuWordNet pay attention to the functional-communicative approach, as well as the key features of speech activity. From these viewpoints, they distinguish between synonyms and non-synonyms, combining words based on semantic commonality into a synset, where lexical units are arranged according to the principle of hierarchical relations based on logical and semantic subordination, that is, when a specific category falls into a general category or, conversely, a universal category includes a specific category. Along with synonymous and hyper-hyponymic relations, this thesaurus also includes causation and succession relationships, relations to the domain, word-formation relations and relations between phrases and their components [10-11].

Ideographic Dictionary of the Russian Language by O. S. Baranov is of the same type as the famous English thesaurus by Roger. It can be defined as a classification type thesaurus with a taxonomic principle of organization of lexical material. A distinctive feature of this dictionary is the representation of the plurality of semantic relations between lexemes in the form of a network of headings that form a single tree structure and are interconnected by a complex of references. Words in the dictionary are collected in groups (semes) by meaning and are clustered around a certain concept (idea), with which they are most often associated by specific relations. The groups, in turn, are arranged into subsections, etc. Most grouped concepts have definitions with an explicit indication of the genus and species (generic and specific semes). The purpose of this dictionary is to trace the connection between the general and the particular, the whole and the part, and to highlight the most significant, dominant concepts (semes), to reveal their interconnections [5].

In Russian Semantic Dictionary the structure of lexical system is also presented in the form of a lexical tree with branches (sets and subsets), diverging from top to bottom, with multi-stage division of macroclasses into conceptual spheres (descriptor zones), which in turn are divided into lexical classes. The dictionary is based on the idea of constructing a thesaurus following the procedures of vocabulary parametrization: the distribution of words by concepts is based on the generality of semantic elements.

When dividing vocabulary into lexical-semantic classes and arranging the material, the authors were guided by the idea “to present their own Russian vocabulary as a level of language that has its own organization, self-created and -regulated by the very language” [4: 20]. First of all, Russian Semantic Dictionary structures lexical-semantic classes in accordance with the part-of-speech belonging, and this is one of the main differences of this dictionary from O. S. Baranov’s “Ideographic Dictionary”, where words of different grammatical classes are combined by semantics. In the information and reference system *National Corpus of the Russian Language* (NCRL), we were interested in taxonomic classification (semantic markup), presented in the form of a metalanguage of lexical-semantic labels, implying a thematic class (taxonomy), mereology (part-whole relationship), topology, evaluation and etc.).

Unlike the dictionaries mentioned above, within the taxonomic classification there is not a tree, but a faceted principle of classification, which implies the belonging of one word to

several classes at once if it is polysemous.

In addition to information on grammatical and semantic parameters in the NCRL, one can also get information about the linguistic behavior of certain classes of lexemes – their compatibility, control model and their types, etc. [3]

Let us trace the peculiarities of the interpretation of hierarchical relations in the above-described sources using the example of a semantic area of names of emotions. In the course of considering nouns with emotive semantics, it was noted the following.

Synoptic schemes of thesauri and dictionaries in accordance with the generic principle of organization of the language system presuppose several levels of generalization, the number of classification headings increases with the transition from a higher level to a lower one. The observations have shown that the upper layer of the classification in the compared sources reveals significant coincidences. So, at the highest level, the names denoting feelings (*страх, восторг, скука / fear, delight, boredom*) have their relevance to the level of *Абстрактная сущность (Abstract Entity)* in RuWordNet (even higher level is the level *Постоянная сущность (Constant Entity)*); in Russian Semantic Dictionary these names are traced from the section “Nouns with Abstract Meaning: Genesis, Spiritual World. Society”; in NCRL these nouns are labeled as “non-objective names”, “the abstract”.

In the lower tiers, there are some differences in the degree of abstraction of concepts. Let us consider them.

RuWordNet in the group *страх (fear)* highlights a synset that contains the words *боязнь (fear)* and *тrepidation (trepidation)* similar in volume and content, represents one-root words, hyperonyms in the form of equivalent variants that are synonyms for each other, and a group of hyponyms, including words and phrases that are in a subordinate relation to the word *страх (fear)* (see Fig. 1):

СТРАХ существительное
Страх страх, боязнь
Синсет
1 БОЯЗНЬ, СТРАХ, ТРЕПЕТ [Понятие Рутез: страх, боязнь]
однокоренные слова
1 РАССТРОИТЬ, НАСТРОЕНИЕ, РАССТРАИВАТЬ, ЗАСТРАЩИВАТЬ, СТРАШИТЬСЯ, СТРАЩАТЬ, СТРАСТНОСТЬ, ЗАСТРАЩАТЬ, НАСТРОИТЬСЯ, РАССТРАИВАНИЕ, НАСТРОПАЛЯТЬ, НАСТРОПАЛИТЬ, СТРАШИТЬ, НАСТРАИВАТЬСЯ, НАСТРОЙ, НЕСТРАШНЫЙ, СТРАСТНЫЙ, НАСТРОИТЬ, НАСТРАИВАНИЕ, НАСТРАИВАТЬ
Interlingual Index
1 i76269 fear.n.01 (fear, fearfulness, fright) [Определение Wordnet: an emotion experienced in anticipation of some specific pain or danger (usually accompanied by a desire to flee or fight)]
гипероним
1 БЕСПОКОЙНОСТЬ, БЕСПОКОЙСТВО, ВСТРЕВОЖЕННОСТЬ, ОБЕСПОКОЕННОСТЬ, ТРЕВОГА, ТРЕВОЖНОСТЬ, ТРЕВОЛНЕНИЕ [Понятие Рутез: тревога, беспокойство]
гипоним
1 ИСПУГ, МАНДРАЖ, ПЕРЕПУГ [Понятие Рутез: испуг]
2 ЖИВОТНЫЙ СТРАХ, ЖИВОТНЫЙ УЖАС, ЖУТЬ, УЖАС [Понятие Рутез: ужас, сильный страх]
3 ПАНИКА [Понятие Рутез: паника]

Fig. 1. The Synset *Cmpax (Fear)* in RuWordNet

To identify the vertical of the word *страх* (*fear*), you need to take a look of the level of hyperonymy. As a result of this procedure, we obtain the following picture:

Fear

- Restlessness. Anxiety. Disquietude. Disquiet. Alarm. Uneasiness. Tribulation↓
- Agitation. Excitement. Emotion. Heartbeat. Deep feeling. Unrest↓
- Inner sense. Stirring. Emotional stress. Sensation. Emotional experience. Feeling. Emotion↓
- Mental object↓
- Abstract entity↓
- Constant entity.

The analysis of the location of the words in question in the “Russian Semantic Dictionary” has shown their relevance to the upper tier of nouns with abstract meaning, including the lexical sets “Бытие. Духовный мир. Общество” / “Existence. Spiritual world. Society”. The further path of the semantic hierarchy assumes a route along the lexical tree from its top to the bottom according to the following scheme:

- Spiritual world: Awareness. Morals. Feelings ↓*
- Feelings, emotional states, their manifestations. Features and personality pattern ↓*
- Different feelings, emotional states, their manifestations↓*
- The very emotions, emotional states↓*

The division ends with the lower lexico-semantic sets “Страх, тревога, нетерпение” / “Fear, anxiety, impatience”; “Восторг, радость, веселье, довольство, умиротворение” / “Delight, joy, fun, contentment, pacification”, and others, each of which includes lexemes united by a common meaning, but representing it in differently in their meanings [4].

Ideographic Dictionary of the Russian Language by O. S. Baranov organizes the description of hierarchical relations of emotive names according to a group principle of semantic commonality of words. So, the entry of the word *страх* (*fear*) gives information in the form of special labels and references about generic concept ▲ *Чувство* (*feeling*), to a differential seme (specific difference) ↑ *Опасность* (*danger*), explanation of the reference *Страх* (*Fear*) via generic and specific semes, indications of partner adjective that forms a collocation (*панический страх* (*panic terror*)), inclusion of the group of synonyms with the derived forms and contexts of use, the notions *Робость*, *Трусость*, which include this entry as a specific difference (↓), information about connotations (colloquial, professional use) (Fig. 2):

страх

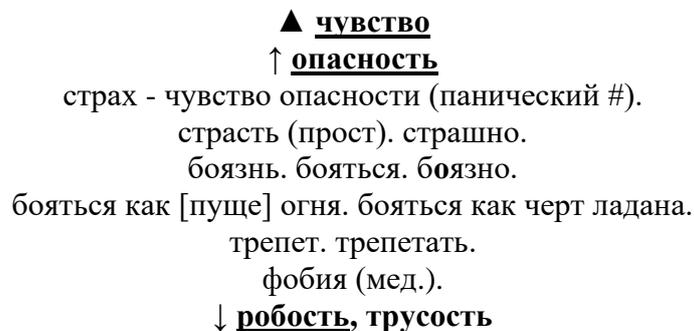


Fig. 2. The Entry of the Lexeme *Страх* in “Ideographic Dictionary” by O. S. Baranov

As you can see, this Ideographic Dictionary presents emotive vocabulary in the form

interrelated concepts-semes via set of references, which reveal various features of a dominant notion. In addition to synonymy and hyperonymy-hyponymy relations, the entry gives information about word-forming, syntagmatic, antonymic, stylistic potentials of a word.

Marking in the NCRL includes the sets of features according to morphological and semantic parameters. The lexeme *страх* (fear) includes the following complex of parameters: *страх* S & (nom | gen | dat | acc | ins | loc) & m & inan r:abstr & (t:psych | t:psych:emot | t:psych:volit), where S stands for belonging to the nouns, the possible case forms are indicated in the brackets, m – masculine, inan – inanimate. From the point of view of semantics, this lexeme indicates the belonging to non-subject names, abstr – the group of abstract names, taxonomic marking – mental sphere, emotions, axiological feature – negative evaluation.

Summary

So, observations on the hierarchical relations of emotive names in the analyzed thesauri, dictionaries, and corpora allow for the following conclusions:

In the upper tier of the classification, there are coincidences in the generalization of these names and their attribution to the abstract sphere, to the world of spiritual culture as a constant essence or a constant of consciousness.

According to Shvedova N. Y., the ideographic approach implies the classification not of words, but lexical units, i.e. meanings, thereby revealing the peculiarities in the interpretation of fragments of reality, refracted in the meanings of words as signs of language, and does not allow to establish a single “ontological” criterion for the classification [4:12].

Our material is proof of this. In the lower tiers there are some differences in the degree of abstraction of concepts.

Differences in the ways of organizing vocabulary in the analyzed sources are observed in the RuWordNet thesaurus, where, for example, the word *страх* (fear), in addition to being related to the level of *Внутреннее ощущение* (Inner Sense), also goes up to the mental sphere in a vertical cut; in the “Russian Semantic Dictionary” and in the “Ideographic Dictionary of the Russian Language” it is referred to the class of words with the general name “Чувства, эмоциональные состояния...” (“Feelings, emotional states ...”).

In the lower tier, at the level of hyponymy, differences in subordination are also observed: *Испуг, мандраж, перепуг, животный страх* (fright, nervous trembling, alarm, bodily fear) in RuWordNet, *робость, трусость* (timidity, cowardice) in the Ideographic Dictionary, etc.

The semantic differentiation of emotive names presented in different dictionaries is determined, in our opinion, by different approaches and methods used by the compilers in the process of systematizing elements.

The evaluation of semantic relations between words in extra-large corpora reveals peculiarities in the systematization of the material. In RuWordNet the Russian Semantic Dictionary, the Ideographic Dictionary lexical units are arranged according to the principle of hierarchical relations based on logical-semantic subordination: when a specific is included in a generic or, on the contrary, a generic includes a specific. At the same time, within the framework of RuWordNet “synset can be considered as a representation of a lexicalized notion (concept)”, in which the connection with the conceptual categories of consciousness is realized; and also as “a union of a word denoting one concept with the meanings of other words

(synonyms), the lexical meanings of which form together the lexical meaning of the very word” [7]. Thus, RuWordNet is not only a lexical resource, but also an ontological one, having the principle of a semantic web with a network of connected nodes and arcs, the purpose of which is automatic information processing.

Conclusion

Anthropocentrism and systemic centricism, noted in modern ideographic dictionaries and thesauri, are embodied in the interpretation of lexical system through the prism of human perception and in the description of not an abstract system, but a living language, given in real use and targeted the the linguistic consciousness of contemporaries.

The anthropocentric vector is observed in considering the needs of society, and in evaluating the addressee’s factor, and in the cognitive approach to describing vocabulary, which makes it possible to get an objective idea of the organization of a person’s lexicon, of the totality of knowledge stored in their memory.

Acknowledgements

This research was financially supported by RFBR, grant № 18-00-01238 and by the Russian Government Program of Competitive Growth of Kazan Federal University.

Bibliography

- [1] Nikitin M. V. The Course of Linguistic Semantics: Manual. SPb: The Publisher of A. I. Hertsen RGPU, 2007. 819 p.
- [2] Thesaurus of the Russian Language RuWordNet. <https://ruwordnet.ru/ru>. (access date 10.26. 2020).
- [3] National Corpus of the Russian Language. <https://ruscorpora.ru/new/corpora-sem.html>. (access date 10.26. 2020).
- [4] Russian Semantic Dictionary. Explanatory Dictionary, systematized according to classes of words and meanings / RAS. Institute of the Russian Language; Edited by N. Y. Shvedova. M.: Azbukovnik, 2003. In 6 volumes. V.3. 720 p.
- [5] Ideographic Dictionary of the Russian Language. Compiled by O. S. Baranov. 1995. <http://rus-yaz.niv.ru/doc/ideographic-dictionary/index-215.htm#215>. (access date 10.26. 2020).
- [6] WordNet: An electronic lexical database. Fellbaum, Ch (ed.). MIT Press, Cambridge, MA (1998).
- [7] M. Kopotev. “Introduction to Corpus Linguistics”. Praga, Animedia Company. 230 p. 2014.
- [8] T. McEnery, A. Hardie. “Corpus Linguistics: Method, Theory and Practice”. Cambridge University Press. 292 p. 2011.
- [9] Graeme Hirst. “Ontology and the Lexicon” // Steffen Staab, Rudi Studer (eds). Handbook on Ontologies (2nd edition). Springer, 2009.
- [10] Lukashovich N. V. Thesauri in Information Search Tasks. M.: The MSU Press, 2011. 510 p.
- [11] V. Solovyev, G. Gimaletdinova, L. Halitova, L. Usmanova. “Expert Assessment of Synonymic Rows in RuWordNet”. In: van der Aalst W. et al. (eds) Analysis of Images, Social Networks and Texts. AIST 2019. Communications in Computer and Information Science, vol. 1086, pp. 174–183. Springer, Cham (2020).

Usmanova Liliya Abrarovna

1. 1977
2. PhD in Philology
3. Kazan State University named after V. I. Ulyanov-Lenin, 2001 г.
4. A defence of PhD thesis *Semantics and Compatibility of the Names of Weathers and Seasons in the Language of I. A. Bunin's Works* in 2004
5. Associate Professor of Russian and Its Teaching Methods, Institute of Philology and Intercultural Communication of Kazan Federal University
6. Semantics and cooccurrence of linguistic units; discursive syntagmatics.

Erofeeva Irina Valerievna

1. 1968
2. Doctor of Philology
3. Kazan State University named after V. I. Ulyanov-Lenin, 1990
4. A defence of doctoral thesis *Nominal Word-Formation in Russian Chronicles* in 2011
5. Professor of Russian and Its Teaching Methods, Institute of Philology and Intercultural Communication of Kazan Federal University
6. The history of Russian, semantic transformations of linguistic units in terms of synchrony and diachrony.