

# On the problem of justification of the term "ethnothesaurus"

Daria Domashevskaya<sup>\*1</sup>, Alexandra Kosheleva<sup>1</sup>, and Ivan Solsoev<sup>1</sup>

<sup>1</sup>Irkutsk State University, 664003 Irkutsk, Russia

**Abstract.** The article presents the main provisions of the theory of thesaurus-type dictionaries, gives definitions of the thesaurus, describes the relationship of the thesaurus with the type of writing, considers the main provisions of the Sapir-Whorf theory of linguistic relativity and their connection with the way of thinking and methods of conceptualizing extra-linguistic reality by representatives of different ethnic cultures. The authors propose the thesaurus of ideographic writing as one of the models of ethnocognitive science and, using the example of the dictionary "Shuo wen jie zi", substantiate the expediency of using the term "ethnothesaurus".

## 1 Introduction

Language, culture and thinking are closely related. Back in the 19th century, W. von Humboldt asserted that "Language is an organ that forms thought ... The activity of thinking and language are therefore an indissoluble unity" [4]. Language determines thinking and at the same time is a reflection of how a person sees the world, that is, it is a kind of reflection of his picture of the world. The language of an ethnic group, accordingly, gives us an idea of how the speakers of this language conceptualize the world. Thesauri are of particular interest in studying and comparing pictures of the world of different ethnic groups from the point of view of reflecting the vocabulary of the language. A thesaurus is a dictionary in which the entire set of thematically grouped vocabulary of a particular language is presented. Reflecting the lexical volume of the language, the thesaurus includes semantic connections and correlations of lexical units and, thereby, makes it possible to analyze and build a model of the picture of the world of a particular ethnic group. That is, we can assume that the thesaurus is one of the models of the systemic representation of the linguistic picture of the world. And here the question arises: is it possible, within the framework of the ethno-linguocultural paradigm, to introduce the term "ethnothesaurus"?

The term "thesaurus" is used in several scientific fields: in computer science, the theory of artificial intelligence, in psychology, in linguistics. In this article, the term "thesaurus" is considered as a linguistic term.

If we turn to the definitions given by dictionaries, then, for example, in Ozhegov's explanatory dictionary, the thesaurus is: "1) a language dictionary, which sets the task of fully reflecting all of its vocabulary; 2) a dictionary or set of data, fully covering the terms, concepts of any special area" [8]. That is, according to this definition, a thesaurus is, first of all, a dictionary in which the entire lexical composition of a particular

language must be represented. The second definition is narrower, here the thesaurus reflects only the terminology of a specific area.

In the Great Soviet Encyclopedia, a thesaurus (from the Greek *thesaurós* - "treasure, treasury") is "a set of semantic-expressing units of a certain language with a system of semantic relations set on it" [14]. That is, the thesaurus actually defines the semantics of the language (national language, the language of a specific science, or a formalized language for an automated control system) [14]. The difference between the definition given in the Great Soviet Encyclopedia and the one considered above is that here the thesaurus is defined as a dictionary that reflects not only the lexical composition of the language, but also the semantic relations between the units of the lexical composition.

The Big Encyclopedic Dictionary defines the thesaurus as "1) a dictionary in which the words of the language are presented as fully as possible with examples of their use in the text; 2) a dictionary in which words related to any area of knowledge are arranged according to a thematic principle and semantic relations (genus-specific, synonymous, etc.) between lexical units are shown" [12]. The first definition given in Big Encyclopedic Dictionary is similar to the definition given in the explanatory dictionary of Ozhegov. Both in the Ozhegov dictionary and in the Big Encyclopedic Dictionary, a thesaurus is a dictionary that reflects the entire vocabulary of a language. The difference is that the Big Encyclopedic Dictionary definition specifies that the thesaurus contains vocabulary with examples of use, that is, the vocabulary of the language is presented in the thesaurus in connection with other lexical units, which also reflects the semantic relations between the lexical units of the language and in this part of this definition close to the definition given in TSB. The second definition of the thesaurus, given in the Big Encyclopedic Dictionary, is narrower, defining a thesaurus as a dictionary in which lexical units related to

<sup>\*</sup>Corresponding author: domashevskayad@gmail.com

any field of knowledge are presented. This definition is similar to the definition given in the Ozhegov dictionary, but only in that part of it, which refers to the thesaurus as a dictionary of words in a certain area of knowledge (terminology). Further, the definition of the thesaurus, given in the Big Encyclopedic Dictionary, has an important clarification: the words are located in the thesaurus according to the thematic principle, indicating the semantic relations.

From the considered definitions, the most fully reflecting the essence of the thesaurus, in our opinion, are those that include the presence in the thesaurus, in addition to lexical units, reflection of semantic connections and relations between these units. At the same time, in all the considered sources, the thesaurus is defined, first of all, as a dictionary.

A dictionary is a collection of words organized in a certain way, usually with comments assigned to them, which describe the peculiarities of their structure and / or functioning [1]. That is, a dictionary is a vocabulary arranged according to a certain principle.

## 2 Methods and material

This study uses linguocultural approach in analyzing different types of thesaurus, methods of observation and interpretation, methods of classification and modeling. Method of linguoculturological reconstruction helps to arrive at the ethnolinguistic level of the ideographic language system. Method of linguoculturological decoding of cultural information within Chinese character helps to prove that lexical meaning of linguistic signs of the language with ideographic type of writing can accumulate a significant extend of cultural information such as ethical, ritual, religious, mythological, philosophical information.

## 3 Results and discussion

Speaking about the approaches to the arrangement of vocabulary in dictionaries, two main approaches in modern lexicography should be noted: 1) by the proximity of sound and 2) by semantic proximity [8]. The first principle of the arrangement of vocabulary - according to the proximity of sound - is the basis of dictionaries, the vocabulary in which is arranged in alphabetical order. According to this principle, vocabulary is located, for example, in explanatory dictionaries, encyclopedic dictionaries, and terminological dictionaries. The second approach to the arrangement of vocabulary is implemented in dictionaries of the ideographic type (from the Greek idea - "concept, idea, image" and grapho - "I write"). The vocabulary in ideographic dictionaries is grouped according to the principle of semantic proximity.

The thesaurus is essentially an ideographic vocabulary. In addition, thesauri refer to the onomasiological type of dictionaries, that is, they represent a connection from concept to word, placing words according to semantic proximity [6]. In particular, an example of a thesaurus can be considered the first ideographic dictionary of the Chinese language - "Shuo

wen jie zi" (說文解字), dated around the 1st century BC. - 1st century AD. The keys are broken down according to formal and semantic principles; signs are sequentially located that are similar to a common graphic element, as well as tracing a common semantic outline. A similar principle for organizing the thesaurus was defined by D. Domashevskaya as semantic and semantic-derivational [13].

The word, being a unit of language, has a sign nature. According to the theory of F. de Saussure, the word-sign is a two-sided unity of the "signifier" and the "signified", which is related to the subject of reality. The sound shell of a word is a plane of expression, the concept that this word reflects is a plane of content. The main function of the word is to highlight and designate objects of the cognizable world. As a consequence, a set of words can be viewed as a kind of reflection or correlate of reality [8]. In addition, the word is an expression of the process of human cognition of the surrounding world, the process of thinking. Language is inextricably linked with thinking.

The connection between language and thinking has long been the subject of research by scientists. Back in the 19th century W. von Humboldt argued that language and thinking, as well as language and culture, are one. Humboldt believed that thinking depends on language, while language, in turn, forms a kind of "intermediate world" between reality and thinking. W. von Humboldt's approach to the study of language was anthropological, language should be studied not as an independent separate system, but in close connection with the consciousness and thinking of a person, with his culture.

Later, W. von Humboldt's ideas about the relationship between language, thinking and culture found their continuation in the works of such linguists as F. Boas and his student E. Sapir. American scientist and anthropologist F. Boas is the founder of ethnolinguistics, the field of linguistics that studies language in its relationship with culture. One of the main problems of a cognitive nature, which scientists are engaged in within the framework of ethnolinguistics, is the study of the process of reflection in the language of the cultural ideas of the people speaking this language. Ethnolinguistics studies how speakers of different languages conceptualize extra-linguistic reality and raises questions: do objects of the world that are universal for all nations and, accordingly, universal concepts exist? To what extent does a specific language affect the perception of the surrounding reality and the formation of a picture of the world of its carriers? In addition, ethnolinguistics also explores the means and forms of communication characteristic of a particular social or ethnic group.

The central concept of ethnolinguistics is linguistic relativity. The hypothesis of linguistic relativity ("the Sapir-Whorf hypothesis") arose on the basis of the ideas of F. Boas and E. Sapir and was formulated by B. Whorf, a student of E. Sapir. According to this hypothesis, the peculiarities of thinking and the system of concepts that exist in the consciousness of a person are determined by the language of which this person is.

That is, the picture of the world of a nation is formed by the language of which this nation is. Accordingly, based on this hypothesis, it can be assumed that speakers of different languages will have different worldviews, systems of concepts, way of thinking and ways of conceptualizing extra-linguistic reality. For example, in English, there are two concepts to designate the parents of the father and the parents of the mother: *grandmother* and *grandfather*. When considering the designation of similar degrees of kinship in Chinese, we find that the grandmother on the father's side is 祖母 (the morpheme 祖 *genus* (丷 *altar of ancestors* + 且 *male genital organ*) + the morpheme 母 *mother*) or 奶奶, where the morpheme 奶 *milk, feed breastfeeding* consists of 女 *female* + 乃 *female breast; yours* (the possessive pronoun of the ancient Chinese language); grandmother on the mother's side - 外祖母 (外 *external* + 祖 *genus* + 母 *mother*) or 姥姥, where the morpheme is 姥 *elderly woman, mother* consists of 女 *woman* + 老 *old*; father's grandfather - 祖父 (祖 *genus* + 父 *father; hand holding the symbol of family power*) or 爷爷, where the morpheme 爷 *father; lord, master* consists of 父 *hand holding the symbol of family power; father* + 丩 is an outdated symbol of power and social status; grandfather on the mother's side - 外祖父 (外 *external* + 祖 *clan* + 父 *father; hand holding the symbol of family power*).

We see that in the English and Russian languages there is no division into kinship along the line of the father and along the line of the mother, and there is also no reference to the clan hierarchy, namely, to the continuation of the clan along the line of the father. Both the grandmother on the father's side and the grandmother on the mother's side are denoted by the same concept – grandmother. In Chinese, different concepts are used to designate relatives on the father's line and on the mother's line, generally marking the mother's blood as "external, foreign", and the blood on the father's line as "the blood of the clan, one's own". The very inner form of words is determined by the ethnocultural scenario that took shape in the course of the historical development of the Chinese ethnoses.

Even at the grammatical level, the level of the Chinese written sign, we can observe how the addition of external blood of other genders is marked in the hieroglyph. As an example, let us give the last of the 540 keys of the dictionary "Shuo wen jie zi", presented in the collective work of representatives of the Irkutsk grammatical school - Etymologic Dictionary of the Basic Chinese Characters:

540. 亥 *hài*<sup>a</sup>

《說文解字》：亥，蓂也。十月，微陽起，接盛陰。從二，二，古文上字。一人男，一人女也。從乙，象裏子咳咳之形。《春秋傳》曰：「亥有二首六身。」

<sup>a</sup> Dictionary entry and commentary are given according to the Etymologic Dictionary of the Basic Chinese Characters [13]

"Sho wen jie zi": *roots of herbs, bulbs. The tenth lunar month, weak yang rises, meets mature yin. The key 二, in guwen 上 top. 一 man, 一 woman. The key 乙, as if holding the laughter of a child. "Chunqiu": "亥 has two heads, six bodies."*

The commentary: 亥: <...> the sources describe

A similar picture is observed, for example, with the designation of brothers and sisters. In Russian, there is the concept of a *sister*, in English - a *sister*, in Chinese, an *older sister* - 姐姐, where the morpheme 姐 consists of morphograms 女 *woman* + 且 *male genital organ*; the first meaning of the sign 姐 is recorded in "Shuo wen jiezi" and reads: 蜀謂母曰姐, 淮南謂之社,<sup>c</sup> which means *the Shu people called their mother 姐, in Huainan they called her 社; a younger sister* - 妹妹, where the morpheme 妹 a *younger sister* consists of morphograms 女 *woman* + 未 *treetop*. To denote a brother in English, the concept of *brother* is used, in Chinese an *older brother* - 哥哥 (哥 *older brother*, an ideogram consisting of two morphograms 可 *happy, contented*; in Shuo wen jiezi" 哥 is defined as 聲也。从二可。古文以為調<sup>d</sup>, i.e. "sound; consists of two 可 *to give voice*, in guwen was used instead of the sign 歌 *a song, to sing*", in the Tang era, the sign 哥 was one of the calls to a word *father*<sup>e</sup>; younger brother - 弟弟: morpheme 弟 the younger brother is interpreted according to "Shuo wen jie zi" as 弟韦束之次第也, which means *dressed skins tied in order*<sup>f</sup>.

Thus, when naming older children in the family, their closeness to the authority of their parents was emphasized, requiring appropriate attitude and respect on the part of the younger generation (a kind of manifestation of one of the Confucian virtues –filial

<sup>b</sup>[13, 650-651]

<sup>c</sup><https://www.zdic.net/hans/%E5%A7%90>

<sup>d</sup><https://www.zdic.net/hans/%E5%93%A5>

<sup>e</sup><https://www.zdic.net/hans/%E5%93%A5>

<sup>f</sup><https://www.zdic.net/hans/%E5%BC%9F>

piety 孝) That is, in addition to indicating the degree of kinship, in the Chinese language there is also an indirect indication of age, involvement in being, inherent in the life of the majority representatives of the Chinese ethnos, ethnocultural characteristics, traditions and ritual. Based on these examples, it can be assumed that the method of categorizing the extra-lingual space of a Chinese speaker is more detailed than, for example, a native speaker of Russian and English. That is, in the process of categorizing the surrounding reality by a person in the presence of uniform initial data (in our case, kinship relations), dissimilar ethnocognitive spaces are formed in different linguistic pictures of the world, which are recorded in the thesauri of a particular language.

Speaking about language as a way of conceptualizing extra-linguistic reality, it is necessary to recall its important component –the material component – writing. It is in writing that a person's ability to fix the processes occurring at the mental level is realized: cognition of the surrounding world, the name of the objects of this world, the formation of an “ethno-linguistic picture of the world” [2, 8].

Being a kind of tool for fixing the human experience of categorizing the extra-linguistic space, writing at the same time contributes to the formation of this space: "Being in general the main way of preserving human thoughts, writing is also a fact of individual ethnic culture, largely shaping its ethnocultural and ethnosemantic appearance" [2, 6].

In addition to the fact that writing is a way of preserving information and takes part in the formation of the ethnocultural image of a particular nation, writing also participates in the “semantic division of the world” [2, 7]. That is, writing helps not only in designating objects in the world, but also carries a meaningful function.

Today, there are several types of writing systems, among which the most common are the alphabet and ideographic writing. The ideographic writing system preceded the emergence of the alphabetical one.

Writing can be viewed as an integral semiotic system, which has a "content plan", "expression plan", "syntagmatics" and "paradigmatics" [10, 30]. The content plan of writing as a semiotic system is the unity of units of different levels of language (sound, syllabic, verbal, numerical, etc.), denoted in various written systems by the corresponding graphic symbols, the specific set of which is the "expression plan" of a particular writing system. The expression plan also includes the specific names of the graphic symbols of a particular system. In addition, the content plan includes semantic associations between the signs (letters) themselves and their names on the one hand, and various cultural phenomena - beliefs, rituals, myths on the other. The syntagmatics of writing is the sequence of characters in the text, paradigmatics is the ratio of graphic characters in the system and their location in the system relative to each other.

Consideration of writing as a semiotic system with a "content plan" and "expression plan" gives grounds for the possibility of a typological classification of writing

[1, 9]. If the classification is based on the "content plan" of writing, then two main types of writing can be distinguished: semasiography (or ideography) and phonography. Semasiographic writing systems are characterized by the designation of a specific concept, and not the sound side of the language. That is, in written systems of this type, the "content plan" of the language, which is expressed in words and phrases, is displayed directly in written signs, which, in addition to being written signs, designate some universal conceptual categories. If we talk about the "plan of expression" of this type of writing systems, then it should be noted that the written signs of these systems can be graphically similar to those objects of the real world that they designate. Ideographic writing is a type of writing in which signs convey certain meanings through images of specific objects associated with these meanings. Such systems are also characterized as “pictographic”. If the connection between the “signifying” sign of the ideographic system and the corresponding conceptual category is conditional, then this type of system is “conventional” or “conditional” [1, 10].

Phonographic types of writing include those in which the "signifying" signs are related to the sound side of a particular language, and not to universal conceptual categories. “Signified” in these types of writing are not concepts, but words, or linguistic units of a lower level – individual syllables or sounds [1, 11]. In phonographic systems, writing is related to speech. In such systems, the "plan of content" is the sound form of the language, that is, a word, syllable, or a separate sound are considered as "signified" graphic signs of the writing system. The phonographic system, in which written signs express individual sounds, is called an "alphabet system" or "alphabet" [1, 11].

## 4 Conclusion

Our attempt to analyze the signs of some kinship terms in the existing spaces of writing (phonographic and ideographic) illustrates the importance of taking into account the grammatological aspect of the existence of a national language thesaurus, and the thesaurus of ideographic writing itself can be considered as one of the models of ethnocognitive science.

The space of writing of this or that nation in all its originality and ability to represent the current ethnolinguocultural vision of the surrounding reality is, in fact, the "treasury" of this or that ethnolinguoculture. The letter itself is an axiological link of culture. If we consider the space of a certain ethnolinguoculture as one of the subject areas of modern linguo-cognitive research, then the writing of this ethnolinguoculture can well be defined as “one of the most effective tools for describing” this area. This applied instrumentality manifests itself in a special way when considering the ideographic type of writing, where the grammatological aspect of the functioning of linguistic units plays an important role in the process of studying ethnolinguocultural signs (myths, rituals, dictionaries, canons, hyponyms, hyperonyms, etc.). Thus, it seems necessary to introduce a term that contains the entire

descriptive potential of one of the tools for studying the ethnolinguocultural space. In our opinion, the most appropriate term would be "ethnothesaurus". By ethnothesaurus, we mean a set of information that fully covers the functioning of one or another ethnolinguocultural space. So, speaking about the ethnothesaurus of the written space, we mean the entire set of written signs of one or another ethnic group, which is already an intra- and interethnic means of communication, more broadly – a tool for studying the processes of linguocognitive activity of representatives of this ethnos.

## References

1. O. S. Akhmanova, *Dictionary of linguistic terms*, 608 (1969)
2. T.V. Gamkrelidze, *Alphabetic writing and ancient Georgian writing (Typology and origin of alphabetic writing systems)*, 349 (1989)
3. O. M. Gotlieb, *Fundamentals of Chinese grammar*, 284 (2007)
4. W. von Humboldt, *Selected Works on Linguistics*, 400 (2000)
5. D. M. Domashevskaya, *The communicative space of writing: phonography vs. ideography / Gottlieb readings: Asia-Pacific region in the context of global development: materials of the Intern. scientific - practical conf.*, 95 – 107 (2017)
6. V.A. Zvegintsev, *On the scientific heritage of Wilhelm von Humboldt* URL: <http://www.philology.ru/linguistics1/zvegintsev-84.htm/> (date of access 13.09.2021)
7. V. A. Zvegintsev, *Theoretical and linguistic premises of the Sapir-Whorf hypothesis. New in linguistics*, URL: <http://www.philology.ru/linguistics1/zvegintsev-60a.htm/> (date of access 13.09.2021)
8. I. Laftimi, *General characteristics and criteria for the classification of thesaurus dictionary type / Young scientist*, **12** (1), 252-255 (2011)
9. A. M. Maksimenkova, *On some aspects of alphabetic systems / Bulletin of Irkutsk State Linguistic University. Series "Philology"*, **4** (21), 87 - 93 (2012)
10. V. V. Morkovkin, *Ideographic dictionaries*, URL: [http://rifmovnik.ru/ideog\\_book.htm/](http://rifmovnik.ru/ideog_book.htm/) (date of access 10.09.2021)
11. S. I. Ozhego, N.Yu. Shvedova, *Explanatory dictionary of the Russian language*, URL: <http://www.ozhegov.com/words/35551.shtml/> (date of access 13.09.2021)
12. Yu. S. Stepanov, S. G. Proskurin, *Constants of world culture. Alphabets and alphabetic texts in the period of dual faith*, 147 (1993)
13. T.E. Shishmareva, E.A. Khamaeva, N.V. Terekhova, D.M. Domashevskaya *Etymologic dictionary of the basic Chinese characters*, 654 (2019)
14. *Great Soviet Encyclopedia*, URL: <http://bse.sci-lib.com/article109420.html/> (date of access 13.09.2021)
15. *Big Encyclopedic Dictionary*, URL: <http://alcala.ru/entsiklopedicheskij-slovar/slovar-T/70083.shtml/> (date of access 13.09.2021)