

PRINCIPLES OF DEVELOPMENT OF THE LINGUISTIC DATABASE OF TATAR COLOUR TERMS

A.M. Galieva
Kazan (Volga region)
Federal University,
Kazan, Russia
Institute of Applied
Semiotics, Tatarstan
Academy of Sciences
amgalieva@gmail.com

A.F. Sitdykova
Institute of Applied
Semiotics, Tatarstan
Academy of Sciences,
Kazan, Russia
sitdikovaa@gmail.com

D. Sh. Suleymanov
Kazan (Volga region)
Federal University,
Kazan, Russia
Institute of Applied
Semiotics, Tatarstan
Academy of Sciences
dvdt.slt@gmail.com

ABSTRACT

This paper presents an approach to the development of the database of Tatar colour terms. The authors describe the main principles of the database building, the criteria of selection of basic and secondary colour terms, and the mapping relations between the words. The list of basic Tatar colour concepts and the scheme of the entry are also provided.

1. Introduction

The construction of a conceptual model of a database is one of the most important stages in the creation of any linguistic database. Therefore, a significant portion of time was spent on the development of the principles that will determine further selection of concepts used by domain experts, and on the search of the correct relationships between different concepts. Designing a conceptual model of a particular database depends on the problem domain and the target audience. Also, revealing these principles plays an important role in the process of filling in the dictionary entries.

In this paper we present the main principles of the construction of the Tatar colour terms database as a structured collection of Tatar colour words, their meanings, derivation potential, and use in different types of collocations and texts.

In the cognitive science and linguistics, the colour is regarded as one of the most important categories in the comprehension of the world around the human being, as well as a significant cultural code. Colour perception, being by nature a priori neutral, in the language becomes the basis for the formation of categories of higher order.

The main objective of the project is to represent to the greatest possible extent full information on the semantics, structure and functioning of Tatar colour terms and, potentially, the colour terms in other Turkic languages, so the database will become a kind of encyclopedia of Turkic colour words.

There are different approaches to classification of colours and study of colour words in the scientific literature. One of the first research works proposing a classification of colour vocabulary was a book by B. Berlin and P. Kay *Basic Color Terms: Their Universality and Evolution* (1969). Basing on the study of 98 different languages, the authors developed principles, which illustrate the formation of colour words. The authors also compiled a universal basic list of colour words appearance: for example, all languages have words for black and white colours, if they have three colour words, then they also have name for the red colour [1].

A polish linguist Anna Wierzbicka, known for her work in the field of semantics, dwells on the question of basic colours and finds their origin in the natural prototypes with which these colours are associated [2]. In the Russian linguistics colour names were studied by V.G. Kulpina [3].

In current Tatar linguistics the study of colour terms is limited to certain papers, as a rule, on the functioning of colour terms in an isolated text and onomastics.

In the XX-XXI centuries various aspects of colour space were examined by the turkologists A.Kononov [4], L.Gabysheva [5], O.Molchanova [6]. Fundamental works of V. Khakov became the basis for the study of colour vocabulary in the Tatar language [7]. Individual studies by other linguists played also an important role in the development of Tatar colour terms research.

2. Principles and sources of building of the colour terms database

A word-list is a kind of skeleton of any linguistic database; it reflects thematic

originality, completeness of terms selection and other significant features of the project. In developing of the word-list of a database word selection plays a key role (which terms are included in the lexicon, and which remain outside of it). In the course of compiling the database of Tatar colour terms we have selected 271 colour words of different types.

There are many approaches to the selection of basic colours and arrangement of colour terms. Colour verbalization in the language has its own peculiarities, and the basic colours from the perspective of linguistics do not correspond in a number of characteristics to the optical basic colours allocated for objective physical (technical) parameters. For example, a spectrally important term *violet* has received very limited use in the Tatar language and is characterized by relative instability (the Tatar language has no single term for this colour, so the synonyms *šämäkhä* (violet) and *miläwšä* (violet) are characterized by irregular use). At the same time, the term for the synthetic colour *aq* (white), which is produced by the combination of all the colours of the spectrum, is associated with a large number of natural and cultural phenomena, and is widely used. Such facts suggest that cognitive processes in the human mind have links to real physical phenomena.

Nevertheless, concept formation in the language is much more closely related to the environment and everyday phenomena than scientific models, and colour concepts, which are often peripheral in terms of optics, are at the centre of the world-view of a particular linguistic system.

The main principles of the word-list compiling for the linguistic database of Tatar colour terms are:

- the optimal reflex of the lexical composition of the group of words denoting colours in the Tatar language; selection of basic and secondary colours from the viewpoint of the Tatar language;
- a close correspondence between the structure of the entry and the structural and semantic features of Tatar colour terms;
- the reflection of the frequency of use of colour terms in the colloquial speech, as well as in the language of folklore and fiction;
- the existence in the Tatar language of concepts reflecting the content of colour terms, which are included in the word-list of the database.

There is quite a large number of dictionaries and thesauri of various kinds in the Tatar lexicography. The glossary of the Tatar colour terms database was formed on the source material of the following dictionaries of the Tatar language:

- the Tatar explanatory dictionary, in 3 volumes, 1977-1981 [8];
- the Tatar explanatory dictionary, in 1 volume, 2005 [9];
- the Tatar-Russian dictionary, in 2 volumes, 2007 [10];
- the Dialectological dictionary of the Tatar language, 2009 [11].

The real frequency and collocability of colour words is tested on the data of the Tatar National Corpus [12].

The glossary of the database contains only forms and metaphorical senses confirmed by the dictionaries and the Tatar National Corpus data (a number of theoretically possible derivatives is much bigger).

The study of the semantic structure and grammatical characteristics of the word is an important stage during the linguistic database development, since such a study promotes the determination of correct

relationships among different concepts and the definition of their semantic boundaries.

3. Selection of basic colour concepts and terms

In the process of compiling the glossary, one of the main problems has been the selection of basic and secondary colours.

The main criteria for selection of the basic colour terms are the following:

- the presence in the minds of the native Tatar speakers of the linguistic concept, directly and immediately identified with a particular colour term, and active use of the respective colour term in the literary and colloquial language;
- the existence in the present-day Tatar literary language of a word producing stem (without additional suffixes) for the colour, which can be used as the basis for the formation of new words;
- the use of colour terms in collocations with a wide range of nouns, and not only with the nouns of the limited group (we have selected a special group of colour words denoting horse colours because of the great significance of the horse in the life and culture of the Tatars);
- comprehensibility of the colour term for the representatives of all the Tatar dialects, in many cases in spite of the presence of equivalent words in the dialects.

A singularity of the Tatar language is the presence of the colour term *al* (pink) among the basic colour terms; this word is commonly used in everyday conversations, folklore and poetry texts. It is not apprehended as a secondary colour (pale red colour, combination of red and white) in the Tatar world-view, but as an independent

colour unit, and the concept *al* has a rich set of associations.

On the basis of the above-mentioned criteria we have identified 12 Tatar basic colour terms: *aq* (white), *al* (pink), *kızıl* (red), *čirän* (ginger red), *sarı* (yellow), *yäşel* (green), *zängär* (blue), *kük* (blue), *šämähä* (violet), *sori* (grey), *körän* (brown), *qara* (black).

Furthermore, in the process of grouping of colour terms selected from the dictionaries we have obtained three additional groups:

- words and word combinations that do not designate a colour in an isolated use; nevertheless, in many contexts such words can express the idea of a certain colour (most of the words and phrases that are included in this group express the colour by means of its comparison with some natural object (*ciyä töäle* – '[dark red] as a cherry'). Such natural objects are usually characterized by a certain colour, however the words denoting them have not formed an isolated colour meaning, regardless of the initial object);

- words denoting a certain colour of animals (by virtue of the way of life and the culture of the Tatar people, most of the words of this group are words denoting the colour of the horse, but some of them are applicable with respect to other animals);

- words denoting not a particular colour, but parti-colouredness, diversity of colour, a mixed character of object's colour consisting of fragments of different colour tones (*cuar* – 'piebald').

The words out of the standard language but widespread in dialects, were also included in the family of basic and additional groups of colour terms. Dialect words have been selected on the basis of the material of the Big dialectological dictionary of the Tatar language compiled by F.S.Bayazitova, D.B.Ramazanova, Z.R.Sadykova and T.Kh. Khayretdinova [11]. Dialect vocabulary has

formed a significant body of the linguistic database.

The derivatives of the basic colour terms were divided into some separate groups according to their meanings: compound colour terms, forms expressing the colour strength, and forms expressing hues.

The Table 1 represents the structure of the entry of the database:

Table 1. The scheme of the entry

1	Colour terms	Examples
1.1	Colour term in the literary language. Definition of the basic terms from explanatory dictionaries	<i>Aq</i> (white)
1.2	Dialectal names of the word	<i>agaylı</i>
2	Derivatives	
2.1	Paired colour words (for equitable colour components)	<i>Aqlı-qaralı</i> (white-and-black)
2.2	Colour strength forms	<i>Ap-aq</i> (very white), <i>aqsıl</i> (whitish)
2.3	Combinations expressing colour tints (hues)	<i>Küksel aq</i> greyish white
2.4	Derivative nouns	<i>Aqlıq</i> (whiteness)
2.5	Verbal derivative	<i>Agartu</i> (whitewash)
3	Scope of denotation	
3.1	Natural phenomena	<i>Aq balcıq</i> (kaolin)
3.2	Social phenomena	<i>Aqlar</i> (the Whites)
4	Proper names	<i>Aqbulat</i>

However, not all the database terms have all three types of entries. For example, hues often absent in dictionaries; the words

šämähä (violet) and *sori* (grey) have no derivatives expressing the colour strength.

The basic colour terms and their derivatives are given alphabetically in the glossary. Most of the words are translated into Russian, which will expand noticeably the target audience of the glossary.

The biggest number of derivatives and dialectic variants are related to the colour term *aq* (white) - 29 units.

4. Conclusion

So, the glossary of the database consists of 271 words from monolingual and bilingual dictionaries and 12 basic colour terms. The database is supposed to accumulate all the cognitive information related to colour concepts in the Tatar language, as well as to provide their equivalents in other Turkic languages. It constitutes a valuable source of information for future studies in the fields of linguistics, translation, cognitive science and culture studies, as well as other peripheral areas such as computer linguistics and pedagogics.

The work is supported by the Russian Foundation for Basic Research (project # 14-06-31362-mol_a).

5. References

- [1] **BERLIN, B.** and **KAY, P.**, Basic Color Terms: Their Universality and Evolution, University of California Press, 1969.
- [2] **WIERZBICKA, A.**, Why there are no 'colour universals' in language and thought? In: Journal of the Royal Anthropological Institute Volume 14, Issue 2, p. 407–425, June 2008.
- [3] **KULPINA, V. G.**, Linguistics of colour: colour terms in the Russian and Polish Languages, Moscow, 2001 (in Russian).
- [4] **KONONOV, A. N.**, The semantics of colour terms in the Turkic languages // Turcological Collection, 1975 (in Russian).
- [5] **GABYSHEVA, L.L.** Traditional colour symbolism of Turkic peoples: The red // Actual problems of modern folklore: Proceedings of the International Scientific-Practical Conference (Kazan, June 29, 2009), Kazan, 2009.
- [6] **MOLCHANOVA, O.T.**, Gray, red, blue, intermediate colours in the Onomastikon of the some Turkic peoples In: Areal research on Bashkir dialectology and onomastics, Ufa, 1988, p. 89-103 (in Russian).
- [7] **KHAKOV, V.** Language is the mirror of history, Kazan, 2003 (in Russian).
- [8] The Tatar explanatory dictionary in 3 volumes, Kazan, 1977-1981 (In Tatar).
- [9] The Tatar explanatory dictionary in 1 volume, Kazan, 2005 (In Tatar).
- [10] The Tatar-Russian dictionary, in 2 volumes, Kazan, 2007.
- [11] **BAYAZITOVA, F.S., RAMAZANOVA, D.B., SADYKOVA, Z.R., KHAYRETDINOVA T.KH.**, Dialectological dictionary of the Tatar language, Kazan, 2009.
- [12] Tatar National Corpus // http://web-corpora.net/TatarCorpus/search/?interface_language=en.