

FORMATION OF THE LEXEME'S CORPUS FOR THE LEVEL MODEL DEVELOPMENT OF LANGUAGE

Irina A. Kuprieva, Stanislava B. Smirnova, Tatyiana V. Perutskaya
Belgorod State National Research University, Belgorod 308015, Russia
stacey33@mail.ru

Article History: Received on 25th April 2019, Revised on 20th June 2019, Published on 4th July 2019

Abstract

Purpose: The article deals with the principles of selection of factual material on the problem of fixing language invariants and variants.

Methodology: This is an analytic-dialectical research in which the data was obtained through the analysis of language and grammar in English texts.

Main Findings: As a result of using the system-object method of knowledge representation, a model for securing language variants is constructed. The UFO model of the organizational system can act as a network, production, and frame models. In conclusion, only the internal flows are of interest at the level of the function description, since the algebra of images of the theory of patterns and the theory of graphs are involved in external flows.

Applications: This article can be used by regional schools, cultural organizations and universities.

Novelty/Originality: In this research, a level model of language invariants and variants was developed and presented. Statistical accounting methods, linguistic and computer modeling methods were used on this model.

Keywords: *Language Variability, Lexemes Selection, UFO-Method.*

INTRODUCTION

Studying the variations of a language lies mostly in a way of studying the interaction between different languages on the same territory, or the influence of one language on another. But the present article tries to study the deep reasons for language variability. The problem, following the advanced postulates of modern psycho-, socio- and cognitive linguistics, can be spoken of as a verbal reflection of the mentality and its changes. Thus, by collecting the actual material for subsequent processing and verification, and through comparative analysis, we gain access to consciousness to the linguacultural aspects and isolate factors that particularly affect the appearance of linguistic variability and are not relevant to invariants. In this case, it becomes possible to track language changes and language universals that do not change under the influence of certain external factors. At the same time, the study of the language, or rather, of its changes during a period that is not long-term in the historical aspect, is a rather laborious and complex process that is most obvious and accessible for linguistic processing on the Internet. This determines the choice of a virtual resource as a source of factual material. The results of the project correlate with the description of the national and regional characteristics of the behavioral stereotypes (mentality) of the inhabitants of the UK, affecting the style of everyday life, the positive and negative perception of the contemporary sociolinguistic situation, and the nationality. We suppose to generalize it as the concept of "Muslim World" ([Zhikharev, 2014](#); [Shirvani et al., 2015](#)).

That is why one of the objectives of the study is to collect the factual material that reflects the results of the language changes. The following task is to find out the principles of linguistic formation of the lexemes' corpus. The units of the research thesaurus will be reflected in the relevant database, which will store information on the semantic shifts of the vocabulary in a synchronous slice. A kind of the similar database will be relevant for comparative analysis in describing language changes. In order to process the actual material, a continuous selection of lexemes from authentic sources of factual material (Internet), lexicographical, contextual, conceptual analysis, statistical counting, as well as computer methods for creating a database acquires special significance. Based on the linguistic and non-linguistic factors obtained, due to the methodology of cognitive linguistics (conceptual analysis, methods of linguistic cognitive modeling, computer modeling), a mental structure will be presented as the primary basis for creating a level model of language invariants and variants ([Piteira et al., 2018](#)).

LITERATURE REVIEW

Linguistic aspects to study the language variants

All concepts can be conveyed "by a combination of linguistic means, with each of them revealing only a part of it, to the entire nominative field of the concept" ([Alefirenko, 2004](#): 189-190). The plan for expressing the concept is represented by a number of language implementations that form the corresponding lexical and semantic paradigm, and the content plan attracts semantic features that are joint with all its language implementations that fix the lexico-semantic paradigm and create its conceptual or prototype basis ([Jerichow & Simansen, 2013](#); [Parvizian et al., 2015](#); Javad et al., 2019; Salimi et al., 2019).

Most consistently, scientists defend the understanding of the concept as a kind of mental representation, which is the container of various structures of knowledge about everything that surrounds a person. Margolis and Lawrence in their monograph "Concepts: Core Readings," offer a generalized analysis of the theory of concepts, which covers the five most common concepts in foreign science: 1) Classical; 2) Prototype; 3) Theoretical; 4) Neoclassical; and 5) Atomic ([Margolis, 1999](#)).

Linguistic cognitive concept is an individual and at the same time an inter-object product of cognitive activity and in its ontology is the embodiment of the body sensory-motor and introspective experience of a person formed within a certain linguistic culture. The conceptual analysis done, including etymological and component analysis based on vocabulary definitions, provides an opportunity to present here the core of the concept - the planet that we live on. Thus, the core of the viewed linguocultural concept is the totality of all those people on the planet, whose religion is Islam, as well as their followers. On the basis of the foregoing, we can say that the interpretation of the concept Muslim World unites concepts that are descriptive to this concept. If we combine the conceptual meanings of both words, we can come to the conclusion that in the explanatory dictionary of the English language, the concept of the Muslim world explains how everything describes the phenomenon of the Muslim world in one way or another ([Gareyev, 2016](#)).

Principles for lexemes selection and database formation

An analysis of the lexicographic definitions shows that the focus is on a certain group of people / community / generation, which is linked together by common ideas (the people or things to particular interest, job (OALD); of activity or work and the people who are involved in it (LAAD); a distinctive class of persons or their sphere of interest (MWD)). From these definitions, it also follows that for the existence of the concept of the Islamic world, it is necessary to have a historical period (a particular period in the history and the society and people of that time (LAAD), a particular period of history and the people of that period (OALD) (MWD)), which in turn presupposes the existence of a certain style of life, as indicated by such definitions as (the life of a particular or group of people, especially the things they do and the people they know (LAAD), the particular type of place or situation in which someone lives or behaves (OALD)). In this case, the concept of 'Sharia' refers to a set of religious, legal and domestic norms (in Islam there is no division into religious and secular), which every Muslim should follow - this is the Muslim way of life. Consequently, the conceptual sign lifestyle / culture of Islamists is a correlate of the Sharia. Sharia is the religious as well as the ethical basis of Islam, which will also include work, life, and the rules of behavior of people. The fact may be discussed that the Muslim way of life cannot be achieved without characterizing the activities and lifestyle of Muhammad, the Prophet. The image of Prophet Muhammad and his works constitute the ideals for all believers.

The Islamist way of life, in turn, makes it possible to judge the existence of such a conceptual area as certain specific conditions of realization for a particular way of life (a particular type of place or situation, the kind of life we have (LAAD)). Here it should be explained that the conceptual criterion conditions of implementation serve as a correlate for all possible situations necessary for the implementation of any actions. In order to create such conditions, it is necessary to have certain knowledge and skills, and knowledge in Islam is regarded as a duty and dignity of a person. Throughout his life, a Muslim has to learn new things and strive for science. Education in Islam is seen as the knowledge of the values of Islam and how to improve a person in terms of his morality. But not all knowledge and aspirations to know are welcome in Islam. In the Sunnah, these are the words of the Prophet concerning knowledge: There are four flaws in knowledge: this boasting of their knowledge and a sense of superiority to recognized scientists; Use of knowledge in competition with fools; attempts to achieve through the knowledge of favors on the part of leaders - people who hold an influential position. Realization of the unity of mind (knowledge), soul (emotions) and body (skills and abilities) in Islam is customary called ADAB. ADAB is not only the result of education, but also the method of cognition, through which learners actualize knowledge, attitudes and activities. ADAB will also correlate the mandatory attribute conditions of implementation.

Next, it should be noted that the Muslim world characterizes the environment of the individual, that is, his family, friends (a person's environment, experiences, friends and family (OALD), the society that we live in (LAAD)). This makes it possible to identify such a conceptual area as a society. The main value for the Muslims is the family and the hierarchy of relationships within the family. The foundations of social justice are laid first in the family, and then spread to the whole society. Hence, the obligatory attribute SOCIETY is a conceptual correlate of the society surrounding the Muslim ([Bonami et al., 2004](#)).

Thus, the analysis of the definitions of the Muslim World concept, given in modern, intelligent English-language dictionaries, makes it possible to judge the existence of certain conceptual areas, namely the Islamist lifestyle / culture, the conditions of implementation, and the society. The content of the concept is formed by cognitive signs that reflect the individual features of the conceptualized object or phenomenon, and is described as the totality of these characteristics. The conceptual meaning of the Muslim World concept is revealed in the definitions of emotion associated with the perception of the phenomenon Muslim world that existed in English scientific concepts and theories, create a fragment of the scientific picture of the world and is realized in an appeal to the individual's sensory memory, primarily to that part of the sensory spectrum, which accumulates unpleasant sensations (tactile, visual, olfactory, acoustic). The source of understanding of the Islamic world can also be a violation of moral and ethical standards.

Understanding the systemic-objective method for knowledge representation (somkr)

This research is of particular value in the aspect of rethinking innovative scientific methods for conducting research, namely, the system-organizational method of representation of knowledge and the model of the functioning of semantic dependencies based on it. In order to create a method for representing organizational knowledge in terms of Unit-Function-Object, it is necessary to adapt the models obtained by the system-object UFO approach to the requirements of models of organizational knowledge. For this, the possibilities of traditional models of knowledge (semantic, production and frame) can be used, since each of them individually corresponds to individual elements of organizational knowledge. In order to create a method for representing organizational knowledge in terms of Unit-Function-Object, it is necessary to adapt the models obtained by the system-object UFO approach for the requirements of models of organizational knowledge. For this, the possibilities of traditional knowledge models (semantic, production and frame) can be used, since each of them individually corresponds to individual elements of organizational knowledge. For this purpose, an algorithm for modeling knowledge is considered. The whole process of building a knowledge model in terms of SOMKR can be divided into several stages ([Wu & Palmer, 1994](#)):

- Construction of a hierarchy of stream objects (similar to the procedure for building hierarchy of links in the UFO analysis);
- Description of the final methods of nodal objects with the use of scripting language;
- Use of the model (implementation of logical inference on the model).

METHODOLOGY

A level model of language invariants and variants will be developed and presented. Verification of information for obtaining this model is to take into account the contextual, lexicographical, and conceptual analysis of lexemes and by adopting statistical accounting methods, linguistic and computer modeling methods.

RESULTS

Lexeme is the second unit of language after phoneme, the smallest unit of language that has a grammatical and semantic role and cannot be broken down into smaller semantic units. The main characteristic of the evangelistic is their incommensurability to grammar and their semantic units are smaller. Another criterion for detecting punishments from other linguistic units is that each proclamation is used in its own sense; therefore, it is necessary to pay attention to the independent or affiliated position in terms of grammar, semantics, spellings and phonetics to determine its form and its types. Apart from impulse, other meaningful constructs of the word, sentence, or group can be divided into meaningful elements, while the lexeme is a meaningful unit and a role of commands that cannot be decomposed. This language unit may be made up of a combination of one or more syllables, so the number of syllables cannot be a suitable criterion for detecting punctures.

The phonemes are also different; although some phrases are composed of only one phoneme, the difference in pronunciation with the phoneme is that phonemes of phonetic units are meaningless but are meaningful to the language,

while many phrases have semantic autonomy. The other difference between them is that the number of lexemes in the language, unlike the phoneme, is unlimited.

Therefore, the analysis of the results of adaptation of the UFO approach to the description of structural, functional and objective characteristics of organizational knowledge on examples of representation of traditional knowledge models in terms of the Unit-Function-Object shows that the UFO model of the organizational system (or organizational knowledge) can act as a network, a production, and a frame model. This allows us to consider a method for representing organizational knowledge in terms of the UFO-approach as an integral system-object knowledge representation method (SOMKR), which integrates both the logical, structural, and procedural paradigms of knowledge representation. The considered method of describing organizational knowledge is universal, since it allows us to describe the object characteristics, structural characteristics and functional characteristics of the studied subject area; however, and in no less an innovative way of realizing the presentation of research results ([Losev, 1984](#)).

Let us consider the basic concepts of the given mathematical theory. In accordance with the theory of processes, a process is understood as a model of the behavior of a dynamic system at a certain level of abstraction. The process can be thought of as a graph P (process graph), whose components have the following meaning: The vertices of graph P are called states, and represent situations (or classes of situations) in which the simulated system can be located at various times of its operation. One of the states is selected, it is called the initial state of the process P . The edges of the graph P have labels indicating the actions that the simulated system can perform. The operation of the process P is described by transitions along the edges of the graph P from one state to another. The operation starts from the initial state.

The label of each edge represents the action of the process, performed when passing from the state at the beginning of the edge to the state at its end. In order to specify the process P , which is a model of the behavior of a certain dynamic system, there must be specified some set of 'Act (P),' i.e. actions that the process P can perform. It is assumed that the actions of all processes are elements of some universal set Act of all possible actions that a process can perform, i.e. for any process P , Act (P) is $\hat{\text{Act}}$. The choice of the set Act (P) of the actions of process P depends on the purposes of modeling. In different situations, different sets of actions can be chosen to represent the model of the analyzed system in the form of a certain process ([Galeev & Habibulina, 2017](#); [Kondubaeva et al., 2018](#)).

In the calculation of processes, the set of Act is divided into the following three classes: Input actions, which are represented by a combination of the form $\alpha?$ An action of the form $\alpha?$ is interpreted as input into the process of some object named α . Output actions, which are represented by a combination of the form $\alpha!$ An action of the form $\alpha!$ is interpreted as the output from the process of some object named α . Internal (or invisible) action, which is denoted by the symbol $\alpha\tau$.

An action of the process P , which is not related to its interaction with the environment (that is, processes that are external to the process P , and with which it can interact) is called internal. For example, an internal action can be associated with the interaction of the components of the process P . Actually, the internal actions can be very diverse, but to denote all the inner actions we will use the same symbol $\alpha\tau$. It is a reflection desire not to distinguish between all internal actions, tk. they are not observable from outside the process P . It should also be noted that objects that are introduced into the process and are derived from it can have a very different nature (both material and non-material) ([Bentley & Bossé, 2018](#)). For example, they can be: material resources, people, money, information, energy, etc. In addition, the very concepts of input and output can be of a virtual nature, i.e. the words input and output can be used only as metaphors, but in reality, no input or output of any real object can and does not occur. Informally speaking, the operation of the process P is considered as:

- * Input, if its initiator is a process external to P ;
- * Output, if it is not internal, and its initiator is the process P .

Proceeding from the above, in the language of the calculus of processes, the process P is a triplet of the form:

$$P = (S, S^\circ, R),$$

The components of which have the following meaning:

- * S is a set whose elements are called the states of the process P .
- * $S^\circ \in S$ is some selected state, called the initial state of the process P .
- * R is a subset of the form $R \subseteq \hat{\text{Act}} \times S$.

Elements of the set R are called transitions. If the transition from R has the form (s_1, τ, s_2) , then:

- * It is assumed that this transition is a transition from the state s_1 to the state s_2 with the fulfillment of the action τ ;

* The states s_1 and s_2 are called the beginning and end of this transition, respectively, and the action of τ is called the label of this transition.

By analogy with the concept of the process P in the calculus of processes, we introduce the notion of the function F of the UV-element. The function F is a triplet:

(S, S°, R) , (where S is the set of sub-processes of the process corresponding to the function F , $S^\circ \subseteq S$ is the set of interface sub-processes (where $S^\circ = S \setminus S!$), R is the set of transitions in the set S , carried out by transferring, inputting and outputting objects).

CONCLUSION

In other words, by analogy with the calculus of processes, the marked system of transitions (S, R) over the set of flows $\text{Act}(F)$ is considered. The elements of the set $\text{Act}(F)$ of flows (input $a?$ output $a!$), corresponding to a set of actions in the calculus of processes, are also interpreted as input, output or transfer of an object with the name of the stream. At the same time, in this case (at the level of the function description) only internal flows are of interest, since the algebra of images of the theory of patterns and the theory of graphs are involved in external (input and output) flows.

ACKNOWLEDGEMENT

Researchers are supported by the President's Grant for the state support of young PhD researchers of the Russian Federation MD-2687.2017.6.

REFERENCES

- Alefirenko, N. (2004). Questions of cognitive linguistics. Wolfgang Wildgen. No 1. pp. 70-81. UK.
- Bentley, B., & Bossé, M. J. (2018). College Students' Understanding of Fraction Operations. *International Electronic Journal of Mathematics Education*, 13(3), 233-247.
- Bonami, O., Boyé, G., & Kerleroux, F. (2004). Stem Selection in Lexeme Formation: Evidence from French. 11th International Morphology Meeting Vienna. Vienna. Austria.
- Galeev, T., & Habibulina, E. (2017). Frequency of Verbal Forms and Language Standard. *Journal of History Culture and Art Research*. Vol. 6, No 5: 90-98. Russia.
- Gareyev, A. (2016). The Necessity of Professionally Oriented Foreign Language Training of Vocational Education Teachers (the Results of the Survey). *Bulletin of Kalashnikov ISTU*. Vol. 19, No 2: 127-128. Russia.
- Jerichow, A., & Simansen, J. (2013). Civilizations: Clash or Cooperation? Islam in a changing world. Europe and the Middle East – Routledge. p. 220. UK.
- Kondubaeva, M. R., Bekalay, N. K., Aubakirova, A. K., Ongarbaeva, A. T., & Tolkinbayev, A. K. (2018). The problem of correctness and reliability of the study in trilingual education. *Opción*, 34(85-2), 517-543.
- Lobão, J., & Pereira, C. (2016). Looking for Psychological Barriers in nine European Stock Market Indices. *Dutch Journal of Finance and Management*, 1(1), 39.
- Losev, A. (1984). The Problem of the Variative Functioning of Poetic Language. *Soviet Studies in Literature*. Vol. 20, N° 2-3: 35-80. Russia.
- Margolis, E. (1999). *Concepts: core readings*. Laurence. MA: MIT Press. Cambridge. p. 652. UK.
- Parvizian, F., Ghojavand, K., & Niknejadi, F. (2015). Effectiveness of Emotional Intelligence on Emotional Alexithymia of Married Women Teachers in Yasuj City. *UCT Journal of Social Sciences and Humanities Research*, 3(1), 32-35.
- Piteira, M., Costa, C. J., & Aparicio, M. (2018). Computer Programming Learning: How to Apply Gamification on Online Courses? *Journal of Information Systems Engineering & Management*, 3(2), 11.
- Shirvani, M., Mohammadi, A., & Shirvani, F. (2015). Comparative study of cultural and social factors affecting urban and rural women's Burnout in Shahrekord Township. *UCT Journal of Management and Accounting Studies*, 3(1), 1-4.
- Wu, Z., & Palmer, M. (1994). Verbs semantics and lexical selection. Paper presented at the Proceedings of the 32nd annual meeting on Association for Computational Linguistics. Mexico.
- Zhikharev, A. (2014). System-object method of knowledge representation. Chernomoretz. Belgorod: Izd-vo NIU BelGU. p. 140. Russia.