

## Artículo de investigación

**The Strategy of Language Corpus Development as a Linguistic Competence of the User of Flex Program**

Стратегия развития языкового корпуса как лингвистическая компетентность пользователя программы Flex

La estrategia de desarrollo del corpus lingüístico como competencia lingüística del programa User of Flex

Recibido: 25 de mayo del 2019

Aceptado: 23 de junio del 2019

Written by:  
Sabrina Kh. Shikhalieva<sup>60</sup>  
Helen Z. Santueva<sup>61</sup>  
Laura P. Fetalieva<sup>62</sup>  
Sapiayt Kh. Gasanova<sup>63</sup>  
Zhavgarat A. Ramazanova<sup>64</sup>**Abstract**

The article considers the first version of the text analysis instructions in the languages of Dagestan. The texts of “the Gospel” in the languages of the peoples of Dagestan are the sets of words and signs, from which it is necessary to single out what is interesting to the user of the computer program FLEx. It is referred to a meaningful tool of the text of “the Gospel”. The text marked with various kinds of information enables the user to have some exposure to a high-quality corpus of electronic information. The division of the components of a computational lexicon is considered as a means of expressing grammatical features “word↑ phrase\* text”. The difficulties of description are determined by the way to develop a website for an educational institution.

**Key Words:** Lexicography, St Luke’s Gospel, text, national dictionary, software program FLEx.

**Аннотация**

В статье рассматривается первая версия инструкций по анализу текста на языках Дагестана. Тексты «Евангелия» на языках народов Дагестана представляют собой наборы слов и знаков, из которых необходимо выделить то, что интересно пользователю компьютерной программы FLEx. Это относится к содержательному инструменту текста «Евангелия». Текст, помеченный различными видами информации, позволяет пользователю получить доступ к высококачественному корпусу электронной информации. Разделение компонентов вычислительной лексики рассматривается как средство выражения грамматических признаков «слово ↑ фраза \* текст». Трудности описания определяются тем, как разработать сайт для учебного заведения.

**Ключевые слова:** лексикография, Евангелие от Святого Луки, текст, национальный словарь, программа FLEx.

<sup>60</sup> Doctor of Philology, the leading researcher of Lexicology and Lexicography Department of Language, Literature and Art Institute of G. Tsadasa D NTS RAN;

<sup>61</sup> Candidate of Philological Sciences, Associate Professor at the Department of Theoretical Foundations and Technologies of Primary Linguistic Education of Dagestan State Pedagogical University;

<sup>62</sup> Candidate of Philological Sciences, Associate Professor at the Department of pedagogy and psychology of primary education of Dagestan State Pedagogical University;

<sup>63</sup> Candidate of Philological Sciences, Associate Professor at the Department of Theoretical Foundations and Technologies of Primary Linguistic Education of Dagestan State Pedagogical University;

<sup>64</sup> Candidate of Pedagogical Sciences, Associate Professor at the Department of English Philology of Dagestan State Pedagogical University.

## Resumen

El artículo considera la primera versión de las instrucciones de análisis de texto en los idiomas de Daguestán. Los textos de "El Evangelio" en los idiomas de los pueblos de Daguestán son los conjuntos de palabras y signos, de los cuales es necesario destacar lo que es interesante para el usuario del programa de computadora FLEx. Se refiere a una herramienta significativa del texto del "Evangelio". El texto marcado con diversos tipos de información permite al usuario tener cierta exposición a un corpus de información electrónica de alta calidad. La división de los componentes de un léxico computacional se considera como un medio de expresar características gramaticales "palabra ↑ frase \* texto". Las dificultades de la descripción están determinadas por la forma de desarrollar un sitio web para una institución educativa.

**Palabras clave:** Lexicografía, Evangelio de San Lucas, texto, diccionario nacional, programa de software FLEx.

## Introduction

The practical effect of any theory depends on its explanatory capacity, and the latter depends on the quality of the corresponding conceptual apparatus of the national literature data. Domestic translation theory currently requires clarification of a number of key concepts of national literature in language data descriptions (Apresyan, 1974). The description of the data of natural language in the form of documents of the national literature as well as any computer modeling of the real-world forms is the result of a compromise of the social phenomenon – corpus-oriented analysis of language data in the word-processing software FLEx. The corpus-oriented analysis of FLEx allows to develop the form of text glossing in the relational database of the electronic dictionary. The purpose of the product description of the software program FLEx is to demonstrate the program shell of the texts and present the profile of the Corpus as declared by the translator (Gasanova, 2014; Yazdekhasti et al, 2015). The electronic resource declared by the manufacturer makes it possible to arrange lexicographic translation steps, thanks to which the user understands where to go, what to do, and returns to the page of structuring the Russian text markup. From the viewpoint of analyzing the information of the dictionary, after the technical task of designing the control of the consumer, the layout of the website is approved. This article considers the text translation tool. Texts are the sets of words and characters from

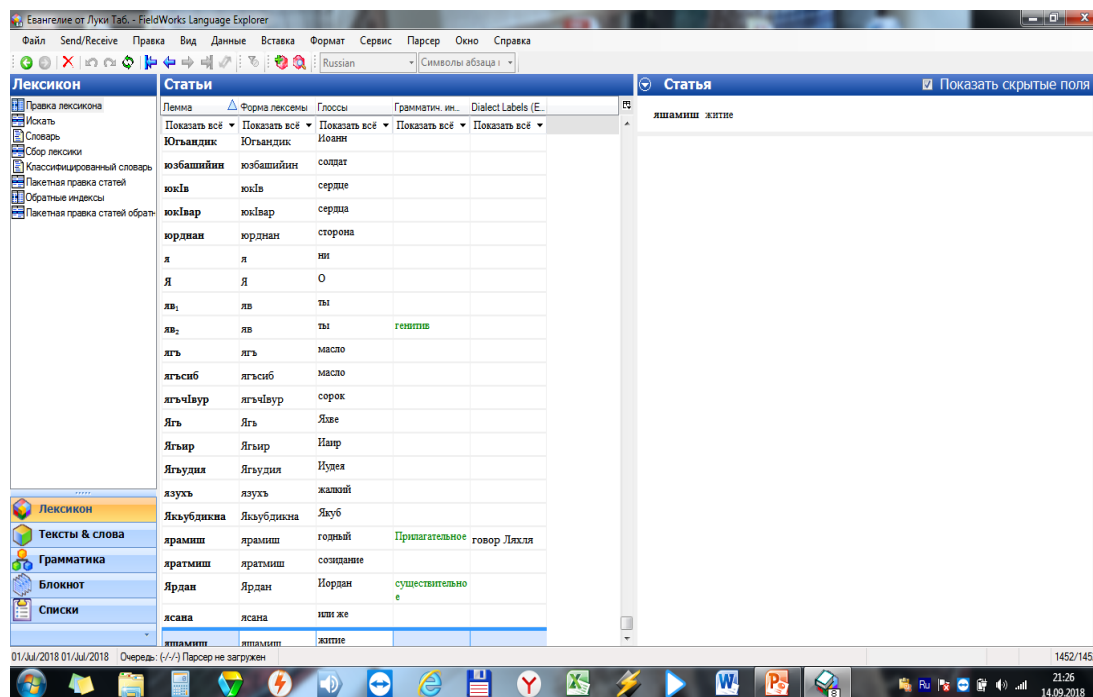
which it is necessary to single out what is interesting to the user of FLEx. It is referred to a meaningful text tool with consideration of regional and national needs of the peoples of the Russian Federation (Gasanova, 2015; Melo et al, 2018). It is shown that such needs of the user who studies Russian as a second native language can serve as a useful material for clarifying and extending the grammatical description of the Russian language, and the very corpus can be a useful tool for describing the national language. The difficulties of description are determined by the peculiarities of the natural language and the level of formalization of the texts in a computer program.

## Review of Literature

At the first stage of the description of translation tools the need of the target audience in the presentation of website data is assessed. The alphabetical paging is determined by the content of the electronic dictionary, and the way the alphabetical pages are structured is included in the website information.

The second stage of the description of the translation information tool is based on the development of technical specifications for specialists. The development of the technical assignment for specialists involves the content of Russian Literature Dictionary (the text of "St Luke's Gospel"). See Table 1.

Table 1. Electronic Dictionary Interface



Translation of the classic work of world literature, i.e., the Bible, is always an important milestone in the development of the national culture of the people. In the course of translation not only the language of the Bible with the form of familiarization of peoples with the cultural terms of “The New Testament” is preserved but also the vocabulary of interlinear translations is extended (Garayev, 2016). The first Greek text with an interlinear translation of the so-called Complutensian Polyglot was published in the city of Alcalá (Spain) in 1514-1517. The first printed edition in Greek in which the text was presented in several languages in adjacent columns with Latin translations, and a literal Latin translation was inserted in the Greek text.

**Methodology**

The author of this electronic dictionary aims to contribute to the development of interlineary Greek-Russian Bible. This Dictionary explores the conceptual domain of “The New Testament” which confirms that the Russian language is a system for creating a gloss for the national languages of the Russian Federation. Today, the system of interlinear translation in the Tabasaran language is defined by the first translation of the text of “The Gospel”. Therefore, the author of this electronic dictionary takes into consideration all editions of the Institute of Bible Translation. The first text of “The Gospel” in the Tabasaran

language was published in Stockholm in 1997 (“The Gospel of Mark” translated by B. G. Khanmagomedov, edited by M. E. Alekseev). The dictionary with parallel translations of the texts of the Gospel makes it possible to understand the nature of the translation of the terms of “The New Testament”. Finally, the work of the Institute of the Translation of the Bible, which is this Dictionary, is targeted at the students who appreciate the translation of “The New Testament” for the peoples of the Caucasus (Shikhaliyeva, 2015).

**Results and discussion**

Actually, the linguistic markup of the text is carried out automatically: the selection of paragraphs, sentences, the glossing of words and type of comment. Then, by the method of word ordering, glosses are marked by their numeric values, the texts and the stages of documenting the language are presented, i. e., preparing and entering data, processing text links, exporting structured data via the file – Upload to webonary – Submit. Frequency connections in the system of representation of the automatic processing of the Tabasaran language data are established by textual references of the Russian language as structured data of translation.

The applied significance of the description of the electronic resource of the national Dictionary is

determined by the method of developing the site for an educational institution (Fetalieva, 2016). The line of research is determined by the feasibility of obtaining the expected results via quantitative calculations. The main linguistic method of studying the language corpus revealed the typologically common features of frequency units (verb forms to which derivational and

inflectional affixes of the Russian language are attached). Further, the glosses are marked by the method of organizing the text data. The stepwise distribution confirms the registration of dictionary entries with the tools of the components of the electronic resource). See Table 2.

**Table 2.** The Sample of Encrypting Terms. That Represent the Resource of Morphological Analysis of Synonyms<sup>65</sup>

| № | Variants of the forms of synonyms with indicators<br>“dialect community ↑ dialect individual” | Encrypting synonymy<br>functions as a form of cultural<br>stereotype “cultural stereotype<br>↑ bookish lexis” |
|---|---|---|
| 1 | conversion of adverbial participle → adverb   | CONV → ADV  |
| 2 | conversion of participle → adjective  | PTCP → ADI  |
| 3 | conversion of participle → noun   | PTCP → N  |
| 4 | conversion of adjective → noun  | ADI → N   |
| 5 | conversion of verb → noun   | V → N   |

Such a principle of understanding the scale of the description of an electronic dictionary allows to get an objective picture of the ratio of materials of various types (the bases of grammatical, lexical, textual markers). The basis of the material description is the concordances of FLEx tool set. The linguistic markup of a document is carried out by the arrangement of units of the national text: paragraph + sentence + comment.

A word usage with the markers of grammatical, lexical and other categories of the Tabasaran language is marked by means of automatic search (Comrie & Khalilov, 2010). The conversion of the Corpus of the national text is carried out by the replenishment of the lexicon in 2 sub-steps: Sub-step 1. Textual level – to systematize the lexicon in the processing of the Corpus Manager.

The basic unit of the Corpus Manager is the use of national texts, their bases (lemmas) and sentences that allow searching for the unit by language metadata (Shikhalieva, 2015). To classify national texts, the main material for requesting predicates was the task to find a verb and its equivalents in the corpus of works of literature (Thomas, 2012). Thanks to the electronic corpus, it became possible to look at the Tabasaran verbs from another viewpoint, namely, to find out the class of the lexemes comprehended in it. The volume of the Corpus being created in terms of the accepted units of

national text is determined by the objective of the second sub-step.

Sub-step 2. Discursive level – to describe the grammatical level of the Corpus with the segmentation of the national text. The conversion of text segmentation is supposed to be performed by the following operations: the division of an input element (words → noun, verb, adjective), the processing of text elements (anthroponym, toponym). The analysis of operations is performed automatically with assigning morphological tags: part of speech → noun, adjective, verb. The description of text elements with the processing of verb classes is comprised of both bibliographic and extralinguistic data. Search for the description of text elements from different sources is made by word forms and morphological tags (onym\*anthroponym\*part of speech → noun, adjective, verb; onym\* toponym\*part of speech → noun, adjective, verb) (Shikhalieva, 2015).

## Conclusion

The concept of describing a national language was primarily defined in the tasks of the scientific background of the website. [www.tabasaran.webonary.org](http://www.tabasaran.webonary.org). The present “Dictionary” aims to increase the students’ computer competence, to extend a linguistic reserve of students, to familiarize them with the

<sup>65</sup> The paper uses the rules of glossing of linguistic terms of Summer Institute of Linguistics (<https://glossary.sil.org/term/SIL>)

culture of the Russian text of “St Luke’s Gospel”. Each element of the Corpus is available as an individual object: any action by the user (clicking the mouse button, moving the mouse or pressing the keyboard key) is documented as an entry of the standard Russian translation procedure. The description of the website of the Dictionary is carried out by the services of creative ideas of the specialists in the field of linguistics. The applied importance of the electronic corpus of the Dictionary is determined by the only and main way – to describe the language of fiction of the Tabasaran language via Russian translations and to develop a website for an educational institution.

### Bibliographic references

- Apresyan Y. D. (1974). Lexical semantics (inflectional language features). M.: Nauka. 346 p.
- Comrie, B., & Khalilov, M. (2010). The Dictionary of Languages and Dialects of the Peoples of the Northern Caucasus.
- Fetalieva, L. P. (2016). Historical stages of development of inclusive education junior high school students. *Herald of the Chelyabinsk State Pedagogical University*. 94-99
- Garayev, S. A. (2016). Proper name common name as the factor of consistency in terminology. *Philology. Questions of theory and practice*, 5-3(59), 166-169
- Gasanova, M. A. (2014). Metaphor as the human factor in the space of culture and language: meaning and interpretation. *World of science, culture and education*, 6(49), 421-423
- Gasanova, M. A. (2015). Memo style "female character": lingvo-folklore units. *Problems of modeling of cognitive linguistics*, 4(45), 129-133.
- Melo, R., Sobrinho, V. D. P. F. M., Filho, I. J. D. M., Feliciano, F., & Maciel, P. R. M. (2018). Redundancy Mechanisms Applied in Video Streaming Service on Eucalyptus Cloud Computing Infrastructures. *Journal of Information Systems Engineering & Management*, 3(3), 25.
- Shikhalieva S. Kh. (2015). Linguistic Mechanism of Dialogue and the Subject Matter of Evidentiality: Palestine, We're with You! *Social Sciences*, 10(6), 1343 - 1346.
- Shikhalieva, S. Kh. (2015). Cult and Culture of the Caucasus in the Palestinian Context: Speech Segment and Substrate Anomaly. *Social Sciences*, 10(3), 226-229.