

TOOLS FOR CREATING A CORPUS OF DICTIONARIES AND AN INSTRUMENT FOR DOCUMENTING LOW-RESOURCE LANGUAGES

FERRAMENTAS PARA CRIAR UM CORPUS DE DICIONÁRIOS E UM INSTRUMENTO PARA DOCUMENTAR LÍNGUAS DE BAIXOS RECURSOS

HERRAMIENTAS PARA CREAR UN CORPUS DE DICCIONARIOS Y UN INSTRUMENTO PARA DOCUMENTAR IDIOMAS DE BAJOS RECURSOS

Zabihat OTSOMIEVA-TAGIROVA¹

Sapiahanum TEMIRBULATOVA²

Magomed MAGOMEDOV³

Zufira KIEVA⁴

Ludmila DUDAROVA⁵

ABSTRACT: The article discusses the concept of creating a corpus of dictionaries and ways to document low-resource languages. Translation of a classic work of world literature is always an important milestone in the development of national culture. Such translations attest to the maturity of small languages, and raise their prestige among the languages of the world. This aspect of the description of bibleisms in the system of small languages of Russia is ordered in the paradigm of contextual synonymy. Although much work has been done in recent years in the study of Caucasian biblical studies and contextual synonymy has been systematized, taking into account their translation, many terms of the classical work have not been sufficiently studied. The concept of translating Biblical texts in the Caucasian languages allows us to understand the linguistic problems of creating a corpus of dictionaries and a means of documenting low-resource languages.

Key words: Low-resource languages. Russian loanwords. Paradigm. Caucasian (Nakh-Dagestan) languages. Contextual synonymy. Historical lexicology. Bibleisms.

RESUMO: *El artículo analiza el concepto de crear un corpus de diccionarios y formas de documentar idiomas de bajos recursos. La traducción de una obra clásica de la literatura mundial es siempre un hito importante en el desarrollo de la cultura nacional. Estas traducciones dan fe de la madurez de las lenguas pequeñas y aumentan su prestigio entre las lenguas del mundo. Este aspecto de la descripción de los bibleismos en el sistema de lenguas pequeñas de Rusia se ordena en el paradigma de la sinonimia contextual. Aunque se ha trabajado mucho en los últimos años en el estudio de los estudios bíblicos caucásicos y se ha sistematizado la sinonimia contextual, teniendo en cuenta su traducción, muchos términos de la obra clásica no han sido suficientemente estudiados. El concepto de traducir textos bíblicos en lenguas caucásicas nos permite comprender los problemas lingüísticos de crear*

¹ Dagestan Federal Research Center Russian Academy of Sciences; ORCID <https://orcid.org/0000-0001-6517-8465>; e-mail: 79034821555@yandex.ru;

² Dagestan Federal Research Center Russian Academy of Sciences; ORCID <https://orcid.org/0000-0001-7526-3831>; e-mail: sapiiakhanum@mail.ru;

³ Dagestan Federal Research Center Russian Academy of Sciences; ORCID <https://orcid.org/0000-0002-2333-6565>; e-mail: dun2@mail.ru;

⁴ Ingush Humanities Research Center by Ch. Akhrieva; Ingush State University; ORCID <https://orcid.org/0000-0002-0802-8337>; e-mail: z-kieva@mail.ru;

⁵ Ingush State University; ORCID <https://orcid.org/0000-0002-6429-2669>; e-mail: lyuda.dudarova@mail.ru.

un corpus de diccionarios y un medio para documentar lenguas de bajos recursos.

Palavras-chave: *Línguas de poucos recursos. Empréstimo russo. Paradigma. Línguas caucasianas (Nakh-Daguestão). Sinonímia contextual. Lexicologia histórica. Bibleismos.*

RESUMEN: *O artigo discute o conceito de criação de um corpus de dicionários e formas de documentar linguagens de poucos recursos. A tradução de uma obra clássica da literatura mundial é sempre um marco importante no desenvolvimento da cultura nacional. Essas traduções atestam a maturidade das pequenas línguas e aumentam seu prestígio entre as línguas do mundo. Este aspecto da descrição dos bibleismos no sistema de pequenas línguas da Rússia é ordenado no paradigma da sinonímia contextual. Embora muito trabalho tenha sido feito nos últimos anos no estudo dos estudos bíblicos caucasianos e a sinonímia contextual tenha sido sistematizada, levando em consideração sua tradução, muitos termos da obra clássica não foram suficientemente estudados. O conceito de tradução de textos bíblicos nas línguas caucasianas nos permite entender os problemas linguísticos de criar um corpus de dicionários e um meio de documentar línguas de poucos recursos.*

Palabras clave: *Lenguas de escasos recursos. Préstamos rusos. Paradigma. Lenguas caucásicas (naj-daguestán). Sinonimia contextual. Lexicología histórica. Bibleismos.*

Introduction

In recent years, significant work has been carried out to study the vocabulary of low-resource Caucasian languages, a classification of terms has been developed, a socio-political terminology has been compiled, and the reference vocabulary of modern literary Caucasian languages has been determined. However, the modern vocabulary of the Caucasian languages has not been sufficiently studied, although there are a number of studies on its study. And as for the lexical features, taking into account the translation of the Bible texts, they are considered in the formal-semantic plan of lexicography. Now there is a need to systematize the lexical-thematic groups of the Caucasian languages in the corpus of the dictionary of terms of the Bible (KRASNIH, 2002, p. 12). The development of the dictionary tool is an urgent and urgent task, since changes in the socio-political, economic and cultural life of society dictate the renewal of the corpus of modern Caucasian languages. Therefore, work is currently underway to compile a corpus of dictionaries with Russian loanwords taken into account (Arnold, 2013, p.31).

The scientific novelty consists in the fact that the article for the first time discusses the problems of compiling a dictionary of Bible terms. See

[<https://www.webonary.org/tabasaran/browse/tabasaran-russian/>]; and also

[<https://www.webonary.org/rutul/?lang=en>]. The development of a dictionary tool will help solve linguistic problems, taking into account the description of the vocabulary of small languages. The dictionary tool will introduce new vocabulary reserves into scientific

circulation, which can be used in the development of various aspects of the synonymy of the national language (Humboldt, 1984). The systematization of texts of different genres and the systematization of the instrument are also important because the language will have an electronic archive.

Research Methods

The corpus of the dictionary of terms of the Bible will include the entire active vocabulary of native words of the national language. The corpus will also include borrowed words from the Russian language (POLAJKOV, 1994). In the dictionary, a special place will be occupied by special vocabulary related to various branches of knowledge, which is used by specialists of different professions: ethnographers, historians, writers, folklorists, etc. When compiling the dictionary tool, it is planned to rely mainly on the material of the Bible texts. The nature and completeness of the dictionary tool requires that each word be given a comprehensive lexical and grammatical description. In the dictionary tool, word forms will be given in the international phonetic transcription (MFA) and grammatical markup of the FLEx computer program.

Creating a tool for the corpus of dictionaries will help solve the most important tasks of distance education, theoretical works on semantic and stylistic study of the vocabulary of Dagestani languages (MUTALOV, 2017).

The general work plan provides for: 1) collection of textual material on the stated topic (texts of the Bible); 2) scientific and technical processing of textual data in the FLEx computer program; 3) electronic typing and markup in the FLEx computer program.

Research Questions

The interpretation and translation of the term is carried out in the assessment of the transformation of stylistic landmarks. The Dictionary of terms of the Bible, which describes the stylistic guidelines of language and speech, distributes the deep structure of paronymy on a semantic basis. It is this initial property of paronymy that determines the parameters of the translation of terms in the national language (mother language).

The objectivity of the term, with the initial theoretical provisions of the ergative construction of the sentence, is provided by the methodological setting of the national language in the continuity of the related Caucasian languages (EFENDIEV, 2000, P. 11):

1) subject orientation – scientific and educational terminology;
2) language orientation-bilingual dictionary;
3) usage orientation – the purpose of the dictionary and what the text of the Dictionary will be read for: to get acquainted with the terminology of science.

4) address orientation – the reader's address *our readers: intelligent people whose interests include reading, mostly living in rural areas (teachers, doctors, etc.); *aged about 16 years and older; by political views-atheists or liberal-minded theists* (this group is expected to make up about 60% of the potential audience); *reading students, young people (about 10% of the audience), philologists, cultural scientists, historians, etc.. In general, the potential audience is about 25% of the total number of carriers. They know the national language, speak it in the family, read both in Russian and in the national language, perhaps listen to the radio and watch TV channels in the national language. At the same time, it is easier for them to read and write in Russian.

5) dictionary function and phraseological level of description of terms: inventory function with levels-phraseological antonymy, phraseological homonymy.

Findings

The terminological concept of the author of the Dictionary of Bible Terms differs from that adopted in the previous academic edition: the standard translations of terms cover, first of all, the levels of description of chronological periods:

I. The pre – national period (with the norm of borrowed texts-Iranisms, Arabisms, Turkisms);

II. National period (with the norm of borrowed Russian texts). This allows, first, to correlate chronologically borrowed facts of Russian texts with the system of the national language, and secondly, to provide an additional context for describing terms.

* According to the semantic type of the national language, broad and narrow names of terms are distinguished.

• According to the structural type of the national language, the terms simple, complex and composite are distinguished.

* According to the word-formation type of the national language, the terms folk (ready-made), scientific (new) and borrowed are distinguished.

New formations in terminology are created on the basis of the resources of the national language or on the basis of Russian loanwords. However, there are some objective criteria for

evaluating the term. One of the main features of the scientific nomination is approbation. In relation to the terminology of the national language, the approbation can be formulated as follows:

It is preferable to form short and one-word scientific terms. In the case of a broad nomination, the properties of folk terminology are used.

It is desirable that the basis (or bases) of the term should be determined by the motivational features of the substantive names. This ensures an international understanding of the term (even with its national design), this way we avoid unnecessary extraneous associations in terminology. Here are examples of such motivational names: hydro -, oko -, cosmo -, astro -, geo -, urban -, pseudo -, info -, photo -, one -. The latter in the system of word formation can exist constantly and serve as the basis for other terms, etc.

An important principle of term construction is the derivational model of word formation.

The scope of national language terms covers the following linguistic aspects:

- 1) constructions with generic names typical for written speech (semasiological aspect);
- 2) special phraseological units (verbal or nominal) typical for written speech;
- 3) special unambiguous arguments typical for written speech.

The methodological criterion of functional style, adopted as the basis for the translation of terms, goes back to the concept of L. Bloomfield and his followers Ch. Morris, C. Peirce, N. Chomsky. The methodological concept of L. Bloomfield, which states that "every utterance is characterized by a form of human behavior", adapts the criterion of translation of the term to the functional correlation of the external environment (Text, 2001).

The meaning of the term is best conveyed when using the language form of the national language as close as possible to the semantic form of the Synodal translation. In addition, the meaning of the Synodal text can be conveyed based on the natural norm of the national language, regardless of how close the grammatical form of the national language is to the form of the Synodal text. Since the same grammatical structure can be associated with several different meanings, the grammatical construction of the New Testament text in the national version can have several different meanings.

Of course, the most important task of a translator is to translate accurately, and sometimes everything else is sacrificed to accuracy. This trend "manifests itself most clearly in the use of borrowing". As a rule, this choice is the result of a mismatch of the resources or structure of the original language and the recipient language to express the desired meaning. Meanwhile, borrowing always violates the naturalness of the translated text. Accordingly, a

translation replete with borrowings does not meet the criterion of clarity. However, for some categories of terms, borrowing is the best way to translate. First of all, this applies to vocabulary that reflects concepts specific to the Biblical cultural and historical context that do not have native correspondences in the national language. Such terms as Pharisee, Sadducee, Levite, Passover, etc. can not be called either clear or natural for the Dagestani reader, so the only way to semanticize them is a commentary (page-by-page or at the end of the book). For example: Sadducees-followers of a religious trend in Judaism; Levites-descendants of Levi, who were destined to serve God.

How we translate these special terms: we adapt them to the special study of the national language (as well as personal names), we fix the endings used in traditional grammar: for the masculine gender, for the feminine gender.

Discussion

From the point of view of the functioning of the linguistic correlates of the national language, common and highly specialized terms differ. Scientific, socio-political, and art history terms of Iranian, Arabic, and Turkic origin have been replaced by variants of synonyms.

In this regard, the work on translation largely determines the expressive capabilities of the Dagestani languages, their lexical and semantic means. Accordingly, the general theoretical guidelines for semantic translation, which ensure the accuracy, clarity and naturalness of the translation language, are of great importance.

So, for example, the words of the Christian lexicon, which go back to the inexhaustible source of Holy Scripture and have a special theological significance, at first glance, have no analogues in the languages of the peoples of Dagestan. Meanwhile, their meaning can be conveyed by: (a) the corresponding Muslim terms. See Table 1

Table 1

Modeling contextual synonymy within the framework of terminological doublets		
<i>National language [international transcription]</i>	<i>English</i>	<i>Russian</i>
<i>малаик [malaik]</i>	<i>angel</i>	ангел
<i>иблис [iblis]</i>	<i>devil</i>	дьявол
<i>иман [iman]</i>	<i>faith</i>	вера

In other words, the modeling of terminological doublets within the context of contextual synonymy is determined by the criterion of discursive phraseology (Kieva, 2017, p. 31-35). The criterion of phraseology reveals the point of terminological units in semantic unity: the nominative function is represented by the unmotivated core of a composite term; the communicative function is represented by the motivated core of a composite term.

Since the terminology of the national language is constantly updated, the viability of a particular term is revealed by the functional doublet of the Russian language (Vinogradov, 1979).

In bringing the text closer to the reader, some biblical common names can play a positive role, which have already found their tradition of pronunciation and spelling in the Caucasian languages. See Table 2

Table 2

Modeling of synonymous doublets in the framework of Russian phraseology		
<i>National language [international transcription]</i>	<i>Russian</i>	<i>English</i>
<i>фарисеѝ [farisej]</i>	<i>фарисеѝ</i>	<i>pharisees</i>
<i>саддукеѝ [saddukej]</i>	<i>саддукеѝ</i>	<i>sadduk</i>
<i>левит [lewit]</i>	<i>левит</i>	<i>levit</i>

Their pronunciation and spelling are determined, on the one hand, by the phonetic features and orthoepic norms of the national language and, on the other hand, by their direct Arabic source, due to the spread of Islam. It was very expedient to consistently observe in the translations the established tradition of writing Biblical personal names using the sound types that are characteristic of small nationalities (Shikhalieva, 2019).

When translating idioms and phraseological units, it is advisable to use equivalent or similar stable phrases and expressions that naturally and harmoniously fit into the context and do not harm its integrity and clarity. At the same time, such concepts as "bride for Christ", "word of God", "good from God", enriching the language, create special difficulties in their interpretation and assimilation. Such expressions in the text always stand apart and stand out in the reader's mind.

Conclusion

The translation language of the Bible texts has significantly enriched the Nakh-Dagestan languages with new lexemes and new grammatical constructions. The nature of their impact corresponds to the current direction of the development of the national language

(Abdullayev, 2015, p. 41). Currently, the process of updating the style of national literature is underway. The translation of the New Testament into the Nakh-Dagestan languages is a complete text of a special genre that actively influences the development of the language (BLYTHE, 1975). Of course, translated literature cannot be perfect in terms of language. In it, we find many shortcomings of grammatical, spelling and stylistic nature. However, it is important to emphasize that the translation language of the Bible texts is generally close to the vernacular.

Translated religious literature, being presented in almost single samples, thus has a serious impact on the further development and improvement of the national language. With careful attention to the problems of translation in general and to the interpretation of certain "particular" moments, it is possible to achieve a completely high-quality translation of Biblical texts into the Nakh-Dagestan languages in the future. This requires a creative approach to the use of the arsenal of the modern national language, a subtle and deep sense of the language, knowledge and use of trends in the further development of its capabilities, both lexical and stylistic. We believe that the main practical requirement for the translation procedure is the involvement of literate people, well-trained linguists, and experts in the national language. A characteristic feature of their activity should be an understanding of the language of the Bible as a special language, an awareness of the specifics of its transmission by means of the national language (SHIKHALIEVA, 2021).

In the language of the Bible translation, there is a tendency to replenish the corpus at the expense of the word-forming capabilities of the national (native) language. This trend should be accessible to the general reader, and contain a minimum number of borrowings from the Russian language, adhering to the general trend of the development of the national (native) language. At this stage, the tools of the national language dictionary will largely determine the answer to the question of whether the Muslim community will accept this style of literary translation language.

Acknowledgments

«The work was supported by RFBR grant 20-012-00380A».

References

ABDULLAYEV, I.H. Interdagestan and Intercaucasian language contacts. Makhachkala: Aleph, 2015. 614 p.

ARNOLD, I.V. Semantics. stylistics. Intertextuality. M., 2013. 320 p.

Blythe R. Translation problems. St. Petersburg: LLI, 1995. V.V. Selected Works. Research on Russian grammar. M., 1975. P. 4 - 9.

Efendiev, I.I. Iranisms in the Lezgi language. Makhachkala, 2000.

HUMBOLDT, W. On the difference between the structure of human languages and its impact on the spiritual development of mankind // Selected works on linguistics. M.: Progress, 1984. P. 24 - 26.

MUTALOV, R.O., YUSUPOV, KH.A. Educational resources of the Dargin language: electronic dictionary // Educational technologies and society. T.20, No. 2. 2017. P 310-322.

POLAJKOV, N. Anglo-Russian dictionary of biblical terms and Christian phrases. United States, 1994. P. 12-15.

[SHIKHALIEVA, S. Semantic Conflict Of The Caucasian Languages And Scenarios For The Development Of Scientific Research // ALRJournal. 2021; 5\(4\): 201-204](#)
[<https://www.alrjournal.com/jvi.aspx?un=ALRJ-90022&volume=5&issue=4>]

[SHIKHALIEVA, S. The Strategy of Language Corpus Development as a Linguistic Competence of the User of Flex Program. Amazonia-investiga. Vol 8 No 21 \(2019\) P.187-191](#) [<https://amazoniainvestiga.info/index.php/amazonia/article/view/89>]

Text: aspects of study: poetics, pragmatism, semantics // Articles. M., 2001. 320 p.

VINOGRADOV, O.I., KLIMOV, G.A. On Armenisms in Dagestan languages // Etymology. 1977. M., 1979. P. 154-159.

KIEVA, Z. Lexicology and phraseology of the Ingush language. Magas, 2017. p. 234

KRASNIH, V. Ethnopsycholinguistics and linguist culture. M., 2002. p. 214