

Научный вестник Бухарского государственного университета Scientific reports of Bukhara State University 1/2023













Ismatullayeva I.I.	Hozirgi zamon tilshunosligida konseptni oʻrganish metodlari,	135
•	usullari va klassifikatsiyasi	
Mehmonova Y. Ch.	Typology of collocations in combinatorial linguistics	139
Nuniyazova Sh. U.,	Comparative analysis in translation of English idioms and set	144
Djumabaeva J.Sh.	expressions into Uzbek and Russian	
Rajabov E.E.	Oʻzbek va ingliz tillaridagi "oila" tushunchasi konseptual maydonining lingvokognitiv tahlili	148
Ruziyeva N. X.	Zamonaviy tilshunoslikda matn va diskurs talqini	153
Xodjayeva D.I., Safarova M.H.	Ingliz tilini oʻqitishda fonetik interferensiyaning oʻrni	157
Samandarova G.Y.	Sodda gap koʻrinishidagi "hasharot" lugʻaviy ma'no guruhlari asosida shakllangan paremalarning oʻziga xos xususiyatlari	161
Sayfullayev B.N.	Noqardosh tillarda o'zlashtirmalik (o'zbek va fransuz tillari misolida)	165
Toʻyboyeva Sh.R.	Fransuz tilida sodda gap tarkibiy qismlarini tarjima qilishda yuzaga keladigan muammolar yechimi	170
Абдурахманова М.У.	"Кўрлик" касаллигининг ўзбек, рус, инглиз тилларидаги лингвистик-семантик тадкики	174
Ахмедова М.Х.	Ўкув жараёнига тааллукли ноадабий сўзларнинг психолингвистик хусусиятлари	180
Вахидова Ф.С.	Дискурсда диний туризм ва зиёрат туризми терминларининг вокеланиши	183
Avezov S.S.	A Corpus-based approach to phraseology research	189
Kobilova A.B.	Ingliz va oʻzbek tibbiy perifrazalarining semantik mikromaydonidagi lingvokulturologik xosliklari	194
Bobokalonov R.R., Narzulaeva D.B.	Vocabulary of the Qur'an in the object of theolinguistics	203
Nazarova G.P.	Muloqotda gender farq muammolari xususida	210
Raxmatov A.P.	Plusquamperfekt zamon shaklining grammatik ma'nolari va uning kontekstual tahlili (nemis tili misolida)	215
Sayidova G.Y.	Alisher Navoiy nazmiy asarlaridagi oʻxshatishlarning leksik- grammatik tabiati	221
Rasulov Z.I., Yuldasheva F.E.	Xushmuomalalikning nutq odobida ifodalanishi	232
Бозоров З.А.	Речевая культура в качестве неотъемлемой части языковой культуры личности	240
Мавлонова М.Д.	Инглизча ахборот технологиялари терминларининг ўзбек тилига ўзлашиш жараёнида маъно торайиши ва кенгайиши ходисалари	245
Шарипова А.А.	Луғат корпусини яратиш технологияси	251
Hayotova D.Z.	Tushuncha va soʻz munosabati	255
Mamatova D.M.	Xitoy tilida ta'm bildiruvchi sifatlar tahlili	259
Сафаров Ф.С.	Маънодошлик қаторидаги сўзлар орасидаги маъновий муносабат хақида	266
Akhmedova D.B.	Semantic comments in an explanatory dictionary and their processing as a semantic extension	272
ADABIYOTSHUNO	SLIK *** LITERARY CRITICISM *** ЛИТЕРАТУРОВЕДЕНИ	1E
Темирова Дж. Х.	Отражение национальной картины мира в русском и узбекском детском фольклоре	277
Насруллаев Э.Ж.	Алишер Навоий "Ҳикмат ва ибрат достони"нинг ғоявий таҳлили	284

UDC 81.004

A CORPUS-BASED APPROACH TO PHRASEOLOGY RESEARCH

Avezov Sukhrob Sobirovich; Senior Lecturer Bukhara State University 1990senigama@gmail.com

Annotatsiya. Maqolada frazeologizmlarni o'rganishda parallel korpuslardan foydalanish imkoniyatlari ko'rib chiqiladi. Tahlil natijalari shuni tasdiqlaydiki, korpus ma'lumotlaridan foydalanish nafaqat tillararo tavsiflarni sezilarli darajada yaxshilaydi, balki tilning leksik tizimining alohida qismi sifatida frazeologizmlarning o'ziga xos xususiyatlari haqidagi g'oyani o'zgartirish imkonini beradi. Xususan, parallel matnlar korpusining materiali an'anaviy ravishda to'liq ekvivalent deb hisoblangan idiomalarning lingvistik o'ziga xos xususiyatlarini aniqlab beradi. Ekvivalentlikning ikki xil turlarini ajratilishini asoslaydi: til tizimi darajasidagi ekvivalentlik va matn darajasidagi ekvivalentlik. Ushbu tadqiqot usuli turli tuzilishga ega tillarning o'zaro bog'liq frazeologik birliklarning alohida juftlarini aniqlash, shu jumladan, semantik, sintaktik, pragmatik va kombinatsion farqlarini aniqlash uchun korpus asosida ketma-ketlikda taqqoslashdan iborat.

Kalit so'zlar: parallel korpus, tarjima, lingvospesifik birliklar, kontrastiv tilshunoslik, semantik kategoriyalar, rus tili, normallashtirish, korpus tarjimashunosligi.

Abstract: This article explores the benefits of using parallel corpora in phraseology research. The findings show that incorporating corpus data enhances cross-linguistic descriptions and modifies our understanding of phraseology as a unique aspect of a language's lexical system. The use of parallel text corpora enables the identification of language-specific idiomatic expressions, which were previously thought to be fully equivalent. It is important to distinguish between two types of equivalence: equivalence in the language system and equivalence in the text. The research method is a corpus-based sequential comparison of individual pairs of correlated phraseologies across different languages, highlighting differences including semantic, syntactic, pragmatic, and combinatorial elements.

*Keywords:*parallel corpus, translation, language-specific units, contrastive linguistics, semantic categories, Russian language, normalization, corpus translation studies.

Аннотация. Статья рассматривает возможности использования параллельных корпусов в исследованиях фразеологии. Анализ, проведённый в работе, подтверждает, что обращение к корпусам существенно улучшает кросс-лингвистическое описание и изменяет само восприятие фразеологии как части языковой лексической системы. В частности, использование материалов параллельных текстов позволяет выявить лингво-специфические особенности идиом, которые ранее считались эквивалентными. Необходимо различать минимум два разных типа эквивалентности: эквивалентность в языковой системе и эквивалентность в тексте. Метод исследования основан на сопоставлении коррелирующих фразеологизмов разных языков, выявление различий между ними, включая семантику, синтаксис, прагматику и сочетаемость, на основе корпусных данных.

Ключевые слова: параллельный корпус, перевод, лингвоспецифичные единицы, контрастивная лингвистика, семантические категории, русский язык, нормализация, корпусное переводоведение.

Introduction. The objective of this research is to present empirical evidence of how the integration of corpus data, particularly parallel corpus data, can modify and enhance the conventional perceptions of linguistic equivalence in the field of phraseology. Our methodology is predicated on the notion, substantiated by prior studies, that incorporating corpus data not only improves contrastive analysis but also reconceptualizes our comprehension of the unique nature of phraseology, particularly idiomatic expressions, as a constituent of a language's lexical system.

The contributions of Valeriy Mikhaelyevich Mokienko to the discipline of comparative phraseology are widely acknowledged. His seminal work, "Slavic Phraseology", remains a seminal reference in the field. The current study, which aims to compare German and Russian phrasemes, aspires to add to the body of research in the area of linguistics and make a meaningful contribution to the field.

The organization of this article is as follows: after introducing the research object, outlining the primary methodology and general corpus analysis principles, we will investigate a case of a German idiom

without a corresponding equivalent in Russian, and compare the systemic representation of this absence of equivalence with the corpus data. In the subsequent section, we will perform a comparative analysis of German and Russian phraseological units, which were previously believed to be equivalent, in order to determine the linguistic peculiarities of these units through a corpus-based approach. Finally, in the last section, we will delve into several theoretical aspects related to the evolution of phraseology, with a focus on their practical implications for lexicography.

Research methodology. To continue, the main focus of the research methodology is a comparison of German and Russian phraseology. The data used in the study was obtained from the National Corpus of the Russian Language (NCRL), specifically through parallel corpora. This study's uniqueness lies in its empirical approach, where the analysis is based primarily on data from NCRL's Russian-German parallel corpora, instead of relying solely on theoretical views about inter-language equivalence in the field of phraseology. According to the literature, there are several key types of corpus-based work that have been identified.

- Research that is directed by the corpus (corpus-driven);

- Research based on the corpus (corpus-based);

- Research that uses the corpus as a source of illustrative material (corpus-illustrated, this principle of using the corpus is also called corpus-informed).

The idea of the diversity of methodologies for working with a corpus is a valid one and helps determine the appropriate method of analysis for each individual case. One such approach is the corpusillustrated approach, where the corpus serves as a source of example material for the researcher. In this case, the task is to select representative contexts that can be considered paradigmatic examples. This approach is particularly useful in dictionary work, as dictionaries that draw their examples from a corpus are considered to be more objective compared to dictionaries relying on linguistic intuition or randomly selected examples from the lexicographer's database [6; p. 45-51].

It's important to note that the use of the corpus principle in selecting examples does not relieve the lexicographer of the responsibility of making decisions in each individual case. Rather, it provides a broader range of options and enables the lexicographer to make informed decisions based on objective data, thus reducing the reliance on introspection.

It's also worth mentioning that the corpus-illustrated approach is just one of many different approaches to working with a corpus and that each individual case will require a specific methodology to be used.

Corpus-driven studies refer to the utilization of a corpus as a primary source of data for linguistic research. This approach views the corpus as defining the field of study and emphasizes the extraction of working hypotheses from the corpus data rather than preconceived theories. In the context of grammar, the identification of linguistic categories is based on the analysis of the corpus. The process of analysis serves as the first step in the research and enables the formation of theoretical generalizations based on the empirical data. This approach follows the principle of "data-driven theory development", rather than "theory-driven data analysis" [2; p. 61].

Corpus-based research represents a middle ground between the two opposing approaches to corpus material discussed earlier. In this type of research, the starting point is a pre-existing understanding of the phenomenon under investigation, often in the form of a hypothesis or assumption. The corpus is then used to validate these representations through empirical analysis. The outcome of this analysis is the acquisition of new insights into the structure of the studied phenomenon.

This type of corpus-based research is being utilized in the current study, although not exclusively. Strict adherence to the corpus-based approach requires quantitative analysis, which is currently limited in its application to phraseology research, particularly when studying idioms, as opposed to collocations or syntactic phrasemes. The Russian-German parallel corpus used in this study has a total of 5213,939 words, while the German-Russian parallel corpus has 3916,879 words. Out of the 2000 most frequently used idioms in modern German, as described in the "Deutsch-russische Idiome online" dictionary under development, only 325 are represented in the German-Russian parallel corpus. The number of contexts in which these idioms are used is limited, with only a few idioms having more than ten occurrences. Therefore, the use of quantitative analysis methods is not meaningful in this study. Despite this, the results obtained are still significant for comparative phraseology research, as demonstrated by the example provided.

The German idiom "von Mund zu Mund gehen" and the Russian idiom "переходить/передаваться из уст в уста" exhibit a close semantic correspondence. Analysis of the German-Russian corpus reveals instances where this idiom pairing is utilized.

Tausend freudige Ausrufungen gingen von Mund zu Mund.

Тысячи радостных восклицаний переходили из уст в уста [3; р. 197-201].

However, it should be noted that there are instances where this idiom pairing is not utilized. This is not necessarily due to a conscious decision on the part of the translator to avoid using the translation "переходить/передаваться из уст в уста" for "von Mund zu Mund gehen", but rather it may be prohibited by the norms of the Russian language.

— da ging es von Mund zu Mund, er habe sich mit einem jungen, bildschönen Fräulein aus zwar armer, aber adliger Familie verlobt.

— в тот самый день пронесся слух, что он обручился с одной молодой и необычайно красивой девушкой из бедной, правда, но аристократической семьи.

The reason why the idiom "переходить/передаваться из уст в уста" may not be suitable for use in a given Russian translation is due to its restrictions on the use of the impersonal form, unlike the German idiom "von Mund zu Mund gehen". By examining corpus data, such differences in the usage of comparable idiomatic expressions can be determined. It is important for bilingual dictionaries to take these features into account in order to accurately reflect the usage of idioms in both languages. The utilization of this approach can therefore be beneficial not only for gaining insight into idiomatic expressions, but also for the compilation of idiomatic dictionaries [1; p. 64-66].

The proposed method can be described as a corpus-based approach that involves the sequential comparison of individual pairs of equivalent idioms in different languages. The objective is to identify variations between these idioms, which may encompass semantic, syntactic, pragmatic, and collocational differences.

Results and its discussion. The Results and Discussion section will examine the issue of translating the German idiom "ein weites Feld" using an example. In the field of phraseology research, non-equivalent units, meaning idioms that do not have a standard equivalent translation in the target language, are a commonly recognized category. However, the concept of equivalence requires clarification. If equivalence is understood as a complete match of the corresponding units in two different languages in all relevant aspects of real usage, then the majority of idioms would be considered non-equivalent. However, if non-equivalence is only defined as instances where the meaning expressed by an idiom in one language in a specific context cannot be accurately conveyed in another language in any manner, then there would be very few cases of non-equivalence [4; p. 72].

Traditionally, the term "untranslatable phrases" refers to expressions in the source language that lack a suitable equivalent in the target language in terms of meaning and form. On the other hand, fully equivalent expressions are those that can be traced back to a single source and display high similarities in terms of semantic and lexical structure, as well as figurative basis. For instance, the Russian expression "метать бисер перед свиньями" and its German equivalent "Perlen vor die Säue werfen". However, such equivalence of idioms is only to a limited extent. Recent research indicates that the German idiomatic expression "Perlen vor die Säue werfen" has undergone significant semantic changes throughout its historical evolution, resulting in a meaning that deviates significantly from the meaning of the Russian expression "метать бисер перед свиньями".

Therefore, the use of corpus data is necessary in order to determine the limits of the concept of equivalence and to develop effective methods for its operationalization.

With regards to the material, the German idiom "ein weites Feld" (literally translated as "a wide field") does not have a standard equivalent in Russian that is recognized by normative dictionaries. Based on traditional perspectives, it can be classified as a non-equivalent idiom. The German-Russian Phraseological Dictionary presents this idiom in two dictionary entries: "das ist ein (zu) weites Feld" meaning "that's too vast a field (it cannot be fully addressed at once)" and "das ist noch ein weites Feld" meaning "there is still much to be accomplished in this field".

The earliest occurrence of the idiom in the National Corpus of the Russian Language is demonstrated in an example and confirms the information provided by the dictionary regarding the lack of a figurative equivalent in the Russian language.

— "Das ist ein weites Feld, von dem ihr da redet", sagte ich, "und da steht der menschli chen Erkenntnis ein nicht unwichtiger Gegenstand gegenüber".

— Вы затронули большую тему, — сказал я. — Тут перед нами явление очень важное.

In the translation of Theodor Fontane's renowned novel "Effi Briest", the German idiom "ein weites Feld" appears seven times. However, the consistent use of the Russian equivalent "темный лес" as its translation raises questions regarding the accuracy of the rendering. To further examine this issue, a comparison of the original contexts in which the idiom is used and its corresponding translations in Russian can be obtained from the Russian National Corpus.

— jedenfalls wollen wir darüber nicht streiten; es ist ein weites Feld. — "Во всяком случае, не будем об этом спорить: это — темный лес".

— Glaube mir, Effi, das ist auch ein weites Feld. "Поверь мне, Эффи, это темный лес".

— Aber lassen wir das, oder wie dein Papa immer sagte: "Das ist ein zu weites Feld". "Ну, хорошо, оставим это. Как говорит твой папа, "это темный лес".

— und Briest sagte ruhig: "Ach, Luise, laß... das ist ein zu weites Feld". — "а Брист спокойно сказал: "Ах, оставь, Луиза... Это уже совсем темный лес".

In "Effi Briest", the idiom "ein weites Feld" is a recurring element that contributes to the linguistic characterization of Effi's father. The repetition of this idiom holds significant artistic value and thus, should be consistently reflected in the translation. The translation's use of the equivalent "темный лес" is considered a successful choice by experts. It is noteworthy that the novel "Effi Briest" popularized the idiom "ein weites Feld". The idiom became even more widely known with the publication of Nobel Prize winner Günter Grass's novel "Ein weites Feld" in 1995. This reference by Grass, who cites Fontane as his favorite writer, to the original novel can be understood as his novel beginning where "Effi Briest" ends. In this sense, Grass's work continues the context established in Fontane's novel.

As of the knowledge cut-off, the novel "Ein weites Feld" has not been translated into Russian. This may be due in part to the challenges in translating its title. A literal translation of "Широкое поле" may not be appropriate as it lacks a corresponding phrase in the Russian language and would not enable the reader to make the connection between the title and its content, as well as the reference to "Effi Briest". Using the idiom "dark forest" as the title could be a more favorable literary solution, as it preserves the intertextual significance. However, this solution would be justified only if the Russian reader has a good understanding of 19th-century German literature. The reader must be familiar with the fact that in Y. Svetlanov and G. Egerman's translation of "Effi Briest", the novel ends with the words "it's already a completely dark forest", in order to fully comprehend the author's intention and appreciate the place of Grass's novel in German literary history. Based on the analysis of the translation of "Effi Briest", the following conclusions can be drawn regarding the concept of interlingual equivalence:

1. Idioms that do not have direct equivalents in another language in terms of meaning and imagery may still find suitable equivalents in a specific text, as demonstrated by the translation of "ein weites Feld" to "dark forest". It is important to recognize that equivalence should be evaluated in the context in which the comparison is being made, which can either be a system-oriented or a text-centered perspective. Corpus data, which is commonly used in research, focuses on text-centered perspectives.

2. A successful match in a specific text does not guarantee success in other contexts of use. Therefore, it is not always appropriate to include such matches in a bilingual dictionary.

The idioms "ein weites Feld" and "темный лес" cannot always be considered as equivalents as their meanings are specific and influenced by their internal structure. The metaphorical components of these idioms play an important role in determining their usage and meaning, and the differences in their metaphorical bases lead to variations in their functions. The idiomatic expressions "ein weites Feld" and "темный лес" both convey the idea of uncertainty and unpredictability, but the sources of this uncertainty differ between the two. "Ein weites Feld" denotes an open and uncertain future due to the lack of landmarks, while "темный лес" suggests opacity and the unknown. These differences are clearly evident when examining corpus examples, with the earliest uses of "темный лес" traced back to the mid-19th century.

Антон также глядел на дряблую избушку свою, бог весть, о чем он думал: чужая голова — темный лес. Using the idiom "ein weites Feld" in the translation of this example is not possible.

The concept of equivalence plays a crucial role in the comparative analysis of phraseological systems, as well as in literary and translation studies. In these fields, equivalence is significant at both the systemic and text levels. Systemic equivalence refers to the comparison of phraseological systems and their composition and structure. On the other hand, text equivalence is crucial in the study of intertextual relationships and translation practices. To achieve an effective bilingual dictionary, it is essential to strike a balance between these two concepts of equivalence and adopt a functional approach that considers both systemic factors and translation methods in different contextual situations. This approach should take into account the dynamic and contextual nature of equivalence and strive to provide a comprehensive understanding of phraseology [5; p. 407-413].

Conclusions. The analysis of the material leads us to some important conclusions regarding interlingual equivalence. Firstly, it is crucial to distinguish between different aspects of equivalence and various levels of analysis. A simple comparison of phrasemes from different languages based on structural parameters and semantic classes provides limited insights into systemic equivalence and only raises broader questions about the phraseme system of a particular language, borrowing mechanisms, and origins. However,

LINGUISTICS

the study of phrasemes within a specific context produces much more nuanced results. When examining the original context and its translation into another language, we can make observations about text-level equivalence. The use of corpus materials is particularly useful in advancing this line of inquiry. While systemic equivalence is concerned with the arrangement of the phraseme system, text-level equivalence focuses on the acceptability and success of the translation in a given context. Additionally, it should be noted that idiomatic expressions, even if they appear to be semantically equivalent when considered individually, may not possess the same level of compatibility in terms of usage context. This phenomenon can be explained through the concept of linguistic specificity, which pertains to the unique properties and characteristics of language elements.

The issue of the degree of linguistic specificity of idiomatic expressions has been a subject of debate in the field of phraseology for an extended period and can be considered a fundamental question in the theory of phraseology. Different perspectives exist on this topic. On one hand, some assert that idioms are inherently unique and reflect the cultural identity of a language community, which is often assumed as selfevident. On the other hand, there is a thesis advocating the prevalence of similarities over differences, which is based on substantial empirical evidence and is reinforced by etymology. Many idioms from various languages (especially those from the European cultural sphere) have a shared historical origin.

As is typical, the reality lies somewhere in the middle. There are a significant number of idioms in different languages (such as Russian and German) that exhibit comparability not only in terms of their core meanings and lexical structures but also in terms of all relevant usage characteristics. To date, no notable differences have been identified in the analyzed materials. Some examples of these comparable idioms include "würde sich im Grab umdrehen" and "перевернулся бы в гробу", "wie Hund und Katze" and "как кошка с собакой", "mit offenem Mund dastehen [dasitzen]" and "стоять [сидеть] разинув рот (от удивления)". However, there are also idioms that do not possess standard equivalents in another language, such as "ein weites Feld".

Finally, a significant portion of linguistic units exhibit significant semantic and formal similarities, but diverge in terms of actual usage. These disparities can be attributed to a range of factors, the most significant of which are:

— incomplete coincidence of semantics (usually referring to semantic features related to the periphery of meaning);

— the degree of wear of the metaphor and the place of the figurative component in terms of content of the idiom;

— compatibility preferences and restrictions;

- restrictions on the formation of grammatical forms, characteristic of units of one of the compared languages;

— degree of usage of the idiom.

It is crucial to identify and thoroughly characterize these discrepancies in order to advance the field of bilingual lexicography.

REFERENCES:

1. Avezov S. S. Machine translation to align parallel texts // International Scientific and Current Research Conferences. – 2022. – C. 64-66.

2. Sharipov S. Таржимавий лексикографиянинг тарихий ва хронологик хусусиятлари // Центр научных публикаций (buxdu. uz). – 2022. – Т. 15. – №. 15.

3. Temirova J. K. Anti-utopic motives in the novel of evgeniy zamyatin" we" // Scientific reports of Bukhara State University. $-2020. - T. 3. - N_{\odot} . 1. - C. 197-201.$

4. Авезов С. О корпусной лингвистике, трудностях перевода и принципах организации параллельных корпусов текстов //«Узбекские национальные образовательные здания теоретическое и практическое создание вопросы" Международная научно-практическая конференция. – 2022. – Т. $1. - N_{\odot}$. 1.

5. Бокарева М. «Игра в текст» как способ построения постреалистического художественного дискурса //Gospodarka i Innowacje. – 2022. – Т. 23. – С. 407-413.

6. Нигматова Л. Х. Язык и культура в лингвокультурологии: https://doi.org/10.53885/edinres. 2022.8. 08.006 Нигматова Лолахон Хамидовна, Доктор филологических наук, доцент. Бухарский государственный университет // Образование и инновационные исследования международный научно-методический журнал. – 2022. – №. 8. – С. 45-51.



UNIVERSITY"

The journal was composed in the Editorial and Publishing Department of Bukhara State University. Bukhara, 200117 Bukhara state university, main building, 2nd floor, room 219. Editorial and Publishing Department. https://buxdu.uz/32-buxorodavlat-universiteti-ilmiyaxboroti/131/131-buxoro-davlatuniversiteti-ilmiy-axboroti/ e-mail: nashriyot_buxdu@buxdu.uz Printing was permitted 27.02.2023 y. Paper format 60x84,1/8. Printed in express printing method. Conditional printing plate – 35,30. Circulation 70. Order № 30. Price is negotiable. Published in the printing house "Sadriddin Salim Buxoriy" LLC Address: Bukhara, M.Ikbol street, 11