ЯЗЫКИ НАРОДОВ ЗАРУБЕЖНЫХ СТРАН (С УКАЗАНИЕМ КОНКРЕТНОГО ЯЗЫКА ИЛИ ГРУППЫ ЯЗЫКОВ)/LANGUAGES OF PEOPLES OF FOREIGN COUNTRIES (INDICATING A SPECIFIC LANGUAGE OR GROUP OF LANGUAGES)

CORPUS APPROACH IN TRANSLATION STUDIES

Research article

Ivanova E.V.[1], *
[1] ORCID : 0000-0002-1990-3061;
[1] Saint Petersburg State University, Saint Petersburg, Russian Federation

* Corresponding author (e.v.ivanova[at]spbu.ru)

Abstract

The article is aimed at examining the advantages gained from the corpus analysis for exploring and choosing the most appropriate translation equivalents for Russian set expressions. The goal of the article resides in the consistent description of English translation equivalents against the background of their frequency in the corpus, as well as the features of the contexts in which they are used. The achievement of this goal results in supplementing some additional strategies for choosing the most suitable translation equivalent. Further research along these lines will contribute to both the more detailed and complete description of the choice principles involved and practical realization of these principles.

Keywords: set expression, translation equivalent, corpus, corpus analysis, frequency.

КОРПУСНЫЙ ПОДХОД В ПЕРЕВОДОВЕДЕНИИ

Научная статья

Иванова Е.В.[1], *
[1] ORCID : 0000-0002-1990-3061;
[1] Санкт-Петербургский государственный университет, Санкт-Петербург, Российская Федерация

* Корреспондирующий автор (e.v.ivanova[at]spbu.ru)

Аннотация

В статье рассматриваются преимущества использования корпусного анализа для исследования и выбора наиболее подходящего переводческого соответствия русским устойчивым словосочетаниям. Цель статьи заключается в последовательном описании возможных переводческих соответствий в английском языке с учетом их частотности в корпусе и характеристик, представленных в корпусе контекстов их употребления. Реализация цели дает возможность добавить некоторые уточнения в стратегии выбора адекватных соответствий. Дальнейшие исследования в подобном направлении будут способствовать как более детализированному и полному описанию принципов отбора соответствий, так и практической реализации этих принципов.

Ключевые слова: устойчивое словосочетание, переводческое соответствие, корпус, корпусный анализ; частотность.

Introduction

The theory and practice of translation have both undergone significant changes in the first quarter of the 21st century due to the impact of cognitive science and cognitive linguistics, on the one hand, and the introduction of corpus data and computer technologies in general into the linguistic research, including translation studies, on the other. Corpus approach in translation studies can be based on parallel corpora, in which texts in source and target languages are lined up sentence by sentence, or on comparable corpora, which incorporate texts in different languages on the same topics but not directly translated. Corpus data can also be employed to select domain-specific terms and their translations, which is useful for compiling dictionaries and glossaries. In this paper, we will consider some aspects of using monolingual corpus data for choosing translation equivalents.

The aim of this paper is to look at the use of corpus data in translating set combinations of words, i.e. set expressions. These expressions are chosen from the sphere of academic studies. To achieve the goal of the research definition analysis, contextual analysis, translation analysis and corpus analysis are used. Corpus analysis in the article is based on the Corpus of Contemporary American English [5].

Main results

The study allows us to outline the following results:

1. Computer technologies are widely used in translation studies, providing researchers and translators with useful tools for theoretical reflection on translation processes and practical implementation of their potential in realizing pragmatic targets of translating and developing translation techniques. The contribution of corpus data in this respect cannot be overestimated.

2. Corpus analysis of the frequencies demonstrated by the set phrases "practice exam / test", "trial exam /test", "preliminary exam/test", "mock exam /test", "simulated exam/test" identifies the most plausible version for translating the Russian set expression "тренировочный экзамен/тест". Nevertheless, though frequency has a huge impact on the choice of the target language unit in translation, other factors, such as the context as a whole should not be ignored.

3. The vast number of variable contexts supplied by corpus data for this or that set phrase helps researchers and translators delineate subtle differences in the semantics and make the appropriate choice in the process of translation. The example of "language use" and "language usage" discussed in this article illustrates the importance of considering these contexts.

Discussion

At the moment, there exists a vast and varied sphere of translation studies based on corpus analysis as well as the implementation of computer technologies in general [4], [6], [8], [9], [10]. In particular, the importance of corpus data for translation was given detailed description in the works of D.O. Dobrovolsky and E.V. Pivovarova [1], [2], [3]. The scholars outlined the irreplaceable value of corpora for examining the contextual suitability of phraseological equivalents in source and target languages. In this article set expressions, characterized by reproducibility, but not imagery, are considered regarding their translation equivalents.

It is a well-known fact that word combining is characterized by idiomaticity and particularity, and these factors are different in different languages. A classical example is "высокий мужчина" — "a tall man", "высокое здание" — "a tall/high building".

For this reason, it is often not possible to substitute words in a combination, especially in a set combination, by their dictionary equivalents in the target language. For example, "тренировочный экзамен/тест" could be translated into English as "practice exam/test", "trial exam/test", "preliminary exam/test", "simulated exam/test" or "mock exam/test".

Let's look at the frequency of these set combinations in the corpus.

The combination "preliminary exam" has the highest frequency — 17, while "preliminary test" scores 49, but only 5 cases out of this number refer to the academic sphere.

The interested students usually take a preliminary test to determine whether they want to attempt the qualifying exam.

The combination "practice exam" on the other hand is not registered at all, while "practice test" achieves the highest frequency among all the analysed combinations — 82, with only a very small number of all cases not referring to the academic sphere (7 cases).

I went online and took the practice test. I knew I struggled with math and science.

The frequency of the next set expressions is as follows: "mock exam" — 3, "mock test" — 4.

Lastly, even though the students participated in a full, eight-hour timed mock exam, it only simulates the real MCAT testing situation,

Alternatively, we'll be happy to mail you an in-home mock test.

The combination "trial exam" is not encountered in the corpus, "trial test" is used 4 times, but not in the academic sense:

This week, the California-based producer of baby carrots launched a trial test of its newest product

There are no examples for "simulated exam", while "simulated test" is registered 5 times, but only 1 example refers to the academic domain:

…undergraduate courses such as general psychology or personality theory. In this situation, simulated test items should be used to demonstrate any given device or technique.

If we go by the corpus data and the frequency of the potential translation equivalents for choosing the best option for translation, we must decide on "preliminary exam" and "practice test". But the combinations "mock exam/test" look more balanced regarding the usage of the same first word. Another factor influencing the decision of the translator could be the meaning of the noun "mock/mocks", designating an exam or test taken for practice. But the corpus data for the frequency of this word requires manual processing because the frequency of the noun and the frequency of the corresponding verb are not differentiated.

The translation equivalents for "оценка научно-исследовательской работы" have similar frequencies in the corpus: "evaluation of research" — 12, "assessment of research" — 10. So going by the frequency, we can choose both.

Another aspect that should be taken into account when looking for the translation equivalents is the context.

Let's look at the Russian frequently encountered set expression "использование языка". The dictionary [7] offers the following definitions for the nouns "use" and "usage":

Use — the act of using something; the state of being used

Usage — the way in which words are used in a language: *current English usage*

It looks at first glance that "language usage" is a more appropriate translation. But the corpus data raise some doubt in this respect.

"language use" — 429

Lying can cause behavioral change in language use because it is cognitively demanding

… problems associated with poor communication between patients and doctors, including issues of language use,

"language usage" — 128

While collocation can reveal new patterns in language usage, it tends to be an exploratory tool

The author notes that the areas in which students struggled were mainly centred on language usage, expressed by the educators as 'the inability of students to express themselves'.

Still, if the goal of proper language usage is to be understood by others, clarity is better than complexity.

It is possible to assume that the contexts with "language usage" are more particular and concrete than those with "language use", which look more generalized, but this difference is very subtle and not traceable in all sentences.

Conclusion

Computer technologies in general and corpus analysis in particular are and will be playing an increasing role in translation, which undoubtedly calls for further theoretical and practical study of various cases of their implementation. All the case studies will provide invaluable material for the development of translation theory and translation techniques. In this paper, only two aspects of using a monolingual corpus for translation are examined.

Конфликт интересов	**Conflict of Interest**
Не указан.	None declared.
Рецензия	**Review**
Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.	All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы / References

1. Добровольский Д.О. Корпусы текстов и двуязычная фразеография. / Д.О. Добровольский // Вестник Новосибирского государственного педагогического университета. — 2015. — 5(27). — С. 23–37.

2. Добровольский Д.О. Корпусный подход к исследованию фразеологии: новые результаты по данным параллельных корпусов. / Д.О. Добровольский // Вестник Санкт-Петербургского университета. Язык и литература. — 2020. — 17(3). — С. 398–411.

3. Пивоварова Е.В. Метод корпусного анализа в изучении фразеологии немецкого языка. Теоретический обзор. / Е.В. Пивоварова // Филологические науки. Вопросы теории и практики. — 2019. — 12(12). — С. 263–268.

4. Baker M. Corpus linguistics and translations studies: Implications and Applications / M. Baker // Text and Technology: In Honour of John Sinclair. — Amsterdam/Philadelphia: John Benjamins, 1993. — P. 233-252.

5. Corpus of Contemporary American English. — 2025. — URL: https://www.english-corpora.org/coca (accessed: 04.09.2025).

6. Ding J. Corpus-based translation studies: Examining media language through a linguistic lens / J. Ding // SHS Web of Conferences. — 2024. — 185. — URL: https://creativecommons.org/licenses/by/4.0/ (accessed: 12.09.2025).

7. Oxford Advanced Learner's dictionary. — 6th. — Oxford : Oxford University Press, 2000. — 1540 p.

8. Saldanha G. Principles of corpus linguistics and their application to translation studies research / G. Saldanha // Revista Tradumatica. — 2009. — 7.

9. Umerova M.V. Parallel corpora in translation studies / M.V. Umerova // Sciences of Europe. — 2018. — 29. — P. 56–59.

10. Wang G. An analytical framework for corpus-based translation studies / G. Wang, Y. Xyn // Humanities and Social Sciences Communications. — 2024. — 11. — URL: https://www.nature.com/articles/s41599-024-04250-4 (accessed: 04.09.2025).

Список литературы на английском языке / References in English

1. Dobrovol'skij D.O. Korpusy' tekstov i dvuyazy'chnaya frazeografiya [Text corpora and two-language phraseology]. / D.O. Dobrovol'skij // Bulletin of Novosibirsk State Pedagogical University. — 2015. — 5(27). — P. 23–37. [in Russian]

2. Dobrovol'skij D.O. Korpusny'j podxod k issledovaniyu frazeologii: novy'e rezul'taty' po danny'm parallel'ny'x korpusov [Corpus approach to phraseological studies: new results based on parallel corpus data]. / D.O. Dobrovol'skij // Bulletin of Saint-Petersburg state university. Language and Literature. — 2020. — 17(3). — P. 398–411. [in Russian]

3. Pivovarova E.V. Metod korpusnogo analiza v izuchenii frazeologii nemeczkogo yazy'ka. Teoreticheskij obzor [Method of corpus analysis in German phraseology studies. Theory review]. / E.V. Pivovarova // Philological Sciences. Theoretical and Practical Issues. — 2019. — 12(12). — P. 263–268. [in Russian]

4. Baker M. Corpus linguistics and translations studies: Implications and Applications / M. Baker // Text and Technology: In Honour of John Sinclair. — Amsterdam/Philadelphia: John Benjamins, 1993. — P. 233-252.

5. Corpus of Contemporary American English. — 2025. — URL: https://www.english-corpora.org/coca (accessed: 04.09.2025).

6. Ding J. Corpus-based translation studies: Examining media language through a linguistic lens / J. Ding // SHS Web of Conferences. — 2024. — 185. — URL: https://creativecommons.org/licenses/by/4.0/ (accessed: 12.09.2025).

7. Oxford Advanced Learner's dictionary. — 6th. — Oxford : Oxford University Press, 2000. — 1540 p.

8. Saldanha G. Principles of corpus linguistics and their application to translation studies research / G. Saldanha // Revista Tradumatica. — 2009. — 7.

9. Umerova M.V. Parallel corpora in translation studies / M.V. Umerova // Sciences of Europe. — 2018. — 29. — P. 56–59.

10. Wang G. An analytical framework for corpus-based translation studies / G. Wang, Y. Xyn // Humanities and Social Sciences Communications. — 2024. — 11. — URL: https://www.nature.com/articles/s41599-024-04250-4 (accessed: 04.09.2025).