

КОГНИТИВНАЯ И КОРПУСНАЯ ЛИНГВИСТИКА COGNITIVE AND CORPUS LINGUISTICS



<https://doi.org/10.15507/2076-2577.26182.170-182>


<https://elibrary.ru/jvwfmv>

УДК / UDC 81'373:621.013:811.511.111:811.161.1




Оригинальная статья / Original article

Разметка финско-русского корпуса фразеологизмов на примере идиом со значением «умереть»

Н. В. Белкина , Д. Ю. Шарешик

*Российский государственный педагогический университет
им. А. И. Герцена, г. Санкт-Петербург, Российская Федерация,
<https://ror.org/01e5ckr65>*

 bel_nat@mail.ru

Аннотация

Введение. Сравнительный языковой анализ картины мира определенных этносов помогает понять различия национальных культур и то, как они дополняют друг друга. Размеченный по определенным критериям финско-русский корпус фразеологических единиц дает возможность выявить общие и национальные черты восприятия окружающего мира. Создание и лингвистическая разметка специализированного двуязычного корпуса фразеологизмов является проблемой, решив которую, возможно получить эффективный инструмент для сопоставительного анализа и машинного перевода. Цель исследования – определение критериев разметки корпуса финско-русских фразеологизмов и апробация ее на тематической группе со значением «умереть».

Материалы и методы. Материалом послужили финские и русские фразеологические единицы тематической группы «умереть», отобранные методом сплошной выборки из финских и русских словарей фразеологизмов, сленга и разговорной речи. Необходимость данного исследования обусловлена отсутствием цифровых параллельных фразеологических корпусов финского и русского языков и, соответственно, критериев разметки подобного корпуса. Пилотный корпус, насчитывающий в настоящее время 123 финских и 136 русских фразеологических единиц, предназначен для сопоставительного лингвокультурологического анализа и прикладных задач машинного перевода. В основе методологии лежит сопоставительный подход, учитывающий тематико-семантические, стилистические и этимологические параметры фразеологизмов.

Результаты исследования и их обсуждение. Предложена детализированная схема разметки, включающая семантическую классификацию, стилистические пометы и указание на происхождение. В процессе разметки фразеологических единиц со значением «умереть» применяется комплекс критериев: семанτικο-тематический (с опорой на универсальную семантическую метафору); семантический (с учетом коннотативных компонентов значения); стилистический; критерий языковой пары; этимологический. Дополнительно при разметке учитываются наличие дериватов и синонимов, а также особенности грамматического оформления фразеологических единиц, что необходимо для последующей генерации корректных контекстов.

Заключение. Разработанный корпус представляет ценность для выявления универсальных и культурно-обусловленных черт в восприятии концепта смерти в финской и русской лингвокультурах, а также может послужить основой для дальнейших исследований в области фразеологии и межъязыковой коммуникации. Создание и разметка корпуса по предлагаемым критериям не требуют специальных знаний в области программирования и доступны при использовании открытого программного обеспечения. Подобный корпус является шагом к созданию компьютерной модели языковой картины мира, которую можно использовать для работы с большим количеством данных и обучения искусственного интеллекта.

Ключевые слова: фразеология, лингвокультурология, параллельный корпус, семантическая разметка, финский язык, русский язык, концепт «смерть»

© Белкина Н. В., Шарешик Д. Ю., 2026



Контент доступен по лицензии Creative Commons Attribution 4.0 License.
This work is licensed under a Creative Commons Attribution 4.0 License.



Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Для цитирования: Белкина Н.В., Шарешик Д.Ю. Разметка финско-русского корпуса фразеологизмов на примере идиом со значением «умереть». *Финно-угорский мир*. 2026;18(2):170–182. <https://doi.org/10.15507/2076-2577.26182.170-182>

Making the Finnish-Russian Corpus of Phraseological Units: A Case Study of Idioms Meaning “to Die”

N. V. Belkina ✉, D. Yu. Shareshik

Herzen State Pedagogical University of Russia,

St. Petersburg, Russian Federation, <https://ror.org/01e5ckr65>

✉ bel_nat@mail.ru

Abstract

Introduction. A comparative linguistic analysis of the worldview of particular ethnic groups makes it possible to understand the differences between national cultures and the ways in which they complement one another. A Finnish-Russian corpus of phraseological units annotated according to specific criteria provides an opportunity to identify both universal and culture-specific features in the perception of the surrounding world. The creation and linguistic annotation of a specialized bilingual corpus of phraseological expressions constitutes a challenge; however, once resolved, it may yield an effective tool for comparative analysis and machine translation. The aim of the study is to determine the criteria for annotating a corpus of Finnish-Russian phraseological units and to test this annotation framework on a thematic group with the meaning “to die”.

Materials and Methods. The material for the study comprised Finnish and Russian phraseological units belonging to the thematic group “to die”, selected through continuous sampling from Finnish and Russian dictionaries of phraseology, slang, and colloquial speech. The relevance of this research stems from the absence of digital parallel phraseological corpora for the Finnish and Russian languages and, consequently, from the lack of annotation criteria for such a corpus. The pilot corpus, which currently includes 123 Finnish and 136 Russian phraseological units, is intended for comparative linguocultural analysis and for applied machine translation tasks. The methodology is based on a comparative approach that takes into account the thematic-semantic, stylistic, and etymological parameters of phraseological units.

Results and Discussion. A detailed tagging scheme is proposed, including semantic classification, stylistic labels, and indications of origin. In the process of tagging phraseological units meaning “to die,” a set of criteria is employed: a semantic-thematic criterion based on universal semantic metaphor; a semantic criterion taking into account connotative components of meaning; a stylistic criterion; a language pair criterion; and an etymological criterion. In addition to semantic, etymological, and stylistic parameters, the tagging process also considers the presence of derivatives and synonyms, as well as the grammatical features of phraseological units, which is necessary for the subsequent generation of contextually appropriate examples.

Conclusion. The developed corpus is valuable for identifying both universal and culture-specific features in the perception of the concept of death within Finnish and Russian linguistic cultures, and it also provides a foundation for further research in the fields of phraseology and cross-linguistic communication. The creation and tagging of the corpus according to the proposed criteria do not require specialized programming knowledge and can be carried out using open-source software. Such a corpus constitutes a step toward the development of a computational model of the linguistic worldview, which can be employed for processing large-scale data and training artificial intelligence systems.

Keywords: phraseology, linguocultural studies, parallel corpus, semantic annotation, Finnish language, Russian language, concept of death

Conflict of interest: The authors declare no conflict of interest.

For citation: Belkina N.V., Shareshik D.Yu. Making the Finnish-Russian Corpus of Phraseological Units: A Case Study of Idioms Meaning “to Die”. *Finno-Ugric World*. 2026;18(2):170–182. <https://doi.org/10.15507/2076-2577.26182.170-182>

Введение

Фразеологизмы представляют собой особый пласт системы языка, отражающий его культурные и исторические особенности. Исследование вопроса о национально-культурной специфике фразеологизмов является традиционным в лингвистике. Этой темой занимались такие ученые, как А. Вежбицкая, В. Н. Телия, В. А. Маслова,

Д. О. Добровольский и др. Так, В. Н. Телия отмечает, что «для идиом характерна образная мотивированность, которая напрямую связана с мировидением народа-носителя языка, идиомы в принципе обладают культурно-национальной коннотацией»¹. Фразеологизм представляет собой уникальную лексическую единицу языка, значение которой не выводится из ее составляющих и позволяет метафорично выразить отношение к происходящему. Кроме того, «фразеология есть фрагмент языковой картины мира, а ее единицы всегда обращены на субъект, то есть они возникают не столько для того, чтобы описывать мир, сколько для того, чтобы его интерпретировать, оценивать и выражать к нему субъективное отношение»². Язык, отражая видение мира, является своего рода ключом к познанию менталитета конкретного народа [1]. Соответственно, сравнительный языковой анализ картины мира того или иного этноса помогает понять различия национальных культур и то, как они дополняют друг друга [2]. Исследователи подчеркивают: культурные коды, содержащиеся во фразеологизмах, могут быть трудны для понимания носителем другого языка, что предопределяет необходимость их детального изучения.

Исследователь Д. О. Добровольский утверждает, что абсолютные межъязыковые эквиваленты фразеологизмов (далее – ФЕ) и метафорических единиц встречаются относительно редко, это обусловлено преимущественно различиями в технике номинации, а не их культурной уникальностью. Также предлагается два подхода к определению национально-культурной специфики фразеологизмов: сопоставительный и интроспективный. В сравнительных исследованиях выделяют полные эквиваленты (единицы, возводимые к одному источнику, тождественные по значению), эквивалентные (максимально совпадающие по значению) и безэквивалентные (не имеют фразеологического аналога в другом языке) [3].

В то же время Н. Ф. Крюкова делает акцент на фразеологической эквивалентности в когнитивно-герменевтическом аспекте: в его основе лежит «не сопоставление лексических единиц, а анализ метафор, их формирующих» [4]. Н. Н. Почуева отмечает, что ФЕ с разной образной составляющей отражают как универсальность законов мышления, так и национальную специфику видения мира разными народами [5].

В данном исследовании мы придерживаемся сопоставительного подхода и обращаемся к семантике, поскольку лексическое и грамматическое наполнение у ФЕ разных языков различно по определению. Специфика образного осмысления действительности в разных языках становится очевидной при сопоставлении фразеологизмов, относящихся к одной тематической группе, также возможно «установить как универсальные, так и культурно-специфичные особенности представления окружающей действительности» при «описании идентичных ситуаций в разных языках» [6, с. 140]. В настоящее время имеется небольшое количество финских фразеологических словарей и не существует двуязычного цифрового финско-русского фразеологического ресурса, что и создает необходимость его создания.

Цель исследования – разработка критериев разметки корпуса финско-русских фразеологизмов и ее апробация на тематической группе фразеологизмов со значением «умереть».

Обращение к данной тематической группе связано с тем, что явление смерти – неизменно интересующий человека факт, который во все времена привлекал внимание исследователей из разных сфер науки. Хотя концепт «смерть» универсален, у каждого народа собственный взгляд на данное явление, своеобразие которого раскрывается в ходе лингвокультурологического анализа фразеологических единиц [2].

¹ Телия В.Н. Русская фразеология. Семантический, прагматический и лингвокультурологический аспекты: моногр. М.: Школа «Языки русской культуры»; 1996. С. 214–215. <https://elibrary.ru/sumhnj>

² Маслова В.А. Лингвокультурология: учеб. пособие. М.: Академия; 2004. 202 с.

Обзор литературы

Проблема исследования стоит на стыке корпусной лингвистики, фразеологии, метафоры и семасиологии. Именно корпусная лингвистика характеризуется наличием специальной разметки по метаданным и делает прикладные лингвистические исследования проще и эффективнее. Рассмотрим основные понятия исследования подробнее. Т. МакЭнери и В. Брезина сформулировали фундаментальные принципы корпусной лингвистики и аннотирования корпусов. Они утверждают, что «корпусная лингвистика в форме аннотирования корпусов – это область, где становятся ясными онтологические предпосылки лингвистики» [7].

Британский лингвист Дж. Лич подчеркивал, что разметка – это «вспомогательный инструмент, а не истина в последней инстанции»³. Он же сформулировал основополагающие принципы разметки корпусов: возможность легкого отделения разметки от текста; полезность разным исследователям; четкое описание правил разметки.

В то время как Д. Синклер акцентировал внимание на репрезентативности и гомогенности разметки на всех уровнях [8], Т. МакЭнери выделял такие преимущества качественной разметки, как удобство эксплуатации и многофункциональность, т. е. возможность ее использования для задач, не предусмотренных при создании [9].

Несмотря на значительный объем существующих научных трудов в области финской и русской фразеологии, сравнительный семантический анализ ФЕ все еще остается неразработанным аспектом. Инструментом для подобного семантического исследования может послужить параллельный корпус, позволяющий «перевод на другой язык рассматривать как свидетельство значения анализируемой единицы языка оригинала» [10, с. 22–23]. Однако поиск ФЕ в текстовом параллельном корпусе сопряжен с большой долей ручного труда, поскольку компьютер может искать отдельные слова, но не всегда распознает устойчивые сочетания как единое целое, и различия в языковых картинах мира действительно обуславливают несовпадения в образности фразеологизмов и приводят к межязыковым несоответствиям [11].

Преодоление подобных трудностей мы видим в создании специальной разметки, предполагающей использование универсальных метафор, опирающихся на схожесть когнитивных механизмов даже для неродственных языков, а именно русского и финского.

Смерть в толковых словарях русского и финского языков рассматривается как физиологический процесс прекращения жизнедеятельности живого существа или конец жизни⁴. В работах по русской и английской фразеологии представлены различные классификации метафор, выражающих понятие «умереть». Здесь и далее мы дополним классификации, составленные для русских и английских ФЕ, финскими примерами из словарей⁵. Так, Е. Середя делит фразеологизмы на три группы: 1. факт смерти без уточнения обстоятельств: рус. – *окончить дни свои*, фин. – *siirtyä ajasta ikuisuuteen* ‘перейти из времени в вечность’; 2. насильственная смерть: рус. – *пасть в бою*, фин. – *saada surmansa* ‘погибнуть’; 3. ненасильственная смерть: рус. – *умереть на своем рабочем месте*, фин. – *kuolla saappaat jalassa* ‘умереть в сапогах’⁶.

Л. М. Пантелеева и С. В. Хоробрых рассматривают смерть как путь в загробное пространство, имеющий три основных этапа: уход из дома, путь к месту захоронения, отправление на «тот свет» [12].

³ Garside R., Leech G., McEnery T. (eds.) *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London; New York: Longman; 1997. 281 p.

⁴ Suomisanakirja «kuoleminen» artikkeli [Электронный ресурс]. URL: <https://www.suomisanakirja.fi/kuoleminen> (дата обращения: 17.10.2025).

⁵ Kielitoimiston sanakirja [Электронный ресурс]. URL: <https://www.kielitoimistonsanakirja.fi/#/kuolla> (дата обращения: 17.09.2025); *Urbaani sanakirja* [Электронный ресурс]. URL: <https://urbanisanakirja.com/search/?q=kuolla> (дата обращения: 16.09.2025).

⁶ Середя Е. Представление концепта «Death» в английской фразеологической картине мира: автореф. дисс. ... канд. филол. наук. СПб.; 2016. 22 с.

С другой стороны, Р. М. Бикмухаметова выделяет в ФЕ изучаемой нами тематической группы следующие метафорические образы: 1. сон (рус. – *заснуть вечным сном*, фин. – *nukkuu pois* ‘уснуть’); 2. конец существования (рус. – *окончить дни свои*, фин. – *viellä viimeinen sivu* ‘перелистнуть последнюю страницу’); 3. живое существо (рус. – *старуха с косой*, фин. – *Tuonenakka* – хозяйка царства мертвых), душа человека (рус. – *испустить дух*, фин. – *lakata hengittämästä* ‘прекратить дышать’); 4. дорога (рус. – *отправляться к праотцам*, фин. – *lähteä manan majoille* ‘отправиться в хижины царства мертвых’); 5. место захоронения (рус. – *ложиться в гроб*, фин. – *tennä matojen valtakuntaan* ‘пойти в царство червей’); 6. прекращение нормального функционирования жизненно важных частей тела (рус. – *протянуть ноги*, фин. – *lähteä jalat edellä* ‘уйти вперед ногами’) [2, с. 349].

Исследователь О. С. Орлова предлагает следующие метафоры: 1. переход из одного пространства в другое, путь, путешествие (рус. – *отойти в мир иной*, фин. – *lähteä viimeiselle matkalle*); 2. воссоединение с близкими и создателем (рус. – *уйти на свидание с апостолом Петром*, фин. – *käydy taivaan porteilla* ‘подойти к небесным вратам’); 3. телесные знаки (рус. – *заснуть непробудным сном*, фин. – *nukkuu ikuiseseen uneen* ‘уснуть вечным сном’). Некоторые метафоры имеют дисфемистичную окраску (рус. – *двинуть кони*, фин. – *heittää lusikka nurkkaan* ‘бросить ложку в угол’) [13, с. 68].

На основе данных классификаций мы предлагаем следующие основные критерии разметки ФЕ: 1. семантика (для рассматриваемой тематической группы это метафоры: приход к богу, вечный покой, физиологический конец, окончание пути, переход в локус, дисфемизмы); 2. принадлежность к определенному стилю; 3. этимология.

Материалы и методы

Методика исследования предполагает последовательное выполнение следующих задач:

- 1) отбор массива ФЕ финского и русского языка определенной тематической группы;
- 2) определение принадлежности ФЕ к определенной семантической метафоре;
- 3) определение их коннотации с помощью контекстов из национальных корпусов языков (Национальный корпус русского языка (НКРЯ), Finnish Web Corpus (fiTenTen));
- 4) определение стилистической принадлежности и происхождения ФЕ;
- 5) сопоставление ФЕ финского и русского языков для определения языковой пары;
- 6) создание корпуса с помощью имеющегося открытого программного инструментария;
- 7) аннотирование корпуса в соответствии с определенными критериями разметки;
- 8) создание комментариев к элементам корпуса;
- 9) апробация корпуса экспертами и внесение возможных правок.

Для создания выборки использовались специализированные словари фразеологизмов, сленга и разговорной речи, тематические форумы и социальные сети⁷. С целью обеспечения однородности корпуса при отборе ФЕ мы придерживались тематического критерия (основное значение «умереть», включая эвфемизмы и дисфеминизмы) и критерия языковой пары: единица должна иметь установленный переводческий эквивалент в другом языке или быть сопоставимой по смыслу, стилю и коннотации.

Выборка составила 136 русских и 123 финских фразеологизма (таблица).

⁷ Фразеологический словарь русского языка. Под ред. А.И. Молоткова. М.: Советская энциклопедия; 1968. 543 с.; Бирих А.К., Мокиенко В.М., Степанова Л.И. Словарь русской фразеологии. Историко-этимологический справочник. СПб.: Фолио-Пресс; 1998. 704 с.; Жуков А.В., Жукова М.Е. Словарь современной русской фразеологии. М.: АСТ-Пресс; 2015. 416 с.; Jarva V., Nurmi T. Oikeeta suomee. Suomen puhekielen sanakirja. Helsinki: Gummerus; 2006. 511 s.; Kari E. Naulan Kantaan: Nykysuomen idiomisanakirja. Helsinki: Otava; 1993. 232 s.; Kielitoimiston sanakirja [Электронный ресурс]. URL: <https://www.kielitoimistonsanakirja.fi/#/kuolla> (дата обращения: 17.09.2025); Muikku-Werner P. Suurella sydämellä ihan sikana. Suomen kielen kuvailu fraasisanakirja. Helsinki: Gummerus; 2008. 467 s.; Suomisanakirja «kuoleminen» artikkeli [Электронный ресурс]. URL: <https://www.suomisanakirja.fi/kuoleminen> (дата обращения: 17.10.2025); Urbaani sanakirja [Электронный ресурс]. URL: <https://urbanisanakirja.com/search/?q=kuolla> (дата обращения: 16.09.2025).

Таблица 1. Структура выборки финских и русских фразеологизмов
Table 1. Structure of the sample of Finnish and Russian phraseological units

Количество ФЕ / Number of Phraseological Units	Всего / Total	Стиль / Style			Происхождение / Origin		Семантика / Semantics					
		книжный	нейтральный	разговорный	исходный	заимствованный	окончание пути	приход к богу	вечный покой	физиологический конец	переход в локус	дисфемизмы
Русские ФЕ / Russian Phraseological Units	136	43	38	71	112	24	17	5	7	57	51	58
Финские ФЕ / Finnish Phraseological Units	123	14	36	66	95	28	29	3	8	37	46	49

Источник: таблица составлена авторами.
Source: Table is compiled by the authors.

Видно, что на уровне стилистики в обоих языках преобладает разговорный стиль (хотя в русском языке книжные идиомы встречаются гораздо чаще). Большинство фразеологизмов со значением «умереть» являются незаимствованными. На семантическом уровне в русском языке преобладают дисфемизмы, единицы со значением «физиологический конец» и «переход в локус». В финском языке дисфемизмы также преобладают, однако «переход в локус» встречается чаще, чем «физиологический конец». Стоит отметить и большее по сравнению с русским языком количество идиом со значением «окончание пути».

Результаты исследования и их обсуждение

Создание специализированных лингвистических корпусов – важный инструмент в современной филологии, переводоведении, лексикографии и компьютерной лингвистике. Принципы составления двуязычных корпусов и описание словарной статьи как формализованного текста даны Д. О. Добровольским, который подчеркивает, что словарная статья «должна неукоснительно следовать особым правилам, предполагающим деление сообщаемой информации на отдельные блоки, каждый из которых содержит сведения определенного типа». Это позволяет «выявить значимые связи как между отдельными его частями, так и внутри конкретных словарных статей». Помимо этого, исследователь предлагает «обращение к различным пластам семантики этой единицы [ФЕ] с учетом ее многозначности, ее прагматическим и стилистическим характеристикам, ограничениям на возможность ее употребления в дискурсе, указание на... потенциал ее варьирования» [14, с. 126–127].

Двуязычный корпус фразеологизмов, сфокусированный на конкретной тематической группе (например, на категории «смерть»), представляет особую ценность для контрастивного анализа, изучения культурных коннотаций и разработки систем автоматического перевода устойчивых выражений. Процесс создания и разметки такого корпуса включает несколько взаимосвязанных этапов.

Первоначальным и фундаментальным этапом является сбор данных. Источники формирования корпуса должны быть репрезентативными и авторитетными, чтобы обеспечить надежность и достоверность включаемых единиц.

Однородность корпуса может быть обеспечена благодаря четко сформулированным параметрам:

1) тематический параметр: включаются только фразеологизмы (идиомы, устойчивые сравнения, фразеосхемы), основным значением которых является обозначение смерти, умирания или состояния после смерти, включая эвфемизмы и дисфемизмы. В данной группе многообразие эвфемизмов и дисфемизмов отражает сложное отношение общества к теме смерти, где переплетаются страх, уважение и протест;

2) параметр устойчивости: отбираются лексически устойчивые сочетания, обладающие признаками воспроизводимости, а не свободно создаваемые в речи;

3) параметр идиоматичности: единицы с переносным значением, не выводимым полностью из суммы значений компонентов (например, фин. *heittää lusikka nurkkaan* – букв. ‘бросить ложку в угол’ или рус. *дать дуба*);

4) параметр частотности и употребительности: включаются фразеологизмы, зафиксированные в словарях или подтвержденные корпусными данными как активно используемые (хотя редкие и архаичные единицы также могут представлять интерес для исторического или диалектологического анализа);

5) параметр языковой пары: для включения в двуязычный корпус фразеологизм должен иметь либо установленный переводческий эквивалент в другом языке (например, фин. *antaa viimeinen hengenveto*, рус. *испустить последний вздох*), либо быть сопоставимым по смыслу, стилю и коннотации (например, фин. *Viedä viimeinen siivu* ‘унести последний кусок’, рус. *перевернуть последнюю страницу*).

Далее следует определение структуры корпуса. Эффективное хранение и поиск информации требуют продуманного формата, наиболее гибким решением является использование реляционной базы данных (например, SQLite, MySQL, PostgreSQL) или структурированных файлов (в формате XML, JSON). Табличные форматы (Excel, CSV) удобны для первоначального сбора и просмотра, но менее эффективны для сложных запросов и связей. Обращение к формату базы данных связано с тем, что это позволяет легко добавлять новые поля разметки, осуществлять сложный поиск по множеству параметров и управлять большими объемами данных.

Каждая словарная статья корпуса содержит следующие обязательные поля:

– уникальный идентификатор. Цифровой или буквенно-цифровой код для однозначной ссылки на запись (например, KLM_FI_001, KLM_RU_045);

– фразеологизм. Точная запись фразеологизма на финском или русском языке (*heittää veivinsä* или *отбросить коньки*, причем для одного фразеологизма может быть несколько эквивалентов с разными стилистическими оттенками);

– буквальный перевод. Дословный перевод компонентов на другой язык, необходимый для понимания внутренней формы (*heittää veivinsä* ‘бросить свои мослы’);

– семантическая помета. Классификация на основе предоставленной разметки: приход к богу, вечный покой, физиологический конец (ненасильственная смерть), физиологический конец (насильственная смерть), окончание жизненного пути, переход в другой локус, дисфемистичная окраска;

- стилистическая помета. Маркировка стилистического регистра (нейтральный, книжный/высокий, разговорный/сниженный, грубый/вульгарный, жаргонный/сленговый);
- помета происхождения. Указание на источник фразеологизма (исконное/заимствованное);

- контекст (опционально). Примеры реального употребления фразеологизма с указанием источника (художественный текст, СМИ, устная речь);

- комментарий. Поле для дополнительных лингвистических или культурологических заметок (калька / полный эквивалент, диалектная единица, жаргон/сленг, литературная, библеизм/мифологизм, синонимические ряды, региональные варианты).

Хотя фразеологизм является устойчивым целым, разметка может учитывать:

- вариативность формы – наличие синонимичных компонентов (рус. *отбросить коньки/копыта/лапти*, фин. *heittää veivinsä/hokkarinsa/femmat*). Варианты должны быть зафиксированы в одной записи с пометой «вариант» или как отдельные, но связанные записи;

- морфологию – особенности грамматического оформления (управление, фиксированный падеж / форма глагола). Это важно для генерации правильных контекстов в будущем;

- синтаксическую функцию: может ли фразеологизм выступать в роли сказуемого, вводной конструкции и др. (чаще всего фразеологизмы со значением «умереть» – предикативные);

- тип дискурса (бытовой разговор, художественная литература, пресса, интернет-форум), в котором преимущественно употребляется фразеологизм. Это уточняет стилистическую помету;

- коннотации (ирония, грубость, почтительность, фамильярность), а также возможная профессиональная принадлежность, что напрямую связано со стилистической и семантической разметкой (дисфемизм/эвфемизм);

- сочетаемость с лексемами или грамматическими конструкциями (например, «внезапно дал дуба», «мирно отдал Богу душу»). Эта информация будет фиксироваться в поле «Контекст» или «Комментарий».

На следующем этапе происходила разметка корпуса – процесс добавления лингвистической информации (аннотаций, метаданных) к собранным фразеологизмам, превращающий простую коллекцию в структурированный исследовательский инструмент.

Методы разметки требуют выбора адекватных инструментов и проработки процедуры. Нами использованы следующие инструменты и программы:

- системы управления базами данных (СУБД) для хранения и первичной разметки (заполнения полей, первоначальной структуризации корпуса) (например, Excel);

- специализированное программное обеспечение для аннотирования: программы, предоставляющие удобный интерфейс для лингвиста-аннотатора (например, UAM CorpusTool, AntPConc для создания собственно параллельного корпуса). Подобные инструменты позволяют эффективно назначать теги, связывать контексты, вести историю изменений;

- скрипты для автоматизации. Простые скрипты на Python (с использованием библиотек Pandas для данных, регулярных выражений для поиска паттернов) помогают в предварительной обработке данных, поиске дубликатов, генерации уникальных ID, массовом присвоении некоторых очевидных помет (например, всем фразеологизмам, содержащим слово *taivas* ‘небо’ или *jumala* ‘бог’, автоматически добавить помету «приход к богу» или «переход в другой локус»). В соответствии с данными целями был разработан скрипт на Python, позволяющий добавлять новые фразеологизмы в параллельный корпус;

- онлайн-платформы для совместной работы: системы типа Google Sheets, Google Documents или специализированные корпусные менеджеры, поддерживающие одновременную работу и контроль версий.

При осуществлении разметки возник ряд сложностей. Так, на этапе подготовки данных потребовалась очистка, включающая удаление дубликатов, приведение к единому регистру, нормализацию орфографии, особенно для разговорных форм.

На этапе лексикографической разметки, т. е. заполнения обязательных полей (каноническая форма, буквальный перевод, значение), мы столкнулись с неоднозначностью интерпретации внутренней формы фразеологизма и точной передачи его значения на другом языке, особенно для уникальных культурно-специфичных образов. Например, в русском языке: *перейти в Могилевскую губернию*; *лечь (уснуть/выпить/поговорить) с Ильичом*; *войти в Политбюро*. В финском языке: *jättää kamppeenasa kirpparille* ‘оставить пожитки для секонд-хенда’; *kasvattaa varpaiden välistä päivänkakkaroina* ‘выращивать ромашки между пальцев ног’; *ottaa kyltti alas* ‘снять табличку’.

На этапе семантической разметки, а именно присвоения фразеологизмам семантических помет на основе разработанной классификации, сложность состояла в разграничении критериев. Например, *купить билет в один конец* – это «окончание пути» или «переход в локус»? В таких случаях обращались к контексту или экспертному решению.

Субъективность разметки прослеживается и на этапе определения стилистического регистра. Разграничение «разговорного» и «сниженного» или «жаргонного» и «вульгарного» требует опоры на словарные пометы, корпусную частотность и экспертные знания. Фразеологизмы из интернет-сленга (urbaanisanakirja.com) особенно сложны для однозначной стилистической классификации.

Указание источника фразеологизма требует не всегда доступных этимологических исследований. Разграничение «калек» и «полных эквивалентов», а также диалектные формы требуют подтверждения источниками или носителями.

На этапе верификации и контроля качества для решения проблемных случаев и неоднозначных толкований потребовалась перекрестная проверка разметки разными аннотаторами, выявление и исправление ошибок и несоответствий, благодаря чему был снят ряд сложностей.

При экспорте и публикации корпуса в формате, удобном для конечного пользователя, используется просмотр корпуса через AntConc, в перспективе предполагается создание веб-интерфейса для поиска.

Приведем несколько примеров размеченных данных из корпуса.

1. Фразеологизм (RU): *отойти в лоно Авраама/Авраамово*:

– букв. перевод (FI): *mennä Aabrahamin helmaan*;

– значение (RU): умереть, отойти в загробный мир (по библейскому представлению);

– значение (FI): *kuolla, päästä tuonelaan*;

– семантическая помета: приход к богу / переход в другой локус;

– стилистическая помета: книжный, высокий (архаичный, библейский);

– помета происхождения: библеизм;

– контекст (RU): «Умер нищий и отнесен был Ангелами на лоно Авраамово»

(источник: Евангелие от Луки, 16:22);

– контекст (FI): «*Kerjäläinen kuoli, ja enkelit kantoivat hänet Aabrahamin helmaan*»

(перевод);

– комментарий: Архаичный и религиозный оттенок. В современной речи используется редко, преимущественно в стилизованных или ироничных контекстах. Единица *mennä* ‘идти’ в составе фразеологизма часто заменяется другими глаголами *siirtyä* ‘переходить’, *päästä* ‘попадать’.

2. Фразеологизм (FI): *siirtyä suuren keskusviraston palvelukseen*:

– букв. перевод (RU): перейти на службу в большое центральное ведомство;

– значение (FI): *kuolla*;

– значение (RU): умереть;

– семантическая помета: дисфемистичная окраска (ироничная) / переход в другой локус (бюрократический «загробный мир»);

- стилистическая помета: разговорный, ироничный, сленговый;
- помета происхождения: жаргон/сленг (образное переосмысление госслужбы);
- контекст (FI): «Eilen sai kuulla, että sedästä on tullut virkamies – siirtyi suuren keskusviraston palvelukseen»;
- контекст (RU): «Вчера узнал, что дядя стал чиновником – перешел на службу в большое центральное ведомство» (прямой перевод, передающий иронию);
- комментарий: Типичный пример современного ироничного финского сленга. Шутливое выражение, отсылающее к неизбежной даже в загробной жизни бюрократии.

Примеры размеченных данных на русском и финском языках представлены на рисунках 1, 2.

Таким образом, создание разметки корпуса по предлагаемым нами критериям не требует специальных знаний в области программирования и доступно при использовании открытого программного обеспечения, однако верификация данных требует перекрестной экспертной оценки для пересмотра спорных случаев, в частности при соотношении ФЕ с универсальной семантической метафорой или при выявлении источника происхождения ФЕ.

Заключение

Создание и разметка двуязычного корпуса фразеологизмов на тему смерти – комплексная задача, требующая тщательного планирования, использования разнообразных лингвистических источников, разработки детальной системы классификации и аннотирования, а также применения адекватных технологических решений. Найденные фразеологизмы демонстрируют богатство финской и русской фразеологии в рассматриваемой сфере, отражая культурные особенности восприятия смерти через эвфемизмы, дисфемизмы, метафоры пути, перехода, вечного покоя и ироничные образы.

Подобный корпус является эффективным инструментом для сопоставительного анализа финской и русской фразеологии. На выборке ФЕ со значением «умереть» предлагаются критерии разметки, позволяющие изучать культурные коннотации,

Отдать богу душу

Буквальный перевод: Antaa sielunsa jumalalle

Семантика: приход к богу

Стилистика: книжный, высокий

Комментарий: Религиозный эвфемизм. Христианское заимствование.

Происхождение: **Заемствованное; религиозное**

Р и с. 1. Пример размеченных данных на русском языке
F i g. 1. Example of labeled data in Russian

Elämän liekki sammui

Буквальный перевод: Пламя жизни погасло

Семантика: окончание пути

Стилистика: книжный, нейтральный

Комментарий: Поэтичная, эвфемистичная метафора.

Происхождение: **Исконное; метафора**

Р и с. 2. Пример размеченных данных на финском языке
F i g. 2. Example of labeled data in Finnish

Источник: <https://github.com/denisshreshik/ru-fi-phrases-corpus/releases/tag/ru-fi-corp-death>.

Source: <https://github.com/denisshreshik/ru-fi-phrases-corpus/releases/tag/ru-fi-corp-death>.

стилистические особенности и проблемы переводческой эквивалентности. Кроме основных указанных критериев семантики, этимологии и стилистики, разметка может учитывать: наличие дериватов, синонимов (рус. – *отбросить коньки/копыта/лапти*, фин. – *heittää veivinsä/hokkarinsa/femmat*), особенности грамматического оформления (управление, фиксированный падеж / форма глагола) для генерации правильных контекстов в будущем [15]. Эта информация фиксируется в поле «Комментарий».

Размеченный корпус становится мощным инструментом для сравнительного анализа, выявления переводческих эквивалентов, изучения культурных коннотаций и стилистических регистров, а также основой для практических приложений в лексикографии и машинном переводе. Учет таких аспектов, как источник происхождения (калька, диалект, сленг) и детальная семантико-стилистическая разметка, значительно повышает научную и практическую ценность проекта создаваемого корпуса. Тем не менее процесс разметки неизбежно сопряжен с трудностями интерпретации и требует высокой квалификации аннотаторов и четких критериев для обеспечения консистентности данных.

Авторами создан и описан экспериментальный финско-русский корпус фразеологизмов тематической группы «умереть». Отличием предлагаемых нами критериев разметки корпуса от уже существующих является учет общих для обоих языков семантических метафор. Для тематической группы «умереть» это: окончание пути, приход к богу, вечный покой, физиологический конец, переход в локус, дисфемизмы.

Основным принципом аннотирования двуязычного корпуса является принцип универсальности для всех тематических групп ФЕ, а также принцип деления на информационные блоки. Исходя из этого, определены основные критерии разметки такого корпуса: семантико-тематический критерий, учитывающий универсальную семантическую метафору; семантический критерий, учитывающий коннотации; стилистический критерий; критерий языковой пары (наличие установленного переводческого эквивалента в другом языке либо возможность отождествления по смыслу, стилю и коннотации); этимологический критерий.

Подобный корпус является одним из шагов к созданию компьютерной модели языковой картины мира, которую можно использовать для работы с большими данными и обучения искусственного интеллекта. В перспективе планируется расширение корпуса для тематической группы ФЕ со значением «неудача» и разработка на его основе веб-интерфейса для исследователей.

СПИСОК ЛИТЕРАТУРЫ

1. Белкина Н.В., Лалым А.С., Осьмак Н.А. К вопросу о пространственных отношениях в финском языке. *Вестник Пермского университета. Российская и зарубежная филология*. 2024;16(2):16–27. <https://doi.org/10.17072/2073-6681-2024-2-16-27>
2. Бикмухаметова Р.М. Языковые образы концепта «смерть» в картине мира разноструктурных языков (на материале русского и башкирского языков). *Современные проблемы науки и образования*. 2014;(1). URL: <https://science-education.ru/ru/article/view?id=12022> (дата обращения: 17.10.2025).
3. Добровольский Д.О. Национально-культурная специфика во фразеологии. *Вопросы языкознания*. 1997;(6):37–58. <https://elibrary.ru/pyiohb>
4. Крюкова Н.Ф., Александрова Е.В., Исакова Е.М. Фразеологическая эквивалентность в когнитивно-герменевтическом аспекте. *Теория языка и межкультурная коммуникация*. 2023;(3):150–158. <https://elibrary.ru/fyiaws>
5. Почуева Н.Н. Понятие эквивалентности во фразеологии (на материале русского и английского языков). *Научные исследования*. 2019;(2):19–21. <https://elibrary.ru/yuraem>
6. Юсупова С.М. Сопоставительный анализ идиом о труде (на материале английского, немецкого, русского и чеченского языков). *Вестник Московского университета. Сер. 19: Лингвистика и межкультурная коммуникация*. 2011;(2):140–146. <https://elibrary.ru/nxpdjb>

7. McEnery T., Brezina V. *Fundamental Principles of Corpus Linguistics*. Cambridge: Cambridge University Press; 2022. 313 p.
8. Sinclair J. (ed.) *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins Publishing; 2004. 307 p.
9. McEnery T., Hardie A. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press; 2012. 294 p.
10. Добровольский Д.О., Зализняк А.А. Об особом типе модальности необходимости: семантика немецкого глагола *sollen* по данным параллельных корпусов. *Вопросы языкознания*. 2021;(6):22–39. <https://doi.org/10.31857/0373-658X.2021.6.22-39>
11. Каримова С.Г. Особенности перевода адъективных фразеологических единиц с использованием потенциала корпусной лингвистики (на материале английского и русского языков). *Казанская наука*. 2023;(9):214–216. <https://elibrary.ru/uuwikq>
12. Пантелеева Л.М., Хоробрых С.В. Пространственные смыслы концепта «смерть» в пермской фразеологии. *Филологические науки. Вопросы теории и практики*. 2018;(10–2):330–337. <https://doi.org/10.30853/filnauki.2018-10-2.24>
13. Орлова О.С. Темы рождения и смерти в русских и английских фразеологизмах: метафора пути. *Вопросы психолингвистики*. 2020;(1):64–74. <https://doi.org/10.30982/2077-5911-2020-41-1-64-74>
14. Добровольский Д.О. Двухязычный корпусный фразеологический словарь: организация текста словарной статьи. *Критика и семиотика*. 2024;(2):124–146. <https://doi.org/10.25205/2307-1753-2024-2-124-146>
15. Белкина Н.В., Лалым А.С., Осьмак Н.А. Валентностная структура глаголов движения в финском языке (на примере идиоматических выражений с глаголом *mennä*). *Ежегодник финно-угорских исследований*. 2025;19(1):6–17. <https://doi.org/10.35634/2224-9443-2025-19-1-6-17>

REFERENCES

1. Belkina N.V., Lalym A.S., Osmak N.A. On Spatial Relationships in the Finnish Language. *Perm University Herald. Russian and Foreign Philology*. 2024;16(2):16–27. (In Russ., abstract in Eng.) <https://doi.org/10.17072/2073-6681-2024-2-16-27>
2. Bikmukhametova R.M. The Language Images of the Concept Death in the World View of Languages of Different Structure (on the Material of the Russian and Bashkir Languages). *Modern Problems of Science and Education*. 2014;(1). (In Russ., abstract in Eng.) Available at: <https://science-education.ru/ru/article/view?id=12022> (accessed 17.10.2025).
3. Dobrovolsky D.O. [National and Cultural Specificity in Phraseology]. *Voprosy jazykoznanija*. 1997;(6):37–58. (In Russ.) <https://elibrary.ru/pyiohb>
4. Kryukova N.F., Aleksandrova E.V., Isakova E.M. [Idiomatic Equivalence in Terms of Cognitive and Hermeneutic Analysis]. *Theory of Language and Intercultural Communication*. 2023;(3):150–158. (In Russ.) <https://elibrary.ru/fyiaws>
5. Pochueva N.N. [The Concept of Equivalence in Phraseology (Based on the Russian and English Languages)]. *Nauchnye issledovanija*. 2019;(2):19–21. (In Russ.) <https://elibrary.ru/yuraem>
6. Yusupova S.M. [Comparative Analysis of Idioms about Work (Based on English, German, Russian, and Chechen Languages)]. *Moscow University Bulletin. Ser. 19: Linguistics and Intercultural Communication*. 2011;(2):140–146. (In Russ.) <https://elibrary.ru/nxpd bj>
7. McEnery T., Brezina V. *Fundamental Principles of Corpus Linguistics*. Cambridge: Cambridge University Press; 2022. 313 p.
8. Sinclair J. (ed.) *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins Publishing; 2004. 307 p.
9. McEnery T., Hardie A. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press; 2012. 294 p.
10. Dobrovolskiy D.O., Zalizniak A.A. On a Special Type of Necessity Modality: The German Verb *Sollen* Viewed from Parallel Corpora. *Voprosy Jazykoznanija*. 2021;(6):22–39. (In Russ., abstract in Eng.) <https://doi.org/10.31857/0373-658X.2021.6.22-39>
11. Karimova S.G. [Features of the Translation of Adjectival Phraseological Units Using the Potential of Corpus Linguistics (Based on the Material of English and Russian Languages)]. *Kazan Science*. 2023;(9):214–216. (In Russ.) <https://elibrary.ru/uuwikq>
12. Panteleeva L.M., Khorobrykh S.V. Spatial Meanings of the Concept “Death” in Perm Phraseology. *Philology. Theory & Practice*. 2018;(10–2):330–337. (In Russ., abstract in Eng.) <https://doi.org/10.30853/filnauki.2018-10-2.24>

13. Orlova O.S. The Topics of Birth and Death in Russian and English Phraseological UNITS. The Metaphor of Way. *Journal of Psycholinguistics*. 2020;(1):64–74. (In Russ., abstract in Eng.) <https://doi.org/10.30982/2077-5911-2020-41-1-64-74>
14. Dobvol'skij D.O. Bilingual Corpus-Based Dictionary of Idioms: Organization of the Dictionary Entry Text. *Critique and Semiotics*. 2024;(2):124–146. (In Russ., abstract in Eng.) <https://doi.org/10.25205/2307-1753-2024-2-124-146>
15. Belkina N.V., Lalyum A.S., Osmak N.A. Valence Structure of Verbs of Motion in Finnish (Using the Example of Idiomatic Expressions with the Verb Mennä). *Yearbook of Finno-Ugric Studies*. 2025;19(1):6–17. (In Russ., abstract in Eng.) <https://doi.org/10.35634/2224-9443-2025-19-1-6-17>

Информация об авторах:

Белкина Наталия Вениаминовна, кандидат педагогических наук, доцент кафедры языков Северной Европы Российского государственного педагогического университета им. А. И. Герцена (191186, Российская Федерация, г. Санкт-Петербург, набережная реки Мойки, д. 48), ORCID: <https://orcid.org/0009-0004-1080-1458>, SPIN-код: 9094-6553, bel_nat@mail.ru

Шарешик Денис Юрьевич, магистрант Института иностранных языков Российского государственного педагогического университета им. А. И. Герцена (191186, Российская Федерация, г. Санкт-Петербург, набережная реки Мойки, д. 48), ORCID: <https://orcid.org/0009-0003-1641-5681>, SPIN-код: 5207-3621, shareshik.denis@yandex.ru

Заявленный вклад авторов:

Н. В. Белкина – формулирование идеи исследования, целей и задач; создание и подготовка рукописи: написание черновика рукописи, включая его перевод на иностранный язык; контроль, лидерство и наставничество в процессе планирования и проведения исследования; разработка методологии исследования.

Д. Ю. Шарешик – осуществление научно-исследовательского процесса, включая сбор данных; проверка воспроизводимости результатов исследования в рамках задач работы.

Все авторы прочитали и одобрили окончательный вариант рукописи.

Поступила 13.11.2025; одобрена после рецензирования 02.12.2025; принята к публикации 13.01.2026.

Information about the authors:

Natalia V. Belkina, Cand.Sci. (Ped.), Associate Professor of the Department of Languages of Northern Europe, Herzen State Pedagogical University of Russia (48 Moyka Embankment, St. Petersburg 191186, Russian Federation), ORCID: <https://orcid.org/0009-0004-1080-1458>, SPIN-code: 9094-6553, bel_nat@mail.ru

Denis Yu. Shareshik, Master's Student of the Institute of Foreign Languages, Herzen State Pedagogical University of Russia (48 Moyka Embankment, St. Petersburg 191186, Russian Federation), ORCID: <https://orcid.org/0009-0003-1641-5681>, SPIN-code: 5207-3621, shareshik.denis@yandex.ru

Authors' contribution:

N. V. Belkina – ideas; formulation or evolution of overarching research goals and aims; preparation and creation of the published work, specifically writing the initial draft (including substantive translation); oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.

D. Yu. Shareshik – conducting a research and investigation process, specifically performing the data; verification of the reproducibility of the research results within the framework of the objectives of the research.

All authors have read and approved the final manuscript.

Submitted 13.11.2025; revised 02.12.2025; accepted 13.01.2026.