



ALEKSANDRA ZIĘBA

 ORCID: <https://orcid.org/0000-0001-897-5777>

Uniwersytet Śląski w Katowicach

## ПАРАЛЛЕЛЬНЫЙ КОРПУС INTERCORP В ПРОЦЕССЕ УСТАНОВЛЕНИЯ ПОЛЬСКО-РУССКИХ ПАР МЕТАОПЕРАТОРОВ

PARALLEL CORPUS INTERCORP IN THE PROCESS OF ESTABLISHING POLISH-RUSSIAN PAIRS OF METAOPERATORS

The article describes the possibilities offered by the parallel corpus InterCorp in the process of translating Polish metaoperators into Russian and searches in the texts the analogous Russian word constructions. The author raises the problem of the interpretation of such concepts as *reproduct*, *multiword unit* and *phrase* in the publications of Polish linguists who more often emphasize the need to define the above-mentioned units by mono- and multilingual dictionaries. The article refers to the metaoperators proposed by Wojciech Chlebda, which are used as an auxiliary tool in the search process from reproducible texts of multiword units. The possibilities and resources of Polish and Russian collections included in the InterCorp corpus were described, and then selected meta-operators of various categories were analyzed. The article describes the possibilities and resources of Polish and Russian collections included in the InterCorp corpus and then selected metaoperators of various categories were analyzed.

Каждый современный язык насыщен бесконечным разнообразием устойчивых и воспроизводимых выражений, которые каждый из нас автоматически употребляет как в письме, так и в повседневной речи, не задумываясь над тем, откуда мы их заимствовали, и как долго данные фразы мы знаем. Они, как будто, укоренились в нашем сознании, сильно активизируясь именно в тех ситуациях, когда их употребление позволяет выразить тот смысл, который автор хотел передать (например, выражение намерения, проявление эмоций и т.д.). Речь здесь не идёт о всем известных фразеологизмах, которым издавна лингвисты и лексикографы уделали немало внимания, результатом чего является обилие доступных одноязычных и переводных фразеологических словарей. Широкое применение хотя бы разговорных

выражений, типа *Ogarnij się, Daj se siana, Cholera wie* или *Krzyż na drogę* безусловно подтвердит каждый носитель польского языка, аргументируя, что *так просто говорится*. В польском языкознании, и прежде всего, лексикографии данного рода конструкции рассматриваются в рамках понятия *репродукт*. Указанный термин, созданный Войцехом Хлебдой, подразумевает под собой извлеченную из текстов единицу языка, которая характеризуется регулярной повторяемостью в этих текстах в функции своеобразного вербализатора<sup>1</sup>. Несмотря на то, что в качестве такой конструкции может выступать и однолексемная единица, польский лингвист сосредотачивается преимущественно на выражениях, состоящих из двух или более элементов. Отсутствие в российском языкознании понятия *репродукт*, дальнейшее использование которого в данной статье может вызывать неоднозначные ассоциации, заставляет нас вместо него употреблять термин *фразама*, который воспринимается синонимично в отношении упомянутого многолексемного образования. При таком подходе фраземы образуют своеобразный набор разнородных единиц, характеризующихся другими структурными и семантическими признаками в отличие от типичных фразеологизмов.

Приступая к работе над подобными конструкциями, необходимо сперва ответить на следующие вопросы: *Где мы можем найти фраземы и при помощи каких критериев следует их выделять? Как объяснить, почему одну многолексемную единицу мы должны воспринимать как фразему, а вторую — нет?*

Фраземы выступают преимущественно в текстах, причём следует учитывать тот факт, что компоненты, влияющие на их полную форму, нередко разделены другими лексемами, вследствие чего многолексемная единица является разброшенной. Извлечение фразем из текстовых ресурсов — это непростой процесс, поскольку, в отличие от однолексемных единиц, фраземы сами по себе графически не выделяются, и нередко требуют сложной формальной и семантической проверки. Не подлежит сомнению то, что именно эта особенность привела к тому, что в словарях отмечена в основном нерепрезентативная часть фразем. Даже

<sup>1</sup> W. Chlebda, *Reprodukty na warsztacie*// W. Chlebda (ред.), *Na tropach reпродукtów. W poszukiwaniu wielowrazowych jednostek języka*, Wydawnictwo Uniwersytetu Opolskiego, Opole 2010, с. 9.

самые большие фразеологические словари, как одноязычные, так и переводные, до сих пор фиксируют лишь определённый набор этих сверхлексемных единиц. Если говорить о переводной лексикографии то, как правило, лингвисты рассматривают два типа эквивалентов: словарные (системные) и функциональные (текстовые)<sup>2</sup>. Причиной тому является то, что практически не существует словарей, учитывающих второй вид эквивалентов. Главная трудность заключается в определении отношения эквивалентности между трансляндом (единица в исходном языке) и транслятном (переводной эквивалент). Принятые критерии извлечения выражений и принципы лексикографической обработки существенно повлияли на факт, что многие конструкции не могли рассматриваться как фраземы и потому не фиксировались в словарях. Только в последние десятилетия лексикографическая ситуация в сфере многолексемных образований начала несколько улучшаться. Применение других методологических решений и, прежде всего, указание других функций как основных, диаметрально изменило взгляд на понятие фраземы. Признание воспроизводимости и метафоричности в качестве ее фундаментальных признаков повлекло за собой резкий рост количества фразем, что, в свою очередь, подтвердило мнение Анджея Богуславского о том, что многолексемные единицы могут исчисляться миллионами<sup>3</sup>. Благодаря обращению к воспроизводимости, стало возможным выделение не имеющих признаков типовых фразеологизмов фразем (метафоричность, образность и экспрессивность).

Необходимость отмечать и исследовать фраземы испытывает каждый, кто более или менее профессионально занимается переводом. Несмотря на то, что всё больше лексикографов обращается к проблематике помещения фразем в словарях, по-прежнему сложно найти практические советы по переводу данного типа конструкции, не говоря уже о готовых достоверных переводных эквивалентах. К другим причинам, которые затрудняют возможность быстро и однозначно идентифицировать в тексте фразему, относят способность фразем к замене одних

<sup>2</sup> A. Charciarek, *Параллельный корпус InterCorp в переводной лексикографии* // „Opera Slavica” 2017, вып. 26, 2, с. 7.

<sup>3</sup> A. Bogusławski, *Uwagi o pracy nad frazeologią* // Z. Saloni (ред.), *Studia z polskiej leksykografii współczesnej III*, Wydawnictwo Uniwersytetu w Białymstoku, Białystok 1989, с. 19.

компонентов другими, т.е. новые элементы могут проникать в состав многолексемной единицы, изменяя тем самым их облик. Таким образом, структура фраземы в определённой степени остаётся открытой — часть выражения может меняться в зависимости от контекста.

Как уже было упомянуто, вербализаторы отдельных понятий имеют многолексемную форму, и, если наша цель заключается в установлении того, к каким заключениям, описывающим определённые срезы действительности, пришли носители языка, необходимо научиться определять границы фраземы. В этом случае обращение к автоматическим процедурам является недопустимым, однако существует немало других решений, которые направляют и поддерживают интуитивную деятельность лингвиста<sup>4</sup>. На начальном этапе работы нужно ознакомиться с методами и техниками извлечения фразем из текста, а также со способами верификации не только их формы, но и смыслового значения. Подтверждение воспроизводимости установленной многолексемной единицы возможно лишь только в том случае, когда лингвист располагает огромным количеством текстов, свидетельствующих о многократном употреблении фраземы, учитывая при этом сохранение ее структуры и семантического значения.

Не будет преувеличением сказать, что немалое количество текстов в некоторых местах содержит своего рода «подсказки», которые могут сигнализировать о присутствии фраземы. В процессе извлечения многолексемных образований чрезвычайно полезными оказываются неавтоматические способы, под которыми следует понимать то, что:

1) Первостепенную роль при выделении фразем играет интуиция лингвиста. Следует учитывать такие способности, как языковая компетентность, умение отделять одни лексемы от других и непосредственный контакт с текстом. В данном случае, это человек считается основным «устройством», анализирующим структуру текста. Компьютер может служить вспомогательным инструментом для верификации обнаруженных лексических единиц.

<sup>4</sup> P. Wierzchoń, *Pięć bardzo skutecznych (sprawdzonych) sposobów na masowe wyodrębnianie wielowyrazowych segmentów podejrzanych o frazematyczność (czyli reproduktów)*// W. Chlebda (ред.), *Na tropach reproduktów...*, с. 87–125.

2) Каждое извлеченное из текстовых ресурсов выражение не может восприниматься как неизменяемый по своей структуре объект. Лингвист может только подозревать, что замеченный комплекс лексем образует потенциальную фразу.

В монографии *Na tropach reproduktów. W poszukiwaniu wielowyrazowych jednostek języka* (2010), которая представляет собой первую попытку описания проблематики извлечения фразем из текстовых ресурсов, Войцех Хлебда представляет термин *potosniki formalne*, под которым понимает следующее:

Przez «pomocniki formalne» rozumiem zawarte w tekście wypowiedzi znaki bądź sformułowania, których autor wypowiedzi użył jako świadectw swej wiedzy lub przeświadczenia o tym, że pewne fragmenty jego narracji, położone najczęściej w bezpośredniej bliskości tych wskaźników, są ponadjednostkowym dobrem wspólnym [...]. Znakiem używanym w tej funkcji jest najczęściej cudzysłów, sformułowaniami zaś — metaoperatory: ogólne i względne<sup>5</sup>.

Наряду с определением «метаоператор» языковед употребляет сокращённую форму данного термина — «оператор». На первый взгляд может показаться, что кавычки представляют собой идеальный и надёжный метод, позволяющий безошибочно обособлять фраземы из текста — между графическими границами кавычек содержится искомая многолексемная единица. Однако, не следует забывать о том, что кавычки — это многофункциональный знак, и его использование и понимание каждый раз требует дополнительной интерпретации. Определяющая функция кавычек — извлекать чужие слова, то есть цитаты, которые обычно заимствованы из чьей-то индивидуальной речи. Поэтому нижеуказанный пример использования кавычек (и того, что находится в границах знака препинания) с точки зрения извлечения многолексемных воспроизводимых образований не будет рассматриваться (примеры из Национального корпуса польского языка):

Papież również zdaje sobie sprawę z nadchodzącej przemiany i uznaje jej konieczność. Andre Frossard, cytując bez cudzysłowu myśli papieża, tak to wyraża: «W chaosie świata widzi on (papież) znaki przyszłego ładu, który nam jeszcze umyka... Ów ład w zrozumieniu i przekonaniu Kościoła może być stworzony jedynie przy uznaniu obecności i potęgi Boga. Ale dążenie do uzyskania lepszego ładu, określane, niejednokrotnie kłamliwie, walką o postęp, towarzyszy człowiekowi od zarania dziejów, różniąc go i w tym od zwierząt» (*Watykan bez tajemnic*)

<sup>5</sup> W. Chlebda, *Reprodukty na warsztacie...*, s. 21.

В работе над фраземами нас должны интересовать такие выражения, которые отличаются обобщённым характером и присутствуют в сознании большинства носителей определённого языка. В свою очередь, в тексте:

Rodzice zaciągają pożyczki. Wynajmują mieszkania, a sami przenoszą się do dziadków. Rezygnują z wakacji, książek, kina. Szukają dodatkowej pracy. Proszą firmy o sponsorowanie studiów córki czy syna (na ogół bezskutecznie). Naruszają oszczędności «na czarną godzinę» («Gazeta Wyborcza»).

– кавычки можно успешно заменить такими операторами, как *jak to się mówi* или *(na) tak zwaną*. Такие конструкции указывают на то, что данная последовательность слов часто используется для обозначения явлений и, следовательно, считается многократной и воспроизводимой. Характеристике операторов и возможностям их обнаружения в тексте и переводе будет посвящена следующая часть статьи, в которой мы постараемся извлечь аналогичные конструкции русского языка.

Как мы уже подчёркивали раньше, зачислить многолексемную единицу в разряд фразем возможно только в том случае, если данная форма и значение выражения присутствуют во многих текстовых ресурсах. Раньше традиционным методом сбора лингвистического материала было ручное создание каталогов в бумажном виде, в которых отмечались одно- или многолексемные единицы в определённых контекстах. Появление национальных корпусов текстов существенно изменило ситуацию и открыло новые возможности в аспекте получения, сортировки и анализа языкового материала. Автоматизация работы лексикографа в сочетании с большими коллекциями текстов повысила не только скорость проведения исследований, но и достоверность полученных данных. Несмотря на то, что решение о концепции лексикографического описания по-прежнему принимает лингвист, статистические данные нередко являются основой для составления конкретных тезисов и выводов. Национальные корпусы постоянно расширяют свои рамки, а в их составе образуются новые подкорпусы, что значительно повышает их репрезентативность, благодаря фиксации текстов из самых разных областей (например, художественные, публицистические, а также записи устной речи). Под понятием *национальный корпус* следует понимать представленное в электронной форме собрание текстов, снабженное научным аппаратом,

т.е. аннотацией<sup>6</sup>. Материалы корпуса, а также дополнительные инструменты, которые можно использовать при анализе языка, в значительной степени относятся к лексическим единицам, что делает данные корпуса наиболее полезными для разработки словарей, тезаурусов, глоссариев или в процессе преподавания иностранного языка<sup>7</sup>.

Для достижения целей, поставленных в настоящей статье, где мы обратимся к методам получения русских операторов и проведения дальнейших исследований, наше внимание главным образом будет сосредоточено на явлении параллельных корпусов (*korpus równoległy*). Параллельный корпус — это собрание текстов и соответствующих им переводов на один (или несколько) иностранных языков. Таким образом, такой корпус состоит из двух, первый из которых включает в себя оригинальные тексты на исходном языке, а второй является его отражением на другом языке. Предполагается, что направление перевода текстов, содержащихся в корпусе, не обязательно должно быть одинаковым для каждого из них. Допускается, чтобы в набор текстов входили некоторые оригинальные тексты на родном языке, а остальные представляют собой переводы с другого языка. Параллельные корпуса в лингвистике, в первую очередь, подходят людям, занимающимся сопоставительными и переводческими исследованиями. Наиболее распространённым типом параллельного корпуса является двуязычный корпус, однако существуют параллельные корпуса, собранные для более чем одного языка. Примером такой текстовой базы является один из лучших славянских корпусов, в состав которого входят переводы для сорока языков. Именно этот корпус и будет служить основным инструментом при анализе польских и русских операторов в дальнейшей части работы.

Параллельный корпус InterCorp включает в себя значительную часть Национального корпуса чешского языка (*Český národní korpus*), т.е. старейшего корпуса славянского языка, который создан в 1994 году. Это, прежде всего, многоязычная база текстов, постоянно обогащаемая новыми текстами (так называе-

<sup>6</sup> В.А. Плунгян, *Зачем нужен Национальный Корпус русского языка? Неформальное введение* // *Национальный корпус русского языка: 2003–2005*, с. 6–20.

<sup>7</sup> В. Lewandowska-Tomaszczyk, *Podstawy językoznawstwa korpusowego*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2005, с. 16.

мый открытый корпус). Корпус InterCorp состоит из двух частей — ядра (польск. *trzon*) и коллекций (польск. *kolekcje*)<sup>8</sup>.

Ядро составляют, прежде всего, художественные тексты, а помимо этого, в корпус также входят автоматически обрабатываемые тексты: вышеупомянутые коллекции. Среди них, в новейшей тринадцатой версии корпуса (2020 год), выделяются публицистические тексты и новости с сайтов Project Syndicate и VoxEurop, юридические тексты корпуса Acquis Communautaire, отчёты с заседаний Европейского парламента Europarl, субтитры к фильмам из платформы OpenSubtitles и переводы Библии<sup>9</sup>. Необходимо помнить, что указанные источники не фиксируют текстов на всех языках, поэтому для каждой отдельной переводной лексикографии (например, польско-русской или русско-польской) InterCorp создаёт лучшие или худшие условия для работы переводчика.

В данном месте следует подчеркнуть, что первоначальное назначение корпуса — это исследование и работа с чешским языком. Также пользователю предоставляются большие текстовые ресурсы английского языка, что повышает вероятность получения достоверных эквивалентов. Ценность данного корпуса с точки зрения надёжности проведения исследований для других переводных лексикографий определяется размерами корпусов (коллекций) отдельных языков. Дифференцированные условия для поиска эквивалентов хорошо видны в случае польского и русского корпусов. Об обстановке данных языков достаточно подробно можем прочитать в статье Анджея Харцярека, посвященной приложению Treq, которое выводит текстовую базу из корпуса InterCorp:

Следует отметить, что польские, чешские и русские тексты, так называемые коллекции, содержащиеся в InterCorp v9, полностью не совпадают. Русский InterCorp не охватывает текстов из корпусов Europarl, Acquis Communautaire и Presseurop; он содержит художественные тексты (так называемое ядро), публицистические тексты на сайте Project Syndicate, а также субтитры на

<sup>8</sup> E. Kaczmarska, *Metody ustalania ekwiwalentów czasownikowych wyrażających stany emocjonalne w przekładzie czesko-polskim na materiale z korpusu równoległego InterCorp*, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa 2019, с. 45.

<sup>9</sup> E. Kaczmarska, A. Rosen, *Praktyczny przewodnik po korpusie InterCorp // M. Hebał-Jezińska (ред.), Praktyczny przewodnik po korpusach języków słowiańskich*, Wydział Polonistyki Uniwersytetu Warszawskiego, Warszawa 2014, с. 207.



сайте OpenSubtitles. Польский InterCorp v9 — один из самых больших по величине иноязычных корпусов и содержит все коллекции InterCorp за исключением публицистических текстов на сайте Project Syndicate<sup>10</sup>.

Как можно заметить, автор описывает содержание польских и русских коллекций в девятой версии корпуса InterCorp, но нынешняя тринадцатая версия никаких переломных изменений в этой области не принесла. Различия будут ещё более заметными, если вспомнить тот факт, что каждая новая версия InterCorp с одной стороны обогащает польский и русский корпусы большим количеством новых текстов, но с другой, увеличивает непропорциональность между этими языками. К примеру, двенадцатая версия польского корпуса насчитывала свыше 87 млн слов, а русского — примерно 19 млн лексических единиц и выражений. В свою очередь, в актуальной тринадцатой версии содержится 88 млн слов для польского языка и 20 млн для русского. Совершенно разные объёмы польского и русского корпусов и существенные различия в доступности предполагаемых параллельным корпусом коллекций не должны останавливать исследователей перед попытками проведения исследований в рамках данных языков при помощи InterCorp. Особого внимания заслуживает тот факт, что как польский, так и русский корпусы содержат коллекцию OpenSubtitles, которую, с одной стороны, не следует воспринимать как достоверную запись устной (разговорной) речи, но с другой — текстовой материал в определённой степени отражает спонтанные высказывания в неофициальных ситуациях. В данном месте неслучайно затрагивается тема разговорного языка, который долгие десятилетия оставался вне поля лингвистического внимания и уступал место анализу литературного языка.

Основная проблема языковых корпусов заключалась в отсутствии в них устных текстов, потому что экскерпция и транскрипция разговорной речи — это трудоёмкий и дорогостоящий процесс. Следовательно, ресурсы данной разновидности национального языка были и, скорее всего, будут дефицитными по сравнению с ресурсами письменных текстов, разработка которых не является таким сложным процессом. Однако возмож-

<sup>10</sup> A. Charciarek, *Программа для установления переводных эквивалентов Treq: преимущества и недостатки*, «Studia Rossica Posnaniensia» vol. XLV/1, Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza w Poznaniu, Poznań 2020, с. 121.

ности, предлагаемые корпусами лексикографических текстов, непрерывно расширяются. Лучшим примером является именно InterCorp, каждый год обновляемый и пополняемый новыми текстовыми (также устными) коллекциями. Следует также отметить, что разговорный язык обычно отождествлялся с «небрежным» разнообразием национального языка, что привело к его отсутствию во многих лексикографических работах.

Возвращаясь к теме операторов (например, *jak przyjęło się nazywać, jak się mówi, jak drzewiej mawiano* и т.д.), служащих своеобразными навигаторами в поиске многолексемных воспроизводимых единиц, с большой вероятностью можно предположить, что в некоторых контекстах они могут сигнализировать присутствие разговорной лексической единицы. В тексте рядом с конструкциями данного типа можно ожидать, что перед лингвистом откроется возможность извлечь довольно распространённую в устной речи фразему. Однако, цель нашего предстоящего анализа не будет заключаться в обнаружении в текстах сверхлексемных единиц при помощи операторов в корпусе InterCorp. Используя текстовые ресурсы указанного корпуса, мы постараемся не столько перевести предложенные Войцехом Хлебдой польские операторы на русский язык, сколько найти в процессе перевода аналогичные конструкции, которые могут оказаться пригодными при поиске фразем в русскоязычных текстах.

Войцех Хлебда в работе, посвященной неавтоматическим методам обнаруживания многолексемных образований в тексте, разделяет операторы (другое название — *formuły wskaźnikowe*) на два типа: безотносительные (*ogólne*) и относительные (*względne*)<sup>11</sup>. Первая разновидность операторов должна передавать информацию о том, что строки слов, к которым она относится, являются частыми и повторяющимися «вообще» выражениями при выполнении своей основной функции, т.е. называть объекты и понятия действительности. Они воспроизводятся всеми носителями данного языка. Функцию таких операторов можно назвать «свидетельствующей» или, по определению Анджея Левицкого, «тестимониальной»<sup>12</sup>. К таким конструкциям следу-

<sup>11</sup> В. Хлебда, *Метаоператоры в функции поисковой системы в обработке ресурсов Рунета для лексикографических целей*, «Rossica Olomucensia» Vol. XLIX, Olomouc 2010, с. 59.

<sup>12</sup> А.М. Lewicki, *Studia z teorii frazeologii*, Oficyna Wydawnictwo Leksem, Łask

ет отнести операторы, такие как: *jak to się mówi, jak to się zwykle mówić, jak to się przyjęło nazywać, tak zwany, przysłowiowy* и т.д. Их перевод на русский язык не должен вызывать больших трудностей, поэтому нашей задачей будет обнаружение возможных вариантов перевода первого из приведённых операторов при помощи параллельного корпуса InterCorp. Тем самым, мы попытаемся установить аналогичные операторы в другом языке, которые при сохранении критерия ситуативности, находят своё применение в текстах и могут помочь лингвистам извлекать многолексемные русские образования. Для этой задачи мы воспользуемся поисковой системой KonText (<https://kontext.korpus.cz>), которая является одним из трёх интерфейсов корпуса InterCorp.

### ЖАК ТО СИЄ МÓВИ

Распространённое употребление данного выражения подтверждают корпусные данные Национального корпуса польского языка (<https://nkjp.uni.lodz.pl>). PELCRA, поисковая система НКПЯ, фиксирует 595 вхождений и примеров использования оператора в 74 разных текстах. Данная фраза присутствует в речи или письме носителей польского языка, когда они убеждены в том, что фраза (стоящая перед или после оператора) считается общеизвестной и, тем самым, они оправдывают применение конкретных слов в определённой речевой ситуации. Обратимся сейчас к поисковой системе KonText с целью установления потенциальных русских транслятов оператора *jak to się mówi*, не ограничиваясь выбором конкретной коллекции (поиск во всех доступных текстовых ресурсах корпуса). Результаты запроса представляются следующим образом:

#### **jak to się mówi**

- jak to się mówi — как говорится (18 вхождений)
- jak to się mówi — что называется (4 вхождения)
- jak to się mówi — как это говорится (3 вхождения)
- jak to się mówi — недаром говорится (1 вхождение)
- jak to się mówi — так называемый (1 вхождение)
- jak to się mówi — как принято говорить (1 вхождение)
- jak to się mówi — как говорят (1 вхождение)
- jak to się mówi — знаешь пословицу...(1 вхождение)

Поисковый инструмент обнаружил 57 русских переводов, из которых только семь репрезентирует коллекцию OpenSubtitles (остальные принадлежат к художественным текстам — ядро корпуса). Мы решили представить и извлечь из документов только те русские переводы, которые, по-нашему мнению, в большей или меньшей степени могут послужить потенциальным оператором при поиске сверхлексемных образований (во многих примерах конструкция *jak to się mówi* не переводилась). Небольшое количество полученных программой примеров, с одной стороны не может служить основой для проведения достоверных исследований и представления надёжных выводов, но с другой — частотность употребления каждой из обнаруженных лексических единиц можно дополнительно проверить в других источниках. Первый из приведённых транслятов, т.е. *как говорится* кажется наиболее адекватным и дословным переводом анализируемого польского оператора. При помощи Национального корпуса русского языка проверим присутствие каждого извлеченного переводного эквивалента в доступных текстовых ресурсах:

Русский оператор	НКРЯ (количество вхождений)
<i>как говорится</i>	6 680
<i>что называется</i>	4 646
<i>как это говорится</i>	108
<i>недаром говорится</i>	71
<i>так называемый</i>	3 588
<i>как принято говорить</i>	201
<i>как говорят</i>	5 118
<i>знаешь пословицу...</i>	36

Оператор *как говорится* по-прежнему остаётся конструкцией с наивысшей частотностью употребления в текстах, а представленные статистические данные подтверждают использование данных выражений в речи русскоговорящих. Компоненты *говорится* и *называется* (или их варианты) представляют собой основные составляющие операторов, причём последний пример скорее предшествует применению пословицы, которая в рамках воспроизводимых многолексемных образований рассматриваться не будет. Таким образом, перевод одного польского оператора открыл доступ к небольшому репертуару потенциальных русских операторов, употребление и обнаружение которых

в тексте каждый раз требует дополнительной интерпретации и верификации значения во избежание неправильных выводов.

В свою очередь, относительные операторы (*operatory względne*) обозначают то, что воспроизводимость последовательности слов оценивается не «вообще», а по отношению к тем или иным «точкам отсчёта», среди которых Войцех Хлебда выделяет три главных оператора: места, времени и сообщества.

Операторы места указывают, какие «обычаи именования» преобладали, или преобладают в конкретной стране, местности, территории и регионе. К таким конструкциям следует отнести выражения, типа *jak to mówią na Śląsku, jak się mawiało na Kresach, jak to się zwykło nazywać w stolicy, jak to się mawia w Rosji, jak mawiano w moich okolicach, jak tu się mówi, jak się u nas mówi, było takie lwowskie powiedzenie* и их различные варианты. Для подавляющего большинства операторов данной категории можно сконструировать следующую схему их построения: *jak* + глагол *mówić/mawiać/nazywać* (в прошедшем или настоящем времени) + предлог *w/na*. Употребление выбранной конструкции может указывать на место жительства говорящего. К примеру, если человек применяет в своей речи оператор *jak to się u nas mówi*, то сигнализирует собеседнику о своём происхождении и демонстрирует использование типичных для жителей данного региона выражений. В свою очередь, употребление оператора *jak to mówią na Śląsku* с одной стороны может подчёркивать дистанцию говорящего, т.е. «они так говорят, не мы», а с другой — помогать жителю Силезии представлять существующие в его регионе устойчивые выражения. Это указывает также на полученные ранее знания о языке, которые человек использует во время ситуации общения. Предварительные попытки установления потенциальных русских транслятов для вышеуказанных операторов в корпусе InterCorp оказались неуспешными, поэтому нами было принято решение найти перевод только для следующих компонентов (конструкций):

- *Jak to mówią w...*,
- *Jak mówią w...*,
- *Jak mówią na...*

Количество зафиксированных корпусом вхождений составляет 5 примеров, что не позволяет составить надёжного списка

русских операторов места. Среди доступных переводов нам удалось извлечь следующие конструкции:

*Jak to mówią w...* — как было принято говорить в...

*Jak mówią w...* — как говорится в..., как говорят в..., как говорят у нас

*Jak mówią na...* — как здесь говорят

Как можно заметить, каждый из приведённых транслятов содержит компоненты «как» и определённую форму глагола «говорить» (некоторые вхождения включают в себя и дополнительные лексемы). Можно предположить, что более обширный репертуар польских операторов позволил бы обнаружить другие русские конструкции, использующих в переводе не только готовые польские трансланды, но и вариации составляющих их элементов. Сопоставим частотность употребления полученных переводных эквивалентов в НКРЯ:

Русский оператор	НКРЯ (количество вхождений)
<i>как было принято говорить в...</i>	3*
<i>как говорится в...</i>	624*
<i>как говорят в...</i>	517*
<i>как говорят у нас</i>	63
<i>как здесь говорят</i>	80

Последние две конструкции действительно указывали в текстах на выражения, употребляемые на определённой территории, хотя количество вхождений не является столь же репрезентативным, как ожидалось. В свою очередь, интересными оказались результаты поиска остальных многолексемных образований, для анализа которых были проверены каждый текстовый пример и контекст, в которых данные конструкции использовались. Выражение *как было принято говорить в...* относилось в корпусных текстах не к конкретному району, а определённому отрезку времени, что позволяет зачислить данную конструкцию в разряд темпоральных операторов, чему подтверждением служит следующий фрагмент: «Есть только одно обстоятельство, которое характеризует, **как было принято говорить в советские годы**, моральный облик противостоящих сторон». Несмотря на то, что формулировка *как говорится в...* отличается наивысшей частотностью употребления в текстах, мы зафиксировали только один пример её применения в контексте территории: «Выгнать из училища,

сделать сюржун, **как говорится в Турции**, то есть изгнать из страны; заморить даже в тюрьме, все это можно было сделать с учителем, своим подданным».

Остальные примеры, в большинстве своём, указывали на место записи определённых сентенций, представляли собой безотносительные операторы или ссылались на анекдоты.

И тогда, **как говорится в Евангелии**, восплачутся все племена земные (Мф. 24: 30), каждый человек станет перед лицом этого последнего Суда.

**Как говорится в народе**, на то и щука в реке, чтобы карась не дремал.

Но, **как говорится в анекдоте**, неприятный осадок остался.

На основании полученных данных можно предположить, что данная конструкция крайне редко употребляется для обозначения места, а данные операторы по своему значению скорее соответствуют таким польским конструкциям, как *jak zostało powiedziane, jak mówią ludzie, jak to było w żarcie*. Таким образом, как следует из вышесказанного, мы не получили подходящих русских аналогов операторов места, но более тщательный анализ каждого вхождения позволил извлечь конструкции, которые можно зачислить к операторам другого типа. Похожая ситуация имела место при поиске выражения *как говорят в...*, где в 134 случаях применения данных слов рядом с конструкцией выступало название государства или слово *деревня*:

Но об этом пока рано, это, **как говорят в Израиле**, другой сипур (другой рассказ). Одевается Бурый, **как говорят в деревне**, под партизан.

В доступных текстовых ресурсах НКРЯ фиксируется употребление указанной фразы преимущественно для описания типичных общеизвестных выражений или употребления поговорок, которые используются определённой группой людей.

**Как говорят в таких случаях** — лиха беда начало

Цель, **как говорят в президентской администрации**, проста — посмотреть, где чиновников избыток, а где дефицит, какие структуры сокращать, какие — расширять

Таким образом, можно сделать вывод, что найденные нами операторы места намного чаще находят своё применение в раз-

ных контекстах, вследствие чего их следует считать многофункциональными (данные конструкции можно отнести к разным операторам). Материалом для проверки частотности употребления послужил исключительно Национальный корпус русского языка, и, возможно, что более расширенный и многофакторный поиск русских конструкций в других электронных источниках (например, Yandex.ru, Google.ru и т.д.) подтвердил бы их применение в контексте местности.

В свою очередь, упомянутые операторы времени (темпоральные) показывают, как было принято называть какой-то отрезок действительности в определённый период: в какую-то эпоху, в какие-то годы или во время конкретного события. Мы имеем ввиду такие конструкции, как *jak drzewiej mawiano, jak to się za komuny nazywało, jak mawiano za okupacji, mówiąc językiem dzisiejszym, jak to się kiedyś mówiło, co określano mianem, jak byśmy dziś powiedzieli* и тому подобное. Параллельный корпус InterCorp при переводе вышеуказанных операторов и их вариантов зафиксировал только один переводной эквивалент, который обнаружился при попытке установления транслята для конструкции *jak mawiano dawniej*:

Twoje teorie to, **jak mawiano dawniej**, spekulacje myślowe.

Это твои соображения, **как прежде говорили**, умственные выкладки.

Очередной раз имеем дело с ситуацией, когда корпус не смог привести как переводы относительных операторов, так и возможные аналогичные конструкции для русского языка, что, вероятнее всего, является следствием недостаточного количества ресурсов русских текстов и диспропорции в доступных для обоих языков коллекциях. В свою очередь, в НКРЯ содержится только семь документов, в которых выступает данная конструкция, что может говорить о низкой употребительности данного выражения в речи русскоговорящих. Необходимо обратиться к другим источникам с целью верификации.

Последними являются операторы сообщества, связывающие соответствующие выражения с той или иной человеческой общиной: социальной средой, профессиональными группами, меньшинствами различных типов, членами обществ и ассоциаций, субкультурами, небольшим кругом лиц, которые создают свой собственный язык (коды) и коммуникативные привычки.



Операторы сообщества, как правило, довольно точно передают информацию о какой именно группе идёт речь. К ним примыкают конструкции, типа *jak mówią sprawozdawcy sportowi, mówiąc językiem polityków, jak zwykli opisywać dziennikarze, mówiąc naszym fachowym żargonem* и т.д. С их помощью не составит труда извлечь из текста лексемы и выражения, характерные для профессиолектов, социолектов, жаргона или идиолектов. Мы решили перевести выбранные компоненты представленных операторов и проанализировать найденные в корпусе тексты. Для запроса *jak mówią...* мы получили и извлекли из источников следующие трансляты:

jak mówią narkomani — как именуют её анашисты  
 jak mówią więźniowie — как говорят арестанты  
 jak mówią chirurgowie — как говорят хирурги  
 jak mówią socjologowie — как это называют социологи  
 jak mówią nianie — как говорят няни  
 jak mówią dzieci — как говорят дети  
 jak mówią ludzie z Bree — как говорят жители Пригорья  
 jak mówią marynarze — как говорят моряки  
 jak mówią Amerykanie — как выражаются американцы

Как можно заметить, способы перевода польского выражения не являются слишком разнообразными — в списке преобладает конструкция *как говорят*. Иногда присутствуют и другие аналогичные по функции образования (*как именуют..., как это называют..., как говорят жители..., как выражаются...*). В случае двух последних конструкций следует задуматься над тем, в рамках какого из операторов мы должны рассматривать следующие фразы, т.е. оператор места или сообщества. Для следующего выражения, т.е. *mówiąc językiem...*, мы нашли только одно вхождение, которое по сути своей указывает на фразу, употребляемую в языке коммерсантов:

Ostatnie lwy wyginęły by z głodu na długo przedtem, zanim zdołały by upolować ostatnią zdolną do rozrodu parę antylop czy zebr, czyli **mówiąc językiem handlowym**, poławiacze wielorybów zbankrutowali by na długo przedtem, zanim udało by im się wyniszczyć ostatnie ich egzemplarze.

Последние львы издохли бы от голода гораздо раньше, чем убили бы последнюю пару антилоп или зебр, способную к продолжению рода, так же как — в переводе на наш коммерческий язык — китобойный флот обанкротился бы задолго до истребления последних китов.

Таким образом, мы получили небольшой список русских операторов сообщества, дальнейшее исследование которых (при наличии других польских конструкции) позволило бы составить более репрезентативный и достоверный список выражений, сигнализирующих о присутствии многолексемной единицы. Стоит ещё вспомнить о одной из разновидностей относительных операторов, так называемых видовых операторах (*operatory gatunkowe*). Это фразы, которые придают соответствующим формулировкам статус пословиц, поговорок, максим и т.д., включая их в разряд языковых единиц, которые нами воспринимаются как общеизвестные, часто повторяемые, и тем самым, воспроизводимые, например, *jak mówi mądrość ludowa, jak głosi znane porzekadło, zgodnie ze starą maksymą*. В ходе нашего анализа нам также удалось обнаружить похожие конструкции типа *znasz posловицу...*, что позволяет предположить, что операторы одной группы могут в определённых контекстах проникать в другие типы операторов и выполнять их функцию.

Подводя итоги, следует отметить, что Параллельный корпус InterCorp лучше подходит для поиска польско-русских пар безотносительных операторов. Небольшое количество содержащихся в русском корпусе текстов и нерепрезентативное количество полученных вхождений вынуждает дополнительно проверять извлеченные конструкции в других источниках с целью верификации не только структуры оператора, но и частотности его употребления. Как показывает исследование, более результативным процессом является перевод отдельной совокупности компонентов оператора (при ручной обработке каждого примера), а не целых конструкций. Актуальность дальнейшего исследования метаоператоров для обоих языков обосновывается необходимостью извлечения и анализа многолексемных образований, которые должны фиксировать актуальные одноязычные и переводные словари. Их основной недостаток заключается в том, что лексема в них трактуется как основная единица языка, без учёта сверхлексемных единиц и ориентации на текст. Поэтому перед современным лексикографом стоит задача выполнить две важные операции: найти фразему в как можно большем количестве текстов и контекстов, а затем деконтекстуализировать многолексемную единицу и поместить её в словарь. Однако, как уже многократно подчёркивалось, проблема часто возникает в самом начале работы — необходимо правильно извлекать многолек-

семные единицы из языка оригинала, метаоператоры же могут служить надёжным инструментом в этой процедуре. Неадекватные пары переводов нередко являются результатом неточного или неполного определения структуры и метафорического значения исходной единицы. Ошибка в определении значения трансланда автоматически приводит к указанию неадекватного эквивалента. Независимо от лингвистической компетентности лексикографа, каждая извлеченная фраза должна быть тщательно проверена в как можно большем количестве источников. Таким образом, при работе с многолексемными образованиями и метаоператорами не нужно ограничиваться только одним корпусом — лингвист должен использовать всю доступную базу текстовых эквивалентов и методов извлечения искомым единиц. Автоматические способы отбора материала, несомненно, облегчают ручную выборку и повышают скорость проведения исследований, но они всегда будут предоставлять материал, языковой статус которого определит решающая инстанция — человек.

## REFERENCES

- Bogusławski, Andrzej. “Uwagi o pracy nad frazeologią.” *Studia z polskiej leksykografii współczesnej III*. Saloni, Zygmunt. (Ed.). Białystok: Wydawnictwo Uniwersytetu w Białymstoku, 1989.
- Charciarek, Andrzej. “Parallel’nyj korpus Intercorp v perevodnoy leksikografii.” *Opera Slavica XXVII*, 2017, vol. 27, 2: 5–17 [Харцярек, Анджей. “Параллельный корпус InterCorp в переводной лексикографии.” *Opera Slavica XXVII*, 2017, vo. 27, 2: 5–17].
- Charciarek, Andrzej. “Programma dlya ustanovleniya perevodnykh Ekvivalentov treq: preimushchestva i nedostatki.” *Studia Rossica Posnaniensia*, 2020, vol. XLV/1: 121–132 [Харцярек, Анджей. “Программа для установления переводных эквивалентов Treq: преимущества и недостатки.” *Studia Rossica Posnaniensia*, 2020, vol. XLV/1: 121–132].
- Lewicki, Andrzej Maria. *Studia z teorii frazeologii*. Łask: Oficyna Wydawnictwo Leksem, 2013.
- Lewandowska-Tomaszczyk, Barbara. *Podstawy językoznawstwa korpusowego*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 2005.
- Kaczmarek, Elżbieta. *Metody ustalania ekwiwalentów czasownikowych wyrażających stany emocjonalne w przekładzie czesko-polskim na materiale z korpusu równoległego InterCorp*. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego, 2019.
- Kaczmarek, Elżbieta, and Rosen, Alexandr. “Praktyczny przewodnik po korpusie InterCorp.” *Praktyczny przewodnik po korpusach języków słowiańskich*. He-

- bal-Jeziarska, Milena. (Ed.). Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego, 2014: 207–231.
- Chlebda, Wojciech. “Reprodukty na warsztacie.” *Na tropach reproduktów. W poszukiwaniu wielowyrazowych jednostek języka*. Chlebda, Wojciech. Ed. Opole: Wydawnictwo Uniwersytetu Opolskiego, 2010: 9–27.
- Plungyan, Vladimir Aleksandrovich. “Zachem nuzhen Natsional’nyj Korpus russkogo yazyka? Neformal’noye vvedeniye.” *Natsional’nyy korpus russkogo yazyka: 2003–2005*: 6–20 [Плунгян, Владимир Александрович. “Зачем нужен Национальный Корпус русского языка? Неформальное введение.” *Национальный корпус русского языка. 2003–2005*: 6–20.]
- Wierzchoń, Piotr. “Pięć bardzo skutecznych (sprawdzonych) sposobów na masowe wyodrębnianie wielowyrazowych segmentów podejrzanych o frazematyczność (czyli reproduktów).” *Na tropach reproduktów. W poszukiwaniu wielowyrazowych jednostek języka*. Chlebda, Wojciech. Ed. Opole: Wydawnictwo Uniwersytetu Opolskiego, 2010: 87–125.
- Khlebda, Voytsekh. “Metaoperatory v funktsii poiskovoy sistema vobrabotke resursov Runeta dlya leksikograficheskikh tseley.” *Rossica Olomucensia*, 2010, vol. XLIX: 57–66 [Хлебда, Войцех. “Метаоператоры в функции поисковой системы в обработке ресурсов Рунета для лексикографических целей.” *Rossica Olomucensia*, 2010, vol. XLIX].