

Юлия Вадимовна Николаева

Италия, Римский университет «Ла Сапьенца»

Корпусный метод и современная паремиография¹

Ключевые слова: паремиография, Национальный корпус русского языка, *Refranero multilingüe*, формирование словника, поиск новых вариантов и неологизмов, активизация/пассивизация пословиц, диахронические изменения

Key words: paremiography, Russian National Corpus, *Refranero multilingüe*, determining the proverb list, researching new paremiological variants and neologisms, updated and outdated proverbs, diachronic changes

Abstract

This paper describes the advantages of using the Russian National Corpus in paremiography and synthesizes the author's experience in the elaboration of the Russian part of *Refranero multilingüe* designed to provide the translation of popular Spanish proverbs into all the languages of the Iberian peninsula and into the main European languages. The use of National Corpus helps to determine the proverb list and to describe diachronic changes and variational processes in paremiology while being useful for researching neologisms and detecting updated/outdated proverbs.

Издавна словари пословиц и поговорок считались частью культурного наследия каждого народа, их составлением занимались фольклористы и гуманисты. Достаточно вспомнить «князя гуманистов» Эразма Роттердамского, чьи *Adagia* 1500 г. дали толчок к описанию и собиранию паремий во многих европейских странах. В русской средневековой традиции паремии также были впервые зарегистрированы не в лингвистических словарях, а в сборниках назидательного содержания.

Лингвистический подход к паремиям стал преобладать в науке лишь в последние десятилетия, до этого времени паремиография воспевала пословицы как самобытное проявление народной мудрости и не уделяла достаточного внимания их языковым особенностям. Многие широко применяемые в лексикографии нормы, такие как обязательное толкование значения, выявление

¹ Настоящая работа выполнена при финансовой поддержке гранта «Paremaistic» (FFI 2011 – 24962 2012-2014), предоставленного Министерством науки и инноваций Испании.

и описание отношений полисемии и омонимии, присутствие продуманной системы помет, начали приживаться в паремиографии и фразеографии лишь в последние годы и пока еще не стали общепринятой практикой. В наши дни еще встречаются многочисленные пословичные словари, подобные телефонным справочникам, в которые в алфавитном порядке вносятся без разбора все паремии, встретившиеся автору в разных источниках. Словник подобных словарей пестр, как лоскутное одеяло, он включает в себя и давно вышедшие из употребления, и существующие в современном узусе пословицы, при этом значимые языковые особенности паремий, предопределяющие их функционирование в текстах, никак не комментируются.

В интервью, данном в 2011 г. международному журналу «Паремия», В. М. Мокиенко очертил основные трудности, с которыми сталкиваются сегодня составители фразеологических и паремиологических словарей: критерии отбора словника, учет и отражение активного/пассивного характера фразеологизмов и паремий, и, наконец, самое главное – выработка системы помет, дающих пользователю исчерпывающую и достоверную информацию о стилистических коннотациях, дискурсивных особенностях, семантических, грамматических ограничениях в реальном узусе, а также о происхождении, частотности, диатопных и диастратных вариантах этих языковых единиц. Разработка многих из этих помет затруднена пока отсутствием соответствующих теоретических исследований [Nikolaeva 2011: 15–16].

В русской национальной фразеографии уже появились словари нового типа, учитывающие потребности современных читателей и представляющие им более полную лингвистическую информацию, т. е. ту, которая помогает избежать коммуникативных неудач, восстановить вкус и запах контекста и служит некоторым «руководством к действию», к использованию паремий в коммуникативном акте, в речи (Бирих и др. 1998; Алефиренко, Золотых 2008; Баранов, Добровольский 2007; Баранов, Добровольский 2009 и другие). Во многом создание подобных словарей стало возможным благодаря достижениям корпусной лингвистики.

Задачи, стоящие перед лексикографом, усложняются, когда речь заходит о двуязычных и многоязычных словарях², тем более о словарях паремиологи-

² Подробнее методологические затруднения, неизбежные при разработке многоязычных словарей, рассматриваются в работах В. П. Беркова и В. В. Дубичинского, которые указывают на пассивный характер таких лексиконов, на невозможность в большинстве случаев использовать их данные при текстопорождении и на то, что традиция издавать подобные словари несколько устарела и относится преимущественно к XVIII и XIX вв. «С несколькими входными языками теперь издаются преимущественно терминологические словари. Из-за присущей основной массе не-терминов многозначности хорошо сделать общие словари такого типа невозможно» [Берков 2004: 7]. Мнение это разделяет и Дубичинский, который видит будущее многоязычной лексикографии «в развитии не всеохватывающих универсальных словарей, а в словарях с наибольшим количеством языков, с описанием лексики достаточно узких отраслей знаний» [Дубичинский 2006: 18].

ческих, ибо лексикографической обработке подвергаются параллельно в разных языках сложные языковые знаки, культурно отмеченные и семантически многомерные, а, значит, по своей семиотической природе предрасположенные к межъязыковым расхождениям по многим критериям³.

Мы столкнулись с этими трудностями при разработке русской части *Refranero multilingüe*, многоязычного словаря пословиц и поговорок, входным языком для которого является испанский. Словарь был задуман Х. Муньос Севильей и М. Т. Сурдо Руис Аюкар как многоязычная база данных, в которой испанские пословицы и поговорки даются с переводом на различные языки и представляются как обширная часть богатейшего провербиального национального пространства⁴ каждого из рассматриваемых языков во всем разнообразии вариантов, синонимов, антонимов и гиперонимов. В микроструктуру словаря входит также указание на типологию паремии и возможность поиска по ключевому слову (для испанской части), толкование значения, лексический и культурологический комментарий, индекс частотности, отсылка к основным источникам и контексты, иллюстрирующие использование этих паремий в художественной литературе и современной публицистике. Сегодня помимо всех языков Иберийского полуострова (испанского, каталанского, галисийского, баскского и португальского) в словаре представлены основные современные европейские языки (английский, немецкий, французский, итальянский, русский, польский, румынский, венгерский и новогреческий), а также древнегреческий, латинский и арабский [*Refranero multilingüe* 1997–2014]. Русскоязычная часть *Refranero multilingüe* создается совместными усилиями преподавателя университета Саламанки Хосефины Веласко Менендес и автора этой статьи.

При разработке русской части провербиального пространства и при поиске эквивалентов мы не только опирались на многочисленные лексикографические источники⁵, но и активно пользовались Национальным корпусом русского языка (далее – НКРЯ). Корпусный анализ может развиваться по двум направлениям: эмпирический поиск или извлечение из корпуса пословиц и фразеологизмов на основе существующей разметки. Первый подход предполагает, что ведется поиск паремий, предварительно известных по словарям или извлеченных из ментального лексикона носителя языка. Именно этим методом мы руководствовались в своей работе. Второй подход сейчас активно разрабатывается, он дал уже свои первые плоды при извлечении из корпусов

³ Традиция составления многоязычных словарей пословиц и новационный характер *Refranero multilingüe* описаны мною подробнее в [Nikolaeva (в печати)].

⁴ Понятие провербиального пространства вошло в паремиологию благодаря исследованиям Ю. И. Левина [Левин 1984].

⁵ Используемые нами источники приведены на сайте *Refranero multilingüe*, http://cvc.cervantes.es/lengua/refranero/bibliografia.htm#dahl_1978 (дата обращения: 26.11.2014).

коллокаций [Захаров, Хохлова 2010; Chiari 2007; Granger, Paquot 2012], однако применить его на материале пословиц пока невозможно.

Использование НКРЯ шло по разным направлениям: от формирования словника до выделения новых значений пословиц. Данные, извлеченные из корпуса, неоднократно предопределяли выбор паремии для заглавной словарной статьи⁶. Например, в традиционных лексикографических источниках наравне регистрируются варианты пословицы библейского происхождения: «Врачу, исцелися сам», «Врач, уврачуй наперед свой недуг» и «Врач, исцели себя самого». По корпусным данным, вариант, начинающийся формой звательного падежа, насчитывает 24 вхождения⁷ и активно употребляется в наши дни (последний раз он зафиксирован в 2011 г.), в то время как другие формы в НКРЯ не встречаются. Эти наблюдения оказались решающими при переводе устаревшей испанской пословицы «*Médico, cúrate a ti mismo*». Подбирая к ней адекватный функциональный эквивалент, соотнесенный по диахронии, мы остановились на вышедшем из употребления варианте «Врач, уврачуй наперед свой недуг», который несет на себе явную печать архаичности.

Благодаря НКРЯ нам удалось существенно расширить вариантную базу словаря. В русской лексикографической традиции на протяжении более двух веков пословица «Без труда не вытащишь и рыбку из пруда» фиксировалась также с глаголами «не вытянешь», «не вынешь», к которым корпус добавил «не выловишь», «не выудишь». К пословице «Плохому танцору и штаны (одежда, яйца) мешают» нашелся вариант – эвфемизм «Плохому танцору всегда что-то мешает». Примеры можно множить. Не ставя перед собой невыполнимую цель отразить все существующие варианты паремий, сама языковая природа которых предполагает высокую степень устойчивости и одновременно широкую вариативность, мы все же сочли необходимым включить в словарь максимальное количество вариантов⁸. *Refranero multilingüe* адресован в первую очередь испанцам, изучающим русский язык как неродной, и исполняет функции предупреждения наиболее распространенных ошибок. Известно, что в процессе изучения иностранного языка даже наличие одного незнакомого слова порой препятствует пониманию и опознанию паремий как таковых, следовательно, фиксация наибольшего числа вариантов отвечает не только критерию лексикографической полноты, но и снижает риск неправильной интерпретации текста.

Серьезному пересмотру подвергся паремиологический минимум, который был разработан в 1970-е гг. Г. Л. Пермяковым и лег в основу многих словарей и схожих социолингвистических опросов в разных странах мира [Пермяков

⁶ Методика применения корпусных методов при составлении словника пословичного словаря активного типа описана в [Беликов 2008].

⁷ Вся статистика НКРЯ приведена по состоянию на 1 ноября 2014 г.

⁸ О варьировании во фразеологии см. [Добровольский 2007].

1985; Котова 2000; Хлебда 2008: 90–92]. Именно этот минимум до сих пор становится основой словарей живых русских пословиц, несмотря на то, что с момента его создания прошло уже более сорока лет. Частотность употребления пословиц быстро меняется во времени и находится в прямой зависимости от социолингвистических факторов, так что целесообразно говорить о паремиологических минимумах *hic et nunc*, а не о «минимумах вообще»⁹. Кардинальные изменения, произошедшие в русском языке в постсоветское время, отодвинули на языковую периферию многие активные в эпоху СССР паремии, и это необходимо учитывать при составлении переводного пословичного словаря, в котором одним из критериев эквивалентности является частотность и соотношенность по диахронии.

Пословицы «На еду – как медведь, а на работу – как заяц», «За дурной головой и ногам непокой», «Одним хлебом жив не будешь», «Каков в еде, таков и в труде», «Поговорка – цветочек, а пословица – ягодка», зарегистрированные в паремиологическом минимуме Г. Л. Пермякова, в НКРЯ не фигурируют. Их отсутствие в корпусе наводит на мысль о том, что эти пословицы, равно как и «Языком болтай, а рукам воли не давай» (последнее вхождение в корпус в 1855 г.), и «Битая посуда два века живет» (датировано 1912 и 1918 гг.), вышли из активного употребления. К некоторым из них в корпусе нашлись варианты, принадлежащие к современному узусу, например, «Дурная голова ногам покоя не дает» (3 вхождения) и пословица библейского происхождения «Не хлебом единым жив человек» (176 вхождений). Именно этот последний вариант «живой» пословицы и вошел как заглавный в словарную статью *Refranero multilingüe*, в качестве эквивалента «живой» испанской паремии «No sólo de pan vive el hombre», в то время как «Одним хлебом жив не будешь» мы отнесли в раздел вариантов с пометой *en desuso* (устаревшее).

Вместе с тем следует отметить, что корпусные данные свидетельствуют о возвращении в современный русский узус паремий, находившихся в 1970-е гг. в пассиве: «Бог терпел и нам велел» (16 вхождений), «Глаза – зеркало души» (26 вхождений), «Глаза боятся, а руки делают» (36 вхождений), «Доверяй, но проверяй» (114 вхождений), «Время не ждет» (134 вхождения), «Время – деньги» (242 вхождения). При подборе эквивалента к активно используемой ныне испанской пословице «El tiempo es oro» (буквально: Время – золото) мы предпочли ориентироваться в первую очередь на узус и выбрали перевод «Время – деньги». Мы сочли, что вариант «Время дороже золота» [Аникин 1988: 51], более близкий к искомой испанской паремии по своему лексическому

⁹ Жизнь и реальное употребление в современной молодежной среде пословиц из минимума Г. Л. Пермякова были описаны Г. Благовой по результатам опроса 2005 г. Это социолингвистическое исследование подтвердило сохранение в живом узусе нескольких десятков единиц, отмеченных Пермяковым, и продемонстрировало, как на базе известных моделей в устной речи рождаются многочисленные новые варианты [Благова 2005].

составу, не может претендовать на роль самого удачного эквивалента в силу принадлежности к пассивному фонду.

С помощью корпуса нам удалось выявить различные диахронические изменения, оставшиеся до сих пор вне поля зрения других исследователей. Например, лексико-синтаксическое оформление пословицы «Береги платье снову, а честь смолоду» зарегистрировано с начала XVII в., именно в такой форме она кочует из словаря в словарь и отмечается в минимуме Г. Л. Пермякова. Корпус, однако, фиксирует ее употребление только в пушкинской повести (1836 г.) и в статье публициста Б. А. Садовского 1909 г., что противоречит традиционным лексикографическим источникам и дает основания полагать, что в XX в. пословица могла обрести другую форму. В более современных текстах и в устном подкорпусе встречается усеченный и видоизмененный вариант «Береги честь смолоду» (всего 24 вхождения), вошедший в канон русской литературы как эпитафия к *Капитанской дочке* А. С. Пушкина и задержавшийся в современном активном узусе.

НКРЯ позволил выделить неологизмы, многие из которых еще не были замечены русской паремиографией. В словник были включены пословицы, отличающиеся абсолютной новизной, что привело к коренному пересмотру традиционно сложившихся отношений эквивалентности и способствовало сокращению количества безэквивалентных паремий. В русскую часть *Refranero multilingüe* вошли неологизмы «Красота требует жертв», «Деньги решают все», «Лучше так, чем никак», известные крылатые слова Жванецкого «Что охраняешь – то и имеешь» и т. д. При фиксации неологизмов мы учитывали присутствующие в текстах метагlossы «давно известно», «говорят», «не зря говорят», «по принципу», «выходит по формуле [...]», «поговорка [...] играет роль», благодаря которым подчеркивается сентенциозный характер изречений и их принадлежность к коллективному языковому сознанию. Вероятно, на эти метагlossы можно будет опереться в дальнейшем при разработке разметки, благодаря которой станут возможными пилотируемый поиск паремий и их автоматизированное извлечение из текста.

В данной статье мы смогли лишь бегло коснуться некоторых аспектов использования национального русского корпуса при составлении паремиологических словарей. Даже пунктирный анализ свидетельствует о том, что ныне работа паремиографа немислима без привлечения современных языковых корпусов. С их помощью можно выявить и описать подлинный узус того или иного периода развития национального языка и собрать все-стороннюю информацию о семантических и прагматических особенностях описываемых языковых единиц. Безусловно, НКРЯ является эффективным средством формирования словаря с учетом основных диахронических процессов и использования паремий в текстах разной функциональной принадлежности. С опорой на огромные массивы текстов стало проще выявлять неологизмы, регистрировать новые значения паремий, наблюдать процессы

их варьирования и активизации/пассивизации в разные периоды развития языка. Все это дает возможность преодолеть ранее непреодолимый барьер между словарем как отражением языковой системы и как текстопорождающим инструментом, и предупредить многие трудности, возникающие при использовании паремий в речи.

Литература

- Алефиренко Н., Золотых Л., 2008, *Фразеологический словарь: Культурно-познавательное пространство русской идиоматики*, Москва: Элпис.
- Аникин В., 1988, *Русские пословицы и поговорки*, Москва: Художественная литература.
- Баранов А., Добровольский Д. (ред.), 2007, *Словарь-тезаурус современной русской идиоматики*, Москва: Мир энциклопедий Аванта+.
- Баранов А., Добровольский Д. (ред.), 2009, *Фразеологический объяснительный словарь русского языка*, Москва: Эксмо.
- Беликов В., 2008, Паремии как объект лексикографии, *Компьютерная лингвистика и интеллектуальные технологии*, № 7 (14), с. 45–49, www.dialog-21.ru/digests/dialog2008/materials/html/7.htm (дата обращения: 26.11.2014).
- Берков В., 2004, *Двуязычная лексикография*, Москва: Астрель Аст.
- Бирих А., Мокиенко В., Степанова Л., 1998, *Словарь русской фразеологии. Историко-этимологический справочник*, Санкт-Петербург: Фолио-Пресс.
- Добровольский Д., 2007, Лексико-синтаксическое варьирование во фразеологии: ввод определения в структуру идиомы, *Русский язык в научном освещении*, № 2 (14), с. 18–47.
- Дубичинский В., 2006, Основная проблематика переводной лексикографии [в:] R. Belentschikow (Hrsg.), *Das Russische in zweisprachigen Wörterbüchern*, Frankfurt am Main: Peter Lang, с. 13–22.
- Дубичинский В., 2008, *Лексикография русского языка*, Москва: Наука – Флинта.
- Захаров В., Хохлова М., 2010, Анализ эффективности статистических методов выявления коллокаций в текстах на русском языке, *Компьютерная лингвистика и интеллектуальные технологии*, № 9 (16), с. 137–144, <http://www.dialog-21.ru/digests/dialog2010/materials/html/22.htm> (дата обращения: 20.11.2014).
- Котова М., 2000, *Русско-славянский словарь пословиц с английскими соответствиями*, Санкт-Петербург: Издательство Санкт-Петербургского университета.
- Левин Ю., 1984, Провербильное пространство [в:] Г. Пермяков (сост.), *Паремиологические исследования: Сборник статей*, Москва: Наука, с. 108–126.
- Национальный корпус русского языка, <http://ruscorpora.ru> (дата обращения: 29.11.2014).
- Пермяков Г., 1985, *300 общеупотребительных русских пословиц и поговорок (для говорящих на немецком языке)*, Москва: Русский язык.
- Хлебда В., 2008, О чем думает русский индюк, или Об эквивалентах пословиц в двуязычном словаре, *Przegląd Rusycystyczny*, № 4 (124), с. 90–105.
- Blagova G., 2005, La vie du proverbe russe dans une mégapole contemporaine, *Revue des études slaves*, t. 76, № 2–3, с. 263–278.
- Chiari I., 2007, *Introduzione alla linguistica computazionale*, Bari: Laterza.

- Granger S., Paquot M. (eds.), 2012, *Electronic Lexicography*, Oxford: Oxford University Press.
- Nikolaeva J., 2011, Valeri Mokienko y la fraseología rusa, *Paremia*, № 20, с. 11–18.
- Nikolaeva J., (в печати), El Refranero multilingüe: una propuesta paremiografica innovativa. Traduzione delle paremie spagnole in russo [в:] F. Martínez de Carnero, L. A. Messina Fajardo (a cura di), *Annuario Phrasis*, Roma: Aracne.
- Refranero multilingüe*, 1997–2014, Madrid: Instituto Cervantes, <http://cvc.cervantes.es/lengua/refranero/Default.aspx> (дата обращения: 28.11.2014).