

RU

Применение параллельного корпуса для перевода этнокультурных коллокаций

Архипова Е. И.

Аннотация. Целью исследования является решение проблем перевода этнокультурных коллокаций с английского на русский язык посредством применения параллельных корпусов. Научная новизна исследования заключается в анализе и оценке возможности применения параллельного корпуса русского языка для перевода этнокультурных коллокаций, характеризующихся наличием национально-специфичного компонента значения и отсутствием регулярных словарных соответствий в рамках пары языков. В результате определено, что применение параллельного корпуса необходимо в исследовании лингвоспецифичной лексики, у которой нет эквивалентов в других языках, но его использование для перевода возможно только для ограниченного количества этнокультурных коллокаций в силу недостаточного охвата данной категории лексики.

EN

Using a Parallel Corpus to Translate Ethnocultural Collocations

Arkhipova E. I.

Abstract. The aim of the research is to solve issues of translating ethnocultural collocations from English into Russian by using parallel corpora. Scientific novelty of the research lies in analysing and assessing the potential use of the Russian language parallel corpus to translate ethnocultural collocations characterised by the presence of a culture-specific component of meaning and the absence of regular dictionary correspondences within a pair of languages. As a result, it has been determined that the use of a parallel corpus is necessary in the study of linguospecific vocabulary, which has no equivalents in other languages, but its application for translation is possible only for a limited number of ethnocultural collocations due to insufficient coverage of this vocabulary category.

Введение

Исключительные ресурсные возможности и специфика систематизирования языкового материала позволяют рассматривать корпус как источник ценного материала в переводческой и обучающей деятельности. Актуальность исследования обусловлена трудностями, с которыми сталкиваются студенты и переводчики, когда речь идет о коллокациях, обладающих не только связанностью компонентов на синтаксическом уровне, но и на уровне семантики. Корпусная лингвистика основывается на накопленном опыте различных лингвистических технологий, в частности социолингвистических – исследующих языковое многообразие. Кроме того, рассматривая корпус как микромодель языка, можно проследить происходящие в нем изменения. Постоянное пополнение базы может послужить инструментом для верификации перевода лексики, вызывающей затруднения. Для достижения цели исследования автору необходимо решить следующие задачи: выявить, насколько полно передано значение английской этнокультурной коллокации в результате поиска по языковым парам в параллельном корпусе русского языка; рассмотреть возможности применения параллельных корпусов текстов для перевода коллокаций в целом и этнокультурных коллокаций в частности; проанализировать целесообразность использования параллельного корпуса в обучении и для перевода безэквивалентных коллокаций.

Для решения обозначенных задач применяются следующие методы исследования: метод контекстуального анализа – анализ фрагмента текста, в котором использована этнокультурная коллокация, и его перевода для выявления полноты или неполноты передачи компонентов значения; сравнительный метод – сопоставление результатов использования параллельного корпуса и контекстного словаря для перевода этнокультурных коллокаций; индуктивный метод – общие выводы основываются на анализе конкретных языковых пар.

Теоретической базой исследования послужили работы отечественных (В. П. Захаров, С. Ю. Богданова и т.д.) и зарубежных авторов (S. Evert, A. Kilgarriff, Q. Zou, T. McEnery, A. Wilson и т.д.), изучающих потенциал использования корпусов текстов в переводческой деятельности.

Практическая значимость работы заключается в том, что полученные результаты исследования могут быть использованы для повышения качества перевода этнокультурных коллокаций и безэквивалентной лексики, а также в работе с изучающими язык: в рамках курса лексикологии и переводоведения.

Основная часть

Согласно данным исследований российских и зарубежных авторов, коллокационная компетенция студентов до сих пор является не до конца сформированной, что показывают результаты письменной части таких экзаменов, как IELTS (Liuqin, Qing, Jiahao, 2021). Из шести основных типов коллокаций: глагол – имя существительное, имя прилагательное – имя существительное, имя существительное – имя существительное, имя существительное – глагол, глагол – имя прилагательное, имя прилагательное – наречие – первые два типа представляют наиболее частотные группы, вызывающие затруднения (Zou, 2019), поскольку на них влияет языковая интерференция и неправильный выбор синонимичных слов.

Подход “data-driven learning” (Gilquin, Granger, 2010) является предпочтительным в силу использования аутентичных примеров из корпусов, а также возможности самостоятельной работы: поиск синонимов для создания разнообразной и аутентично звучащей речи. Умение пользоваться различными технологиями с обучающим потенциалом получает еще большую значимость в условиях дистанционной работы студентов.

Индуктивный метод мышления, необходимый для поиска коллокации и шаблона ее использования, позволяет студенту занять активную позицию в процессе изучения языка и самостоятельно ориентироваться в широком массиве информации.

Предпосылками для создания и использования в переводческой практике корпусов текстов являются объем и полнота представления типизированного языкового материала, объективность исследований, основанная на контекстной демонстрации данных, возможность постоянного пополнения базы. Исходя из предназначения лингвистического корпуса текстов – решения определенных лингвистических задач, из широкого списка определений корпуса остановимся на варианте Т. МакЭнери и Э. Вилсона (McEnery, Wilson, 2001), которые описывают корпус как собрание языковых фрагментов, отобранных в соответствии с четкими языковыми критериями для использования в качестве модели языка.

Обратимся к принципам деления корпусов по параллельности: к первому типу относятся одноязычные корпусы, в которых внимание уделяется противопоставлению диалектов и вариантов языка. Второй тип представлен многоязычными корпусами текстов, которые полезны, в частности, переводчикам в силу специфики содержащегося материала: независимо написанные тексты одной тематической области на разных языках, тексты-оригиналы и тексты-переводы.

Метатекстовая и статистическая информация о совместной встречаемости слов может быть использована при исследовании устойчивых словосочетаний в рамках комбинаторной лингвистики, лексикографии, переводоведения и лингвокультурологии в силу особенностей данных лексических единиц. Помимо этого, в корпусе содержится экстралингвистическая информация о жанре, происхождении текста, его авторе и т.д. Следует отметить, что коллокации в таких корпусах после формулировки запроса и фильтра результатов выделяются цветом, что облегчает процесс поиска и запоминания.

Для определения степени спаянности компонентов коллокации используют статистические методы, например меры ассоциации, вычисляющие линейную близость компонентов биграммы (Захаров, Богданова, 2020), что может послужить базой для создания различных лексикографических произведений, выявляющих особенности сочетаемости.

В рамках нашего исследования необходимо остановиться на понятии коллокации, которое подразумевает сочетание слов, имеющих тенденцию совместно встречаться в дискурсе: дантист – стоматология, school – teacher (Firth, 1957; Evert, 2007). Для лексикографии и фразеологии характерно рассматривать данный термин как неделимую комбинацию слов с различной степенью спаянности компонентов: to deliver a lecture (Hausmann, 1985; Cowie, 1994). А. Килгаррифф (Kilgarriff, 2006) определяет коллокации как “restricted co-occurrences of two syntactically bound lexical items” (совместная встречаемость двух синтаксически связанных лексических единиц, ограниченная определенными условиями).

Этнокультурные коллокации – тип коллокаций, свойственных определенному языковому коллективу и отражающих социально значимые явления действительности (например, heat pollution – тепловое загрязнение окружающей среды, сброс отработанной горячей воды в водоем; quality paper – солидная газета, рассчитанная на осведомленного читателя, и т.д.) (Arkhipova, Vlavatskaya, 2019). Для перевода этнокультурных коллокаций корпус может быть использован как репрезентативный и сбалансированный эмпирический банк языкового материала.

Национальный корпус русского языка обладает следующими характеристиками: жанровое разнообразие, социологически разнообразный состав авторов, различные периоды создания текстов, что помогает проследить динамику изменений языковых явлений. Подкорпусы представлены следующими разделами: глубоко аннотированный корпус, параллельный русско-английский корпус – именно этот раздел необходим

для анализа перевода этнокультурных коллокаций, корпус диалектных текстов, корпус поэтических текстов, обучающий корпус и корпус устной речи.

Возможность использования корпуса вместо нескольких двуязычных словарей делает процесс перевода более быстрым и качественным за счет репрезентативности картины употребления той или иной коллокации. Выбор окружения данного слова можно проверить, используя информацию о частотности, вероятности употребления, эксклюзивности (встречаемости только в этом составе) и порядке следования компонентов.

Рассмотрим примеры перевода этнокультурной коллокации “snake oil” в результате поиска в параллельном корпусе Национального корпуса русского языка: It was imported from the Middle East at prohibitive prices and used sparingly as a secret ingredient in delicacies and **snake-oil** medicines (Harari, 2014). / Его ввозили с Ближнего Востока. Цена была неподъемной для простых людей. Сахар по чуть-чуть подмешивали (секретный ингредиент!) в деликатесы и **шарлатанские зелья** (Харари, 2017).

The **snake-oil** salesmen were gradually driven out of business and mainstream medicine was forced to show that its treatments were both safe and effective before they could be employed (Singh, Ernst, 2008). / Торговцев **змеиным жиром** постепенно вытесняли из бизнеса, а общепринятую медицину принудили доказывать, что ее методы безопасны и эффективны, до того, как их позволялось применять (Сингх, Эрнст, 2016). Данные примеры наглядно демонстрируют затруднения, с которыми сталкивается переводчик. В исходном тексте коллокация употребляется в переносном смысле, описывая недействующее лекарство, продавец которого считался шарлатаном. Первый пример перевода сохраняет необходимые компоненты значения, а второй упускает подтекст псевдонаучности, что не в полной мере отражает задуманное автором исходного текста.

Рассмотрим примеры перевода этнокультурной коллокации “jet lag” в результате поиска в параллельном корпусе Национального корпуса русского языка: In the long run, the strains on human performance of these unusual schedules, or ‘planetary **jet lag**’, actually began to pace the rovers’ scientific output (Mindell, 2015). / В конечном итоге напряжение, которое испытывали люди, работавшие по необычным графикам, или «синдром межпланетной **смены часовых поясов**», действительно стал негативно сказываться на темпах научной работы с данными роверов (Минделл, 2017).

Suddenly all I wanted to do was crawl into a nice, soft bed and sleep off my **jet lag**, but Renaud quickly quashed my hopes (Weisberger, 2003). / Все, чего мне в тот момент хотелось, – это заползти в уютную мягкую постель и отоспаться как следует, отдохнуть от **тягостного перелета** (Вайсбергер, 2006).

Before Debbie went to sleep – for about twelve hours straight, a logical consequence of jet lag, her revels with friends in New York on the eve of her departure, and a healthy nervous system – they drank vodka downstairs in the bar (Ulitskaia, 2014). / Прежде чем Деби завалилась спать – часов на двенадцать, с учетом **сдвига во времени**, большой гульбы с друзьями в Нью-Йорке накануне отъезда и хорошей нервной системы, – они выпили водки внизу, в баре (Улицкая, 2011).

Данные примеры подчеркивают, что в отличие от двуязычных словарей параллельный корпус представляет стратегии и варианты перевода, освоенные другими переводчиками, которым уже приходилось сталкиваться с данным материалом. Если речь идет о безэквивалентном сочетании слов, словарь дает только описательный эквивалент, а корпус, в свою очередь, предоставляет готовые варианты перевода. Таким образом, обращение к корпусу может помочь при проверке и отборе наиболее полно передающего значение исходного сочетания соответствия.

В то же время следует отметить ряд трудностей, с которыми можно столкнуться при переводе этнокультурных коллокаций с помощью корпуса. Электронный онлайн-корпус русского языка начал свою работу в 2004 году, возможно, поэтому корпус параллельных текстов не отличается полнотой. К примеру, при поиске соответствий коллокации “cabin fever” и многих других мы столкнулись с проблемой, т.к. они не были представлены. В этом случае возможно прибегнуть к помощи контекстного словаря как инструмента для перевода на основе искусственного интеллекта, где используются двуязычные реальные тексты. Одним из наиболее популярных контекстных словарей является сервис ReversoContext. Рассмотрим результаты, полученные в этом приложении. Этнокультурная коллокация “cabin fever” означает состояние, возникающее у человека, если он проводит слишком много времени в помещении из-за плохой погоды, выражается в крайней раздражительности и странностях поведения. Контекстный словарь (<https://context.reverso.net/>) дает следующие варианты перевода: *I’m getting a little cabin fever here.* / Меня тут немного лихорадит от плотной полицейской охраны. *Okay, I’m getting a little cabin fever in here.* / Ладно, у меня тут небольшая клаустрофобия. *The police thought that it was what the old-timers used to call cabin fever.* / По мнению полиции, произошло то, что старики называют «буйством в клетке». *Did you have a cure for cabin fever?* / У тебя есть лекарство от депрессии? *Sometimes our kids get cabin fever and decide to go exploring.* / Иногда наши дети устают от замкнутого пространства и решаются на путешествия. *He has cabin fever.* / Он просто спятил в четырех стенах. *Call it cabin fever.* / Назовите это крайней раздражительностью. *OK, Tony, I think you’re getting a little cabin fever.* / Тони, по моему, ты в этом лифте перегрелся, но ты не волнуйся.

Как мы видим, преимуществом контекстного словаря является большее количество примеров использования коллокации, но в отличие от параллельного корпуса можно выделить следующие недостатки: отсутствие информации об источнике примера, соответственно отсутствие возможности проверить контекст и достоверность перевода. Некачественный перевод можно отметить в последнем примере, где составные элементы коллокации переведены по отдельности, что демонстрирует нарушение семантической целостности и утерю необходимых компонентов значения.

Заключение

Таким образом, мы приходим к следующим выводам. Правильное использование коллокаций является фактором знания языка, а понимание этнокультурных коллокаций выводит неносителя на новый уровень, поэтому следует отметить, что изучение языка на определенном этапе требует более широкого охвата культурной информации. Параллельный корпус русского языка может способствовать этому, обеспечивая возможность осознать и увидеть то, что человек, не знакомый с культурно-специфичными единицами, закрепленными в языке, не видит.

Настоящее исследование продемонстрировало, что для оптимизации подбора подходящего соответствия этнокультурной коллокации можно использовать параллельный корпус Национального корпуса русского языка. Проанализированные примеры исходного текста и текста перевода показали, что в некоторых случаях этнокультурные коллокации переданы без потерь в плане значения.

Изучение и перевод коллокаций требуют применения технологий с использованием аутентичного материала для изучения. Репрезентативный банк, включающий экстралингвистическую информацию об особенностях текстов, в которых они встречаются, может служить неоценимым подспорьем для работы. Параллельный русско-английский корпус отмечен следующими характеристиками: разнообразие жанров текстов, состава авторов, различные периоды создания текстов, достаточно богатый материал. Однако в силу своих особенностей этнокультурные коллокации как тип сочетаний слов, обладающих синтактико-семантическим единством и связанных традицией носителей языка, представлены в данной базе в недостаточном объеме. Придя к этому выводу, мы обратились к контекстным словарям, которые, в свою очередь, не обладают таким же уровнем достоверности и точности.

Проведенное исследование способствует более широкому видению возможностей использования параллельного корпуса русского языка, а также анализу эффективности вышеперечисленных технологий в обучении языку и переводе, что составляет перспективы дальнейшего исследования.

Источники | References

1. Вайсбергер Л. Дьявол носит Prada / пер. М. Малкова, Т. Шабаевой. М.: АСТ, 2006.
2. Захаров В., Богданова С. Корпусная лингвистика. СПб.: Изд-во СПбГУ, 2020.
3. Минделл Д. Восстание машин отменяется! Мифы о роботизации / пер. В. Краснянской. М.: Альпина нон-фикшн, 2017.
4. Сингх С., Эрнст Э. Ни кошелек, ни жизни. Нетрадиционная медицина под следствием / пер. А. Бродоцкой. М.: Corpus, 2016.
5. Улицкая Л. Е. Зеленый шатер. М.: АСТ, 2011.
6. Харари Ю. Н. Sapiens. Краткая история человечества. М.: Синдбад, 2017.
7. Arkhipova E. I., Vlavatskaya M. V. Ethnocultural Collocations in the Context of Combinatorial Linguistics // Advances in Social Science, Education and Humanities Research. 2019. Vol. 333. Humanities and Social Sciences: Novations, Problems, Prospects (HSSNPP 2019).
8. Cowie A. P. Phraseology. The Encyclopedia of Language and Linguistics. Oxford: Pergamon, 1994.
9. Evert S. Corpora and Collocations. Berlin: Mouton de Gruyter, 2007.
10. Firth J. R. Modes of Meaning. Papers in Linguistics, 1934–1951. Oxford: University Press, 1957.
11. Gilquin G., Granger S. How Can Data-Driven Learning Be Used in Language Teaching. L., 2010.
12. Harari Y. N. Sapiens: A Brief History of Humankind. L.: Vintage Books, 2014.
13. Hausmann F. J. Kollokationen im Deutschen Wörterbuch: ein Beitrag zur Theorie des lexikographischen Beispiels. Lexikographie und Grammatik. Tübingen: Niemeyer, 1985.
14. Kilgarriff A. Collocationality (and How to Measure It) // Proceedings of the 12th Euralex International Congress on Lexicography (EURALEX). Turin: Springer-Verlag, 2006. URL: [https://www.semanticscholar.org/paper/Collocationality-\(and-how-to-measure-it\)-Kilgarriff/ae813a14bdfcf8d33d1846a11c5b7e6896bac9a7](https://www.semanticscholar.org/paper/Collocationality-(and-how-to-measure-it)-Kilgarriff/ae813a14bdfcf8d33d1846a11c5b7e6896bac9a7)
15. Liuqin F., Qing M., Jiahao Y. The Effectiveness of Corpus-Based Training on Collocation Use in L2 Writing for Chinese Senior Secondary School Students // Journal of China Computer-Assisted Language Learning. 2021. Vol. 1. Iss. 1.
16. McEnery T., Wilson A. Corpus Linguistics. Edinburgh: Edinburgh University Press, 2001.
17. Mindell D. A. Our Robots, Ourselves: Robotics and the Myths of Autonomy. N. Y.: Penguin Random House LLC, 2015.
18. Singh S., Ernst E. Trick or Treatment: The Undeniable Facts about Alternative Medicine. N. Y. - L.: Bantam Press, 2008.
19. Ulitskaia L. E. The Big Green Tent. N. Y.: Picador, 2014.
20. Weisberger L. The Devil Wears Prada. N. Y.: Anchor Books, 2003.
21. Zou Q. A Corpus-Based Study of Verb-Noun Collocation Errors in Chinese Non-English Majors' Writings // Proceedings of the 4th International Conference on Contemporary Education, Social Sciences and Humanities (ICCESSH 2019). P.: Atlantis Press, 2019.

Информация об авторах | Author information

Архипова Елена Ивановна¹, к. филол. н.

¹ Новосибирский государственный технический университет



Arkhipova Elena Ivanovna¹, PhD

¹ Novosibirsk State Technical University

¹ elena1503@inbox.ru

Информация о статье | About this article

Дата поступления рукописи (received): 14.11.2021; опубликовано (published): 28.02.2022.

Ключевые слова (keywords): параллельный корпус; этнокультурная коллокация; безэквивалентная лексика; контекстный словарь; коллокация; parallel corpus; ethnocultural collocation; culture-bound vocabulary; contextual dictionary; collocation.