

УДК 81'255:004.5

КОРПУС ТЕКСТОВ КАК АЛЬТЕРНАТИВА СЛОВАРЮ В ПЕРЕВОДЕ И КОРПУСНАЯ ПРАГМАТИКА ПЕРЕВОДА

А. Ю. Мордовин

Иркутский государственный лингвистический университет, Иркутск

На основе прагматической и функциональной схожести словаря и корпуса текстов для целей перевода текстов предпринята попытка сопоставить два указанных вида ресурсов с целью определить их принципиально важные методологические и прагматические свойства, выявить преимущества и ограничения, присущие словарю и корпусу применительно к переводу, а также описать прагматический аспект взаимодействия переводчика с корпусом.

Ключевые слова: корпус текстов, словарь, перевод, прагматика перевода.

The Text Corpus as an Alternative to the Dictionary, and Corpus Pragmatics of Translation

Alexey Mordovin

Accepting the pragmatic and functional similarity of the dictionary and the text corpus for the purpose of text translation, the article compares the two types of resources in order to define their key methodological and pragmatic properties, to identify their inherent advantages and restrictions, as applicable to translation, and to describe the pragmatics of “translator-corpus” interaction.

Key words: text corpus, dictionary, translation, translation pragmatics.

Данное исследование посвящено анализу потенциальной «конкуренции», которую корпусы текстов не только могут, но уже начинают оказывать традиционным словарям, и представляет собой попытку осмыслить причины и сущность происходящих в этой связи процессов в области профессионального перевода.

С целью максимально точного обозначения поля и основных объектов исследования необходимо уточнить объем обсуждаемой прототипической ситуации и уровень вовлеченных языковых компетенций.

Суть рассматриваемого вопроса в следующей триаде: «переводчик – словарь – корпус». Исключительно для целей нашего размышления, за пределами этой рабочей связки, организованной по типу «исполнитель (творец) – инструмент», мы оставляем исходный текст (автора) и текст перевода (реципиента). Иначе говоря, нас интересует исключительно внутренний механизм процесса перевода,

поэтому мы вынужденно отвлекаемся от вопросов его соотнесенности с исходным замыслом автора и степени удовлетворенности потребителя перевода (реципиента).

Определим основные методологически важные признаки элементов нашей триады.

Говоря о переводчике, мы подразумеваем профессионального переводчика, имеющего необходимое образование, навыки, опыт работы, а также компетенции по взаимодействию с соответствующими электронными интерфейсами.

Наше представление о словаре определяется двумя принципами: прототипом переводчика и его переводческой задачей, а также необходимостью объяснить природу постулируемой методологической схожести словаря и корпуса текстов.

С одной стороны, поскольку речь идет о профессиональном переводчике, необходимо придавать максимально возможный приоритет эргономичности словаря и его «энергоэффективности». Иначе говоря, при обращении переводчика к словарю в дополнение к базовым ожиданиям любого пользователя (наличие достаточного выбора языковых единиц, системы помет, точность определений и т.д.), для профессионального переводчика большое значение имеет субъективная эффективность словаря, т.е. отношение затрат времени и усилий на навигацию по словарю к удовлетворенности полученным результатом. Такая эффективность словаря оценивается переводчиком субъективно и обязательно соотносится с текущими требованиями в отношении сроков, качества, лингвокультурной точности и иных параметров перевода.

Кроме того, сфера деятельности профессионального переводчика, как правило, не ограничивается узким кругом специализации. Поэтому большое значение уделяется также и принципу эргономики словаря, т.е. удобства и привычности в использовании несмотря на смену тематики или формата перевода.

С учетом текущего развития компьютерных технологий можно констатировать, что всем указанным требованиям к эргономичности и эффективности словаря в полной мере соответствуют лишь профессиональные электронные словари. Поскольку данное исследование не ставит задачи составления таксономии электронных словарей, ограничим наше внимание двумя наиболее

распространенными на рынке продуктами – словарями Lingvo и Multitran.

Эти словари обладают следующими общими свойствами, которые позволяют значительно повысить уровень субъективно воспринимаемой эффективности словаря: многоязычность, однообразие интерфейса и основных принципов взаимодействия с ним, удобный способ организации переводов в тематики (подсловари), возможность индексации, подшивки, пометки выбранного пользователем варианта перевода и т.д. Кроме того, словарь Multitran позволяет всем желающим пользователям добавлять в словарь собственные варианты переводов, которые далее становятся доступными для всех остальных пользователей словаря при минимальной «цензуре» (за что Multitran на сленге переводчиков достаточно быстро снискал прозвище «свалки»).

Основные шаги обращения переводчика к такому словарю следующие: копирование искомого слова (словосочетания) в интерфейс словаря – отправка поискового запроса – вывод большого числа вариантов, разбитых по тематикам – выбор необходимой тематики и варианта – быстрое копирование перевода в переводимый документ.

Нетрудно заметить, что такой порядок действий отнюдь не является уникальным для электронного словаря, но свойствен любой электронной поисковой системе. Таким образом, на уровне прагматики словарь постепенно перестает восприниматься как авторский труд, или как продукт «ручного» анализа языковых компетенций информантов. В особенности, если переводчику известно, что интересующий его вариант перевода добавлен в словарь другим таким же переводчиком, то, в сущности, обращение к словарю превращается в обращение к прочим участникам (двуязычного) языкового сообщества, организованного в оболочке словаря, и в целом мало чем отличается от запроса в поисковую систему по сети Интернет, на профильный форум и т.д.

Различие заключается лишь в том, что в электронном словаре искомым ответ доступен мгновенно, частично «легитимизирован» (если внесен пользователем с «авторитетным» именем пользователя, например «Dzherko» в словаре Multitran). Электронные словари тесно увязаны (особенно в онлайн-версии) с профильными форумами, где

всегда присутствуют компетентные переводчики, также способные дать оперативный ответ.

Ссылки на исходные бумажные словари, послужившие в качестве исходника при разработке тематик словаря (подсловарей), как правило, не приводятся; впрочем, и не вызывают интереса пользователя.

Для случаев, когда от переводчика требуется сознательное обращение к определенному изданному словарю, когда нужны его выходные или авторские данные, существуют иные специализированные онлайн-сервисы, например «Яндекс.Словари».

Все описанное выше необходимо для того, чтобы проиллюстрировать, что с точки зрения практикующего переводчика словарь – всего лишь один из удобных способов обратиться к опыту соответственного языкового сообщества.

В этом смысле и оказывается интересным его сходство с корпусом текстов. Строго говоря, между корпусом текстов и словарем, применительно к переводу существует фундаментальная разница: словарь указывает на межъязыковые эквиваленты, тогда как перед корпусом такой задачи не стоит, он всего лишь более или менее адекватно воспроизводит текущее состояние иностранного (родного) языка переводчика и оборудован системой для построения конкордансов и иных видов поиска.

Однако во взятой нами за основу прототипической ситуации, в силу высокого профессионального уровня переводчик нередко знаком с большинством вариантов перевода и нуждается в информации для совершения мотивированного выбора. Для этого он может выбрать обращение либо к словарю, либо к корпусу.

Обращение к словарю возможно по двум сценариям:

1) направление соответственного поискового запроса в электронный словарь с дальнейшим выбором варианта на основании того, в какую тематику (подсловарь) он будет включен,

2) использование комбинаторного словаря.

На наш взгляд, корпус позволяет решить такой тип задачи гораздо эффективнее и убедительнее.

Справедливо будет отметить, что несвязанные с корпусами текстов численные методы для выбора варианта сочетаемости/выбора из группы синонимов известны достаточно давно, со времени появления крупных поисковых сервисов в сети Интернет. Это очень

разнообразные по способу и форме алгоритмы принятия решения на основании данных о количестве найденных поисковой системой веб-страниц, содержащих необходимую речевую единицу (единицы). Могут использоваться различные операторы поисковых запросов, фильтры и т.д., что позволяет приспособить метод к решению конкретной задачи.

Этот метод страдает рядом недостатков, как будто бы устраненных в корпусах текстов. Во-первых, поиск осуществляется по неоправданно зауженному сектору дискурса – только по тем жанрам, которые доминируют в сети Интернет (именно доминируют, а не просто присутствуют, ведь этот метод – строго статистический по своей сути). Во-вторых, при таком поиске только частично решена проблема численной адекватности результата, особенно, когда полученные результаты численно близки. Дело в том, что в сети нередко существует большое количество «клонов» текстов, и чем текст популярнее, тем больше у него «клонов» на различных ресурсах. В предельном выражении это может привести к тому, что какое-либо языковое/речевое явление, будучи практически идиосинкратическим по сути, за счет популярности текста может создавать видимость высокой частотности употребления в языке. Частично проблема «клонов» решается с помощью интерфейса самой поисковой системы.

Формально корпуса текстов лишены присущих поисковым системам недостатков, хотя существует точка зрения, что сеть Интернет необходимо также считать корпусом («веб как корпус»).

Корпусы текстов сочетают в себе методологические преимущества и сети Интернет, и классических словарей одновременно. Корпусы составляются таким образом, чтобы наиболее полно и непредвзято отобразить весь объем дискурса на данном языке. При этом применительно к целям перевода они превосходят словари по объективности, ведь корпуса не являются, в отличие от словаря, «опосредованным» продуктом. Взамен корпуса текстов переключают на пользователя задачу, уже выполненную в рамках словаря, а именно – смысловую интерпретацию количественных результатов поиска в корпусе.

Корпус не может рекомендовать какие-либо варианты перевода, он лишь указывает на их частотность, тогда как словарь не только указывает, но и рекомендует. При этом переводчик может счесть

рекомендации словаря искаженными, устаревшими, авторскими и т.д. А вот с корпусом уже не поспоришь, правда, и ответственность за принятое на основании корпуса решение переводчик несет лично, и сослаться в случае ошибки практически не на кого. Это связано с тем, что численное доминирование никоим образом не трансформируется в смысловую правильность употребления.

Собственно говоря, корпус – это не такой уж непосредственный языковой ресурс, каким он представляется самим составителям корпуса. Корпус также подвергается опосредованию в момент составления по ряду причин: социолингвистический «срез» корпуса, его жанровое наполнение, интерференция коллективной языковой личности авторского коллектива, та или иная форма «цензуры» корпуса. Другое дело, что в силу несопоставимо большого для словаря размера «словарной статьи» в корпусе, т.е., как минимум фразы, или текста в целом, в корпусе эти погрешности нивелируются значительно сильнее, чем в словаре.

Следует понимать, что корпус в какой-то мере – это недословарь. Это – словарь, в котором авторов попросили «отойти в сторону» сразу после того, как необходимый иллюстративный языковой материал был подобран, и вместо языковой компетенции авторов или их информантов эту базу текстов проиндексировали и снабдили поисковым механизмом. Можно считать, что обращение именно к корпусу, а не к словарю – это определенная реакция недоверия, либо проявление жажды самостоятельного языкового исследования со стороны пользователя.

Между тем, существуют также и частотные словари, которые, по мнению автора, неспособны конкурировать с корпусами. Не предлагая ничего качественно нового в сравнении с корпусами текстов, частотные словари не снабжены прозрачной или легкодоступной базой текстов, на основании которой составлены, и, соответственно, вызывают меньше доверия.

Если взглянуть на ситуацию с иной стороны, можно утверждать, что переводчику всегда хочется наблюдать языковую компетенцию носителя иностранного языка в как можно более непосредственном и удобном виде. Тем не менее, на начальном профессиональном уровне переводчик может оказаться не готовым столкнуться с необработанным, метафорически говоря, «непереваренным» массивом текстов на иностранном (родном) языке, поэтому он

предпочтет обратиться к словарю, предлагающему определенные гарантии уверенности. По мере профессионального роста переводчик, с одной стороны, все более нуждается в основаниях для выбора среди синонимов, чем в собственно переводе, а с другой, – все менее склонен слепо доверять словарю, и все более готов к «бунту» против него.

Помимо прочего, привлекательность корпусов текстов для совершенствующегося переводчика объясняется осознанным или подсознательным пониманием того, что «содержание текста или его фрагмента приравнивается к ситуации в реальной действительности, которая описывается в данном тексте или фрагменте текста» [2. С. 75] Очевидно, что чем выше требования к качеству перевода устанавливает для себя переводчик, тем в большей степени ему необходимо вникнуть в коммуникативную ситуацию и коммуникативное намерение автора. В случаях, когда это невозможно или затруднительно, любой вариант, основанный на количественной корпусной оценке, представляет собой убедительный аргумент.

Таким образом, если существует общий принцип, гласящий, что «сверх»-переводчику интересен «недо»-словарь, то в крайнем выражении мы имеем бесконечно профессионального переводчика, которого абсолютно не интересует какой-либо словарь, он полагается исключительно на собственное непосредственное восприятие иностранного языка с минимальной количественной и без абсолютно какой-либо качественной обработки. С другой стороны – наихудший переводчик избегает «прямого» контакта с языком в виде корпуса и полностью доверяет словарю.

Очевидно, что в действительности ни один из типов не существует в чистом виде, однако можно постулировать, что готовность обращения к корпусу является диагностическим критерием профессионального роста переводчика, манифестацией когнитивного диссонанса переводчика [1].

Наконец, необходимо указать на возможные риски и трудности, поджидающие переводчика при работе с корпусом. Метафорически выражаясь, можно сказать, что корпус не адаптирован под человека. Если угодно, корпус – это некоторая глыба текстов, которая может легко подавить языковую эрудицию отдельного индивида. Неохватность корпуса – это его и преимущество, и недостаток одновременно. С одной стороны, она гарантирует невиданную ранее в

лингвистике объективность результата, но, с другой стороны, значительно повышает планку для переводчика в части исследовательской интуиции и смекалки, т.е. предъявляет к переводчику дополнительные и достаточно высокие требования.

В самом деле, как убедиться в уместности/правильности определенного синонима/варианта сочетаемости в определенном контексте с помощью корпуса, без какого-либо обращения к словарю? Необходимо составить конкорданс искомого слова относительно иных возможных кандидатов, однако в качестве одного из условий поиска задать дополнительный параметр: указать жанровую, временную или иную (гендерную, социальную и т.д.) характеристику текста, либо указать слово или ряд слова, на определенном расстоянии от которых может находиться искомая единица в тексте. Последний вариант уже требует от переводчика определенных знаний в области особенностей синтаксиса переводимого языка (например, нужно знать, располагать ли в запросе сопутствующие слова слева или права от искомого, и на каком расстоянии), его семантики (как правильно подобрать такой комплект концептов, который наиболее вероятно укажет на искомый дискурсивный контекст), или прагматики (в случае, если корпус оборудован широким диапазоном дифференциации авторов, жанров или формата текстов).

В целом, на начальном этапе обращение с такой целью к корпусу очень может напоминать известную задачу «пойди туда, не знаю куда, найди то, не знаю что». Переводчик изначально совсем не обязательно будет знать, в едином концептуальном поле с какими единицами должен соседствовать искомый вариант. Поэтому в дополнение к глубокому знанию языка, его теоретических принципов и лингвострановедческой компетенции, при работе с корпусами текстов переводчик обязан проявить высокий уровень исследовательской интуиции.

Сложность работы с корпусом в целом и в целях перевода в частности обусловлена тем, что мы предлагаем называть «неорганичностью» машинного языкового массива. Корпус – это техногенный по своей сути объект, поэтому взаимодействие с ним сопровождается неестественной, можно даже сказать «патологической» прагматикой.

Изначально при рождении человек обладает лишь способностью непосредственно работать только с тем объемом языкового материала, который умещается в его краткосрочной и долгосрочной памяти. До изобретения письменности на этом уровне человек и оставался на всю жизнь. После ее изобретения и распространения, по мере социализации, человек научается работать со все большим объемом языкового материала одновременно, т.е., грубо говоря, с книгой.

Тем не менее, в отсутствие электронного поискового интерфейса происходило лишь «усиление» человеческой памяти, повышалась готовность информации к извлечению в нужный момент.

В отличие от книги электронный поисковый интерфейс – это гигантский качественный методологический и философский прорыв, значение которого еще не было оценено по достоинству.

Сохраняя функцию «ускоренной книги», электронный поиск способен генерировать новое объективное знание там, где это было ранее невозможно и немислимо. С методологической точки зрения, выводы, полученные путем количественного анализа языка на столь высоком уровне и в столь большом объеме, как в случае с корпусом, уникальны.

По сути, сама постановка вопроса о том, можно ли использовать количественные корпусные данные при принятии переводчиком решений в отношении языкового узуса – это микрореволюция, потому что только теперь, с появлением корпусов текстов, человек смог по-настоящему, хоть и не полностью, абстрагироваться от языка.

Представляется, что все человечество в целом еще пока находится в процессе осознания наступившей машинной реальности, и опосредование языка машиной находится лишь в зачаточном состоянии. Однако уже сейчас возникают принципиально новые вопросы взаимодействия человека и оцифрованного языка, такие, например, как вопрос взаимоотношения переводчика и корпуса. Время покажет, в какой степени машинные массивы текстов, построенные по принципу корпуса, смогут вытеснить словари.

В заключение отметим, что корпуса текстов представляют собой ценный языковой ресурс, функционал, эффективность и область применения которого не аналогичны словарю. Ряд переводческих задач, таких как выбор синонима/альтернатив сочетаемости, может быть решен с помощью корпуса более

эффективно, и тому имеются все необходимые методологические обоснования. Тем не менее, корпус, являясь техногенным объектом, обладает рядом онтологически непривычных для человеческой языковой способности аспектов, что требует от переводчика особых навыков, подготовки и интуиции, а также налагает на переводчика дополнительную ответственность и уменьшает возможности опоры последнего на авторитет словаря.

Библиографический список

1. Воскобойник Г.Д. Лингвофилософские основания общей когнитивной теории перевода: Дис. ... д-ра филол. наук : 10.02.20 : Иркутск, 2004. – 296 с.
2. Сдобников В.В. Мир за текстом (к вопросу о восприятии оригинала в процессе перевода) // Вестник Нижегородского государственного лингвистического университета им. Н.А. Добролюбова. Вып.2. Лингвистика и межкультурная коммуникация. Нижний Новгород: НГЛУ им. Н.А. Добролюбова, 2008. С. 72-83.

Сведения об авторе

Мордовин Алексей Юрьевич
кандидат филологических наук,
докторант кафедры прикладной лингвистики
Иркутского государственного лингвистического университета
E-mail: alexmordovin@mail.ru