

Б. В. Орехов

*Параллельный
корпус переводов
«Слова о полку
Игореве»:
итоги
и перспективы¹*

Корпусная лингвистика стимулирует создание разнообразных исследовательских инструментов, среди которых один из самых любопытных—параллельный корпус: сложно организованная система текстов, интуитивно ощущаемых как «идентичные», но созданных на разных языках, то есть—в обиходном понимании—сопоставленные друг другу текст-оригинал и текст-перевод на другой язык. Ещё в процессе создания корпуса между единицами текстов устанавливается соответствие, а дальнейшая пользовательская работа направлена на выявление общих и различных характеристик сопоставленных друг другу текстов.

Имеющиеся на сегодняшний день параллельные корпуса, в основном, содержат прозаический текст и небольшое количество его переводов (в подавляющем числе случаев—один). См. The Regensburg Parallel Corpus, а также Корпус параллельных текстов

¹ Работа выполнена при финансовой поддержке Российского гуманитарного научного фонда (грант № 08-04-12104В)

в составе Национального корпуса русского языка. На сайте Национального корпуса русского языка имеется ссылка на не совсем обычный параллельный корпус, который, во-первых, организует не вполне прозаический текст (поэтическая природа «Слова о полку Игореве» остаётся дискуссионной, но принципы организации этого произведения выходят за рамки традиционной прозы), а во-вторых, выступает способом хранения сразу многих (в перспективе—всех) переводов. Ниже излагаются некоторые концептуальные и технические подробности этого проекта.

С февраля 2007 года в Интернете по адресу <http://nevmenandr.net/slovo> действует Параллельный корпус переводов «Слова о полку Игореве». Его появление стало попыткой ответа на вопрос, как следовало бы в удобной для читателя и исследователя форме собирать и хранить многочисленные переводы «Слова о полку Игореве» (далее—СПИ).

СПИ в этом смысле совершенно уникально. Если можно попробовать найти произведение русской литературы, которое бы поспорило со СПИ в количестве переводов на другие языки, то уж, во всяком случае, никакой другой текст (включая знаменитую оду Горация III, 30) на русский язык столько раз не переводился. Точно-го количества хотя бы русских переводов не в силах назвать никто, тем более что их число постоянно растёт, однако можно сказать, что попыток перевести СПИ на современный русский язык было предпринято не меньше ста. Переводов на другие языки в сумме больше этого числа в полтора раза.

Основных причин появления такого количества переводов две, хотя полный перечень, конечно, должен быть гораздо длиннее.

Даже среди древних памятников, как правило, имеющих драматичную судьбу, СПИ завоевало особую репутацию необычайно проблемного текста. Если не принимать во внимание спор о его подлинности², останется с трудом поддающееся счёту количество «тёмных» мест, каждое из которых имеет целую серию конкурирующих версий прочтения. Если какой-то перевод не учитывает принятое исследователем чтение, он автоматически становится

² После книги академика А. А. Зализняка [Зализняк 2004] версия о позднейшем происхождении СПИ фактически перешла из поля научного обсуждения в ряд суеверий.

неправильным и, стало быть, требуется создание нового перевода. В отношении СПИ срабатывает формула «сколько интерпретаторов, столько и переводов».

Но попыток передать содержание древнего текста средствами современного языка было бы гораздо меньше, не будь СПИ признанным поэтическим образцом. Его художественные достоинства не ставились под сомнение даже теми, кто высказывался против его подлинности.

В последнее время к двум очевидным причинам добавилась ещё одна: поэтический перевод СПИ превратился в особый жанр литературного упражнения, почти столь же традиционный, сколь и строгие стихотворные формы. Освящённая именами Жуковского, Пушкина (который собирался сделать свой перевод СПИ), Майкова, Заболоцкого традиция перевода СПИ настолько сильна, что привлекает любителей словесности, зачастую далёких в своих профессиональных занятиях от чтения древних текстов.

Таким образом, текст СПИ, из-за своей древности недоступный среднему читателю, существует в русской культуре сразу во многих переводах. Эту ситуацию принято называть поливариантным функционированием. «Каждый перевод, сколь бы он ни был превосходен, проецирует многомерную сложность подлинника на плоскость, делает оригинал упрощённым и представляет его односторонне. Сопоставляя два или несколько переводов, читатель может получить как бы стереоскопическое изображение оригинала, увидеть его с разных сторон» [Гаспаров 2006, с. 5]. Позиция Гаспарова, хотя и нуждается в перепроверке (в особенности уточнения заслуживает метафорический термин «стереоскопический»), отражает довольно стойкий и, по всей видимости, интуитивно обоснованный взгляд на предмет поливариантного функционирования иноязычного текста.

В любом случае, задача собрания переводов в одном месте довольно актуальна до сих пор, несмотря на то, что решать её пытались давно. Дело в том, что формат книги даёт для решения этой задачи очень ограниченные возможности. Издания, включающие сразу несколько переводов СПИ, разумеется, предпринимались, такую эдиционную практику можно даже назвать вполне устоявшейся. Но как расположить переводы в книге? Читателю удобнее всего иметь нужные тексты перед глазами, но определить, какие именно

тексты и в каком порядке понадобятся, а какие окажутся лишними, заранее невозможно; статический же характер бумажного издания предопределяет, что этот выбор должен быть осуществлен раз и навсегда. Неудобства для пользователя здесь очевидны. Вторая проблема в этом ряду—расположение текстов. Книжный формат даёт возможность предложить читателю для одновременного ознакомления два, максимум—четыре текста, которые размещаются слева и справа на развороте или втиснуты в две-три колонки (большого не позволит ширина страницы) на одном листе. Но, как уже отмечалось, в случае со СПИ количество переводов совсем другого порядка.

Можно поместить другие переводы на следующих страницах, выдерживая линейную композицию книги, хотя трудно представить себе объём издания, включающего хотя бы половину общего числа переводов. Чаще всего, публикуя несколько переводов, издатели склоняются именно к этому решению. Но и такую форму трудно назвать удобной. Владимир Набоков (один из переводчиков СПИ), устами своего героя, настаивавшего на параллельном чтении двух частей романа, предлагал, чтобы обойтись без хлопотного перелистывания вперёд-назад, либо разрезать книгу и скрепить вместе соответствующие страницы произведения, либо купить сразу два экземпляра, которые можно будет положить перед собой: «I find it wise in such cases as this to eliminate the bother of back-and-forth leafings by either cutting out and clipping together the pages with the text of the thing, or, even more simply, purchasing two copies of the same work which can then be placed in adjacent positions on a comfortable table» [Nabokov 1962, p. 28].

Весьма утомительно даже сопоставление одного текста и одного перевода на развороте, так как это требует от читателя постоянных усилий по поиску соответствий. Задача становится проще, если текст хорошо структурирован, как, например, в летописи, где мы наблюдаем естественное погодное членение материала, но другие жанры не предоставляют такого удобства. Увеличение числа переводов создало бы читателю дополнительные сложности.

Всё это довольно ясно свидетельствует не в пользу традиционного книжного, а в пользу электронного собрания текстов, в пользу корпуса. Поскольку речь идёт не просто об электронной коллекции с надстроенной возможностью поиска, а о форме представления

связанных друг с другом текстов, самым логичным решением предстало создание параллельного корпуса.

Параллельный корпус переводов СПИ решает перечисленные трудности следующим образом.

Во-первых, изменено направление представления текстов, которые расположены не в привычных колонках, а в строках. Соответствующие фрагменты текста, таким образом, оказываются друг под другом и дают пользователю полную и наглядную информацию о сходствах и различиях вариантов, переводческих вольностях, разнообразии трактовок. В большинстве случаев это приводит к появлению горизонтальной прокрутки экрана, но традиционные принципы HTML-вёрстки приходится приносить в жертву наглядности.

«Вытянутые» в строку поэтические переводы также лишаются своего традиционного облика колонки, но такое фундаментальное свойство стихотворной речи, как деление на строки, в корпусе сохраняется, отмеченное специальным знаком—вертикальной чертой: «|». Строфы отграничиваются друг от друга двумя вертикальными чертами: «||». Деление на абзацы в прозаических переводах не оговаривается. Деление на логические части, предпринятое рядом переводчиков, также не учитывается, их названия опущены. Не воспроизводится пагинация Первого издания и Екатерининской копии: пользователь может обратиться к любому дипломатическому их воспроизведению. Пользователю также дана возможность по его желанию просматривать текст без символа «|», если граница стиха для него не важна, а вертикальная черта мешает восприятию. Кстати, по одним только расставленным в хронологическом порядке переводам СПИ можно было бы построить довольно внятную картину как стилистической эволюции русского художественного языка, так и ритмической эволюции русского стиха. Другое дело, что целесообразность создания именно такой модели сомнительна, а подобное упражнение вряд ли можно было бы расценить иначе, чем игру ума исследователя, но сама возможность такого упражнения даёт представление о богатстве содержащейся в переводах информации.

Теоретически можно представить и такое бумажное издание, в котором каждый следующий фрагмент текста для всех переводов

располагался бы на следующем развороте. Разница между книгой и корпусом здесь в том, что в книге количество переводов было бы ограничено длиной листа (около 40–45 строк), а в корпусе оно ограничено только размером базы данных, которая может содержать миллионы записей. Как бы там ни было, воображаемому верстальщику, который вынужден был бы взяться за такую работу, можно только посочувствовать. Среди прочего ему, например, пришлось бы решать нетривиальную задачу сопоставления на бумажном листе лаконичных прозаических переводов и зачастую пространных поэтических. Для «растяжимого» экранного представления это проблемой не является.

Во-вторых, поскольку на экран теперь возможно вывести сколько угодно текстов, предусмотрена возможность их отбора самим пользователем. В соответствующей форме можно галочками отметить требуемые переводы и таким образом сформировать своего рода подкорпус. В случае, если пользователь нажал на ввод, не отметив ни одного перевода, на экран будут выведены все имеющиеся в базе тексты. На момент написания этой статьи их около восьмидесяти, и в ближайшее время благодаря поддержке РГНФ это количество должно возрасти вдвое.

Параллельный корпус СПИ—пример неаннотированного корпуса. Вся метатекстовая информация, которая на данный момент доступна пользователю,—это сопоставленность данного отрывка текста отрывку как минимум ещё одного текста, то есть перед помещением в корпус тексты проходят предварительный процесс ручного выравнивания.

Текст СПИ не слишком велик, но всё же в горизонтальном представлении вывести его на экран целиком было бы не лучшим решением, так как создало бы пользователю значительные неудобства. Поэтому каждый перевод в корпусе разбит на 218 фрагментов («звеньев») в соответствии с членением СПИ, предложенным Р. О. Якобсоном. В процессе работы выяснилось, что это вполне универсальное членение, которое подходит не только для оригинала, но и для его переводов. На экран текст выводится по фрагментам, номер фрагмента можно задать в специальном поле в главной форме одновременно с выбором переводов. Предусмотрена возможность листать текст от фрагмента к фрагменту,

сохраняя выбор переводов, то есть в рамках сформированного подкорпуса. Если в поле ввода номера фрагмента пользователь не ввёл ничего, по умолчанию будет выведен первый фрагмент. При вводе нечислового значения или значения не в диапазоне 1–218, система сообщит об ошибке. Допустим ввод как в виде «001», так и «01», и «1».

В отдельных случаях—разумеется, это в большей степени касается поэтических переводов—адекватное разбиение текста на фрагменты (которое представляет собой процесс выравнивания) связано с некоторыми сложностями. В основном, это случаи инверсии или исключения некоторого фрагмента переводчиком. В случае, если по причине инверсии эквивалентный эталону отрывок в переводе попадает в другой фрагмент, даётся ссылка на этот фрагмент «См. фрагмент п» и во всплывающей подсказке воспроизводится текст. Например, в переводе Н. А. Заболоцкого содержание фрагмента 86 оказалось внутри текста, относящегося к фрагменту 87. В соответствующей выдаче, таким образом, показывается надпись «См. фрагмент 87», а при наведении курсора на эту строку высвечивается текст, который было бы логично сопоставить с фрагментом 86: «И растет крамола меж князьями, | И не видно от князей добра. | |»

В случае, когда в переводе отсутствует эквивалент выбранного фрагмента, в строке перевода выводится тройной прочерк: «—», как, например, во фрагменте 51 перевода С. В. Шервинского.

Отдельную сложность представляет обоснованная в своё время А. И. Соболевским и принятая в некоторых переводах перестановка отрывка «Тогда Игорь възрѣ на свѣтлое солнце ... съ вами, русици, хошу главу свою приложити, а любо испити шеломомь Дону» (соответствует фрагментам 8–14) дальше от начала. Сообразно с нуждами корпуса эта перестановка элиминируется, текст приводится в последовательности Первого издания, а место перестановки обозначается специальной пиктограммой.

Как именно СПИ разбито на фрагменты, можно посмотреть на отдельной странице, и там же осуществить выбор требуемых переводов.

В ходе дискуссий уже после появления корпуса (например, на совместном заседании семинара «Информационные технологии

в исследованиях российских филологов» и Общества исследователей Древней Руси в Институте мировой литературы им. А. М. Горького (16 мая 2007 года) высказывались сомнения в целесообразности выбора разбиения текста «по Якобсону» как основы для выравнивания. Однако разработчикам этот выбор позже сослужил хорошую службу, так как многие западные переводы СПИ публикуются уже разбитыми на фрагменты именно таким образом, что, безусловно, облегчило их подготовку к помещению в корпус. Это закономерно объясняется весом Якобсона в среде западных славистов, а в России предложенное им членение фактически стало стандартом благодаря авторитету А. А. Зализняка после выхода его книги о СПИ [Зализняк 2004].

Сейчас тексты в меню корпуса распределены по пяти категориям: тексты и издания, переводы на современный русский язык, переводы на славянские языки, переводы на другие зарубежные языки, переводы на языки народов России. Это деление, однако, чисто условное и не мешает вызвать сопоставление любого текста из одной категории с любым текстом из другой категории. В планах разработчиков также создание динамического меню, в котором пользователь мог бы выстраивать переводы по времени создания, в алфавитном порядке по фамилии переводчика и т.д.

В параллельном корпусе, коль скоро он параллельный, должны выводиться на экран минимум два текста, в нашем случае это—эталонный (по всей видимости, древнерусский текст памятника) и какой-то другой, избранный пользователем.

В качестве эталонного был избран текст, помещённый в Энциклопедии «Слова о полку Игореве». Такой выбор был сделан благодаря гармоничному учёту составителями Энциклопедии двух разнонаправленных тенденций в подходе к памятнику. С одной стороны, корпусу требовался вариант, максимально аккуратно (в рамках здравого смысла) воспроизводящий Первое издание, с другой,—избавленный от очевидных опечаток и признанных ошибок издателей в чтении древней рукописи. Характеризуя свой вариант, составители Энциклопедии отмечают, что «в его основе текст издания 1800 г., в который внесено лишь минимальное число смысловых и буквенных исправлений: по преимуществу это исправления явных опечаток или недосмотра издателей. Иные чтения, имеющие

различные толкования в науке, оставлены без изменений» [Энциклопедия 1995, с. 8].

Первоначально ресурс задумывался как средство хранения и представления в электронном виде только переводов СПИ. Однако довольно быстро стало ясно, что такая концепция нуждается в уточнении. Ресурс, располагающий переводами СПИ, но исключающий доступ к Екатерининской и Щукинской копиям, разным изданиям (среди которых Первое издание, разумеется, обладает особым статусом сверхважного источника) и реконструкциям памятника, выглядит обеднённым. Вполне логично, что серьёзного исследователя вряд ли мог удовлетворить единственный текст Энциклопедии, такой исследователь непременно захотел бы сравнить данные, представляемые другими релевантными источниками. Поэтому, несмотря на некоторую несогласованность такого подхода с названием ресурса, было принято принципиальное решение о включении в корпус текстов Первого издания, транскрипций существующих списков памятника и наиболее авторитетных реконструкций древнего текста. В будущем, возможно, в корпусе появится и выровненная с остальными текстами прорисовка Екатерининской копии.

Важным для идеологии проекта стало такое качество, как удобство для пользователя (то, что на жаргоне разработчиков называется «user friendly»). Именно оно зачастую оказывалось определяющим при выборе того или иного решения задачи. Открытый формат ресурса—его общедоступность через Интернет—с самого начала ставил разработчика перед тем, что в роли пользователя может оказаться вовсе не специалист-филолог, готовый к преодолению трудностей для получения нужной ему информации, а придирчивый посетитель, которому дополнительные усилия могут показаться излишними и отталкивающими.

Такая позиция имеет несколько принципиальных следствий. Например, представлять информацию на экране необходимо только заведомо имеющимися в компьютере пользователя средствами: загрузка дополнительных шрифтов или плагинов должна быть исключена как вариант решения технических трудностей. В результате обработки запроса на компьютер к пользователю должна поступить простая HTML-страница, по возможности свободная от утяжеляю-

щих скриптов или кодов, неодинаково обрабатываемых разными программами просмотра.

Например, серьёзные трудности возникли с формой представления использовавшейся в рукописях буквы *а-йотированное*, до последнего времени (до версии 5.1) отсутствовавшей в стандарте Unicode, следствием чего, естественно, было и отсутствие начертания этой буквы во всех распространённых шрифтах.

В ходе эксплуатации и пополнения корпуса пришлось отказаться от использования гарнитуры Palatino Lynotype, в своих распространённых версиях не включающей знаков еврейского алфавита (в корпусе должны быть представлены два перевода СПИ на идиш и один перевод на иврит), буквы *юс мальй* и т.д.

Предполагаемое удобство использования также обусловило ряд технических особенностей. Так, запрос к системе осуществляется методом GET. Его отличие от другого распространённого метода (POST) в том, что программа, с помощью которой пользователь просматривает интернет-страничку (юзер-агент) генерирует уникальную ссылку, в которой закодированы все параметры запроса (при POST-запросе адресная строка браузера остаётся пустой), например, такого вида: [http://nevmenandr.net/slovo/slovo.php?fragm=4&str=1&it\[\]=a3&it\[\]=b2&l=ru](http://nevmenandr.net/slovo/slovo.php?fragm=4&str=1&it[]=a3&it[]=b2&l=ru). В нашем случае это язык интерфейса, идентификаторы выбранных переводов, номер фрагмента, необходимость выводить строковые разделители в тексте. По этой ссылке в любой момент можно вызвать именно то, что в данный момент пользователь видит на экране. Аналогичная технология используется во всех распространённых поисковых системах, в том числе Яндекс и Google, характерна она и для Национального корпуса русского языка. Многие опытные пользователи уже знакомы с эффектом, который это даёт: ссылка позволяет сохранять найденную информацию и обмениваться ею с другими пользователями. Для удобства сохранения ссылки на выбранное сопоставление и фрагмент в интерфейсе корпуса также предусмотрено несколько небольших усовершенствований.

В ближайших перспективах развития корпуса—создание поиска по текстовой базе. Вероятно, без поисковой надстройки ресурс и не может с полной правотой называться корпусом, так что поиск—это

одно из приоритетных направлений, в котором будут двигаться разработчики.

Здесь тоже необходимо будет предусмотреть ряд нюансов. Ясно, что поиск должен быть не по всей базе сразу, а дифференцированно по языкам. Вероятно, нет особой необходимости специально разрабатывать поисковую систему для языков, для которых существует только один перевод (как, например, китайский или осетинский). Если пользователь пожелает найти какую-то лексему в этих текстах, он сможет вывести их на экран и отыскать то, что ему нужно, стандартным способом поиска по документу. Иное дело—большой массив текстов, вроде десятков русских переводов или же собрания украинских переводов, количество которых тоже внушительно.

Зачем нужен такой корпус? Предполагается, что в большей степени он окажется полезен специалисту, нежели рядовому читателю, однако принципиальная простота интерфейса и ориентация на дружелюбность к пользователю располагает и к медленному чтению, не ради достижения академических исследовательских целей.

Очевидно, что корпус может оказаться полезен при преподавании русского языка как иностранного. Отчасти из-за этого у корпуса предусмотрен интерфейс на других языках, прежде всего на тех, для которых имеется более одного перевода. Кроме русского это английский, немецкий, французский, украинский, польский, хорватский и идиш. Разумеется, в корпусе представлен довольно специфичный материал, но он вполне может вписаться отдельным блоком в учебную программу курса.

Особый потенциал содержится в корпусе как в хранилище интуитивно близких, «рассказывающих об одном и том же» текстов. В нём легко можно найти иллюстрации разных грамматических и лексических вариантов языкового кодирования для учебных и научных целей.

Естественно, что всё названное дополняет главное—возможность исследования традиции художественного перевода (не только в отношении русского, но и таких языков, на которые СПИ переводилось на протяжении последних двух веков: французский, немецкий, английский) и истории рецепции СПИ.

Надо признать, что сама по себе идея такого параллельного корпуса довольно проста, а техническая реализация не требует осо-

бренных усилий. Однако представляется, что именно такая форма хранения и представления текстов будет наиболее удобна конечному пользователю, кем бы он ни оказался—заинтересованным читателем или исследователем.

ЛИТЕРАТУРА

- Гаспаров 2006—*М. Л. Гаспаров*. О новом переводе «Ада» Данте, выполненном В. Г. Маранцманом // Данте Алигьери. Божественная комедия: Ад. Чистилище. Рай. СПб., 2006. С. 5–8.
- Зализняк 2004—*А. А. Зализняк*. «Слово о полку Игореве»: Взгляд лингвиста. М, 2004.
- Энциклопедия 1995—[Преамбула: Текст «Слова о полку Игореве»] // Энциклопедия «Слова о полку Игореве»: В 5 т. СПб., 1995. Т. 1. С. 8–16.
- Nabokov 1962—*V. V. Nabokov*. Pale Fire. London, 1962.