

НЕКОТОРЫЕ ПРОБЛЕМЫ СОЗДАНИЯ ПАРАЛЛЕЛЬНОГО РУССКО-АНГЛИЙСКОГО КОРПУСА

Е. Г. Андреева

Насколько известно автору, в России до сих пор не существует параллельных компьютерных корпусов, в том числе и англо-русских. Создание такого корпуса требует решения нескольких непростых задач. Первым шагом должно быть определение соотношения оригинальных и переводных текстов: поскольку нас интересуют «переходы в обе стороны», корпус, по существу, должен состоять из двух подкорпусов – англо-русского и русско-английского. По-видимому, в обоих подкорпусах должно быть одинаковое соотношение текстов разных жанров. В настоящее время мы ориентируемся на примерно равную представленность художественной литературы, драмы (как определенного приближения к разговорному языку) и публицистики.

Практическое применение параллельного корпуса распространяется на самые разнообразные области и может представлять весьма значительный интерес для лингвистов, переводчиков, речевых технологов, преподавателей русского и английского языка и других специалистов. Небезынтересны даже такие внешние, казалось бы, параметры, как соотношение числа слов, предложений, абзацев.

Из собственно грамматических проблем, которые лучше всего изучать также на материале репрезентативного корпуса, можно назвать соотношение таких типов конструкций, как активная и пассивная. Данная грамматическая категория характеризует оба исследуемых языка, но из этого вовсе не следует, что мы обнаружим в параллельных текстах взаимно-однозначное соответствие между обсуждаемыми конструкциями оригинала и перевода. Даже если отвлечься от того, что сфера доступности актантов для изменения их синтаксического ранга при переходе к пассиву шире в английском, чем в русском, существует достаточно много случаев, когда актив передается пассивом и наоборот. Уже О.Есперсен отмечал склонность английского к пассиву, но только репрезентативный корпус способен дать надежные данные в сопоставительном плане.

Не менее интересным будет и выяснение того, каковы типичные английские соответствия русских безличных конструкций, а также то, к каким переводческим процедурам прибегают при переводе английских безличных конструкций (с вводящим *it* в частности). Также ожидается, что параллельный русско-английский корпус поможет установить типичные стратегии при переводе разного рода безэквивалентных грамматических явлений, таких, например, как конструкции с определенными/неопределенными артиклями, адекватная передача простых и перфектных времен.

Уже, скорее, для литературоведения и теории коммуникации репрезентативный корпус позволит выявить, насколько и каким образом переводчик реструктурирует текст, чтобы добиться, в разумной степени, эффекта “своего” (в противоположность “чуждому”) текста (как известно, существуют специальные процедуры, способствующие превращению “чуждого” текста в “свой”). Анализ разного рода замен (лексических и синтаксических, замен частей речи и форм слова), приемов компенсации, добавления и опущения, а также возможности антонимического перевода продемонстрирует одновременно реальное соотношение между грамматическими структурами и между лексическими единицами английского и русского языков.

Излагаемые здесь результаты представляют собой самые первые шаги по работе с параллельным корпусом. На данном этапе был обследован массив из четырех текстов: оригинальный - «Alice in Wonderland» Льюиса Кэрролла и три его перевода на русский язык (В.Набокова, Б.Заходера и Н.Демуровой). Данные тексты подверглись процедуре выравнивания, в результате которой каждый из них представлен в виде двуязычного текста. В этом представлении значительно легче сопоставлять фрагменты оригинального и переводного текстов и проводить параллели между ними.

Любопытно, что уже по числу словоупотреблений разные варианты перевода значительно отличаются друг от друга. В наибольшей степени к оригиналу в этом отношении приближен перевод Заходера (34236 по сравнению с 37250 в оригинале), в переводе Демуровой использовано 29777 словоупотреблений, тогда как в набокковском варианте всего лишь 27561. Как можно видеть, число словоупотреблений еще не отражает степень следования перевода за оригиналом: как известно, перевод Заходера в наибольшей степени является «вольным», во многих отношениях являясь фактически пересказом – тем не менее, именно вариант Заходера по «длине» в наибольшей степени приближается к оригиналу (что в принципе могло бы говорить – если бы это вообще было возможно – о близком к пословному переводу).

Полученный результат интересен также тем, что он противоречит общепринятому мнению о большей «многословности» русского языка, которое до некоторой степени подтверждается тем, что переведенные с английского на русский язык тексты деловых документов, как правило, оказываются более объемными, нежели оригинал.

Еще одним, также достаточно формальным, параметром рассматриваемых текстов может служить их структурирование в терминах предложений. Соотношение числа предложений в исходном и переводном тексте может отражать разные явления: предпочтение к использованию простых/сложных предложений, предпочтения конкретного переводчика к использованию определенных трансформаций и т.д.

По техническим причинам мы можем пока говорить лишь о приблизительных результатах. В английском тексте содержится 1601 предложение. И в данном случае наиболее близкие – практически совпадающие – параметры обнаруживаем в тексте Заходера: в нем 1605 предложений. Перевод Демуровой содержит несколько больше таких единиц – 1669, тогда как в варианте Набокова их меньше – 1419. Таким образом, по обоим учтенным нами параметрам перевод Заходера представляется наиболее приближенным к оригиналу, перевод Набокова – наиболее отдаленным и в некотором роде сокращенным (и словоупотреблений и предложений в нем значительно меньше, чем в оригинале).

Таковы первые, самые приблизительные и скромные результаты использования англо-русского мини-корпуса параллельных текстов, но даже они, думается, представляют определенный интерес.