

Национальный корпус в практике  
лингвистических исследований и преподавании языка

Марданшина Р.М.

В процессе развития иноязычной коммуникативной компетенции во всем многообразии ее компонентов особую актуальность на современном этапе приобретает развитие информационной компетенции субъектов образовательного процесса. Информационная компетенция, по нашему мнению, включает владение новыми информационными технологиями, понимание диапазона их применения и умения пользования интернет-ресурсами. Особое место в этом смысле занимают *корпусные технологии*. Обучение языку с помощью корпуса сегодня это огромная область современной лингвистики. По твердому убеждению российского лингвиста В.А. Плунгяна, для овладения языком человеку нужны три вещи, словарь, грамматика и корпус текстов данного языка, потому что словарь и грамматика, в общем-то, бесполезны вне живого пространства, где язык, собственно, и функционирует [2].

Новый тип языковых ресурсов, именуемый национальный корпус (НК), – это информационно-справочная система, основанная на собрании текстов на некотором языке в электронной форме [5]. НК представляет данный язык на определенном этапе (или этапах) его существования во всем многообразии жанров, стилей, территориальных и социальных вариантов и т.п. НК создается специалистами по корпусной лингвистике (быстро развивающейся современной области языкознания) для научных исследований, преподавания филологических дисциплин, составления словарей, учебников, справочных материалов. Сегодня практически все крупные издательства учебной литературы по английскому языку задействуют корпусные данные. Примером может послужить незаменимый в процессе изучения английского языка и переводческой практике толковый словарь *the Longman Dictionary of Contemporary English*, который создан на основе компьютерного корпуса

British National Corpus, что оговаривается в предисловии и фиксируется на обложке словаря.

Большинство крупных языков мира имеют свои НК. Общеизвестным образцом является упомянутый выше Британский национальный корпус, на него ориентированы многие современные корпуса. Среди славянских языков выделяется Чешский национальный корпус, созданный в Карловом университете Праги. Существуют несколько корпусов русского языка, различающихся объемом текстов (их принято считать в количестве словоупотреблений), способом аннотации (соответственно, удобством поискового аппарата) и доступностью. Наибольшим объемом в настоящий момент обладает Национальный корпус русского языка (НКРЯ). Это бесплатный, открытый для всех сайт компании Яндекс, который начал действовать три года назад благодаря усилиям группы ученых-лингвистов из Института русского языка РАН. НКРЯ содержит несколько подкорпусов (около 120 млн. словоупотреблений, планируемый объем – 200 млн.), включая параллельный русско-английский подкорпус текстов, в котором можно найти все переводы для определенного русского или английского слова или словосочетания [1].

НК имеет две важные особенности. Во-первых, он характеризуется представительностью, или сбалансированным составом текстов. Это означает, что корпус содержит по возможности все типы письменных и устных текстов, представленных в данном языке (художественные разных жанров, публицистические, учебные, научные, деловые, разговорные и т.п.), и все эти тексты входят в корпус по возможности пропорционально их доле в языке соответствующего периода. Хорошая представительность достигается только при значительном объеме корпуса (десятки и сотни миллионов словоупотреблений). Во-вторых, корпус снабжен особой дополнительной информацией о свойствах входящих в него текстов, так называемую разметку или аннотацию. Разметка – главная характеристика корпуса, она отличает корпус от простых коллекций текстов или библиотек, в изобилии

представленных в современном интернете. Для составителей НК увлекательность или полезность книги, ее высокие художественные или научные достоинства являются важными, но не первостепенными. НК, в отличие от электронной библиотеки, - это не собрание «интересных» или «полезных» текстов, это собрание текстов, интересных или полезных для изучения языка. Чем богаче и разнообразнее разметка, тем выше научная и учебная ценность корпуса. В Национальном корпусе русского языка в настоящее время используется четыре типа разметки: метатекстовая, морфологическая, акцентная и семантическая [5].

Проведенный анализ ряда исследований по информационным технологиям в гуманитарном образовании [1, 3, 4] позволяет выделить следующие преимущества и особенности работы с НК.

Во-первых, НК позволяет обратиться к современной живой речи, реальному языковому материалу, решая тем самым проблему функциональной грамотности, которая позволяет человеку быть эффективным в жизни.

Во-вторых, корпус предоставляет возможность за считанные минуты получить примеры того или языкового явления в современной речи в статистически релевантном количестве и использовать их при составлении практически любых заданий: по морфологии, синтаксису, лексике и пунктуации. Все больше учебников и учебных программ ориентированы на корпус, при этом процесс составления многих видов заданий и упражнений ускоряется в разы.

В-третьих, частотность того или иного языкового явления может и должна быть учтена при определении объема обязательного образовательного стандарта.

В-четвертых, тексты, имеющиеся в корпусе, можно сортировать, то есть получать примеры не из всех текстов, а только из тех, которые интересуют вас. Например, можно ограничить корпус тем или иным периодом времени (вплоть до конкретного года), определенным типом текстов – рекламными текстами, деловыми письмами или просто выбрать одного или несколько авторов и т.д.

Наличие всех возможных жанров, стилей и тем – это принципиальная установка составителей Корпуса, поэтому он постоянно пополняется. Благодаря этой опции можно задать профессиональную направленность процессу обучения.

Проблема стилистических норм также может быть решена с помощью НК. Например, по предмету Культура делового общения в разделе «Языковые нормы в официально-деловом стиле речи» сообщается, что неправильной является конструкция «обязательства по договору»; рекомендуется использовать выражение «договорные обязательства» [1]. Если обратиться к данным НК, мы получим сорок примеров на конструкцию «обязательства по договору»; «договорные обязательства» встречаются три раза. Стоит ли навязывать нежизнеспособную нормативную рекомендацию, и может ли такая рекомендация возыметь какой-то эффект?

Наконец, национальный корпус – ресурс, использовать который особенно уместно в процессе преподавания языка, особенно при обучении переводчиков. Самый очевидный способ работы с корпусом – самопроверка слушателей, сомневающих в выборе оборота. Потеря ориентации в родном языке – ощущение, знакомое каждому, кто пытался передать в чужеродной для них грамматической и семантической системе даже самые простые конструкции. Обращаясь к корпусу, переводчик может попробовать поискать выражение, которое собирается употребить: отсутствие примеров, скорее всего, укажет на недопустимость их использования. Другая возможность – найти существительное (глагол, прилагательное и т.д.), наиболее часто сочетающиеся с данным словом. Результатом такого поиска, скорее всего, станет клише; именно его, как правило, ищут переводчики публицистических или научных произведений.

Перевод финансовой литературы становится заметно эффективнее благодаря корпусу. Финансовый язык в России находится на стадии становления, поскольку банковская система в ее современном виде и фондовый рынок появились у нас относительно недавно. Участники российского

фондового рынка, как правило, черпают знание об этом институте из книг, написанных в США, где такой рынок давно и хорошо развит. Проблема распространения полученного знания оказывается, тем самым, связанной с вопросом о возможности адаптации англоязычных понятий в русской языковой среде. Переводчику неоднократно приходится решать вопрос об уместности перевода: русский аналог, не получивший еще статус термина, может исказить присущий оригиналу смысл, в то время как оставленное в тексте заимствование, возможно, окажется непонятным. В последнее время было издано несколько качественных англо-русских словарей, посвященных лексике фондовых рынков, однако многие статьи в них, не предлагая перевода, ограничиваются толкованием понятия. Переводчику, в особенности начинающему и не искушенному в специальной терминологии, стоит проверять себя на материале корпуса (точнее, подкорпуса экономической и финансовой лексики). Например, базовое для фондового рынка понятие «market maker». В корпусе «маркетмейкер» встретился всего 13 раз, однако, на фоне 327 прочих «мейкеров» («фильммейкер», «имиджмейкер») заимствование выглядит вполне уместным [1].

Таким образом, корпусные технологии в практике лингвистических исследований и преподавании языков, включая профильно-ориентированное, можно позиционировать как высокоэффективные инновационные дополнения к традиционным образовательным технологиям в силу их эмпирической адекватности, аутентичности, гибкости и адаптации к конкретным задачам и целевым группам пользователей.

#### Литература

1. Добрушина Н.Р. Информационные технологии в гуманитарном образовании: Национальный корпус русского языка / Н.Р. Добрушина, А.И. Левинзон // Вопросы образования. – 2006. - № 4 [Электронный ресурс]. – Режим доступа: [http://vo.hse.ru/arthiv.aspx?catid=252&z=378&t\\_no=379&ob\\_no=394](http://vo.hse.ru/arthiv.aspx?catid=252&z=378&t_no=379&ob_no=394), свободный.

2. ПОЛИТ.РУ\ПУБЛИЧНЫЕ ЛЕКЦИИ \ Почему современная лингвистика должна быть лингвистикой корпусов [Электронный ресурс]. – Режим доступа: <http://polit.ru/lectures/2009/10/23/corpus.html>, свободный.
3. Синклер Д. Предисловие к книге «Как использовать корпуса в преподавании иностранного языка»/ Д. Синклер [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru/corpora-infro.html>, свободный.
4. Соснина Е.П. Корпусная лингвистика и корпусный подход в обучении иностранному языку/ Е.П. Соснина [Электронный ресурс]. – Режим доступа: <http://ling.ulstu.ru/>, свободный.
5. Что такое корпус? Национальный корпус русского языка [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru/corpora-infro.html>, свободный.