

## PECULIARITIES OF TRAINING OF THE FOREIGN STUDENTS-INTERPRETERS WHILE RECEIVING ADDITIONAL EDUCATION

T. V. Samosenkova

*The peculiarities of training of the foreign students studying to get supplementary education with a specialization «An interpreter in the sphere of professional activity» are observed at the following article. The author of the article considers the specifics of teaching such an activity since training of the interpreters is connected with some difficulties.*

*Key words: interpretation, interpreter, professional activity, lexical unit, semantic conformity, without equivalent vocabulary.*

*Samosenkova Tatiana Vladimirovna, Doctor of Pedagogical Sciences, Professor, samosenkova@bsu.edu.ru, Russia, Belgorod, National Scientific Research University Belgorod State University.*

УДК 81'42

## ОБ ИСПОЛЬЗОВАНИИ ДАННЫХ НАЦИОНАЛЬНОГО ЯЗЫКОВОГО КОРПУСА ПРИ ПЕРЕВОДЕ

О.Ю.Семина

*Описываются возможности использования национальных языковых корпусов в переводе; подчеркивается значимость корпусов как дополнительного по отношению к традиционным словарям источника лингвистической информации.*

*Ключевые слова: национальный языковой корпус, перевод*

Национальный языковой корпус можно назвать одним из важнейших нововведений в лингвистике последних десятилетий. Без справочной информации, полученной из национального корпуса, в наши дни сложно представить себе сколько-нибудь серьезное лингвистическое исследование. Обращение к корпусным данным позволяет разнообразить и вывести на принципиально новый уровень процесс изучения иностранных языков, создания учебно-методических материалов и, конечно же, процесс перевода. Вместе с тем, приходится констатировать, что многие практикующие переводчики пренебрегают столь ценным источником лингвистической информации, немного или вообще ничего не зная о национальных языковых корпусах и предлагаемых ими возможностях.

Прежде всего нужно уточнить, что представляет собой национальный языковой корпус. В широком смысле корпус – это любая совокупность текстов, связанных одной характеристикой: авторством, жанром и т. д. [2]. Как корпус текстов иногда может рассматриваться и все информационное пространство сети Интернет [1, с. 11]. В связи с этим нужно признать, что результат запроса в Интернет с помощью одной из поисковых систем

вполне может использоваться для поверхностного изучения сочетаемости или частотности той или иной лексемы или фраземы в исследовательских или переводческих целях. Тем не менее, нужно понимать, что массив данных, в который при этом делается запрос, не контролируется с точки зрения содержания (в том числе, имеет место неограниченное дублирование информации на разных сайтах) и не структурируется с точки зрения жанра.

Что касается языкового корпуса в узком смысле, под ним понимают унифицированный, структурированный и размеченный массив языковых (речевых) данных в электронном виде [1]. Принятое в корпусной лингвистике определение корпуса основывается на четырех основных признаках: расположение корпуса на машинном носителе, стандартизованное представление материала на этом носителе, конечный размер, репрезентативность как результат особого отбора [10]. В. Рыков подчеркивает, что надежные результаты в исследовании может дать только универсальный, или эталонный, корпус текстов [5], каковым и является национальный корпус. Национальный корпус – это собрание текстов в электронной форме, представляющих данный язык на определенном этапе его существования [4]. Система поиска, которой обычно снабжен национальный корпус, позволяет варьировать параметры запроса: выбирать грамматические и семантические характеристики, определять хронологические границы интересующих пользователя текстов. Немаловажно, что многие современные корпуса (в том числе, British National Corpus (далее – BNC), Corpus of Contemporary American English (COCA), Национальный корпус русского языка (НКРЯ)) размечены в соответствии с жанровыми характеристиками текста; таким образом, запрос по тому или иному слову или словосочетанию может быть ограничен конкретным типом дискурса.

Несомненно, ценным будет обращение к корпусному материалу, если требуется прояснить значение слова или словосочетания, не обнаруженного в словарях. Так, выражение *Generation O* не зарегистрировано в словарях, но в COCA оно встречается, причем только в источниках последнего десятилетия. Еще более существенно, однако, что данные национальных языковых корпусов могут эффективно использоваться для уточнения значения и закономерностей функционирования единиц языка, уже зафиксированных в словарях (подробнее об этом см. [6]). Практика показывает, что даже при наличии хороших словарей у начинающих переводчиков нередко возникают сложности, связанные с выбором правильного переводного эквивалента из ряда представленных в словарной статье. Не менее важным аспектом использования корпусных данных в переводе является и прояснение сочетаемости слов в языке перевода. Даже очень авторитетные словари сочетаемости не могут включать в себя всех лексем языка или всех возможных коллокатов той или иной лексемы. Они дают лишь общие рекомендации по комбинированию лексических единиц и не всегда содержат достаточно полную и соответствующую современному состоянию языка инфор-

мацию. Море контекстов, составляющих национальный корпус, дает гораздо больше шансов объективно оценить реальные закономерности сочетаемости слова в современной речи. Поясним вышесказанное на примере слова «общежитие». В словаре АБВУД Lingvo 12 (а на сегодняшний день это один из наиболее полных и широко используемых при переводе текстов общей тематики словарей) для его перевода предлагаются варианты residence hall; hostel (с пометой *брит.*); dormitory (с пометой *амер.*). Сделав запрос на каждый из трех вариантов в BNC и COCA, получаем следующие результаты по употребительности: dormitory - BNC 153, COCA 978; hostel – BNC 530, COCA 464; residence hall BNC 0, COCA 161. Анализируя расхождения в частотности каждой единицы в двух корпусах, сразу отметим, что

1) объем корпуса COCA в 4,5 раза превышает объем BNC; 2) в состав COCA входят тексты, датируемые разными годами вплоть до настоящего времени, в то время как состав BNC ограничен текстами периода начала 1990-х гг. Оба этих факта, безусловно, сказываются на достоверности данных BNC с точки зрения современного состояния языка, однако не умаляют достоинств корпуса в целом как эталонного и тщательно сбалансированного с точки зрения жанрового состава. Учитывая это, для получения наиболее точного представления о положении дел в современном английском языке, данные обоих корпусов рекомендуется использовать в сочетании (при условии отсутствия заинтересованности конкретно в том или ином варианте английского языка). Как видно из приведенных выше данных, в Британском национальном корпусе наиболее частотной из трех единиц является hostel, в американском – dormitory, что вполне соотносится с пометами в словаре. Но так как конкордансы (конкорданс – список всех употреблений заданного языкового выражения в контексте) слова dormitory в английском варианте и слов hostel и residence hall в американском варианте языка также оказались довольно объемными, возникает вопрос о том, есть ли существенные различия в употреблении этих двух лексем. Для этого воспользуемся функцией создания списка слов, сочетающихся с заданным словом в найденных контекстах (опция COLLOCATES в строке поиска). Согласно результатам поиска в BNC, наиболее часто со словом dormitory в британском варианте английского языка используются следующие лексемы: suburb (10), bed (7), town (7), accommodation, rooms, boys, girls, village (по 4). В то же время, более популярное слово hostel наиболее часто используется в сочетании со словами youth (73), bail (16), homeless (14), salvation (14), accommodation (14). Как показывает анализ контекстов, в которых встречаются эти слова в BNC, для обозначения студенческого общежития больше подойдет все же dormitory, в то время как hostel более часто используется для обозначения чего-то вроде недорогой гостиницы (ср. одно из наиболее распространенных сочетаний – bail hostel). Примерно такая же ситуация наблюдается и в COCA – слово hostel чаще всего комбинируется со словами youth, dwellers, township, workers, residents, elder, а в списке наиболее ча-

стотных коллокатов для слова dormitory (и для образованного от него dorm) первые места занимают room, college, university, living, girls, campus, freshman. Что касается варианта residence hall, он встречается исключительно в американском национальном корпусе; более того, 161 употребление в СОСА объясняется тем, что эта единица много раз повторяется в пределах того или иного текста.

Неплохим подспорьем в выборе эквивалента при переводе является и параллельный корпус, например, корпус текстов на английском языке с параллельным переводом, размещенный в свободном доступе на сайте проекта Национальный корпус русского языка (<http://ruscorpora.ru/search-parallel.html>). Параллельный корпус дает возможность обратиться к опыту других переводчиков, его основу составляют уже опубликованные (и во многих случаях ставшие классическими) переводы как англоязычных, так и русскоязычных произведений, поэтому полученная информация может стать ключевым фактором при выборе эквивалента. В качестве примера проанализируем запрос в параллельный корпус для слова «сало», передача которого на английский язык может вызвать затруднение у переводчика. Словари предлагают для его перевода следующие эквиваленты: (*животный жир*) fat; (*нутряное*) suet; (*топлёное свиное*) lard; (*топлёное для свечей*) tallow. При этом игнорируется тот факт, что в нашей стране, употребляя данное слово, часто подразумевают соленое сало, нарезанное кусками и употребляемое в пищу в сыром виде. Сделав запрос в параллельный англо-русский корпус, получаем целый ряд контекстов, в которых встречается эта лексема, с параллельными англоязычными вариантами. Изучение этого списка показывает, что для перевода слова «сало» в тех случаях, когда этот продукт употребляется в пищу соленым, сырым, нарезанным кусками, используется эквивалент bacon. Вместе с тем, контексты из англоязычных произведений, переведенных на русский язык, наглядно демонстрируют существующее между этими двумя словами принципиальное культурологическое различие: bacon носители английского языка преимущественно используют в жареном виде, употребление его в пищу в сыром виде остается чем-то «экстремальным». Например: *Bacon could be eaten raw on a pinch, and flour, stirred in hot water, could keep men going.* – Сало в крайнем случае можно есть и сырым, а болтушка из муки и горячей воды тоже поддерживает силы (НКРЯ, Jack London. *The Burning Daylight* (1910)). Следует признать, что параллельный англо-русский подкорпус в НКРЯ пока не отличается очень большим объемом (на 14 января 2014 г. это всего 54 млн словоупотреблений), в связи с чем получаемые данные не могут отражать в полной мере всего многообразие существующих вариантов перевода. Вместе с тем, эти данные подтверждают, что многие единицы требуют при переводе гораздо большего внимания к контексту, чем можно было бы предположить, исходя из словарных статей.

Современные языковые корпуса подразделяются на подкорпусы, ограниченные жанром текста (в COCA выделяются следующие жанры: SPOKEN, FICTION, MAGAZINE, NEWSPAPER, ACADEMIC). Поскольку поисковая система корпуса делает возможным проанализировать употребительность слова или фразы в каждом из жанров, это позволяет получить более полное представление о стилистических характеристиках языковых единиц и может оказать неоценимую помощь переводчику в тех случаях, когда словари не отражают стилистическую заряженность той или иной единицы. Например, лексем *well-nigh* в большинстве толковых словарей не имеет стилистических помет; в англо-русских словарях (БАРС, Lingvo Universal) переводится как «почти, приблизительно», также без указания на стилистические ограничения. Вместе с тем, запрос в подкорпусы разных жанров показывает следующие результаты: SPOKEN – 1 контекст, FICTION – 21, MAGAZINE – 13, NEWSPAPER – 5, ACADEMIC – 24, что свидетельствует о явной принадлежности лексемы к числу слов высокого, книжного стиля и неприемлемости для разговорного языка.

Из сказанного выше очевидно, что при умелом использовании корпусные данные могут применяться переводчиком для уточнения значения слова и его комбинаторики; для выбора наиболее адекватного в данном контексте переводческого эквивалента; для определения жанрово-стилистических особенностей функционирования той или иной единицы языка – и это лишь некоторые из многочисленных путей применения национального корпуса в качестве источника справочной информации при переводе.

### Список литературы

1. Захаров В.П. Веб-пространство как языковой корпус. Электрон. изд. URL: <http://www.dialog21.ru/Archive/2005/Zakharov%20V/ZakharovV.htm> (дата обращения: 04.02.2014).
2. Клименко С.В.; Рыков, В.В. Корпус текстов как принцип самоорганизации предметной области // Труды конференции Диалог-2000. Электрон. изд. URL: <http://rykov-cl.narod.ru/t.html> (дата обращения: 04.02.2014)
3. Национальный корпус русского языка 2003-2012. Электрон. изд. URL: <http://www.ruscorpora.ru> (дата обращения: 22.01.2014)
4. Подлесская В.И. Современные компьютерные методы в изучении и преподавании лингвистических дисциплин: корпусная лингвистика. Электрон. изд.: <http://www.rsuh.ru/article.html?id=398> (дата обращения: 20.01.2014).
5. Рыков В.В. Корпусная лингвистика и лексикография – проблема репрезентативности // Труды конференции по лексикографии. Гродно,

2002. Электрон. изд. URL: <http://rykov-cl.narod.ru/t32.html> (дата обращения: 20.01.2014).

6. Семина О.Ю. Верификация данных переводных словарей с помощью языковых корпусов // Известия ТулГУ. Гуманитарные науки. 2012. Вып. 3. С. 572-576.

7. АБВУ Lingvo 12.0: Многоязычный электронный словарь – М.: АБВУ Software House, 2007. 1 эл. опт. диск (CD-ROM). Загл. с экрана.

8. Davies, Mark. (2008-) The Corpus of Contemporary American English: 450 million words, 1990-present. Available online at <http://corpus.byu.edu/coca/> (дата обращения: 02.02.2014).

9. Davies Mark. (2004-) BYU-BNC. Based on the British National Corpus from Oxford University Press. Available online at <http://corpus.byu.edu/bnc/> (дата обращения: 02.02.2014).

10. McEnery T.; Wilson A. Corpus Linguistics. Edinburgh: Edinburgh University Press, 1999. Электрон. изд. URL:: <http://bowland-files.lancs.ac.uk/monkey/ihe/linguistics/contents.htm> (дата обращения: 04.02.2014).

11. Renouf A. WebCorp: providing a renewable energy source for corpus linguistics // ICAME 2001 Future Challenges in Corpus Linguistics. – Louvain-la-Neuve, 2001. Электрон. изд. URL: <http://www.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/Events/icamepr.htm> (дата обращения: 20.01.2014).

*Семина Ольга Юрьевна, канд. филол. наук, доц., [olga@rambler.ru](mailto:olga@rambler.ru), Россия, Тула, Тульский государственный университет.*

#### ON USING NATIONAL LANGUAGE CORPUS DATA IN TRANSLATION

*O.Yu. Syomina*

*The article dwells upon the possibility of using national language corpora in translation; the author emphasizes the significance of national corpora as a source of linguistic information to be used as complementary to dictionaries.*

*Key words: national language corpus, translation.*

*Syomina Olga Yurievna, PhD (Philological Sciences), Associate Professor, [olga@rambler.ru](mailto:olga@rambler.ru), Russia, Tula, Tula State University*