

Ольга Юрьевна САВИНА —
ст. преподаватель
кафедры немецкой филологии
Тюменского государственного университета
osawina@mail.ru

УДК 811.112.2

**О НЕКОТОРЫХ ВОЗМОЖНОСТЯХ ИСПОЛЬЗОВАНИЯ
КУЛЬТУРНОСПЕЦИФИЧЕСКИХ ЛИНГВИСТИЧЕСКИХ
ИНФОРМАЦИОННЫХ РЕСУРСОВ ПРИ РЕШЕНИИ
ЭКСТРАЛИНГВИСТИЧЕСКИХ ЗАДАЧ**

**SOME POSSIBILITIES OF USING THE SPECIFIC CULTURAL
LINGUISTIC INFORMATION RESOURCES IN SOLVING
THE EXTRALINGUISTIC PROBLEMS**

АННОТАЦИЯ. В статье представлены основные типы и важнейшие характеристики текстового корпуса как одного из видов лингвистических информационных ресурсов, приведены примеры культурноспецифических корпусов текстов и показан вариант использования тематического корпуса для решения экстралингвистической задачи.

SUMMARY. The article presents the main types and the essential characteristics of a text corpus as one of the types of linguistic information resources, contains examples of the specific cultural text corpus and shows the option of using a thematic corpus to deal with extralinguistic tasks.

КЛЮЧЕВЫЕ СЛОВА. Лингвистические информационные ресурсы, лингвистический корпус, «Берлинский корпус, посвященный объединению Германии», роль России в процессе объединения Германии.

KEY WORDS. Linguistic Information Resources, Linguistic Corpus, «Berliner Wendekorpus», the role of Russia in the process of Germany reunification.

Современные лингвистические исследования редко обходятся без привлечения лингвистических информационных ресурсов. Среди них важное место занимают текстовые корпуса. Функционирование одного из них, «Берлинского корпуса, посвященного объединению Германии», рассматривается в данной статье. Материал корпуса помогает решить экстралингвистический вопрос об оценке роли России в процессе объединения Германии немецким обывателем.

Информационные ресурсы являются сегодня неотъемлемым инструментом при решении лингвистических задач разного характера. А.В. Зубов определяет лингвистические информационные ресурсы как «множество определенным образом организованных речевых и языковых данных, находящихся на машинных носителях информации и используемых в различных сферах практической деятельности» [1; 156].

Разнообразие принципов создания и функционирования подобных ресурсов привело к формированию следующих их видов [2]: 1) корпуса: корпуса текстов и речевые корпуса; 2) лингвистические базы данных: грамматические ресурсы, лексико-семантические ресурсы, морфологические базы данных, семантико-синтаксические ресурсы; 3) словари и тезаурусы.

А.В. Зубов и И.И. Зубова выделяют: 1) письменный лексикон (конкорданс, контекстологический словарь, энциклопедия и тезаурус), 2) терминологический словарь, 3) письменные текстовые массивы и 4) фонетические ресурсы [1; 157-162].

Одним из наиболее активно развивающихся направлений лингвистических информационных ресурсов является на сегодняшний день корпусная лингвистика. Ярким свидетельством этого могут служить многочисленные промежуточные и окончательные результаты исследований, опубликованные в разные годы в рамках Международной конференции по компьютерной лингвистике и интеллектуальным технологиям «Диалог» [3].

Вслед за В.П. Захаровым определим лингвистический корпус как «большой, представленный в электронном виде, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач» [4; 3].

Устоявшейся типологии корпусов на данный момент еще не существует, разные исследователи приводят следующие возможные типы корпусов* [1], [4], [5]: по типу данных (письменные и устные, в т.ч. звуковые), по языку (корпусы национальных языков, одно- и многоязычные), по «параллельности» (корпусы параллельных текстов на языке-оригинале и языке-переводе, в т.ч. корпусы устных переводов), по специфичности (литературные, диалектные, разговорные, терминологические, поэтические и др.), по жанру (литературные, фольклорные, драматургические, публицистические и др.), по доступности (закрытые, свободного доступа или коммерческие), по назначению (исследовательские, предназначенные для изучения языка, и иллюстративные, содержащие отобранные примеры, подтверждающие языковые и речевые факты), по наличию разметки и ее характеру (морфологические, синтаксические, семантические, просодические и др.), по объему текстов (н-р, полнотекстовые), по хронологии (синхронические и диахронические), по динамичности (статические и динамические, или мониторные (в зависимости от отражаемого временного промежутка)).

Важнейшая черта, отличающая корпус от библиотеки текстов, — разметка, т.е. структурное и лингвистическое метаописание единиц хранения корпуса. Основными характеристиками корпуса принято считать:

1) репрезентативность, т.е. способность отражать все свойства проблемной области, важные для определенного лингвистического исследования, в пропорции, определяемой частотой явления;

2) полноту, т.е. учет релевантных явлений, даже если при этом невозможно пропорциональное сужение проблемной области (т.е. нарушается репрезентативность);

3) экономичность, т.е. корпус должен существенно отличаться по объему от генеральной совокупности текстов (но не в ущерб репрезентативности);

4) структуризацию материала, т.е. единицы хранения должны быть охарактеризованы по понятным и необходимым для пользователя параметрам;

5) комплексную компьютерную поддержку, т.е. наличие т.н. корпусных менеджеров, позволяющих получать из корпуса необходимую информацию [5; 118-119].

Все ученые, создающие и/или исследующие текстовые корпуса, подчеркивают тот факт, что подобные массивы текстов могут служить источником ин-

* Даже форму множественного числа термина «корпус» трудно считать устоявшейся: «корпуса» и «корпусы» одинаково часто встречаются в публикациях разных исследователей, подробнее см. [4; 14].

формации для решения конкретных лингвистических задач и, таким образом, обеспечивать надежные лингвистические выводы.

Следует отметить, что часть корпусов, например, динамические, или мониторинговые, т.е. постоянно пополняющиеся, могут адекватно отражать не только изменения в языковой системе, но и развитие политических, культурных и других характеристик социума.

В качестве примера культурноспецифических лингвистических ресурсов можно привести коллекцию тематических корпусов текстов, так или иначе отражающих социальные проблемы разных периодов истории Германии, собранных в *Архиве устного немецкого языка* (Archiv für gesprochenes Deutsch) [6] — крупнейшем в мире речевом корпусе на немецком языке [7]. Часть этого архива доступна в сети через *Банк данных устного немецкого языка* (Datenbank gesprochenes Deutsch) [8]. Некоторые из них посвящены территориальным различиям немецкой речи как внутри Германии, так и за ее пределами: *диалекты района Böblingen; диалекты юго-западной Германии; диалекты района Vorarlberg; диалекты района Schwarzwald; Литературное немецкое произношение; немецкие диалекты в Бразилии; язык немецких эмигрантов в Израиле; немецкие диалекты в Румынии; славянские диалекты в Рурской области* и др.

Множество корпусов составлены с точки зрения социокультурных различий немецкого языка: *диалекты ГДР, эмигранты; Берлин после объединения Германии; язык детей* и даже *конфликты между мамами и дочерьми-подростками* и многие другие.

В преддверии двадцатилетнего юбилея объединения Германии, которое состоялось 9 ноября 1989 г., наибольший интерес лингвистов вновь вызывает отражение этих исторических событий в речи носителей немецкого языка. В связи с этим следует более подробно представить Берлинский корпус, посвященный объединению Германии — устный одноязычный аннотированный исследовательский корпус [9].

Данный речевой корпус создан Свободным берлинским университетом, в частности, Институтом немецкой и нидерландской филологии. Руководители проекта — профессор Норберт Диттмар и Урсула Бредель. Все записи, собранные в проекте, были созданы в Берлине с 1992 по 1996 год. Большая часть записей была сделана 9 ноября 1993 года. Информантами стали 32 жителя Восточного Берлина и 25 жителей Западного Берлина. Возраст участников — от 19 до 55 лет, из них — 28 мужчин и 29 женщин. Продолжительность отдельных высказываний составляет от 5 до 90 минут. Общая продолжительность аудиозаписи — 27 часов 41 минута. Таким образом, в проекте собрано 51 интервью (некоторые из них — с двумя информантами), которые практически в полном объеме оцифрованы, транскрипции доступны для исследования, также как и аудиозаписи. Корпус оснащен экстралингвистической метаразметкой (информация об авторах высказываний (пол, возраст, профессия, место проживания) и о ситуациях, в которых состоялись интервью). Транскрипции текстов снабжены просодической (метки ударения и интонации) и дискурсной разметкой (обозначения пауз, повторов, оговорок и т.д.)

Содержанием высказываний стали свободные интервью, описывающие события 9 ноября 1989 г. и личные события в жизни участников, так или иначе связанные с объединением Германии. При этом участники рассказывали о своих жизненных обстоятельствах, мечтах и заботах, и об актуальной житейской ситуации в момент записи (о работе, жилье, покупках, путешествиях, социальном климате, о воспитании, образовании и т.д.).

Настоящий корпус формировался с целью создания коллекции личного опыта участия в социальном переломе, известном под названием «Разрушение стены 1989 года» («Mauerfall 1989»), т.е. для документации коллективной памяти. Предполагалось, что на основании «переломных» дискурсов могут быть разработаны модели устной речи, типичной для жителей Восточного и Западного Берлина. При этом авторы корпуса руководствовались гипотезой о том, что решающие необратимые общественные изменения обязательно влекут за собой языковую нестабильность отдельных представителей социума и целых общественных групп [10].

Результаты проекта были обобщены его авторами в книге «Die Sprachmauer. Die Verarbeitung der Wende und ihre Folgen in Gesprachen mit Ost- und WestberlinerInnen», изданной в год десятилетнего юбилея указанных событий — в 1999 году [11].

В 1998 г. все записи были оцифрованы в Институте немецкого языка в Маннхайме, так что практически весь корпус находится сейчас в свободном доступе в интернете, например, через *Цифровой словарь немецкого языка 20 века DWDS (Digitales Worterbuch der deutschen Sprache)* [12].

Воспользовавшись представленным речевым корпусом, мы решили проверить гипотезу о том, что немцы отмечают существенную роль России в процессе объединения Германии, возможно, расценивая ее либо положительно, либо отрицательно. При этом нам хотелось увидеть, есть ли различия в оценке роли России с точки зрения жителей Западного и Восточного Берлина.

Для проверки сформулированной гипотезы из «Берлинского корпуса» были отобраны все контексты, в которых встречалось любое слово, содержащее корень *russ*. Таких слов оказалось четыре: *Rußland* (5 словоупотреблений), *Russen* (3 словоупотребления), *russisch* (3 словоупотребления) и *Russischlehrer* (1 словоупотребление). Полная выборка контекстов приведена в табл. 1.

Таблица 1

**Выборка контекстов, включающих леммы с корнем *russ*
(на основании речевого корпуса *Berliner Wendekorpus*)**

Лемма	Информант: псевдоним (возраст), дата записи и продол- жительность интервью	Место проживания	Контекст словоупотребления*
<i>Rußland</i>	Yvonne (53), 13.11.1993 22 мин. 37 сек.	Покинула Восточный Берлин до 1989 г., сейчас живет в Западном Берлине.	und als dann die mauer war^ ham-wa uns so'fort hier gesehen die kam hierher^ wir habm groe fete gefeiert und das is dann eingeschlafn__ ja weil also sie ham wenich zeit gehabt^ der mann is denn nach ruland gegang hat da noch weiter gearbeitet^ zum beispiel aber  wir ham-se dann auch noch besucht

* Примеры приведены в том виде, как они затранскрибированы в «Берлинском корпусе» (т.н. нормализованный вариант).

Продолжение табл. 1.

	Karin (29), 02.12.1993 38 мин. 24 сек.	Восточный Берлин	letztendlich sagen auch wenn ich mit diesem oder jenem anecke und mir gehts persönlich nich schlecht^ sagen immer noch und passiert is es ja im letzten sommer^ aber du hast doch einen guten schnitt gemacht^ mit diesem system guck dir <i>rußland</i> an rumdnien oder sonstwas vergessen aber völlig daß da auch moral^ ürgendwie eine rolle spielt
	Alfred (57), 18.01.1995 28 мин. 01 сек.	Западный Берлин	wir wolln ja das deutschland wieder eins is und dis is ja nu och eins und dit war find ick war doch janz gut das es schnell eins jeworden is wenn man sich mal überlegt wenn vielleicht in <i>rußland</i> ne umwälzung gekommen wär ja^ und die hättn drüben in <i>rußland</i> alles dicht jemacht denn wär die wiedervereinigung gar nich gekommen
	Piet (37), 12.03.1996 29 мин. 18 сек.	Западный Берлин	und also wie ham leute aus jugoslawien aus polen aus <i>rußland</i>
Russen	Micha (33), 28.04.1993 1 час 24 мин. 08 сек.	Восточный Берлин	da gabs nur eene zeitung bloß wer ebend weiter sich wissen beschaffen wollte sone typen wie sie und wie ich der hat sich ebend weil die <i>russen</i> ebend schneller waren als wir durch gorbatschow und so weiter hat man sich ebend den verbotenen sputnik organisiert oder die neue zeit ick weeß nich ob se die kennen
	Alfred (57), 18.01.1995 28 мин. 01 сек.	Западный Берлин	ja und darum hat der kohl ja dit richtige denn wenn wir eins sind dann sind wir stark und dann wird der <i>russe</i> niemals wieder sagn also wir eh wolln da wieder rein oder wir ham noch truppen drin und dat war von dem kohl nach meiner meinung dit beste wat er gemacht hat
	Jens (52), 9.11.1995 1 час 09 мин. 16 сек.	Западный Берлин	von den mathematikern damals da war n kongreß in versailles und da warn auch viele ausm ostblock <i>russen</i> und sowas und die ham manchmal warn paar große leute bei ham gute vorträge gehalten und bei andern hatte man wirklich das gefühl die mußten irgendnen kleinen vortrag halten damit se da nun hinkonnten

Окончание табл. 1

<i>russisch</i>	Ведущий интервью 1, 11.1993 46 мин. 34 сек.	нет информации	wir trafen uns zufällig anner rezeption^ er hatte sich erkundigt und sprach englisch und ick konnt ja nun keen englisch mit meim <i>russisch</i> bin ich da nich weiter gekomm und der erkun- digte sich und ich fragte natürlich gleich so nach den preisen und so man mußte ja doch ma n bißchen gucken
	Ведущий интервью 2, 9.11.1995 1 час 09 мин. 16 сек.	нет информации	da nee die mochten ja aber <i>russich</i> nich__ also bei uns war eher eng- lisch inn^ aber <i>russisch</i> war nich inn
<i>Rus- sisch- lehrer</i>	Kira (36), 10.11.1993 1 час 03 мин. 47 сек.	Восточный Берлин	also ick hatte ümma noch die ruhe weg neununachzich übahaupt__ weil ick dachte also hast-e aba wa bist keen <i>russischlehra</i> __ bist keen phülloso:ph nüscht ja: nüscht nüscht du bist

Анализ словоупотреблений показывает, что лишь немногие из информантов (9 из 57 опрошенных), принявших участие в «Берлинском корпусе», упоминали Россию или что-либо, с нею связанное (в общей сложности лишь 12 словоупотреблений*). При этом из девяти информантов четверо проживают (или выросли) в Восточном Берлине (далее ВБ), трое — в Западном Берлине (далее ЗБ), об авторах двух высказываний информации нет.

Контексты употребления лемм с корнем *russ* можно большей частью интерпретировать как нейтральные. Результаты анализа представлены в табл. 2.

Таблица 2

**Интерпретация контекстов употребления лемм с корнем *russ*
(на основании речевого корпуса Berliner Wendekorpus)**

Лемма	Информант		Число словоупотреблений с нейтральным отношением к России	Число словоупотреблений с положительной оценкой роли России в объединении Германии	Число словоупотреблений с отрицательной оценкой роли России в объединении Германии
	ЗБ	ВБ			
<i>Rußland</i>	Yvonne		1		
	Karin		1		
	Alfred			2	
	Piet		1		
<i>Russen</i>	Micha			1	
	Alfred				1
	Jens		1		
<i>russisch</i>	ведущий интервью 1		1		
	ведущий интервью 2				2
<i>Russisch-lehrer</i>	Kira		1		

* У некоторых информантов искомая лемма встретилась дважды — в одном и том же или в разных контекстах, см. табл. 1.

Одно из высказываний (Micha, ВБ) можно интерпретировать как положительную оценку роли России в объединении Германии: «... weil die *Russen* eben schneller waren als wir, durch Gorbatschow...» («потому что русские оказались быстрее нас, благодаря Горбачеву»).

Интересно, что один из информантов (Alfred, ЗБ) в двух разных контекстах дает разную оценку роли России: «wir wolln ja das deutschland wieder eins is und dis is ja nu och eins und dit war find ick war doch janz gut das es schnell eins geworden is wenn man sich mal überlegt wenn vielleicht in *rußland* ne umwälzung gekommen wär ja^ und die hättn drüben in *rußland* alles dicht jemacht denn wär die wiedervereinigung gar nich gekommen» («мы же хотим, чтобы Германия снова стала единой, и теперь она стала единой, и мне кажется, это очень хорошо, что она быстро объединилась; а если подумать, если бы в России произошел какой-то переворот, и Россия бы закрыла границы, тогда объединения бы вообще не было»).

В другом контексте этого же информанта явно звучит отрицательная оценка агрессивной позиции России: «... dann wird der *Russe* niemals wieder sagen, also wir wolln da wieder rein oder wir ham noch Truppen drin ...» («... тогда русские никогда уже не смогут сказать, мы снова хотим ввести свои войска или наши воинские подразделения еще на вашей территории»).

Еще один информант косвенно указывает на степень непопулярности всего русского: «... also bei uns war eher Englisch in, aber *Russisch* war nicht in ...» («у нас в моде был скорее английский, русский был не в моде»). Но в данном случае невозможно определить принадлежность автора высказывания к Западному или Восточному Берлину, поскольку сделано оно ведущим интервью, информации о котором корпус, к сожалению, не предоставляет.

Таким образом, на основании проведенного анализа невозможно с уверенностью утверждать, что восточные и западные немцы по-разному оценивали роль России в объединении Германии. Более того, относительно небольшой объем выборки с искомой леммой *russ* свидетельствует о том, что жители как Западного, так и Восточного Берлина, оценивая исторические события 1989 г. с точки зрения рядового гражданина Германии и жителя Берлина, не придавали особого значения вопросу о роли России в этих событиях.

Очевидно, что представленный здесь анализ может быть лишь частным примером функционирования культурноспецифического речевого корпуса. Возможности его использования гораздо шире и они ждут своих исследователей, для того чтобы в полной мере реализовать потенциал лингвокультурной оценки общественных событий.

СПИСОК ЛИТЕРАТУРЫ

1. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике: Учебное пособие. М.: Academia, 2004. 208 с.
2. Загорулько Ю.А., Боровикова О.И., Загорулько Г.Б., Кононенко И.С., Соколова Е.Г. Разработка портала знаний по компьютерной лингвистике. http://www.gaai.org/cai-08/files/cai-08_exhibition_24.doc (от 15.05.2009).
3. Международная конференция по компьютерной лингвистике. <http://www.dialog-21.ru/conference/> (от 04.10.2009).
4. Захаров В.П. Корпусная лингвистика: Учебно-метод. пособие. СПб.: Санкт-Петербургский государственный университет, 2005. 48 с.

5. Баранов А.Н. Введение в прикладную лингвистику. М.: Издательство ЛКИ, 2007. 360 с.
6. Archiv für gesprochenes Deutsch. <http://agd.ids-mannheim.de> (от 26.05.2009).
7. Fiehler, R., Wagener, P. Die Datenbank Gesprochenes Deutsch (DGD). <http://www.gespraechsforschung-ozs.de/heft2005/px-fiehler.pdf/> (от 15.05.2009).
8. Datenbank gesprochenes Deutsch. <http://dsav-wiss.ids-mannheim.de> (от 26.05.2009).
9. Berliner Wendekorpus. <http://agd.ids-mannheim.de/html/korpora/korpus-bw.shtml> (от 04.10.2009).
10. Dittmar, N. Informationen zu meinen Korpora. http://userpage.fu-berlin.de/~nordit/HP/Ditt_Korpora.html (от 20.09.2005).
11. Dittmar, N., Bredel, U. Die Sprachmauer. Die Verarbeitung der Wende und ihre Folgen in Gesprächen mit Ost- und WestberlinerInnen. Berlin: Weidler Buchverlag, 1999. 208 S.
12. Digitales Wörterbuch der deutschen Sprache DWDS. <http://www.dwds.de> (от 26.05.2009).

*Андрей Борисович КУТУЗОВ —
ст. преподаватель кафедры перевода и переводоведения
Тюменского государственного университета,
кандидат филологических наук
akutuzov72@gmail.com*

УДК 81'255.4

ЧАСТОТНЫЕ ХАРАКТЕРИСТИКИ ЛЕКСИКИ КАК ИНВАРИАНТ ПЕРЕВОДА (НА МАТЕРИАЛЕ ПЕРЕВОДА РОМАНА К. ВОННЕГУТА «БОЙНЯ НОМЕР ПЯТЬ, ИЛИ КРЕСТОВЫЙ ПОХОД ДЕТЕЙ»)*

LEXICAL FREQUENCIES AS TRANSLATIONAL INVARIANCE (BASED ON RUSSIAN TRANSLATION OF K. VONNEGUT'S SLAUGHTERHOUSE-FIVE OR THE CHILDREN'S CRUSADE)

АННОТАЦИЯ. На основе лексико-статистического анализа оригинала и русского перевода романа К. Воннегута Slaughterhouse-Five была выявлена линейная корреляция между динамикой изменения соотношения числа словоформ и словоупотреблений в оригинале и в переводе. Автор делает предположение о том, что эта динамика является инвариантом перевода.

SUMMARY. The article provides lexical statistical analysis of K. Vonnegut's Slaughterhouse-Five and its Russian translation. It is found out that there exists a linear correlation between the speed of types and tokens ratio change in the source and target texts. The author hypothesizes that this dynamics is a translational invariance.

КЛЮЧЕВЫЕ СЛОВА. Переводоведение, корпус, лингвостатистика, словоформа, словоупотребление, переводческий инвариант.

KEY WORDS. Translation studies, corpora, computational linguistics, type, token, translational invariance.

Переводоведение всегда испытывало некоторые трудности с определением своего предмета. Тем более неоднозначным является вопрос о методах этой науки. Как можно объективно и надежно оценить адекватность или эквивалентность перевода? Есть ли в процессе перевода некие четкие зависимости, влия-

* Работа выполнена при поддержке Федерального агентства по науке и инновациям (государственный контракт № 02.740.11.0594).