

УДК 811.111

## КОРПУС ПАРАЛЛЕЛЬНЫХ ТЕКСТОВ В ОБУЧЕНИИ ИНОСТРАННОМУ ЯЗЫКУ

© **Анастасия Александровна КОКОРЕВА**

Тамбовский государственный университет им. Г.Р. Державина, г. Тамбов,  
Российская Федерация, аспирант, кафедра лингвистики и лингводидактики,  
e-mail: kokoreva81@yandex.ru

Рассмотрен вопрос использования корпуса параллельных текстов в обучении иностранному языку. Дается определение термину «корпус параллельных текстов», рассматриваются дидактические свойства и методические функции корпуса параллельных текстов, предлагается типология корпусов параллельных текстов.

*Ключевые слова:* лингвистический корпус; корпус параллельных текстов; корпусная лингвистика.

Стремительное развитие новых информационных и коммуникационных технологий способствовало появлению ряда методических работ, в которых авторами разрабатывались методики обучения видам речевой деятельности и формирования языковых навыков речи обучающихся на основе современных ИКТ [1–7]. Лингвистический корпус является одной из технологий, на основе которой можно формировать лексические и грамматические навыки речи учащихся и студентов.

Отметим, что история использования лингвистического корпуса в преподавательской и исследовательской работе уходит корнями в 60-е гг. XX в., когда в Брауновском университете (США) исследователи У. Френсис и Г. Кучера создали первый в мире лингвистический корпус. Их лингвистический корпус включал 500 печатных произведений, каждое из которых по объему было не более 2000 слов. Все тексты принадлежали наиболее распространенным жанрам печатной прозы, написанным и опубликованным в США. Отобранные для корпуса тексты должны были представлять 15 жанров, из которых было сделано от 6 до 80 выборок:

- 1) пресса: репортаж;
- 2) пресса: передовица;
- 3) пресса: обзоры;
- 4) религиозные тексты;
- 5) навыки, занятия, хобби;
- 6) научно-популярная литература;
- 7) беллетристика, биографии, эссе;

8) разное (правительственные документы, отчеты предприятий, промышленные отчеты, каталоги колледжей);

9) научные сочинения;

10) художественная литература;

11) мистика и детективы;

12) научная проза;

13) приключенческая литература и вестерны;

14) любовные романы;

15) юмористические произведения.

Именно в 60-е гг. XX в. У. Френсис впервые использовал термин «корпус», под которым понималось «совокупность наиболее представительных текстов для лингвистического анализа». Как утверждает в своей работе П.В. Сысоев, появление первого лингвистического корпуса мгновенно привлекло к нему внимание научной общественности, что отразилось в целой научной дискуссии [8]. Одним из основных аспектов обсуждения были принципы отбора текстов и перечень решаемых на основе данного лингвистического корпуса исследовательских задач. Согласно У. Френсису и Г. Кучеру, при создании своего лингвистического корпуса они руководствовались рядом критериев, к которым относятся:

1) происхождение и состав текста (автор должен был быть урожденным носителем американского варианта английского языка, диалог должен был занимать менее половины объема текста);

2) синхронизация (включены были тексты, впервые изданные в 1961 г.);

3) продуманное соотношение численной представленности различных жанров и отбор

отдельных текстов при помощи особой вероятностной процедуры;

4) доступность для компьютерной обработки (специальные пометы для передачи графических особенностей текста и т. п.) [8].

Основная задача авторов любого лингвистического корпуса – собрать как можно большее количество текстов, относящихся к тому подмножеству языка, для изучения которого корпус создается. В этой связи можно заключить, что корпус – это уменьшенная модель конкретного языка. Основным характерным признаком корпуса является его *репрезентативность*, т. е. обязательное пропорциональное представление в корпусе текстов различных периодов, жанров, стилей, авторов и т. д. Таким образом, *лингвистический корпус* – это система текстов, расположенных на машинном источнике и объединенных по различным параметрам (язык, жанр, стиль, отрасль). Одна из характеристик корпуса – это его размещение на электронном носителе для компьютерной обработки данных. **Конкорданс** является одной из программ, позволяющих обрабатывать и анализировать большие массивы текстов, а также выявлять в них языковые закономерности. Конкорданс осуществляет поиск конкретного слова или словосочетания в корпусе. После соответствующей команды программа выдаст заданное количество фрагментов текстов, содержащие запрашиваемые единицы. На основе полученных сведений можно сделать вывод о функциях, употреблении и применении заданной нами единицы в конкретном языковом пространстве [8; 9].

Управление текстовыми и лингвистическими данными в корпусе часто называется корпусным менеджером. Корпусный менеджер – это специализированная поисковая система, включающая программные средства для поиска данных в корпусе, получения статистической информации и предоставления результатов пользователю в удобной форме [10, с. 17].

Корпусный менеджер позволяет осуществлять следующее:

- строить как KWIC (Key Word In Context), так и полные конкордансные списки;
- искать не только отдельные слова, но и словосочетания;
- осуществлять поиск по шаблонам (сложные запросы);

- сортировать списки по нескольким критериям, выбираемым пользователем;

- давать возможность отображать найденные словоформы в неограниченном контексте;

- давать статистическую информацию по отдельным элементам корпуса;

- отображать леммы, морфологические характеристики словоформ и метаданные (библиографические, типологические), что зависит от степени размеченности корпуса;

- сохранять и распечатывать результаты;

- работать как с отдельными файлами, так и с корпусами, неограниченными по размеру;

- быстро обрабатывать запросы и выдавать результаты;

- поддерживать различные форматы текстовых данных (txt, doc, rtf, html, xml и др.);

- быть легким (интуитивно понятным) в использовании как для опытного, так и для начинающего пользователя [10, с. 18].

Лингвистические корпуса могут использоваться для получения разнообразных справок и статистических данных о языковых и речевых единицах. В частности, на основе корпусов можно получить данные о частоте словоформ, лексем, грамматических категорий, проследить изменение частот и контекстов в различные периоды времени, получить данные о совместной встречаемости лексических единиц и т. д. Представительный массив языковых данных за определенный период позволяет изучать динамику процессов изменения лексического состава языка, проводить анализ лексико-грамматических характеристик в разных жанрах и у разных авторов. Корпусы призваны служить также источником и инструментом многоаспектных лексикографических работ по подготовке разнообразных исторических и современных словарей [10]. Данные корпусов могут быть использованы для построения и уточнения грамматик и в целях обучения языку.

Одним из самых известных и широко используемых является Британский национальный корпус (British National Corpus). Именно он лег в основу разработки современных словарей по английскому языку. Национальный лингвистический корпус обладает двумя важными особенностями. Во-первых, корпус характеризуется представительностью, или сбалансированным составом

текстов. Иными словами, в нем содержатся все типы письменных и устных текстов, представленные в данном языке (художественные разных жанров, публицистические, учебные, научные, деловые, разговорные, диалектные и т. п.). Кроме того, все данные типы текстов 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, в корпусе есть разметка, позволяющая отследить конкретные слова или словосочетания в тексте.

В научной литературе можно встретить различные типологии лингвистических корпусов, в основу каждой из которых легли те или иные типологические признаки.

В центре внимания нашего исследования выступают корпусы параллельных текстов, поэтому видится целесообразным подробнее остановиться на таком типологическом признаке, как «параллельность». Отметим, что большая часть методических исследований была посвящена изучению вопросов формирования грамматических и лексических навыков речи на основе монологических одноязычных корпусов [8; 9; 11; 12]. Однако кроме одноязычных лингвистических корпусов существуют еще и двуязычные и многоязычные корпусы, т. е. корпусы, в которых представлены тексты на двух и более языках. Многоязычные корпусы можно условно разделить на несколько типов:

1) корпус, в котором представлен текст оригинала и его перевод;

2) одноязычный корпус, в котором представлен текст оригинала и его толкование на этом же языке;

3) корпус, в котором представлен текст оригинала и его толкование на другом языке.

Необходимо отметить, что исследователи так и не пришли к единому мнению относительно терминологии в рамках данного типологического признака. Одни ученые считают первый тип корпуса – корпусом переводов, а второй тип корпуса – корпусом параллельных текстов [13]. Другие же, наоборот, первый тип корпуса – корпусом параллельных текстов, а второй – сравнительным корпусом [14]. Для многих других авторов и первый, и второй типы корпусов являются корпусами параллельных текстов [15]. В нашем исследовании под **корпусом параллельных текстов** мы будем понимать тип **лингвистического корпуса, состоящий из исходного текста на одном языке и его перевода на другой или другие языки**.

Корпусы параллельных текстов могут также разделяться по следующим признакам (табл. 1.).

Анализ приведенных выше типов лингвистических корпусов и типов корпусов параллельных текстов позволяют сформулировать характеристики корпуса параллельных текстов, на основе которого можно обучать студентов профессионально-ориентированной лексике (табл. 2).

Таблица 1

Типы корпусов параллельных текстов

№	Типологический признак	Типы корпусов параллельных текстов
1	Языки корпуса	Корпусы параллельных текстов могут быть двуязычные и многоязычные. Двуязычные корпусы включают исходные тексты на одном языке и их переводы на другом. Многоязычные корпусы включают исходные тексты на одном языке и их соответствующие переводы на другие языки
2	Направленность перевода	Корпусы могут быть однонаправленные, двунаправленные и многонаправленные. Однонаправленные корпусы позволяют осуществлять перевод с одного языка на другой (например, с английского на русский) без возможности обратного перевода (с русского на английский). Двунаправленные корпусы включают параллельные тексты на двух языках (английском и русском) и позволяют осуществлять перевод как с английского на русский, так и наоборот). Многонаправленные корпусы, включающие тексты на более чем двух языках, позволяют осуществлять перевод с любого языка на любой (в рамках существующих языков корпуса)

Таблица 2

Характеристики корпуса параллельных текстов, на основе которого  
можно обучать студентов профессиональной лексике

№	Типологические признаки	Характеристики корпуса параллельных текстов
1	Тип языковых данных	Ввиду того, что в процессе обучения в вузе студенты в большей мере будут использовать академические статьи профессиональной направленности, должен быть использован <b>корпус письменных текстов профессиональной направленности</b>
2	«Параллельность»	В процессе обучения используются двуязычные корпуса, включающие исходные тексты на одном языке и их переводы на другом
3	Направленность перевода	Используются двунаправленные корпуса, включающие параллельные тексты на двух языках (английском и русском) и позволяющие осуществлять перевод как с английского на русский, так и наоборот
4	«Литературность»	В процессе обучения используется терминологический корпус письменных текстов по выбранной специальности
5	Цель	В учебном процессе используются специализированные корпуса, созданные для учебных или научных целей
6	Жанр	В обучении используются научные тексты профессиональной направленности
7	Доступность	В процессе обучения могут быть использованы разные корпуса (коммерческие, открытые и закрытые), которые будут доступны студентам
8	Назначение	В процессе обучения должны использоваться «иллюстративные» корпуса, отчетливо показывающие особенности словоупотребления профессиональной лексики в изучаемом языке
9	Динамичность	В процессе обучения должны использоваться корпуса, включающие современные тексты по научной проблематике
10	Разметка	В процессе обучения должны использоваться корпуса с морфологической, синтаксической, семантической разметкой
11	Объем текстов	В процессе обучения могут использоваться как полнотекстовые, так и фрагментотекстовые корпуса

Таблица 3

Дидактические свойства и методические функции корпуса параллельных текстов

Дидактические свойства	Методические функции
<b>Многоязычие</b> (в корпусе параллельных текстов представлены тексты на одном языке с их переводами на другой язык или несколько языков)	Корпус параллельных текстов может быть использован для формирования лексических навыков речи обучающихся посредством перевода
<b>Контекстность результатов поиска</b> (при осуществлении поискового запроса конкретного слова в корпусе программа конкорданс предложит несколько вариантов использования данного слова в контекстах на языке оригинала с возможными переводами на другой язык)	Корпус параллельных текстов позволяет более точно определять значения новых слов и показывать особенности словоупотребления в реальных языковых ситуациях
<b>Возможность ограничения области использования языка</b> при осуществлении поискового запроса (корпус параллельных текстов позволяет значительно ограничить область поиска конкретного слова или выражения до определенной сферы общения (бизнес, образование, медицина, повседневное общение и т. п.)	Корпус параллельных текстов позволяет более точно определять значения новых слов и показывать словоупотребление в конкретной запрашиваемой сфере общения. Данное свойство позволяет использовать корпусы параллельных текстов для формирования лексических навыков речи обучающихся при изучении иностранного языка для специальных целей

Корпус параллельных текстов, как и любое средство ИКТ, обладает отличительными дидактическими свойствами и методическими функциями. Вслед за П.В. Сыроевым под дидактическими свойствами мы понимаем «основные характеристики, признаки конкретных технологий, отличающие одни от других, существенные для дидактики (включая лингводидактику) как в плане теории, так и в плане практики» [16, с. 16], а под методическими функциями понимаются внешние проявления средств ИКТ, используемые в учебно-воспитательном процессе для реализации поставленных целей [16, с. 16]. Корпус параллельных текстов обладает несколькими характерными дидактическими свойствами, отличающими данный тип корпуса от других корпусов, и обуславливающими методику формирования лексических навыков речи студентов на его основе. **К таким дидактическим свойствам относятся: а) многоязычие; б) контекстность результатов поиска; в) возможность ограничения области использования языка.** Подробно дидактические свойства и методические функции корпуса параллельных текстов представлены в табл. 3.

Таким образом, в отличие от других типов корпусов, отличительным дидактическим свойством корпусов параллельных текстов является многоязычие. Это позволит на его основе формировать лексические навыки обучающихся на основе перевода.

Подводя определенный итог, отметим, что под **корпусом параллельных текстов** понимается тип **лингвистического корпуса, состоящий из исходного текста на одном языке и его перевода на другой или другие языки. К характеристикам корпуса параллельных текстов**, на основе которого можно обучать студентов профессиональной лексике, относятся следующие:

- а) в корпусе используются письменные тексты профессиональной направленности;
- б) в корпусе используются исходные тексты на одном языке и их переводы на другом;
- в) в корпусе можно переводить как с изучаемого языка на родной, так и наоборот;
- г) в корпусе представлены тексты исключительно профессиональной направленности студентов;

д) в корпусе представлены тексты научного жанра;

е) доступ к ресурсам корпуса может быть открытым или ограниченным (при использовании коммерческих продуктов);

ж) в процессе обучения должны использоваться «иллюстративные» корпуса, отчетливо показывающие особенности словоупотребления профессиональной лексики в изучаемом языке;

з) в корпус должны быть включены современные тексты профессиональной направленности;

и) в процессе обучения могут быть использованы как полнотекстовые, так и фрагментотекстовые корпуса.

Корпус параллельных текстов обладает отличительными дидактическими свойствами и методическими функциями. К дидактическим свойствам относятся: а) многоязычие; б) контекстность результатов поиска; в) возможность ограничения области использования языка. Эти дидактические свойства корпуса параллельных текстов позволяют формировать у студентов лексические навыки речи посредством перевода при изучении иностранного языка для специальных целей, а также более точно определять значения новых слов и показывать особенности словоупотребления в реальных языковых ситуациях.

1. Сыроев П.В., Евстигнеев М.Н. Современные учебные интернет-ресурсы в обучении иностранному языку // Иностранные языки в школе. 2008. № 6. С. 1-10.
2. Сыроев П.В., Евстигнеев М.Н. Разработка авторских учебных интернет-ресурсов по иностранному языку // Иностранные языки в школе. 2009. № 2. С. 8-16.
3. Сыроев П.В., Евстигнеев М.Н. Технологии Веб 2.0: Социальный сервис блогов в обучении иностранному языку // Иностранные языки в школе. 2009. № 4. С. 12-18.
4. Сыроев П.В., Евстигнеев М.Н. Технологии Веб 2.0: Социальный сервис вики в обучении иностранному языку // Иностранные языки в школе. 2009. № 5. С. 2-8.
5. Сыроев П.В., Евстигнеев М.Н. Технологии Веб 2.0: Социальный сервис подкастов в обучении иностранному языку // Иностранные языки в школе. 2009. № 6. С. 8-11.



6. *Соломатина А.Г.* Развитие умений говорения и аудирования посредством учебных подкастов // *Иностранные языки в школе.* 2012. № 9. С. 71-74.
7. *Маркова Ю.Ю.* Методика развития умений письменной речи студентов на основе вики-технологии: автореф. дис. ... канд. пед. наук. М., 2011.
8. *Сысоев П.В.* Лингвистический корпус, корпусная лингвистика и методика обучения иностранным языкам // *Иностранные языки в школе.* 2010. № 5. С. 12-21.
9. *Сысоев П.В.* Лингвистический корпус в методике обучения иностранным языкам // *Язык и культура.* 2010. № 1 (9). С. 99-111.
10. *Захаров В.П.* Корпусная лингвистика. СПб., 2005.
11. *Чернякова Т.А.* Методика формирования лексических навыков студентов на основе лингвистического корпуса: автореф. дис. ... канд. пед. наук. М., 2012.
12. *Рязанова Е.А.* Методика формирования грамматических навыков речи студентов на основе лингвистического корпуса: автореф. дис. ... канд. пед. наук. М., 2012.
13. *Aijmer K., Altenberg B., Johansson M.* Language in contrast: Papers from a symposium on text-based cross-cultural studies. Lund, 1996.
14. *Baker M.* The role of corpora in investigating the linguistic behavior of professional translators // *International Journal of Corpus Linguistics.* 1999. № 4. P. 281-298.
15. *Johansson S.* On the role of corpora in cross-linguistic research // *Corpora and cross-linguistic research.* Amsterdam, 1999. P. 3-24.
16. *Сысоев П.В.* Дидактические свойства и функции современных информационных и коммуникационных технологий // *Иностранные языки в школе.* 2012. № 6.

Поступила в редакцию 25.10.2012 г.

UDC 811.111

PARALLEL CORPUS IN FOREIGN LANGUAGE TEACHING

Anastasia Aleksandrovna KOKOREVA, Tambov State University named after G.R. Derzhavin, Tambov, Russian Federation, Post-graduate Student, Linguistics and Linguodidactics Department, e-mail: kokoreva81@yandex.ru

The issue of the use of parallel corpus in teaching a foreign language is considered. The term "parallel corpus" is defined, didactic features and pedagogical functions of the parallel corpus are studied, a typology of parallel corpus is offered.

*Key words:* language corpus; parallel corpus; corpus linguistics.