

А.Н. Овешкова

Институт философии и права УрО РАН,

Уральский федеральный университет имени первого Президента  
России Б.Н. Ельцина

Екатеринбург, Россия

## **КОРПУС GLOWBE КАК УЧЕБНЫЙ РЕСУРС И ИНСТРУМЕНТ ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЫ СТУДЕНТОВ-ЛИНГВИСТОВ**

**Аннотация:** В статье рассматриваются два возможных подхода к использованию Интернет-корпуса GloWbE в обучении студентов-лингвистов. С позиций первого подхода корпус рассматривается как справочная и информационно-поисковая система. Во-вторых, GloWbE может служить инструментом для изучения варьирования английского языка. Научно-исследовательская работа студентов на материалах GloWbE в этом случае будет заключаться в конструировании собственных исследовательских корпусов и их анализе.

**Ключевые слова:** корпусная лингвистика, языковой корпус, корпусные данные, варианты английского языка, научно-исследовательская работа студентов-лингвистов.

Интерес к использованию корпусных инструментов в преподавании английского (или иного) языка как иностранного подтверждается значительным объемом литературы по данной проблематике, см., например, [6; 8; 10]. Неуклонно растет количество работ и русскоязычных авторов (так, по поисковому запросу «корпусная лингвистика» в электронной библиотеке e-LIBRARY.ru в феврале 2018 г. было обнаружено 1767 публикаций). Российские исследователи проводят изыскания как на материалах представительных корпусных онлайн-ресурсов, так и на специализированных корпусах, разработанных самими авторами или студентами в экспериментальной группе [2; 7]. Корпуса (вариант множественного числа: корпусы) в современных условиях – это электронные собрания текстов или базы данных, обладающие, с

точки зрения лингводидактики, ярко выраженным методическим потенциалом (назначением). Так, О.Г. Горина в своем диссертационном исследовании доказывает, что «...корпусные технологии – это педагогически релевантные корпусные ресурсы: лингвостатистика, поисковые возможности и лингвистическая и статистическая наглядность корпуса» [2]. Иными словами, речь идет о справочных и информационно-поисковых возможностях данной информационно-коммуникационной технологии, которые, на наш взгляд, должны быть использованы в образовательном процессе.

Цель данной статьи заключается в описании достоинств и недостатков интернет-корпуса английского языка GloWbE (the Corpus of Global Web-Based English), созданного в 2013 г. в Университете Бригама Янга (США), с точки зрения обучения студентов-лингвистов. К сожалению, данный ресурс пока недостаточно описан в литературе, хотя в соответствии с информацией, представленной на сайте корпуса по адресу <<https://corpus.byu.edu/glowbe/>>, и данными в статье М. Дэвиса, Р. Фухса [5], корпус включает в себя 1,9 млрд словоупотреблений (или слов), полученных с 1,8 млн веб-страниц. Главная особенность GloWbE заключается в том, что в нем представлены 20 вариантов английского языка. Среди них шесть вариантов внутреннего и 14 вариантов внешнего круга по теории концентрических кругов Б. Качру [3]. Источником материала всех региональных подкорпусов стал Интернет, а именно, корпус представляет собой в основном два сегмента англоязычного Интернет-дискурса: блоги и сайты. Данный ресурс может быть кратко описан следующим образом: одноязычный, свободно доступный, морфологически размеченный мега-корпус с региональными подкорпусами.

Разработчики GloWbE ставили перед собой три основные задачи:

- создать корпус такого размера, который позволил бы проводить сопоставительные исследования «английских языков мира» или мировых вариантов английского языка (World Englishes);
- обеспечить баланс между письменной и устной составляющими корпуса (т.е. создать смешанный корпус), что проявляется в наполнении GloWbE. 60% текстов корпуса относятся к такому Интернет-жанру как блог. Блоги в компьютерно-опосредованной коммуникации представляют собой устно-

письменный гибрид, близкий разговорному жанру. Однако в своей статье разработчики оговариваются, что данные материалы служат источником «fairly informal language» и не являются собственно транскриптами разговорной речи (Davies & Fuchs 2015: 26; ср. также Агапова, Полоян 2016: 55). Необходимо все же отметить, что Интернет-жанр веб-сайтов тяготеет к письменным, а жанр блогов – к устным текстам, таким образом, все вышесказанное не снижает ценности рассматриваемого Интернет-корпуса английского языка.

- гарантировать точность данных, касающихся определения тех географических регионов, где тексты были созданы [5].

Несмотря на ограничения, связанные с репрезентативностью корпуса, которая определяется соотношением «объем текстов + разнообразие текстов», GloWbE достаточно репрезентативен, т.к. естественные тексты объективно отражают бытование английского языка в Интернете. Эти живые языковые данные могут служить источником эмпирического материала в лингвистических исследованиях студентов, а также использоваться в процессе преподавания английского языка, иными словами, корпус обладает образовательным и научным потенциалом. Так, метаданные, касающиеся региональной принадлежности автора текста, могут стать базой для контрастивных лингвистических исследований различных вариантов английского языка. Данный корпус позволяет изучать вариативность, проявляющуюся в лексической, морфологической, синтаксической, орфографической системах 20 вариантов английского языка, при этом результаты подобных исследований будут статистически обоснованы. Дополнительно можно проводить дискурсивные контрастивные исследования. Таким образом, GloWbE как инструмент для изучения варьирования языка в зависимости от географии распространения английского, педагогически значим для развития навыков исследовательской работы студентов-лингвистов.

Одним из вариантов научно-исследовательской работы студентов-лингвистов может стать конструирование исследовательских, а в дальнейшем – и иллюстративных корпусов. Например, на создание исследовательского корпуса некоторых категориальных форм выражения нереальности в британском варианте современного английского языке у нас ушло около года. Всего было собрано 103 контекста, полученных методом сплошной

выборки из 15 оригинальных произведений современной художественной литературы Великобритании объемом 4857 страниц [4]. Для сравнения: благодаря такому свойству корпуса как экономность, на компьютеризированный поиск уходит несколько секунд, анализ 100 строк конкорданса GloWbE занимает примерно полчаса.

Еще одним плюсом корпусного анализа является возможность получения более актуальных данных чем те, которые представлены в печатных грамматиках английского языка. Наше исследование дополнительных придаточных предложений, вводимых, например, глаголом *suggest*, выявило несоответствие между явлениями живого языка и сведениями, представленными в аутентичных грамматиках, см., например [9]. Анализ материалов корпуса полностью подтверждает полученные нами результаты [4]: на поисковый запрос *suggest \_p\**, где *\_p\** – это любая форма местоимения, GloWbE не выдал ни одного контекста с *Suppositional Mood* (предположительное наклонение, т.е. *should* + инфинитив глагола) в дополнительном придаточном предложении в первых 200 контекстах. В качестве иллюстрации приводим строки конкорданса 12-14 (Таблица 1):

Таблица 1 - Строки конкорданса 12-14

|   |   |                                                  |   |                                                                                                                                                                            |
|---|---|--------------------------------------------------|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2 | B | <a href="http://hetimes.co.uk">hetimes.co.uk</a> | f | fall in behind Romney. # 00.21 If you want exit poll rumours, I <b>suggest you</b> head to Twitter. There are plenty confidently predicting great success for both parties |
| 3 | B | <a href="http://uardian.co.uk">uardian.co.uk</a> | g | word," true "? # And before you answer, can I just <b>suggest you</b> read previous threads on Jung because certain people are going to vomit if we                        |
| 4 | B | <a href="http://uardian.co.uk">uardian.co.uk</a> | g | both perspective on the problem and ideas on how to solve it. # I <b>suggest you</b> look at alcohol as an example of a drug. It's widely available,                       |

Как демонстрируют примеры выдачи 12-14 из корпуса GloWbE, придаточное дополнительное содержит сказуемое в форме the Present Subjunctive, которая в литературе обычно описывается как черта, характерная для американского варианта английского языка или для формальных стилей британского варианта. Однако мета-разметка позволяет нам судить о пространственном происхождении текстов (GB = Great Britain) и регистре (G = general). Тексты блогосферы вряд ли можно отнести к формальным функциональным стилям. Из 3416 контекстов, полученных методом сплошной выборки и содержащих

suggest \_p\*\_v\* (\_v\* = любой глагол), лишь в 144 (4%) из них имеется вспомогательный глагол should.

Все вышесказанное позволяет нам заключить, что современные корпусные технологии представляют значительный интерес как для формирования навыков исследовательской работы, так и для учебной деятельности студентов-лингвистов.

A.N. Oveshkova

Institute of Philosophy and Law, Ural Branch of the Russian Academy of Sciences

Ural Federal University named after the first President of Russia B.N. Yeltsin

Yekaterinburg, Russia

## THE GLOWBE CORPUS AS A LEARNING AND RESEARCH TOOL FOR LANGUAGE STUDENTS

**Abstract:** the paper focuses on the two ways the potential of the Corpus of Global Web-based English (GloWbE) can be exploited: on the one hand, GloWbE should become a learning resource for language students. On the other hand, it is important that the corpus become a tool in students' theoretical and practical research.

**Key words:** corpus linguistics, language corpus, corpus-based data, varieties of English, language students' research work.

## СПИСОК ЛИТЕРАТУРЫ:

1. Агапова С.Г., Полоян А.В. Интернет-дискурс: основные жанры и особенности их исследования // Известия Южного федерального университета. Филологические науки. 2016. № 4.
2. Горина О.Г. Использование технологий корпусной лингвистики для развития лексических навыков студентов-регионоведов в профессионально-ориентированном общении на

английском языке / Дисс. на соискание степени канд. пед. наук. М, 2014.

3. Качру Б.Б. Модели вариантов английского языка, неродного для его пользователей // Личность. Культура. Общество. Международный журнал социальных и гуманитарных наук. М. Том XII. Вып. 1. 2010. № 53–54.

4. Овешкова А.Н. Некоторые формы сослагательного наклонения (the Present Subjunctive и the Suppositional Mood) в британском варианте современного английского языка // Актуальные проблемы германистики, романистики и русистики: Материалы ежегодной международной конференции. Екатеринбург, 1–2 февраля 2013 г. Екатеринбург, 2013. Ч. III.

5. Davis M., Fuchs R. Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE) // English World-Wide. 2015. 36 (1).

6. Gavioli L., Aston G. Enriching reality: language corpora in language pedagogy // ELT Journal. 2001. 55 (3).

7. Levchenko V. Use of Corpus-Based Classroom Activities in Developing Academic Awareness in Doctoral Students // The New Educational Review. 2017. 48 (2).

8. Römer U. Using general and specialized corpora in English language teaching: past, present and future // In M.C. Campoy-Cubillo, B. Bellés-Fortuño, M.L. Gea-Valor (Eds.). Corpus-Based Approaches to English Language Teaching / London, New York: Continuum, 2010.

9. Vince M., Sunderland P. Advanced language practice: with key. English grammar and vocabulary / Macmillan Publishers, 2003.

10. Zanettin F. Corpus-based translation activities for language learners // The Interpreter and Translator Trainer. 2009. 3 (2).