ИСПОЛЬЗОВАНИЕ ЛИНГВИСТИЧЕСКИХ ОНЛАЙН РЕСУРСОВ ДЛЯ РАССМОТРЕНИЯ ЛОКАЛЬНЫХ И ГЛОБАЛЬНЫХ СЛУЧАЕВ ВАРИАТИВНОСТИ НОРМЫ В АНГЛИЙСКОМ ЯЗЫКЕ

У.М. Прошина

В статье делается обзор современных лингвистических онлайн ресурсов, доступных широкой аудитории, на которые стоит обратить внимание при рассмотрении отдельных употреблений на соответствие общепринятой норме и для определения узуса отдельных локальных вариантов, рассматривается ряд толковых словарей и корпусов английского языка. Обращается внимание на представление информации о вариантах английского языка в данных ресурсах. В качестве примеров рассматриваются слова и выражения *handphone*, *data* и *big data*. Делаются выводы о необходимости комплексного использования ресурсов разного плана при проведении исследований и для уточнения информации о соответствии отдельного употребления языковой норме.

Ключевые слова: варианты английского языка, вариативность, нормы, словари, корпусы.

USE OF LINGUISTIC ONLINE RESOURCES FOR CONSIDERING NORM VARIANTS IN LOCAL AND GLOBAL ENGLISHES

U.M. Proshina

The paper addresses useful linguistic resources that are freely accessible online and should be paid attention to when considering norm variants in local and global Englishes. A number of popular dictionaries and corpora of contemporary English and Englishes are reviewed. The inclusion of Englishes and information about English norm and standards in the resources is analyzed. The words and phrases *handphone*, *data* and *big data* are given as examples. The paper concludes that comprehensive use of online resources is needed when considering norm variants.

Key words: world Englishes, varieties of English, norms, dictionaries, corpora.

Введение

В эпоху глобализации английский язык выступает в качестве языка международного общения для людей разных национальностей, являясь неродным языком для большинства коммуникантов. В результате увеличения количества языковых контактов современный английский язык динамично меняется, появляются новые варианты английского языка, представляющие большой интерес для исследований в области вариантологии. Новые нестандартные употребления находят отражение в публикуемых текстах и живом общении, влияют на языковую норму и в ряде случаев закрепляются в языке.

Традиционными лингвистическими ресурсами, кодифицирующими нормативное употребление, являются словари. В эпоху бумажных словарей процесс фиксации новых слов, новых значений, новых форм был сильно растянут во времени, но и темпы глобализации были несколько ниже, чем в современном мире. Повсеместное использование сети Интернет способствовало не только наращи-

ванию темпов международного общения, но и дало возможность работы с постоянно обновляемыми и увеличивающимися в объеме текстовыми данными. Значительная часть новых текстов появляется онлайн в электронном виде, опубликованные ранее бумажные тексты оцифровываются и переводятся в электронный формат. Формируются корпусы текстов большого объема, а компьютерная поддержка дает возможность поиска в текстовых массивах большого объема. Целый ряд лингвистических корпусов находится в открытом доступе, а большинство современных толковых словарей включают в словарные статьи «живые» примеры из собственных мониторинговых корпусов.

В рамках данной статьи рассмотрим ряд лингвистических онлайн ресурсов, которые могут быть использованы для разбора отдельных случаев вариативности нормы в английском языке, приведем примеры и попытаемся определить, являются ли данные случаи локальными вариантами употребления или же свидетельствуют о более глобальных

Прошина Ульяна Михайловна – преподаватель кафедры лингвистики и информационных технологий факультета иностранных языков и регионоведения Московского государственного университета им. М.В. Ломоносова.

Proshina Uliana Mikhailovna – Lecturer, Department of Linguistics and Information Technologies, Faculty of Foreign Languages and Area Studies, Lomonosov Moscow State University.

E-mail: uproshina@yandex.ru

© Прошина У.М., 2021

изменениях в английском языке. Рассмотрение начнем с онлайн словарей, далее обратимся к корпусам и сервисам на основе корпусных данных, приведем примеры.

Современные онлайн словари

В относительно недавнем прошлом словарь был первым источником, к которому обращались преподаватели, переводчики и студенты-лингвисты при возникновении вопросов о соотношении отдельного употребления и языковой нормы в английском языке. Появление фундаментальных лексикографических работ было большим событием в мире лексикографии и лингвистики. Новые издания являлись предметом обсуждений, споров или же представляли собой новый эталон, фиксировавший изменившееся нормативное употребление.

В распоряжении современного лингвиста в настоящее время есть целый ряд инструментов, которые помогают сделать выводы о функционировании и использовании отдельных единиц на современном этапе развития языка: это корпусы, сервисы на основе корпусов и даже информационно-поисковые системы. Тем не менее, в отличие от словаря, никакой даже самый сбалансированный корпус с детальной разметкой не даст готовых рекомендаций о соответствии отдельного употребления языковой норме. На основании данных лингвистических корпусов исследователь может провести собственный анализ и сделать выводы о степени распространения отдельных вариантов употреблений языковых элементов, но именно словари ведущих издательств продолжают сохранять статус авторитетных источников, кодифицирующих нормативное употребление.

Современные словари в большинстве случаев уже не издаются в бумажном виде. В 2012 г. издательство Макмиллан объявило о том, что прекращает издавать словарь в бумажном виде, так как цифровое пространство является идеальной средой для современного словаря. Были обозначены следующие преимущества словаря в онлайн формате перед его бумажной версией:

- снимаются ограничения, связанные с книжным форматом (добавлены тезаурус, звуковое воспроизведение слов, звуковые эффекты, интерактивные игры и т.д.);
- словарь регулярно обновляется и исправляется;
- в онлайн формате издательство может добавить целый ряд дополнительных ресурсов, например блог об английском языке, в котором затрагиваются такие вопросы, как изменение языка, типичные ошибки и т.д.;
- есть колонка о сленге и модных словах Buzzword;

– пользователи могут внести свой вклад в развитие словаря, добавив самостоятельно новые слова, термины и т.д. в открытый словарь Open Dictionary [18].

К перечисленным преимуществам можно добавить еще такие возможности современных онлайн словарей, как наличие большого количества гиперссылок, позволяющих пользователю в один клик получать дополнительную информацию, наличие иллюстраций, примечаний, большое количество примеров, в том числе примеры из лингвистических корпусов. Значительно увеличилось и количество словарных статей. Словари английского языка теперь более оперативно и гибко реагируют на появление неологизмов, активно пополняются словами из локальных вариантов английского языка. Теперь лексикографам нужно решать вопрос не о критериях для включения отдельных единиц в словарь, а скорее о критериях для исключения отдельных элементов. Согласимся со словами А.Н. Баранова о том, что современная лексикография «существенно расширила и усилила свой инструментарий компьютерными технологиями создания и эксплуатации словарей» [1: с. 90].

Рассмотрим ряд популярных английских толковых словарей, которые доступны пользователям онлайн в открытом доступе. Для сравнения словарных статей в качестве примера возьмем существительное *handphone*, употребляющееся в азиатских вариантах английского языка. Зафиксировано ли данное локальное употребление в популярных английских толковых словарях? Какую информацию включает в себя словарная статья для данного существительного?

Обратимся прежде всего к одному из самых авторитетных английских словарей – **Oxford English Dictionary (OED)**. Отметим, что на сайте OED есть блог, в котором публикуются посты про английский язык, в том числе есть публикации про включение локальных вариантов английского языка в толковый словарь. К сожалению, словарь OED (<https://oed.com>) доступен только при наличии платной подписки. Тем не менее есть открытый ресурс Lexico (<https://www.lexico.com>), разработанный Oxford University Press в сотрудничестве с Dictionary.com. В разделе «о словаре» есть информация, что все дефиниции подготовлены лексикографами издательства Oxford [7]. Рассмотрим словарную статью для существительного *handphone* в английском толковом словаре **Lexico**. Данное существительное включено в словарь, есть транскрипция, возможность прослушать произношение, дефиниция ‘a mobile phone’, пометка *South East Asian*, пример использования данного существительного в предложении, кнопка для просмотра большего ко-

личества примеров [12]. Таким образом, существительное *handphone* включено в общий толковый словарь с пометкой, указывающей на локальное употребление данного слова в районе Юго-Восточной Азии.

Словарь **Cambridge Dictionary** ориентирован прежде всего на людей, изучающих английский язык как иностранный. Словарной статьи для существительного *handphone* в данном словаре нет [9].

Ресурс **Macmillan Dictionary** уже приводился в нашей статье в качестве примера современного обновляемого словаря большого объема. Более того, данный ресурс включает блог, в котором есть публикации про изменения в английском языке, в том числе про пополнение словаря словами из разных вариантов английского языка. В одном из постов упоминается существительное *handphone* [17]. Рассмотрим словарную статью для данного существительного. В словарной статье *handphone* есть информация о том, что данное исчисляемое существительное относится к малазийскому варианту английского языка, приведена транскрипция, есть возможность прослушать произношение, словоформы, дефиниция ‘a mobile phone’, один пример, ссылка на тезаурус [14]. Таким образом, пользователи словаря получают информацию о слове *handphone* как о существительном, относящемся к малазийскому варианту английского языка.

Еще одним популярным толковым словарем, доступным в режиме онлайн, является словарь **Longman**. Данный словарь хорошо демонстрирует разницу между британским и американским вариантами английского языка, но является учебным словарем и не включает примеры из вариантов английского языка внешнего и расширяющегося кругов (о теории трех кругов Б. Качру см.: [11]). Словарная статья *handphone* в данном словаре отсутствует [13].

Collins Dictionary является еще одним популярным словарем и открытым онлайн ресурсом, где можно найти интересные публикации об изменениях в английском языке, новых словах, добавленных пользователями, и словах, недавно добавленных в общий толковый словарь. Данный словарь включает словарную статью для существительного *handphone*, в которой содержится следующая информация про использование *handphone* “in British English”: транскрипция, аудиопроизношение, частеречная принадлежность, дефиниция ‘the usual SE Asian English term for mobile phone’, синоним *mobile phone* с дефиницией. В качестве примеров приведены два предложения из википедии со словом *handphone* [10]. Таким образом, пользователи словаря могут узнать, что данное слово вошло в английский язык из азиатских вариантов английского языка.

Словарь **Merriam-Webster** является авторитетным американским лексикографическим трудом на протяжении многих лет. Онлайн версия словаря представляет собой современный лексикографический ресурс, дающий доступ не только к бесплатной версии словаря, но и к различным материалам о языке, в том числе регулярно публикуются сообщения о новых словах, включенных в словарь. К сожалению, при запросе *handphone* открывается словарная статья *hand phone* с дефиницией-ссылкой на словарную статью *handset* [15]. Никаких помет о локальном употреблении данного существительного нет. Невнимательный пользователь может сделать выводы, что данное слово вошло в словарь, является синонимом слова *handset* и активно используется в американском варианте английского языка, но если внимательно прочитать сообщение под данной словарной статьей, то становится ясно, что данное слово включено в полную версию словаря, которая доступна по подписке, т.е. открытая бесплатная версия словаря не дает детальной информации об употреблении слова *handphone* в английском языке.

Таким образом, современные онлайн словари, находящиеся в открытом доступе, динамично развиваются, увеличиваются в объеме, взаимодействуют с пользователями из разных уголков мира и ориентируются не только на варианты внутреннего круга, но и включают информацию о языковых элементах внешнего круга [11]. Слово *handphone*, относящееся к малазийскому варианту английского языка, представлено в словарях Lexico, Macmillan Dictionary и Collins Dictionary, но данный вариант не отражен в учебных словарях Cambridge Dictionary и Longman. Что касается слов расширяющегося круга, то в статье в соавторстве с З.Г. Прошиной мы анализировали примеры из онлайн корпусов для выражения *homework*, которое иногда встречается в российском образовательном контексте. Мы пришли к выводу, что данное выражение не является уникальным русицизмом и обнаруживается в других вариантах английского языка [5]. В упомянутых выше толковых английских словарях данное выражение не представлено. Несколько примеров не дают возможности сделать общие выводы, и для полной картины необходимы дополнительные исследования на предмет объема включения в словари языковых элементов внешнего круга и расширяющегося круга. Вхождение лексем в общий толковый словарь может свидетельствовать о значительном распространении слов и выражений из данных вариантов. Словарь дает готовые рекомендации по их использованию, но отражает отнюдь не все изменения, происходящие в языке. Для иллюстрации и анализа употреблений, вызываю-

щих вопросы, современные исследователи и студенты обращаются к лингвистическим корпусам и сервисам на основе корпусов.

Корпусы и сервисы на основе корпусов

В эпоху онлайн коммуникации и компьютерных технологий словари уже не являются единственным источником информации о современном языке. В распоряжении лингвиста появились новые инструменты, позволяющие просматривать и анализировать данные большого объема: лингвистические корпусы и сервисы на основе корпусов. Более того, информацию о современном употреблении отдельных единиц можно получить с помощью информационно-поисковых систем (Google, Яндекс), но данные системы не являются специальными лингвистическими инструментами, поэтому подобный поиск, как правило, характеризуется большим количеством «информационного шума» в результатах поиска. В отличие от информационно-поисковых систем лингвистический корпус характеризуется сбалансированным составом текстов, т.е. представляет собой уменьшенную модель языка и содержит тексты разных жанров, стилей, авторов в определенных пропорциях. Поиск нужной лингвистической информации упрощается благодаря наличию разметки, содержащей «особую дополнительную информацию о свойствах входящих в него текстов» [6].

Лингвистический корпус данных представляет собой «унифицированный, структурированный, размеченный, филологически компетентный массив данных, предназначенный для решения конкретных лингвистических задач» [3: с. 11], т.е. идеальный корпус является сбалансированным, размеченным и прагматически ориентированным. Для рассмотрения вариативности нормы в современном английском языке нужен прежде всего мониторинг корпус английского языка, т.е. постоянно развивающийся обновляемый и дополняемый корпус. Для разведения употреблений по вариантам нужен корпус, включающий данные из разных вариантов английского языка с соответствующими метками.

В распоряжении современного лингвиста есть ряд корпусов английского языка, но не все ресурсы являются открытыми и доступными для широкого круга пользователей. Часть корпусов, находящихся в открытом доступе, содержат тексты определенного временного промежутка и не подходят для анализа нормы в современном английском языке (Brown, LOB, BNC). Прежде всего обратим внимание на соответствие отдельных употреблений общепринятой норме и определим узус отдельных локальных вариантов.

Очень удобным ресурсом, содержащим целый ряд корпусов, является проект Марка Дэвиса, про-

фессора университета Бригама Янга. На главной странице проекта (www.english-corpora.org) на момент написания статьи представлено 19 корпусов английского языка. Корпусы доступны всем пользователям после бесплатной регистрации. Ресурс включает корпусный менеджер и предоставляет удобный интерфейс для онлайн работы с корпусами. Рассмотрим некоторые из корпусов, размещенных в рамках данного проекта.

Для наблюдения за изменениями в американском варианте английского языка следует прежде всего обратить внимание на один из самых популярных корпусов проекта М. Дэвиса. Корпус современного американского английского языка (**Corpus of Contemporary American English**, сокращенно **COCA**) является самым большим сбалансированным корпусом американского варианта английского языка. Объем корпуса составляет более 1 млрд слов. Корпус является мониторинг, пополняется несколько раз в год, на сегодняшний день включает тексты разных жанров с 1990 по 2020 г., в том числе в корпус включена устная речь. По запросу *handphone* корпус дает 7 употреблений, есть информация о времени фиксации (год), жанре, к которому относится найденный пример; результатом запроса *hand phone* являются 5 употреблений, отличных от результатов по запросу *handphone*.

Другой корпус **Global Web-Based English** (сокращенно **GloWbE**) в рамках проекта М. Дэвиса содержит 1,9 млрд слов и представляет 20 вариантов английского языка (США, Канада, Великобритания, Ирландия, Австралия, Новая Зеландия, Индия, Шри-Ланка, Пакистан, Бангладеш, Сингапур, Малайзия, Филиппины, Гонконг, Южная Африка, Нигерия, Гана, Кения, Танзания, Ямайка). Корпус содержит тексты из сети Интернет (в том числе тексты блогов), данные относятся к 2012–2013 гг. Данный корпус позволяет просмотреть употребление языковых элементов в обозначенных вариантах английского языка в указанном временном промежутке. По запросу *handphone* максимальное количество употреблений характерно для сингапурского варианта английского языка (196), затем следует малазийский вариант английского языка (170 употреблений). На третьем месте бангладешский вариант (10 употреблений). В остальных вариантах английского языка данное выражение в корпусе GloWbE встречается менее 10 раз. По запросу *hand phone* в результатах поиска в корпусе получено 54 примера из малазийского варианта английского языка, 10 примеров из сингапурского варианта английского языка. Корпусные данные иллюстрируют употребление выражений *handphone* и *hand phone* в малазийском варианте английского языка, что соответствует данным английских толковых

словарей. Дополнительно можно сделать выводы об употреблении данных выражений в сингапурском варианте английского языка.

Еще одним корпусом, позволяющим рассмотреть отдельные употребления по вариантам английского языка, является корпус **News on the Web** (сокращенно **NOW**). Данный корпус также размещен на сайте проекта М. Дэвиса. Корпус NOW в несколько раз превосходит корпус GloWbE по объему (более 11,9 млрд слов), включает новостные тексты из сети Интернет, относящиеся к 20 вариантам английского языка. Корпус ежедневно обновляется, включает тексты с 2010 г. По запросу *handphone* в корпусе найден 1621 пример. Есть возможность ограничить поиск определенным вариантом английского языка или временем публикации текстов. Если по запросу *handphone* ограничить выдачу малазийским вариантом английского языка, то на сегодняшний день сервис выдает 1219 примеров; при этом есть распределение употреблений по годам. Наибольшее количество раз выражение *handphone* встречалось в новостных малазийских текстах в 2017 г. (326 примеров). Для сравнения в 2020 г. данное выражение встречалось 136 раз. Раздельное написание выражения *hand phone* во всем корпусе встречается 132 раза (все варианты). Данный корпус подходит для диахронических исследований во временном промежутке с 2010 г. по настоящее время на материале новостных ресурсов сети Интернет и для исследований в области вариатологии.

Среди корпусов проекта English-Corpora.org отметим еще один проект, корпус **Coronavirus Corpus**. Как следует из названия, тематика корпуса ограничена новостными текстами сети Интернет о новой коронавирусной инфекции. В корпусе собраны данные из 20 англоговорящих стран с января 2020 г., объем корпуса на момент написания статьи составляет более 840 млн слов.

Мы уже упомянули возможность проведения диахронических исследований на материале некоторых корпусов проекта English-Corpora.org. Следует отметить крупный и оригинальный диахронический проект **Google Ngram Viewer** (<https://books.google.com/ngrams>), созданный на материале оцифрованных и размеченных текстов книг проекта Google Books. Были сгенерированы корпуса для разных языков. Для английского языка в настоящее время доступно 13 корпусов. Корпус **English 2019** позволяет просмотреть частоту употреблений языковых единиц с 1800 по 2019 г. в текстах книг. Есть возможность сравнить частоту употреблений в американском и британском вариантах английского языка; есть отдельные корпуса **American English 2019** и **British English 2019**. Результаты поиска выводятся в

виде графиков. Если графический интерфейс Google Ngram Viewer окажется неудобным для проведения исследования, то в проекте под руководством М. Дэвиса представлена часть корпуса проекта Google Books Ngram с альтернативным интерфейсом.

Еще одним очень крупным корпусным проектом является совместный проект Братиславского университета имени Коменского и Словацкой академии наук (<http://unesco.uniba.sk/guest/index.html>). В рамках проекта создано семейство корпусов **Aranea**, состоящее из сопоставимых веб-корпусов, созданных по технологии WaC. Суть технологии заключается в объединении двух методологий извлечения данных для корпуса из сети Интернет. Сначала с помощью сервиса BootCaT путем нескольких итераций собирались URL-адреса (более 200 тысяч). Потом эти адреса интеракционно подавались на вход программы SpiderLing. Этот процесс потребовал мощной вычислительной техники и занял несколько месяцев непрерывной работы вычислительного комплекса [3: с. 85]. В настоящее время на сайте проекта для английского языка пользователям доступны корпуса Araneum Anglicum II (данные для корпуса были собраны из сети Интернет с 2013 по 2017 г.), Araneum Anglicum Africanum (английский язык Африки) и Araneum Anglicum Asiaticum (английский язык Азии). Без регистрации пользователи могут осуществлять поиск по уменьшенной версии корпусов (120 млн токенов), после бесплатной регистрации предоставляется доступ к сверхбольшим корпусам. Поиск по корпусам осуществляется с помощью корпусного менеджера NoSketch Engine. По запросу *handphone* в корпусе Araneum Anglicum II Minus 120 M найдено 16 совпадений, в корпусе Araneum Anglicum Asiaticum Minus 120 M – 115 совпадений, в корпусе Araneum Anglicum Africanum Minus 120 M – всего одно совпадение. Данные корпуса могут быть использованы для сопоставительных исследований, рассмотрения употреблений языковых единиц в регионах Азии и Африки.

Таким образом, в распоряжении современного исследователя есть целый ряд корпусов, находящихся в свободном доступе. Сверхбольшие корпуса собраны из текстов сети Интернет, но есть сбалансированные корпуса (например, COCA). Ряд корпусов позволяет проводить диахронические и сопоставительные исследования.

Пример использования лингвистических онлайн ресурсов

Несколько лет назад развернулся спор вокруг существительного *data*. В 2012 г. на страницах газеты “The Guardian” был опубликован пост *Data are or data is?* [16]. Ресурс Merriam-Webster в отдельном сообщении обращает внимание пользовате-

лей словаря на категорию числа существительных, заимствованных в английский язык из латыни [8]. В 2016 г. Н.К. Власко была проанализирована специфика функционирования формы *data* в научном тексте: «...несмотря на включение формы *data* в значении единственного числа в словари ... и на "благоклонное" отношение к ней авторитетных авторов, утверждать, что данный вариант получил широкое употребление в научном тексте, нельзя» [2].

Действительно, категория числа является одним из участков нестабильности у существительных, где нередко наблюдается варьирование и отклонение от нормы. Рассмотрим, какую информацию о форме существительного *data* содержат современные лексикографические ресурсы, проанализируем частотность употреблений *data is* и *data are* в корпусах COCA, GloWbE и Coronavirus Corpus, в сервисе Google Ngram Viewer, сделаем выводы об употреблении популярного на сегодняшний день выражения *big data*.

Оксфордские лексикографы в словаре Lexico относят *data* к неисчисляемым существительным (пометка *mass noun*), есть примечание по использованию данного слова, в котором указывается, что *data* в настоящее время воспринимается как неисчисляемое существительное и сочетается с глаголом в единственном числе, но в научном контексте исторически является формой множественного числа существительного *datum* [12]. В словаре Macmillan находим аналогичное указание на возможное сочетание с глаголом во множественном числе в научных текстах [14]. Словарь Collins разграничивает употребление данного существительного в британском и американском вариантах английского языка: в американском варианте английского языка *data* обычно является существительным во множественном числе, а в британском варианте английского языка в технических текстах и официальном стиле иногда используется как форма множественного числа, но в остальных случаях является неисчисляемым существительным [10]. В американском словаре Merriam-Webster в примечании для пользователя (*usage guide*) указывается, что *data* может употребляться и как существительное во множественном числе, и как абстрактное неисчисляемое существительное. Отмечается, что обе формы являются стандартными, но употребление *data* в качестве существительного во множественном числе больше характерно для печатных изданий [15]. Учебный словарь Cambridge отмечает возможное употребление *data* как с глаголом в единственном числе, так и с глаголом во множественном числе [9]. Словарь Longman советует использовать *data* с глаголом в единственном числе в повседневном общении, а с глаголом во множе-

ственном числе – в научных и официальных текстах [13]. Таким образом, согласно информации из словарей *data is* используется в повседневном общении, *data are* – в научных текстах; есть информация о том, что выражение *data is* больше характерно для британского варианта английского языка, чем для американского английского.

В корпусе американского английского языка COCA по запросу *data is* получаем 5457 совпадений, из них 1196 относятся к научному стилю. Выражение *data is* со временем используется чаще: 332 совпадения в период с 2005 по 2009 г., 533 – с 2010 по 2014 г., 755 – с 2015 по 2019 г. По запросу *data are* в корпусе на сегодняшний день получаем 4749 совпадений, из них 2882 относятся к научному стилю. Со временем выражение *data are* встречается в корпусе чаще: 460 совпадений в период с 2005 по 2009 г., 706 – с 2010 по 2014 г., 780 – с 2015 по 2019 г. Действительно, наблюдается предпочтение использования *data are* в научных текстах, но при рассмотрении целого корпуса оба выражения употребляются в американском варианте английского языка, есть небольшое преобладание в использовании *data is*. Увеличение количества совпадений по обоим запросам со временем свидетельствует о цифровизации современного общества, о важности данных в современном мире. Если рассмотрим популярное на сегодняшний день выражение *big data*, то в корпусе COCA количество совпадений каждые пять лет увеличивалось: 4 совпадения в период с 2005 по 2009 г., 176 – с 2010 по 2014 г., 481 – с 2015 по 2019 г. При этом *big data is* встречается в корпусе значительно чаще, чем *big data are* (112 и 8 совпадений соответственно).

В корпусе GloWbE *data is* встречается 17 779 раз, из них 4304 совпадения получено для американского варианта английского языка, 4083 – для британского варианта английского языка, но если принять во внимание неравность подкорпусов по объему и рассмотреть частоту на миллион (коэффициент *per mil*), то получаем следующие данные: 12.98 Индия, 11.64 Гонконг, 11.13 США, 10.96 Австралия, 10.53 Великобритания. Выражение *data are* встречается 6094 раза, 1839 раз в американской части корпуса, 1292 – в британских текстах, но при рассмотрении частоты на миллион получаем следующие данные: 4.75 США, 4.35 Гонконг, 3.62 Бангладеш, 3.34 Канада и 3.33 Великобритания. Данный корпус не является сбалансированным в отличие от COCA и состоит из текстов 2012–2013 гг., но можно сделать выводы, что в повседневном английском языке выражение *data is* является наиболее часто используемым.

Можно предположить, что слово *data* будет встречаться довольно часто в текстах новостей о

новой коронавирусной инфекции. Сравним частоту совпадений в результатах поиска *data is* и *data are* по корпусу Coronavirus Corpus. Количество совпадений для *data is* в результатах поиска значительно превышает количество совпадений для *data are* (16 146 и 3861 соответственно). При этом для *data is* частота на миллион значительно выше для британского варианта английского языка, чем для американского варианта (41.13 Великобритания и 16.55 США). Сочетание *data are*, согласно частоте на миллион, больше характерно для американских новостных изданий, чем для британской прессы (6.07 США, 3.56 Великобритания).

Рассмотрим употребление существительного *data* в сервисе Google Ngram Viewer. В корпусе English 2019 в книгах Google *data are* встречается чаще, чем *data is*. В 1981 г. наблюдается максимальное количество употреблений *data are*, а после 1981 г. кривая *data are* начинает спускаться к кривой *data is*; при этом значительного уменьшения употреблений *data is* не наблюдается. В 2019 г. кривые *data is* и *data are* находятся очень близко друг к другу: встречаемость *data are* немного выше, чем *data is*. Встречаемость выражений *big data is* и *big data are* в корпусе English 2019 заметно растет с 2008 г. В 2019 г. количество употреблений *big data is* превышает употребление *big data are* более чем в два раза. Если сравнить графики *data is* и *data are* в корпусах American English 2019 и British English 2019, то в 2019 г. *data are* встречается чаще, чем *data is* в обоих вариантах, при этом разница между употреблениями немного меньше в британском варианте английского языка. Таким образом, данные Google Ngram Viewer говорят об увеличении употреблений существительного *data* в качестве неисчисляемого существительного, в британском варианте английского языка данное употребление на сегодняшний день является более распространенным, чем в американском варианте английского языка. Количество употреблений термина *big data* постепенно увеличивается, при этом *data* в большинстве случаев в составе данного термина рассматривается как неисчисляемое существительное и сочетается с глаголом в единственном числе, что может свидетельствовать о глобальной тенденции переосмысления формы существительного *data*, но прогнозировать, выйдет ли из употребления *data* как форма множественного числа данного существительного, на данном этапе развития языка еще рано.

Заключение

Современные онлайн словари предоставляют пользователям гораздо больше информации, чем словари в эпоху бумажных изданий. Словари постоянно обновляются как лексикографами, так и пользователями; ряд лексикографических ресурсов

включает лексемы, относящиеся не только к вариантам внутреннего круга, но в современных популярных онлайн словарях можно найти примеры лексем из вариантов внешнего круга. Заметно увеличилось количество и качество приводимых иллюстративных примеров. Тем не менее не существует одного идеального словаря, поэтому пользователям необходимо обращаться к разным лексикографическим ресурсам, а в случае наличия противоречивой информации и с целью проведения исследований у современного лингвиста есть возможность обращаться к корпусам, находящимся в свободном доступе. В отличие от словаря корпус не дает готовых рекомендаций, но пользователи корпусов получают возможность рассмотреть употребление языковых элементов на материале большого количества примеров, в текстах разных временных периодов, разных жанров, рассмотреть употребление данных единиц в разных вариантах английского языка, что является важным для определения «локальности» отдельных употреблений и для выводов о «глобальности» отдельных процессов. Современные лингвистические онлайн ресурсы играют важную роль в работе лингвиста, при этом важно не только иметь представление о наиболее актуальных ресурсах, но и выбрать наиболее подходящие ресурсы для решения конкретной задачи, уметь использовать выбранные ресурсы. В заключение приведем цитату из монографии под ред. Л.Н. Беляевой о важности и необходимости работы специалиста: «Никакая лингвистическая технология ни сейчас, ни в будущем не отменит необходимость работы специалиста» [4: с. 157].

ЛИТЕРАТУРА

1. Баранов, А.Н. Введение в прикладную лингвистику / А.Н. Баранов. – Изд. 5-е. – Москва : Ленанд, 2017. – 368 с.
2. Власко, Н.К. Проблемы грамматической вариативности и нормы в практике преподавания английского языка / Н.К. Власко // Актуальные проблемы гуманитарных и естественных наук. – 2016. – № 3-2. – С. 129–132. – URL: https://publikacia.net/archive/uploads/pages/2016_3_2/35.pdf (дата обращения: 05.12.2020).
3. Захаров, В.П. Корпусная лингвистика / В.П. Захаров, С.Ю. Богданова. – Санкт-Петербург : Изд-во С.-Петерб. ун-та, 2020. – 234 с.
4. Лексикографический потенциал современных лингвистических технологий : монография / Л.Н. Беляева, О.А. Данилова, Т.Л. Джепа [и др.] ; под ред. Л.Н. Беляевой. – Санкт-Петербург : Книжный Дом, 2014. – 168 с.
5. Прошина, З.Г. Вариативность, норма и корпусная лингвистика в учебном дискурсе / З.Г. Прошина, У.М. Прошина // Социальные и гуманитарные науки на Дальнем Востоке. – 2019. – Т. 16, Вып. 1. – С. 36–43.
6. Что такое Корпус? – URL: <https://ruscorpora.ru/new/corpora-intro.html> (дата обращения: 21.12.2020).
7. About // Lexico. Powered by Oxford. – 2020. – URL: <https://www.lexico.com/about> (дата обращения: 09.12.2020).

8. All About Latin Plurals // Merriam-Webster. – 2020. – URL: <https://www.merriam-webster.com/words-at-play/formation-of-latin-plurals> (accessed: 22.12.2020).
9. Cambridge Dictionary. URL: <https://dictionary.cambridge.org/dictionary/english/> (accessed: 22.12.2020).
10. Collins. – 2020. URL: <https://www.collinsdictionary.com/dictionary/english/> (accessed: 22.12.2020).
11. *Kachru, B.B.* Standards, codification and sociolinguistic realism. The English language in the outer circle / B.B. Kachru // *English in the World: Teaching and Learning the Language and Literatures* / Ed. by Randolph Quirk and H.G. Widdowson. – Cambridge : Cambridge University Press, 1985. – P. 11–30.
12. Lexico. Powered by Oxford. – 2020. – URL: <https://www.lexico.com/definition/> (accessed: 22.12.2020).
13. Longman. – 2020. – URL: <https://www.ldoceonline.com/dictionary/> (accessed: 22.12.2020).
14. Macmillan Dictionary. – 2020. – URL: <https://www.macmillandictionary.com/dictionary/british/> (accessed: 22.12.2020).
15. Merriam-Webster. Since 1828. – URL: <https://www.merriam-webster.com/dictionary/> (accessed: 22.12.2020).
16. *Rogers, S.* Data are or data is? / S. Rogers // *The Guardian*. – 8 Jul. 2012. – URL: <https://www.theguardian.com/news/datablog/2010/jul/16/data-plural-singular> (accessed: 22.12.2020).
17. *Rundell, M.* New year, new words: Macmillan Dictionary's latest update / M. Rundell // *Macmillan Dictionary Blog*. – 2017. – URL: <http://www.macmillandictionaryblog.com/new-year-new-words-macmillan-dictionary-latest-update> (accessed: 22.12.2020).
18. *Rundell, M.* Stop the presses – the end of the printed dictionary / M. Rundell // *Macmillan Dictionary Blog*. 2013. – URL: <http://www.macmillandictionaryblog.com/bye-print-dictionary> (accessed: 20.12.2020).