
УДК 81-25

ББК 81.1

DOI: <https://doi.org/10.17308/lic.2020.2/2843>

ПРИМЕНЕНИЕ ЦИФРОВЫХ ГУМАНИТАРНЫХ ТЕХНОЛОГИЙ ДЛЯ АНАЛИЗА ВАРИАТИВНОСТИ АНГЛИЙСКОГО ЯЗЫКА (НА МАТЕРИАЛЕ ТЕЛЕСЕРИАЛОВ)

Н. В. Фомина

Финансовый университет при Правительстве Российской Федерации

APPLICATION OF DIGITAL HUMANITIES FOR THE ANALYSIS OF ENGLISH LANGUAGE VARIABILITY (CASE STUDY OF TV SERIES)

N. V. Fomina

Financial University under the Government of the Russian Federation

Аннотация: в настоящее время наблюдаются тенденции утверждать, что не существует единого варианта английского языка. Вместо этого есть несколько региональных разновидностей английского языка, каждая из которых имеет свои специфические черты. Целью данного исследования является демонстрация различий и сходств между английским и американским вариантами английского языка в отношении словарного запаса с помощью цифровых гуманитарных технологий. Исследование проводилось на материале анализа сценариев телесериалов. Для проведения настоящего исследования мы рассматривали сценарии двух телесериалов по мотивам романов “House of Cards” (в русском переводе: «Карточный домик») британского писателя Майкла Доббса. Методы, используемые в данной статье, – это цифровые гуманитарные технологии, такой способ исследования, где компьютер используется для анализа текста и его структуры. Как оказывается, любая современная наука, как практическая, так и теоретическая, построена на вычислениях. В данной работе мы опишем эволюцию применения цифровой гуманитарной науки как метода лингвистического исследования. Итак, мы сравнили и сопоставили наиболее распространенные слова в обоих сценариях; например, были проанализированы фразеологизмы, фразовые глаголы и идиомы. Рамки статьи не позволили рассмотреть также и американские диалекты. Результаты исследования подчеркивают, что, хотя британский и американский варианты английского языка во многом разнообразны, их словарные запасы не сильно отличаются. Исследование также показывает, что в настоящее время применение цифровых гуманитарных технологий стало действительно необходимым для лингвистического анализа.

Ключевые слова: британский английский, американский английский язык, цифровые гуманитарные технологии, словарный запас, телесериал.

Abstract: nowadays there is a tendency to argue that there is no single version of the English language. Instead, there are several regional variations of the English language and each of them has its own specific features. The purpose of this article is to demonstrate the differences and similarities between the English and American versions of the English language in relation to vocabulary by means of Digital Humanities technologies. The study was based on the analysis of the television series' script. In order to conduct this study we examined the scenarios of two television series based on the novel “House of Cards” by British writer Michael Dobbs. The method used in this article is Digital Humanities, a method of research where a computer is used to analyze a text and its structure. As it turns out, any modern science, either practical or theoretical, is built on computing. In this paper we describe the evolution of the application of Digital Humanities as a method of linguistic research. We compared and contrasted the most common words in both series; for example, phraseological units, phrasal verbs and idioms were analyzed. The scope of the article did not allow us to take American dialects as well. The results of the study

© Фомина Н. В., 2020



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.

emphasize that, although the British and American versions of the English language are largely diverse, their vocabulary is not very different. The study also shows that the use of Digital Humanities has now become truly necessary for linguistic analysis.

Key words: *British English, American English, digital humanities, vocabulary, television series.*

Введение

Для проведения настоящего исследования мы рассматривали сценарии двух телесериалов по мотивам романов “House of Cards” (в русском переводе: «Карточный домик») британского писателя Майкла Доббса, чтобы поддержать или опровергнуть мысль, что не существует единого варианта английского языка, а просто есть его несколько региональных разновидностей, каждая из которых имеет свои особенности. Методы, используемые в данной статье, – это цифровые гуманитарные технологии, такой способ исследования, где компьютер используется для анализа текста и его структуры. Таким образом, мы сравнили и сопоставили наиболее распространенные слова в обоих сценариях; например, были проанализированы фразеологизмы, фразовые глаголы и некоторые другие.

Применение технологий цифровой гуманитарной науки как метода лингвистического исследования

История гуманитарных наук сильно отличается от истории естественных наук. Во время первой научной революции (XVI–XVIII вв.) в естественных науках появился новый метод математического анализа, который с тех пор полностью изменил понятие науки. Любая современная наука, как практическая, так и теоретическая, построена на вычислениях.

Гуманитарные науки давно испытывают неудобства из-за отсутствия точного математического метода анализа. Будучи чрезвычайно субъективными, гуманитарные ученые не могли ни проверить, ни опровергнуть результаты своих коллег по каким-либо объективным критериям. Такие предметы, как лингвистика, филология, литературоведение, история и т. д., основывались исключительно на мнении автора.

Первые попытки формализовать гуманитарные исследования появились в XIX в. Идея о том, что разные авторы склонны использовать слова одинаковой длины, была впервые предложена британским ученым Августом де Морганом в 1851 г. [1]. В 1887 г. американский ученый Томас Менденхолл предложил метод определения авторства текста путем анализа длины слова, используемой различными авторами. Т. Менденхолл провел некоторые исследования, основанные на своем методе [2], однако составление статистики текста вручную было чрезвычайно утомительным занятием, в то время как счетчики допускали ошибки.

До появления высокоскоростных цифровых компьютеров в 1950-х гг. методы цифровой гуманитарной науки были в значительной степени неясны. Вскоре после того, как компьютеры стали широко использоваться, появились первые проекты компьютерного анализа текста. В 1949 г. Роберто Буса совместно с IBM Co. запустил проект, который считается первым в истории компьютерной лингвистики. Благодаря механизму перфокарты IBM Р. Бусе удалось проанализировать словарный состав произведений, содержащихся в книге работ Фомы Аквинского [3], и поставить данные слова в алфавитном порядке [4].

Алфавитный указатель слов Р. Бусы считается первым проектом цифровых гуманитарных наук. Однако настоящий цифровой компьютер для этого не использовался. Первым использованием такого компьютера для изучения литературы было составление алфавитного указателя изречений Библии Дж. У. Эллисона, выполненное в 1957 г. компьютером UNIVAC [5].

Таким образом, технологии цифровой гуманитарной науки запустили быстро растущий арсенал методов для каждой области исследований. Первая конференция по обработке литературных данных была организована в 1964 г. [6], первый журнал «Компьютеры и гуманитарные науки» был опубликован уже в 1966 г. [7]. Цифровые гуманитарные науки вскоре переросли область литературоведения, и первый нелитературный проект, интерактивная модель международных отношений, был разработан в 1976 г. [8].

В настоящее время технологии цифровой гуманитарной науки используются во множестве дисциплин, от лингвистики и литературоведения до истории, географии и т. д. С появлением персональных компьютеров в 1980-х гг. и Интернета в 1990-х гг. цифровые исследования стали популярными во всем мире, начиная от старомодных исследовательских институтов до частных учебных заведений. Это, в свою очередь, позитивно повлияло на вычислительные исследования в гуманитарной сфере как в качественном, так и в количественном отношении.

Телесериалы как объект анализа

Современная массовая культура существует в самых разных формах и жанрах. Возможно, сегодня мы сталкиваемся с большим количеством форм культуры, чем любое предыдущее поколение. Почему исследователь должен обращать внимание на такое явление, как телесериал?

Понятие серийного продукта культуры появилось еще в XVIII в. [9]. Художественная литература разного объема и качества публиковалась в журналах и газетах. Такие известные авторы, как Ч. Диккенс, М. Твен и Н. Г. Чернышевский, одними из первых опубликовали свои работы в серийном виде, когда одна глава следовала за другой, но уже в другом номере газеты или журнала [10]. В 1910-х гг. появилась концепция фильмов-сериалов, включающая до десяти серий; а в 1920-х гг. были выпущены в эфир первые серийные радиопередачи [11].

Телевидение вышло из экспериментальной стадии и стало применяться на практике в конце 1930-х гг., но его развитие было остановлено Второй мировой войной. После войны регулярные передачи начались как в США, так и в Европе. «Faraway Hill», как полагают, является первым телесериалом, который когда-либо транслировался, начиная с 1 октября 1946 г. [12].

С тех пор популярность телесериалов значительно возросла. Однако их разнообразие оставляло желать лучшего, и поэтому их аудитория состояла преимущественно из домохозяйек и детей, а сюжеты были в основном примитивными, состоящими из клише. Первое изменение в этой ситуации произошло в 1960-х гг., когда впервые вышли в эфир более сложные и запутанные сериалы, такие как «Доктор Кто» и «Звездный путь». Общее качество телесериалов неуклонно росло до конца XX в.

Главный прорыв произошел в конце 2000-х, когда стал доступен высокоскоростной доступ в Интернет. Теперь зрители могли выбрать контент и лучшее время для его просмотра, без необходимости подчиняться графику телевизионного вещания. Одновременно произошел важный культурный сдвиг – телесериал стал самым потребляемым продуктом средств массовой информации, объединив развлекательные свойства фильма с увлекательностью сюжета «беллетристики».

Высококачественные телесериалы сформировали новую модель потребления массовой культуры, превращая предыдущий продукт массовой культуры с ограниченным количеством вызываемых эмоций, такой как фильм или книга, в новый продукт непрерывных эмоций. Мало того что сами телесериалы стали чрезвычайно популярными, они также показали пример другим формам культуры следовать той же самой модели, приводя к появлению фильмов-продолжений и приквелов.

Таким образом, выбор телесериалов в качестве объекта исследования обусловит новое понимание человеческой культуры в таких аспектах, как эмоции, восприятие реальности, принятие и нетерпимость, а также английский язык и его использование, что мы попытаемся показать в данной статье.

История английского языка в Северной Америке

История английского языка в Северной Америке начинается в начале XVII в. с появлением британских колоний Джеймстаун и Плимут [13]. Эти колонии существовали почти двести лет. Ко времени войны за независимость в американском обществе сохранялись две противоположные тенденции: первая должна была сохранить английскую идентичность, другая – отделиться от всего британского.

В то время различия между двумя вариантами английского не были актуальны, поскольку не было единой стандартной формы языка. Ни орфография, ни речь данной эпохи не были последовательны по обе стороны Атлантики. Если фонетическое несоответствие можно было легко понять, то различия письменного и напечатанного текста необходимо было объяснять. Иногда они были связаны с отсутствием языкового стандарта, а в некоторых случаях – с методом печати того времени, так как ручной набор текста был монотонным, трудоемким и склонным к ошибкам [14].

Первое обсуждение двух вариантов английского языка можно найти в работах двух знаменитых лексикографов, британского профессора Сэмюэля Джонсона и американца Ноя Вебстера. Оба считаются первыми современными лексикографами. В то время как словарь С. Джонсона был впервые опубликован в 1755 г. [15], словарь Н. Вебстера был напечатан в 1828 г. [16].

Будучи жестоким консерватором, профессор С. Джонсон был категорически против любых изменений, включая разделение британских колоний. Он начал свою работу в 1746 г. и в течение девяти лет сумел собрать обширный корпус английских высказываний – в основном от таких авторов, как В. Шекспир, Дж. Милтон и Дж. Свифт. Конечно, С. Джонсон был не первым на пути составления словаря английского языка, но он был первым, кто систематически проанализировал широкий спектр источников [17]. Позиция С. Джонсона может рассматриваться как компромисс между описательным и предписывающим подходами к анализу языка, поскольку его попытка установить стандарт английского языка была основана на реальных примерах использования языка.

Н. Вебстер, с другой стороны, был сильным сторонником американской независимости. В 1789 г. он написал: «Как независимая нация, наша честь требует, чтобы у нас была своя система как в языке, так и в правительстве» [16]. В том же году он предложил последовательную реформу английского языка, основанную на принципах аналогии, этимологии, разума и использования. Первый словарь Н. Вебстера был опубликован в 1806 г. под названием «Сокращен-

ный словарь английского языка». Тем не менее он работал еще двадцать лет, чтобы создать новый, совершенно другой «Американский словарь английского языка», в котором он представил более 5000 новых слов, значительно различающихся в британском и американском вариантах английского языка [16].

Таким образом, словарь Н. Вебстера поставил под сомнение авторитет С. Джонсона (а также его последователей) в установлении стандарта английского языка.

Тем временем разделение между британским и американским английским росло. Для этого было несколько причин. Наиболее очевидным является обычная эволюция двух вариантов как отдельных языков. Другая причина, которую следует отметить, – это технический прогресс. В течение XIX в. новые технологии развивались быстрее, чем когда-либо прежде. Для обозначения новых изобретений общепринятой практикой среди ученых и инженеров всего мира было использование древних языков, таких как греческий или латинский. Это можно заметить в таких словах, как *telegraph* (два греческих корня) [18], *radio* (латинский корень) [Ibid.] или *automobile* (греческий и латинский корень) [Ibid.]. Тем не менее для обозначения некоторых реалий использовался английский, а не греческий или латинский язык, поэтому эти слова отличаются в британском и американском вариантах. (Сравним, например, британский завод *factory* [Ibid.] и грузовик *lorry* [Ibid.] и американский завод *plant* [Ibid.] и грузовик *truck* [Ibid.].)

Тем не менее основное разделение между британским и американским языками того времени было отображено в контексте заимствованных слов. С расширением Британской империи британский английский заимствовал слова из местных языков, таких как индийский *curry* («карри») [18] или австралийский *kangaroo* («кенгуру») [Ibid.]. В то же время в американский английский вошли слова из разных языков:

– индийского. Национальные языки обеспечивали английский главным образом словами для описания уникальных видов американской природы, таких как *raccoon* («енот») [18];

– французского. Французы были одними из первых поселенцев в Новом Свете, владея обширной территорией Луизианы (купленной США в 1803 г. и разделенной на 15 штатов США и 2 канадские провинции) [19];

– немецкого. Немецкие иммигранты были более чем многочисленны в первой половине XIX в. (в 1860 г. предполагаемое количество немцев в США составляло 1,3 миллиона) [20].

Другие влияния были вызваны испанским языком и идишем.

К 1870 г. американский английский стал независимым от стандарта английского языка. Более того, в американском английском появилось несколько региональных диалектов [21]. После гражданской войны территориальная экспансия расширила границы диалекта. Так как наше исследование ограничено рамками статьи, то американские диалекты не будут являться предметом рассмотрения.

В конце XIX и в течение XX в. вышеупомянутые процессы продолжались. Начав с одного источника, американский и британский варианты английского языка развивались отдельно, поэтому даже такое слово, как «осень», представляя из себя германо-заимствованное понятие *fall* [18], используется в американском английском, тогда как британский английский предпочитает заимствованное французское *autumn* [Ibid.]. Американский английский был первым диалектом, который развивался так независимо от оригинальной версии. В фонетической структуре, правописании, лексике, грамматике и синтаксисе американский английский приобрел огромное количество особенностей.

Различия между американским и британским вариантами английского языка

Первое и наиболее известное различие между двумя вариантами – это произношение. Звуки с различным произношением хорошо известны и классифицированы:

– британский английский не является рутическим, это означает, что звук «г» не произносится перед согласным или в конце слова. Тем не менее он был рутическим до конца 1800-х гг. Поскольку разделение между двумя вариантами произошло в 1600-х гг., американский английский оставался более рутическим;

– в британском английском «t» обычно произносится четко, в то время как в американском английском оно может произноситься как «d» или исчезать. Когда между двумя гласными звучит «t», оно часто произносится как «d». «t» в американском варианте имеет тенденцию исчезать после назальных звуков, таких как «m», «n» и «ng»;

– с XVIII в. «a» в британском английском изменилось с плоского на удлиненное (плоское «a», как у *map*, удлиненное «a», как у *father*). Это изменение повлияло на слова, в которых звуки «a» появились перед «f», «sk», «sp», «st», «ss», «th» и «n», за которыми следуют определенные согласные. Как и в случае с «r», разделение между вариантами произошло раньше, и американский английский по-прежнему пользуется плоским «a». Тем не менее эта особенность влияет менее чем на 250 слов, используемых ежедневно;

– произношение «о» также отличается. В Британии «о» открыто и произносится с округлыми губами и языком в задней части рта. В Америке оно утратило свою округлость и стало звучать очень похоже на «а» в *father*. Это пример общей американской тенденции к нейтрализации гласных звуков, в то время как английские гласные имеют тенденцию сохранять свою резкость;

– «и» произносится одинаково по обе стороны Атлантики, т. е. с воображаемым «у», вставленным перед ним. В США, однако, обычное произношение без звука «у».

Эти различия дополняются списком слов, которые в обоих вариантах произносятся по-разному, но не подчиняются какому-либо определенному правилу. Однако число этих слов мало, и вариации в произношении можно понимать, как случайные изменения языка [22].

Несоответствия в правописании двух вариантов английского языка так же общеизвестны, как и произношение. Уже упоминалось, что различия в правописании были основой конфликта между С. Джонсоном и Н. Вебстером. С тех пор британский и американский английский приняли разные системы орфографии слов, которые позволяют отличать писателей соответствующих стран.

Суффиксы

Суффиксы, заимствованные у французов, такие как *-ouг* (*-eug* по-французски) и *-tre*, достаточно хорошо распространены в британском английском. Американский английский заменил их на *-ог* и *-ter* соответственно.

Суффикс *-ise* британского английского соответствует *-ize* американского английского. Тем не менее нет четких критериев для различия. Некоторые британцы считают, что такие слова, как *apologise* или *legalise*, могут быть написаны как *apologize* и *legalize*, в то время как другие осуждают это написание. Кроме того, есть слова, которые пишутся с *-ize* в обоих вариантах (*seize*), а также некоторые слова, которые могут быть написаны с *-ise* и *-ize* в обоих вариантах (*advertise/advertize*).

Слова греческого происхождения, содержащие группу «ое», обычно пишутся американцами с одной буквой «е» (*encyclopedia/encyclopaedia*). Аналогично сочетание «ае» британского английского пишется просто «а» в Америке (*anemia/anaemia*).

Несколько глаголов, состоящих из двух слогов и имеющих ударение на втором слоге, пишутся с одним «l» в Великобритании, но с двумя «ll» в Америке (*fulfil/fulfill*). В американском правописании, когда к слову добавляется суффикс типа *-ing*, *-ed* или *-er*, конечный согласный удваивается только в том случае,

если ударение падает на второй слог корневого слова (*to patrol* дает *patrolling* и *patrolled*, но *to travel* становится *traveling* и *traveler*).

Другие варианты также включают дополнительные или разные буквы в слове, наиболее заметный случай – американский *aluminum* против британского *aluminium*.

Последнее, на что следует обратить внимание, это различие дефисов и пробелов в сложных словах. Как правило, американский английский имеет тенденцию отбрасывать дефис намного быстрее, чем британский английский (*bookkeeper/book-keeper* [23], *makeup/make-up*, *breakdown/break-down*).

Нарушения грамматики британского и американского английского языка не слишком серьезны, чтобы вызвать недопонимание среди говорящих. Тем не менее они значительно превышают несоответствия, имеющиеся только в британском или только в американском английском.

Глаголы

Определенное количество глаголов в простом прошедшем времени могут быть как правильными, так и неправильными. В американском английском обычно предпочитается правильная форма, а в британском английском – неправильная. Такие глаголы как *fit*, *quit* и *wet* являются неправильными в британском английском, но правильными в американском.

Времена

В американском английском времена прошедшего простого (Past Simple) и настоящего совершенного времени (Present Perfect) часто взаимозаменяемы, в то время как в британском английском можно использовать только настоящее совершенное время. Когда действие в прошлом имеет результат в настоящем, т. е. сейчас, используется настоящее совершенное время. То же самое верно для слов *just*, *already*, *yet*, *ever* и *never*, когда речь идет о периоде времени, который продолжается до сих пор.

Вспомогательные и модальные глаголы

В британском английском языке глагол *shall* для местоимений «я» и «мы» можно заменить на *will*, указывая на будущее, а в американском английском это не принято.

В обоих вариантах английского языка можно использовать глаголы *can* и *could* с глаголами восприятия, но такое явление гораздо чаще встречается в британском английском.

В сослагательных наклонениях, например, после глаголов, таких как «*suggest*», «*recommend*», «*demand*»,

«insist» и т. д., в британском английском *should* используется чаще, чем в американском.

В британском английском, но не в американском языке, «do» может использоваться в качестве замены глагола после вспомогательного глагола.

Определенные наречия, такие как *sometimes, always, never, often, definitely, certainly*, обычно помещаются между вспомогательными и другими глаголами. Но если необходимо подчеркнуть вспомогательный глагол, перед ним ставится большинство наречий среднего положения. В британском английском эта вторая конструкция всегда выделена голосом, в то время как в американском английском наречие часто помещается перед вспомогательным глаголом, даже когда нет намерения выделять.

В британском английском собирательные существительные могут употребляться с глаголом как в единственном, так и во множественном числе. В американском английском такие существительные обычно принимают глагол в единственном числе. То же самое относится к определенным личным именам, таким как названия стран или компаний [24].

Исследование

Для проведения данного исследования нам потребовалось найти такой материал, который, будучи посвящен одной теме, был бы выпущен как в американской, так и в британской версиях. На наш взгляд, наиболее удачным примером является сериал «Карточный домик».

Британский политик и писатель Майкл Доббс написал серию из трех романов «Карточный домик» (1989), «Играть в короля» (1992) и «Окончательный вариант» (1994), в которых мы встречаем персонажа по имени Фрэнсис Уркварт. Он – безжалостный британский политик, который борется за пост премьер-министра Соединенного Королевства. Эти романы получили большой успех, став бестселлерами [25].

В 1990 г. компания Би-би-си снимала мини-сериал по мотивам первого романа карточной трилогии, который носил то же название, что и роман, и состоял из четырех эпизодов по девяносто минут каждый. В 1993 г. вышел фильм «*To Play the King*», снятый в том же стиле, а за ним последовал «*The Final Cut*» в 1995-м. Как и книги, сериал имел огромный успех у британской аудитории [26].

Почти двадцать пять лет спустя, в 2013 г., американская компания Netflix приобрела права на экранизацию американской версии «*House of Cards*». В отличие от британской версии, новый сериал был лишь частично основан на оригинальной трилогии: место действия было изменено на США, главные герои получили новые имена, сюжет также эволюционировал.

На сегодняшний день было выпущено пять полных сезонов американской версии «*House of Cards*», шестой сезон начался осенью 2018 г. Этот сериал приобрел еще больший успех, чем английская версия, выиграв семь премий Эмми в номинации Primetime, два Золотых глобуса и другие награды [27].

Главной задачей настоящего исследования является нахождение различий между английским и американским вариантами английского языка. Поскольку методы, использованные нами для проведения исследования, были цифровые гуманитарные науки, нам потребовалось проанализировать две коллекции текстов достаточного объема. Таким образом, были собраны два корпуса, далее называемые корпус А (американский вариант) и корпус В (британский вариант).

Первая часть исследования была проведена по общей статистике корпусов, представленной в табл. 1.

Т а б л и ц а 1

Общая статистика двух корпусов

Характеристика	Корпус А	Корпус В
Общее количество слов	331 942	67 289
Общее количество предложений	54 769	11 444
Средняя длина слова	4,19	4,20
Средняя длина предложения	6,06	5,87
Общее количество существительных	46 271	9455
Общее количество глаголов	91 918	8332
Общее количество прилагательных	13 839	4335

Примечание 1: поскольку корпуса значительно различаются по объему, другие данные для сравнения представлены в процентах, а не в целых числах.

Как показано в табл. 1, если брать предварительные характеристики – средняя длина слова и предложения, то эти два варианта английского языка очень близки. Несмотря на то, что многие слова в этих двух вариантах различаются, тем не менее средняя длина слова остается неизменной. Это доказывает, что общий объем словарного запаса не изменился.

Эта идея была дополнительно проверена на следующем этапе исследования – составлении списков частотности для наиболее часто используемых слов в корпусах А и В. Список приведен в табл. 2.

Таблица 2 ясно демонстрирует сходство между американским и британским вариантами английского языка. Не только 6 из 10 слов совпадают в обоих списках, но и слово «мой» одинаково пишется и произносится в американской и британской версиях.

Стоит отметить, что, несмотря на сходство сюжетов двух телесериалов, единственное слово, касающееся политики, – «Президент» – встречается наиболее часто в словах корпуса А.

Т а б л и ц а 2

Список частотности употребления для 10 наиболее часто используемых слов каждого корпуса

	Корпус А	Корпус В
1	Have	What
2	What	Have
3	Can	All
4	My	My
5	If	Yes
6	Know	Can
7	President	Know
8	Just	Think
9	All	Very
10	Well	As

Примечание 2: список исключений функциональных слов, таких как «I», «is», «to» и т. д., использовался, чтобы убрать их из списка частотности, поскольку эти слова чаще всего встречаются в обоих вариантах и не дают возможности изучить значимые слова. Полный список исключений приведен в приложении.

Третий этап исследования заключался в составлении списка, в котором длина слов сравнивалась с частотой их использования. Результаты представлены в табл. 3 и на рис. 1.

Т а б л и ц а 3

Сравнение отношения длины к частотности употребления слова

Количество букв в слове	Корпус А		Корпус В	
	Число	Процент	Число	Процент
2	63 614	19,16 %	13 075	19,43 %
3	74 126	22,33 %	14 638	21,75 %
4	68 299	20,58 %	12 238	18,19 %
5	34 016	10,25 %	7244	10,77 %
6	21 697	6,54 %	4660	6,93 %
7	16 874	5,08 %	3780	5,62 %
8	10 252	3,09 %	2554	3,80 %
9	8721	2,63 %	1515	2,25 %
10	4096	1,23 %	1088	1,62 %
11	1837	0,55 %	446	0,66 %
12	964	0,29 %	299	0,44 %
13	501	0,15 %	130	0,19 %
14	228	0,07 %	44	0,07 %
15	51	0,02 %	12	0,02 %
16	19	0,01 %	3	0,00 %

Примечание 3: однобуквенные слова не были включены в список, так как в большинстве случаев это был бы неопределенный артикль «a». Большое количество трехбуквенных слов связано с большим объемом определенного артикля «the», который, к сожалению, невозможно удалить с помощью существующих методов исследования.

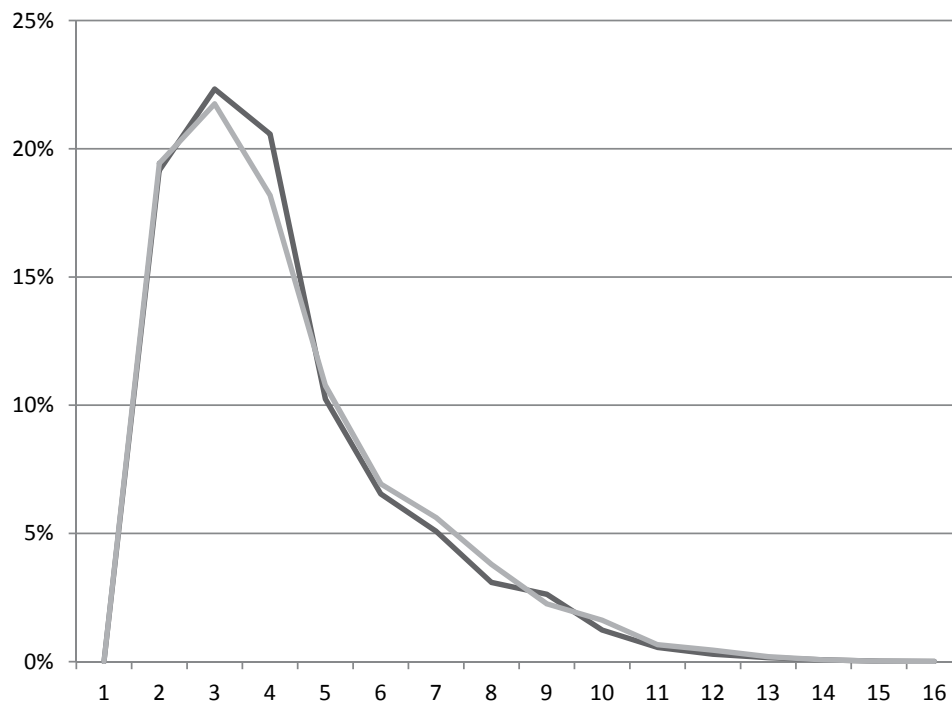


Рис. 1. Сравнение соотношения длины и частотности употребления слова:
черная кривая – корпус А; серая кривая – корпус Б

На рис. 1 хорошо видно, что обе кривые очень близки друг к другу, а также сходны данные частотности употребления, когда они представлены в процентах. Таким образом, можно с уверенностью сделать вывод о том, что при анализе распределения частотности слов нет существенной разницы между английским и американским вариантами английского языка.

Четвертый этап исследования касался частей речи. Части речи были автоматически помечены в обоих корпусах, а затем полученные результаты были проанализированы статистически. Данные представлены в табл. 4 и на рис. 2.

Т а б л и ц а 4

Сравнение частей речи, используемых в двух корпусах

Характеристика	Корпус А		Корпус В	
	Число	Процент	Число	Процент
Общее количество существительных	46 271	14 %	9455	14 %
Общее количество глаголов	45 681	14 %	8332	12 %
Общее количество прилагательных	13 839	4 %	4335	6 %

Как и в предыдущих случаях, полученные данные демонстрируют удивительное сходство между двумя вариантами английского языка.

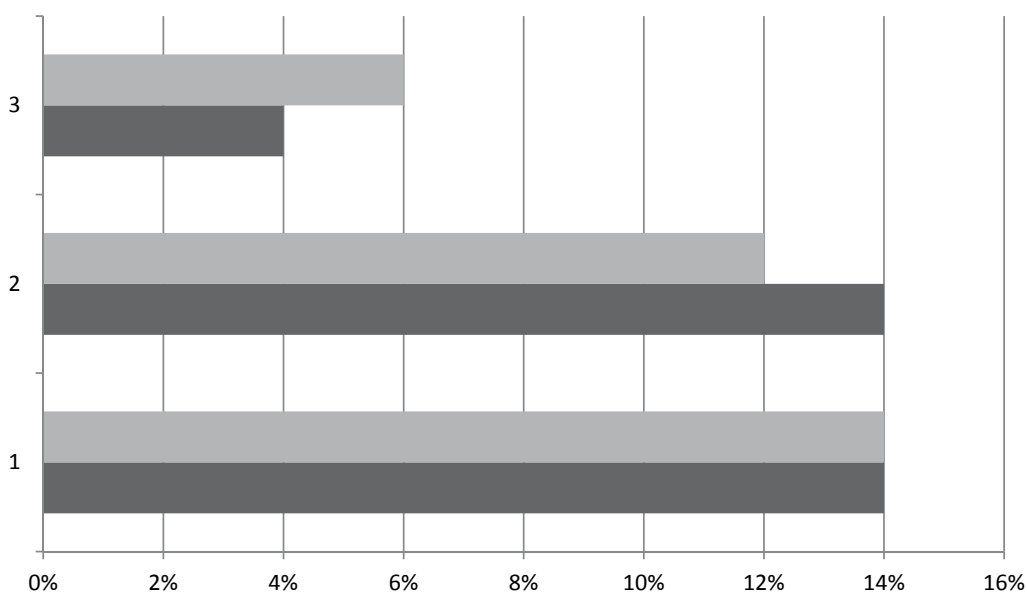


Рис. 2. Сравнение частей речи, используемых в двух корпусах:

1 – существительные; 2 – глаголы; 3 – прилагательные; черный – корпус А; серый – корпус Б

В ходе исследования мы также имели возможность проверить предположение ученого Д. Байбера, указывает ли распространенность прилагательных в британском английском языке на то, что этот вариант является более сложным и богатым, чем американский [28]. Однако полученная разница в 2 % между английским и американским вариантами английского языка не позволяет нам считать эту идею доказуемой на данном этапе исследования.

Заключительная часть исследования была посвящена анализу артиклей в двух вариантах английского языка. Согласно исследованию М. Джокерса, американский английский отдает предпочтение определенному артиклю «the», в то время как британский английский избегает его во многих случаях [29]. Полученная нами статистика по двум корпусам представлена в табл. 5 и на рис. 3.

Т а б л и ц а 5

Статистика использования артиклей в двух корпусах

Артикль	Корпус А		Корпус В	
	Число	Процент	Число	Процент
«the»	18 102	5 %	2269	3 %
«a»	7420	2 %	1484	2 %

Однако, как и в случае с частями речи, полученный процент разницы слишком низок, чтобы сделать однозначный вывод о том, предпочитает ли американский английский определенный артикль «the», и избегает ли его в таких же случаях британский английский. Полученные результаты не могут быть достаточными для утверждения, является ли это тенденцией или статистической аномалией. Нужны дальнейшие исследования с большим количеством текстов, чтобы это прояснить, так как на данный момент выводы будут преждевременными.

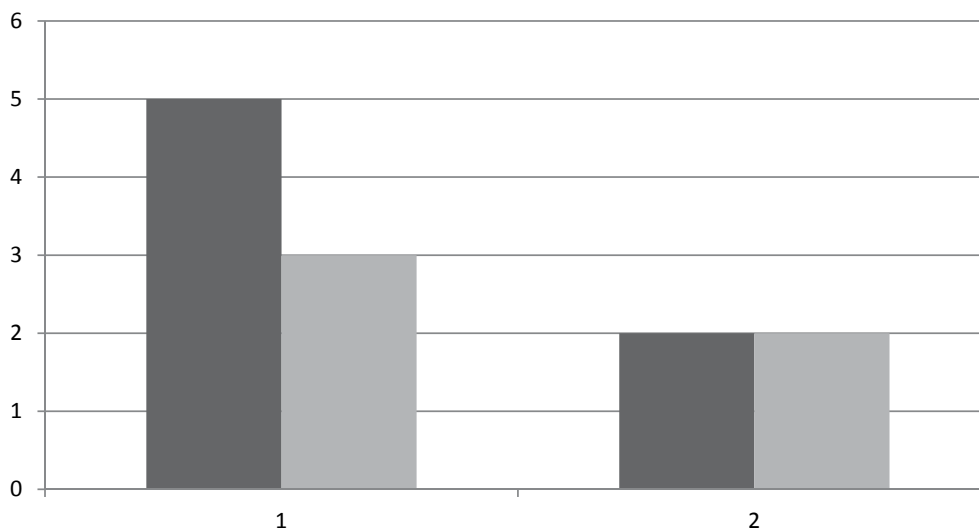


Рис. 3. Статистика использования артиклей в двух корпусах: 1 – «the»; 2 – «a»; черный – корпус А; серый – корпус Б

Заключение

В этом исследовании мы рассмотрели проблему цифровой методологии, которая будет в дальнейшем широко использоваться в лингвистических исследованиях. Мы проанализировали историю и общее современное состояние технологии цифровой гуманитарной науки, представили схему американского варианта истории английского языка, искали теоретическую основу для исследования при изучении результатов нецифровых исследований в этой конкретной области. Наконец, мы провели на наш взгляд, обширный практический анализ двух лингвистических корпусов.

Из результатов, описанных в данной статье, мы можем сделать вывод, что нет существенной разницы между словарными составами английского и американского вариантов английского языка ни в одном из рассмотренных аспектов. Некоторые результаты оставляют место для обсуждения; однако требуются дальнейшие исследования для определения правильности какой-либо гипотезы.

Мы также старались наглядно продемонстрировать, что цифровые гуманитарные методы вполне применимы к проблемам изучения языковой вариативности. Мы считаем это чрезвычайно важным для дальнейших исследований.

В заключение хотелось бы отметить, что данная работа является одним из первых шагов исследования, которое еще не закончено. Считаем, что эти результаты пока не могут и не должны применяться ко всему английскому языку, так как нами был проанализирован только жанр телесериала и только один его вид. К тому же, несмотря на то, что сериал объединяет словарный запас литературного произведения с выразительными методами кино, звучащие в нем

фразы включают в себя устные высказывания, строго предписанные авторами текста.

Нам видится несколько направлений для дальнейшего исследования. Во-первых, данные корпусов можно было бы расширить, включив в них различные телесериалы, произведенные в Америке и Великобритании. Это позволило бы нам изучить вопросы этого исследования, но на гораздо более широкой основе. Во-вторых, мы считаем, что телесериал каждой страны можно сравнить с корпусом других текстов. В этом случае мы полагаем, что много интересных результатов можно было бы получить из сравнения сценариев сериалов как в устной, так и в письменной форме. Затем можно провести сравнение британских и американских телесериалов еще раз, демонстрируя влияние устной или письменной формы языка.

Мы также считаем, что можно было бы сравнить британские и другие диалекты английского языка с сериалами и другими текстами из таких стран, как Австралия, Канада, Новая Зеландия и т. д.

Наконец, мы хотели бы отметить, что это исследование стало возможным из-за появления технологии цифровой гуманитарной науки, так как частотный анализ текста был почти невозможен до появления высокоскоростных электронных компьютеров. Это исследование было проведено достаточно быстро и эффективно только потому, что в нашем распоряжении были такие мощные инструменты. Мы хотели бы также заметить, что в сегодняшнем ускоряющемся мире принятые методы гуманитарных исследований почти полностью устарели. Современные ученые не должны игнорировать инструмент такой силы и универсальности, лежащий прямо перед глазами.

Приложение

Список слов-исключений, используемых при частотном анализе корпусов А и В

you	on	re	ll	were
we	s	are	will	with
they	t	is	shall	at
he	your	and	do	about
she	his	of	don	get
it	her	that	me	so
a	no	this	for	
the	not	mine	but	
to	am	their	be	
in	m	our	was	

ЛИТЕРАТУРА

1. Mendenhall T. C. The Characteristic Curves of Composition // Science. 1887. IX (214). Pp. 237–248.
2. Mendenhall T. C. A Mechanical Solution of a Literary Problem // The Popular Science Monthly. 1901. LX (7). Pp. 97–105.
3. Concordance // The Oxford English Dictionary. URL: <http://www.oed.com/>
4. Busa R. IBM Adapt Punched Card Tabulating to Sort Words in a Literary Text // The Origins of Humanities Computing. The History of Information. URL: <http://www.historyofinformation.com/>
5. J. W. Ellison Issues the First Computerized Concordance of the Bible. The History of Information. URL: <http://www.historyofinformation.com/>
6. Arader, Parrish & Bessinger Organize the First Humanities Computing or Digital Humanities Conference The History of Information. URL: <http://www.historyofinformation.com/>
7. Joseph Raben Finds “Computers and the Humanities”, the First Humanities Computing Journal The History of Information. URL: <http://www.historyofinformation.com/>
8. The World Event/Interaction Survey: A Pioneering Application of Systems Theory to International Relations. The History of Information. URL: <http://www.historyofinformation.com/>
9. Law G. Serials and the Nineteenth-Century Publishing Industry // Dictionary of Nineteenth-Century Journalism. London : Academia Press, 2009.
10. Patten R. Serial literature Oxford Reader’s Companion to Dickens. URL: <http://www.oxfordreference.com/view/10.1093/acref/9780198662532.001.0001/acref-9780198662532-e-0391/>
11. Lahue K., Kalton C. Continued Next Week : A History of the Moving Picture Serial. Norman : University of Oklahoma Press, 1969.
12. Weinstein D. The Forgotten Network : DuMont and the Birth of American Television. Philadelphia : Temple University Press, 2004.
13. Crystal D. English as a global language. Second edition. New York : Cambridge University Press, 2003.

14. Simpson D. American English to 1865 // A Companion to the History of the English Language / edited by Momma H., Matto M. Chichester : Wiley-Blackwell, 2008.
15. Johnson S. A Dictionary of the English Language. London : W. Strahan, 1755.
16. Webster N. American Dictionary of the English Language. New York : S. Converse, 1828.
17. Hitchings H. Dr. Johnson’s Dictionary : The Extraordinary Story of the Book That Defined the World. London : John Murray, 2005.
18. The Oxford English Dictionary. URL: <http://www.oed.com/>
19. Baugh A. C., Cable T. A History of the English Language. London : Routledge, 2002.
20. Germans in America. European Reading Room, Library of Congress. URL: <https://www.loc.gov/rr/european/imde/germchro.html/>
21. Wolfram W. American English since 1865 // A Companion to the History of the English Language / edited by Momma H., Matto M. Chichester : Wiley-Blackwell, 2008.
22. Darragh G. A to Zed, A to Zee. A Guide to the Differences between British and American English. Imprenta Berekintza, 2000.
23. Davies C. Divided by Common Language. A Guide to British and American English. New York : Cambridge University Press, 2003.
24. One language, two grammars : differences between British and American English / edited by G. Rohdenburg, J. Schluter. New York : Cambridge University Press, 2009.
25. House of Cards. Website of Michael Dobbs – Author and British politician. URL: <http://www.michaeldobbs.com/>
26. House of Cards. BBC One. URL: <http://www.bbc.co.uk/programmes/articles/1F1TT8ng10xQn3fTrdgy2RM/house-of-cards-trilogy/>
27. House of Cards. Netflix Official Site. URL: <https://www.netflix.com/ru/title/70178217/>
28. Biber D. A. Textual Comparison of British and American Writing // American Speech. 1987. Vol. 62, No. 2. Pp. 99–119.
29. Jockers M. L. Macroanalysis : Digital Methods and Literary History. Chicago : University of Illinois Press, 2013.

REFERENCES

1. Mendenhall T. C. The Characteristic Curves of Composition. Science. 1887. IX (214). Pp. 237–248.
2. Mendenhall T. C. A Mechanical Solution of a Literary Problem. The Popular Science Monthly. 1901. LX (7). Pp. 97–105.
3. Concordance. The Oxford English Dictionary. Available at: <http://www.oed.com/>
4. Busa R. IBM Adapt Punched Card Tabulating to Sort Words in a Literary Text. The Origins of Humanities Computing. The History of Information. Available at: <http://www.historyofinformation.com/>
5. J. W. Ellison. Issues the First Computerized Concordance of the Bible. The History of Information. Available at: <http://www.historyofinformation.com/>

6. Arader, Parrish & Bessinger Organize the First Humanities Computing or Digital Humanities Conference The History of Information. Available at: <http://www.historyofinformation.com/>
7. Joseph Raben Finds “Computers and the Humanities”, the First Humanities Computing Journal The History of Information. Available at: <http://www.historyofinformation.com/>
8. The World Event/Interaction Survey: A Pioneering Application of Systems Theory to International Relations. The History of Information. Available at: <http://www.historyofinformation.com/>
9. Law G. Serials and the Nineteenth-Century Publishing Industry. G. Law. Dictionary of Nineteenth-Century Journalism. London: Academia Press, 2009.
10. Patten R. Serial literature Oxford Reader’s Companion to Dickens. R. Patten. Available at: <http://www.oxfordreference.com/view/10.1093/acref/9780198662532.001.0001/acref-9780198662532-e-0391/>
11. Lahue K., Kalton C. Continued Next Week: A History of the Moving Picture Serial. Norman: University of Oklahoma Press, 1969.
12. Weinstein D. The Forgotten Network: DuMont and the Birth of American Television. Philadelphia: Temple University Press, 2004.
13. Crystal D. English as a global language. Second edition. New York: Cambridge University Press, 2003.
14. Simpson D. American English to 1865. D. Simpson. A Companion to the History of the English Language. Ed. by Momma H., Matto M. Chichester: Wiley-Blackwell, 2008.
15. Johnson S. A Dictionary of the English Language. S. Johnson. London: W. Strahan, 1755.
16. Webster N. American Dictionary of the English Language. N. Webster. New York: S. Converse, 1828.
17. Hitchings H. Dr. Johnson’s Dictionary: The Extraordinary Story of the Book That Defined the World. H. Hitchings. London: John Murray, 2005.
18. The Oxford English Dictionary. Available at: <http://www.oed.com/>
19. Baugh A. C. A History of the English Language. A. C. Baugh and T. Cable. London, Routledge, 2002.
20. Germans in America. European Reading Room, Library of Congress. Available at: <https://www.loc.gov/rr/european/imde/germchro.html/>
21. Wolfram W. American English since 1865. W. Wolfram. A Companion to the History of the English Language. Ed. by Momma H., Matto M. Chichester: Wiley-Blackwell, 2008.
22. Darragh G. A to Zed, A to Zee. A Guide to the Differences between British and American English. G. Darragh. Imprenta Berehintza, 2000.
23. Davies C. Divided by Common Language. A Guide to British and American English. C. Davies. New York, Cambridge University Press, 2003.
24. One language, two grammars: differences between British and American English. Ed. by G. Rohdenburg, J. Schluter. New York, Cambridge University Press, 2009.
25. House of Cards. Website of Michael Dobbs. Available at: <http://www.michaeldobbs.com/>
26. House of Cards. BBC One. Available at: <http://www.bbc.co.uk/programmes/articles/1F1TT8ng10xQn3fTrdgy2RM/house-of-cards-trilogy/>
27. House of Cards. Netflix Official Site. Available at: <https://www.netflix.com/ru/title/70178217/>
28. Biber D. A. Textual Comparison of British and American Writing. American Speech. 1987. Vol. 62, No. 2. Pp. 99–119.
29. Jockers M. L. Macroanalysis: Digital Methods and Literary History. M. L. Jockers. Chicago, University of Illinois Press, 2013.

Финансовый университет при Правительстве Российской Федерации

Фомина Н. В., кандидат филологических наук, старший преподаватель Департамента языковой подготовки

E-mail: magnitnadya@gmail.com

Поступила в редакцию 21 января 2020 г.

Принята к публикации 25 марта 2020 г.

Для цитирования:

Фомина Н. В. Применение цифровых гуманитарных технологий для анализа вариативности английского языка (на материале телесериалов) // Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация. 2020. № 2. С. 75–85. DOI: <https://doi.org/10.17308/lic.2020.2/2843>

Financial University under the Government of the Russian Federation

Fomina N. V., Candidate of Philology, Senior Lecturer of the Language Training Department

E-mail: magnitnadya@gmail.com

Received: 21 January 2020

Accepted: 25 March 2020

For citation:

Fomina N. V. Application of digital humanities for the analysis of english language variability (case study of TV series). Proceedings of Voronezh State University. Series: Linguistics and Intercultural Communication. 2020. No. 2. Pp. 75–85. DOI: <https://doi.org/10.17308/lic.2020.2/2843>