

О.О. БОРИСКИНА, К.М. ШИЛИХИНА*
КОРПУСНЫЕ ИССЛЕДОВАНИЯ
ПОЛИТИЧЕСКОГО ДИСКУРСА В ЛИНГВИСТИКЕ

Аннотация. В статье обсуждается возможность использования корпусных методов для исследования современной политической коммуникации. В работе рассматриваются пять вариантов применения корпусов различного объема: на материале корпуса NOW (News on the Web) описывается использование центрального для политического дискурса понятия *politics* в 20 вариантах английского языка, проводится анализ метафорической сочетаемости слова *politics* в английском языке. Третий пример использования корпуса NOW – изучение неологизмов, образованных от имен и фамилий двух кандидатов на пост президента США. Целью анализа является изучение тональности новостного политического дискурса, в котором освещалась предвыборная президентская кампания в США в 2016 г.

На материале двух микрокорпусов, включающих тексты предвыборных выступлений Хиллари Клинтон и Дональда Трампа, показаны возможности выделения и анализа наиболее частотной лексики и ключевых слов. Цель подобного анализа – выявление риторических способов воздействия на аудиторию.

Ключевые слова: политический дискурс; языковой корпус; метафора; неологизм; частотность; ключевое слово.

* **Борискина Ольга Олеговна**, доктор филологических наук, доцент, профессор кафедры теоретической и прикладной лингвистики Воронежского государственного университета, e-mail: olboriskina@gmail.com.

Шилихина Ксения Михайловна, доктор филологических наук, доцент, профессор кафедры теории перевода и межкультурной коммуникации, Воронежский государственный университет, e-mail: shilikhina@rgph.vsu.ru.

Boriskina Olga, Voronezh state university (Voronezh, Russia), e-mail: olboriskina@gmail.com.

Shilikhina Ksenia, Voronezh state university (Voronezh, Russia), e-mail: shilikhina@rgph.vsu.ru.

O.O. Boriskina, K.M. Shilikhina
Corpus methods in the political discussion in linguistics

Abstract. The paper describes the potential of corpus methods in the political discourse studies. The paper illustrates the advantages of corpus methods by providing five examples of corpus-based studies. The data extracted from The Corpus of NOW (News on the Web) is processed in order to compare the use and linguistic categorization of the concept *politics* in twenty varieties of the English language. The third example covers the emergence and usage of neologisms derived from Donald Trump and Hillary Clinton's names during 2016 presidential campaign in the USA. The study of the new words can contribute to sentiment analysis of the mass-media political discourse during election campaign in the USA.

Two small-size corpora – transcripts of D. Trump and H. Clinton's 2016 election campaign speeches – are used for the analysis of the most frequent words and key words that reveal rhetorical strategies used by the two politicians to influence the audiences.

Keywords: political discourse; language corpus; metaphor; neologism; frequency; key word.

Введение

Политическая коммуникация, равно как и политика вообще, традиционно исследуется с использованием инструментария и методов политической науки. Однако в эпоху цифровых гуманитарных наук (Digital Humanities) актуальным становится вопрос о необходимости соединения практики изучения социально-политических реалий с новыми информационно-коммуникационными технологиями сбора и анализа данных, с одной стороны, с другой – с результатами корпусных лингвистических исследований в области политического дискурса. Эти три составляющие – политическая наука, лингвистика и IT-методология – дают возможность взглянуть на, казалось бы, достаточно изученное явление в ином ракурсе, что будет способствовать приращению знания междисциплинарного характера.

Изучение политического дискурса на материале макрокорпусов

Политологи, социологи и лингвисты все чаще обращаются к макрокорпусам¹ за материалом для своих исследований, и это по-

¹ Под макрокорпусом в данной работе понимается объемный массив текстов, содержащий сотни миллионов словоупотреблений, в отличие от специали-

степенно становится частью исследовательской практики в гуманитарных науках [Wiedemann, 2013; Khokhlova, 2016], поскольку корпусный подход делает результаты анализа более эмпирически релевантными.

Ярким примером языкового макрокорпуса является ежедневно пополняемая коллекция новостных текстов, «обитающих» в интернет-СМИ (News on the Web, или сокращенно NOW Corpus) от создателя языковых корпусов М. Дэвиса [Davis, 2013]. В корпусе, который на 01/2017 насчитывает около 3,8 млрд слов, представлены 20 региональных или национальных вариантов английского языка: британский (GB), ирландский (IE), американский (US), канадский (CA), австралийский (AU), новозеландский (NZ), ямайский (JM), индийский (IN), пакистанский (PK), филиппинский (PH), сингапурский (SG), шриланкийский (LK), гонконгский (HK), малайзийский (MY), бангладешский (BG), кенийский (KE), южноафриканский (ZA), нигерийский (NG), ганский (GH), танзанийский (TZ). Такая палитра языковых вариантов придает дополнительное измерение исследованию политического дискурса: в результате можно наблюдать специфику функционирования языковых единиц в различных регионах мира.

Рассмотрим базовое понятие политической науки – политику. Под политикой в широком смысле этого слова понимают управленческую деятельность государственного уровня. Какие новые возможности открывают корпусные технологии для лингвистической оценки базового понятия политической науки?

Во-первых, сравнение частоты употребления слова *politics* в англоязычном новостном дискурсе (NOW Corpus) за последние 6 лет по полугодиям (табл. 1, рис. 1) показывает, что коммуникативное обращение слова не только находится на стабильно высоком уровне, но и имеет тенденцию к росту независимо от варианта языка, что вполне ожидаемо в свете современных политических процессов. Однако при этом неожиданным представляется тот факт, что чаще о политике говорят не в Америке или Европе, а в Африке.

зированных микрокорпусов, небольших по объему коллекций текстов объемом несколько тысяч словоупотреблений.

Таблица 1

**Количество употреблений имени *politics*
по полугодиям с 2010 по 2016 г.**

Полугодия	Общая частота употребления слова в корпусе	Size размер к орпуса	Относительная частота употребления слова (количество словоупотреблений на миллион)
2010 A	9328	115,2	81,00
2010 B	11639	129,2	90,09
2011 A	12656	145,1	87,25
2011 B	13126	160,0	82,05
2012 A	16476	185,1	89,00
2012 B	17005	186,4	91,23
2013 A	17364	196,9	88,20
2013 B	19262	204,9	94,00
2014 A	20463	209,8	97,51
2014 B	20164	219,9	91,68
2015 A	20407	223,8	91,19
2015 B	20088	289,1	90,24
2016 A	70435	684,5	102,90
2016 B	81469	759,0	107,34
Total Bcero	355882		

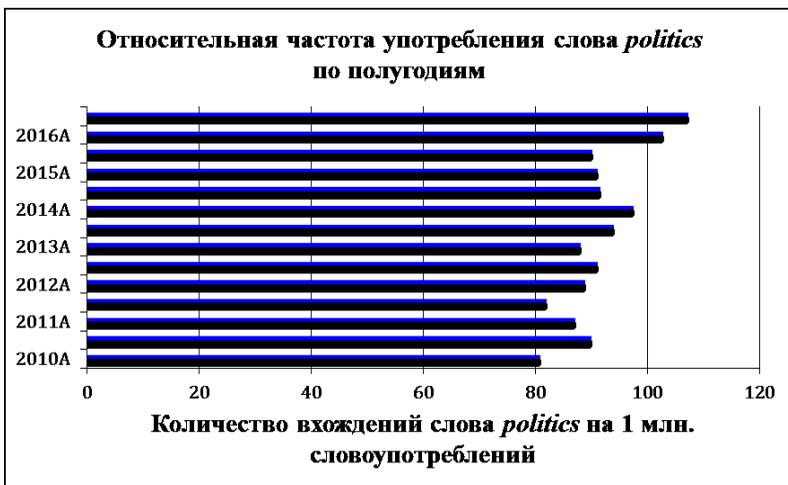


Рис. 1
**Количество употреблений имени *politics*
по полугодиям с 2010 по 2016 г.**

Действительно, несмотря на то что в абсолютных величинах максимальное число примеров с именем *politics* характерно для английского языка в США (65 124 млн словоупотреблений, рис. 2, вторая колонка), наибольшее количество словоупотреблений в относительных величинах (рис. 2, четвертая колонка, встречаемость словоформы в 1 млн словоупотреблений) было обнаружено в вариантах английского языка Нигерии, Кении и Ганы.

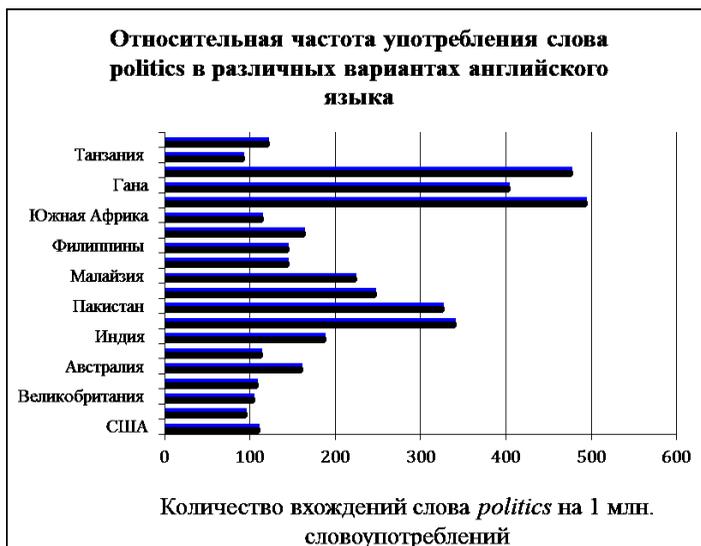


Рис. 2
Соотношение употребления имени *politics*
в вариантах английского языка

Таблица 2

Соотношение употребления имени *politics*
в вариантах английского языка

Страны	Общая частота употребления слова в корпусе	Size размер корпуса	Относительная частота употребления слова (количество словоупотреблений на миллион)
1	2	3	4
США	65 024	585,8	110,99
Канада	39 624	409,7	96,72
Великобритания	45 470	431,1	105,48
Ирландия	16 153	147,3	109,67
Австралия	28 788	178,7	161,07

Продолжение таблицы 1

1	2	3	4
Новая Зеландия	6561	57,3	114,57
Индия	45 453	240,8	188,76
Шри-Ланка	3919	11,5	341,08
Пакистан	15 473	47,2	327,65
Бангладеш	2901	11,7	247,88
Малайзия	7621	33,9	224,67
Сингапур	6663	45,8	145,43
Филиппины	9770	67,1	145,62
Гонконг	1304	7,9	164,64
Южная Африка	18 994	165,4	114,87
Нигерия	22 648	45,7	495,21
Гана	7801	19,3	404,86
Кения	8584	17,9	478,32
Танзания	595	6,4	93,19
Ямайка	2536	20,7	122,38

На следующем этапе исследования варианты английского языка корпуса NOW были сгруппированы в шесть кластеров, соответствующих крупным географическим ареалам: американскому, австралийскому, европейскому, азиатскому, африканскому и карибскому.

Средняя частота употреблений имени *politics*, рассчитанная по ареалам (рис. 3, где по вертикальной оси отложено среднее количество словоупотреблений в корпусе NOW по ареалам в относительных величинах), свидетельствует о том, что по частотности коммуникативного обращения рассматриваемого слова «лидирует» новостной дискурс африканского ареала, далее в порядке убывания расположились азиатский, австралийский ареалы, вариант английского языка Ямайки. Заключает данное множество интернет-СМИ англоязычных стран европейского и американского ареалов, где слово *politics* употребляется значительно реже.

Такое распределение частоты использования слова *politics* позволяет сделать предположение о высокой степени актуальности политики как темы для публичного обсуждения в новостном дискурсе Африки и Азии.

Как видим, корпусные технологии позволяют отслеживать активность функционирования языковых единиц на малом временном отрезке, а также сопоставлять данные разных сегментов англоязычного пространства, что в свою очередь способствует

адекватной интерпретации и оценке той роли, которая отводится политике как социальному институту в сознании разных языковых сообществ, исторически объединенных одним языком.

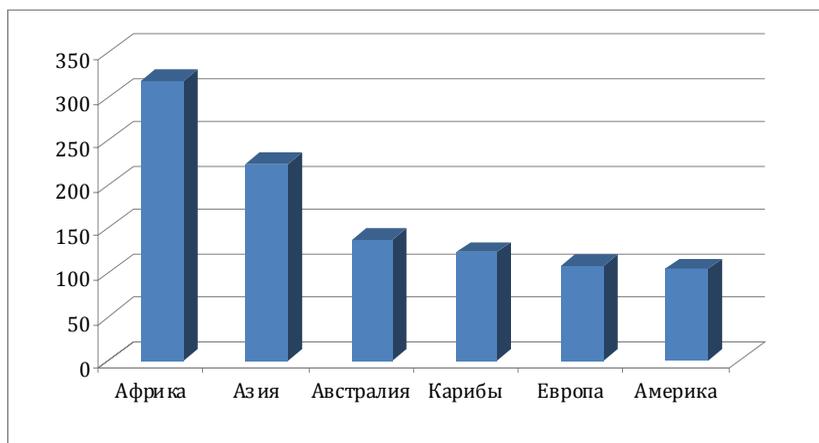


Рис. 3

Соотношение употребления имени *politics* по ареалам

Еще одной возможностью макрокорпусных технологий является анализ метафоричности политического дискурса. Метафоры в политической коммуникации играют особую роль в создании реальности: они не только влияют на то, как воспринимается ситуация большими группами людей, но и позволяют формировать идеологии, которыми оправдываются различные стратегии решения проблем, в частности оккупационная политика США (см. подробнее в работе [Lakoff 1991]). При этом наблюдаются зависимости между идеологическими следствиями метафор и действиями государства.

Исследование метафорической категоризации понятий, обозначающих социальные явления и институты, позволяет проследить, какие образы помогают носителям языка осмысливать эти явления. С этой точки зрения интересно посмотреть, как в современном английском языке осуществляется категоризация *политики* как социального института. Для этого был проведен анализ метафорической сочетаемости существительного *politics*. Результаты количественной обработки данных корпуса NOW показали, что в первую очередь политика категоризируется во всех вариантах английского языка как небольшой предмет стабильной формы, который можно

взять в руку (пример 1) и удерживать (пример 2), бросить (пример 3), отдать или забрать (пример 4), при этом во всех вариантах английского языка, кроме кенийского, такая метафорическая модель имени *politics* является доминирующей.

(1) *We cannot **grasp** the underlying **politics** which bounds all of us together (LK);*

(2) *«Radical feminism» has an iron **grip** on both Swedish **politics** and jurisprudence (GB);*

(3) *Actually, it was Hitler who **threw** race **politics** into fascism (US);*

*It would be a mistake to **cast** Chinese **politics** at an elite level as wholly predictable (HK);*

(4) *Regional government is not a way of **taking politics** to the people; it is a way of **taking it away from** the people (IE).*

Во вторую очередь политика категоризуется в английском языке как нить, которую вплетают в ткань национальной идентичности (пример 5), которой связывают людей (пример 6), переплетают с религией, национальностью (пример 7) и другими реалиями. Образ «нитевидной» политики встречается во всех 20 вариантах английского языка, при этом в кенийском английском он доминирует.

(5) *The underlying **thread** of Tamil **politics** became political emancipation (LK);*

*Umno **weaved** its race-based **politics** into the **fabric** of the nation (MY);*

(6) *The American public has become deaf to the **politics** that **bind** us (US);*

(7) *These elections have been an eye opener for me as far as ethnicity and **politics** are **intertwined** in Kenya (KE).*

Третьим приоритетом метафорической категоризации политики выступает образ длинно-тонкого предмета стабильной формы типа копья, которым можно проткнуть нечто, например жизнь, а также который можно использовать в функции «указки» (примеры 8). Как длинно-тонкий предмет политика категоризуется в девяти вариантах английского языка, при этом данная метафорическая модель не характерна для ямайского английского и африканского ареала в целом.

(8) *In Taiwan, according to our research, **politics** already **penetrated** almost every layer of their daily lives (AU);*

*For more than a decade, **politics** have **pointed** toward one direction and evidence has pointed toward another (US).*

Также политика осмысляется по образу и подобию текущей (пример 9), льющейся или струящейся жидкости (примеры 10), которой свойственно вытекать и сочиться (примеры 11) (ср.: рус. *просочиться на страницы газет*), что подтверждается корпусными данными 13 вариантов английского языка. Интересно, что такое метафорическое осмысление политики не свойственно большинству азиатских лингвокультур.

(9) *Mindful of the **flow** of politics in Nigeria, we may not have seen the last of this sort of unjustifiable...* (NG);

(10) *However, the Buguma crisis and **stream** of politics at the time stalled the continuity of the scheme* (GH);

(11) *The employees under me would surely reply in tersely written emails **leaking politics** and reeking of something more hidden away* (US);

***Politics spilled** into the public realm* (CA).

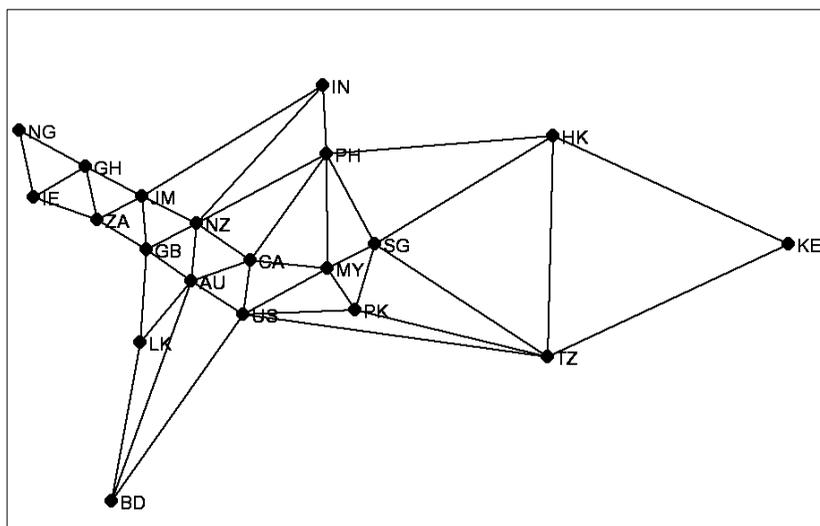


Рис. 4

Карта *politics* по данным корпуса

На карте имени *politics* (рис. 4), построенной из расчета тесноты связи между вариантами языка при помощи статистических методов, наблюдается следующая зависимость: чем короче линия, соединяющая варианты, тем выше коэффициент корреляции между ними и теснее связь. Как видим, кенийский английский (KE)

находится на максимальном удалении от других вариантов, часть которых, связанных достаточно тесно, образуют кластер с доминирующим при метафоризации политики образом мелкого предмета (примеры 1–4). В новостном дискурсе Кении доминирует метафора «нитевидной» политики, при этом чем ближе национальный вариант английского языка располагается на карте к варианту Кении, тем более значимым данный образ политики является для данного языкового сообщества. В новостном англоязычном дискурсе Бангладеш (BD) лучше, чем в других странах, реализуется метафорический образ «длинно-тонкой» политики, в вариантах Нигерии (NG) и Ирландии (IE) наиболее ярко представлен метафорический образ «жидкой» политики.

Еще одной возможностью интегрированной методики анализа метафорической сочетаемости является компьютерно-когнитивное графическое представление лингвистического объекта. В результате проведенного исследования нам удалось смоделировать когнитивный континуум метафоризации политики по данным электронного ресурса NOW. При помощи методов Data Mining и самоорганизующихся карт Кохонена, являющихся разновидностью нейросетевых алгоритмов, была получена многомерная диаграмма (рис. 5), построенная по данным анализа метафорической сочетаемости имени *politics*¹. Самые высокие точки графа соответствуют метафорическому образу «мелкой», «помещающейся в руке» политики, которой можно *мани*-пулировать (удерживать, передавать или отбирать), максимально коммуникативно востребованному в политическом дискурсе британского и американского вариантов языка. Таким образом, корпусные технологии позволяют получить данные о том, как носители языка выражают идею о том, в *чьих руках* находится политика, и как осуществляются политические действия.

¹ Подробнее с процедурой получения многомерной диаграммы можно познакомиться в работе [Донина, Борискина, 2016]

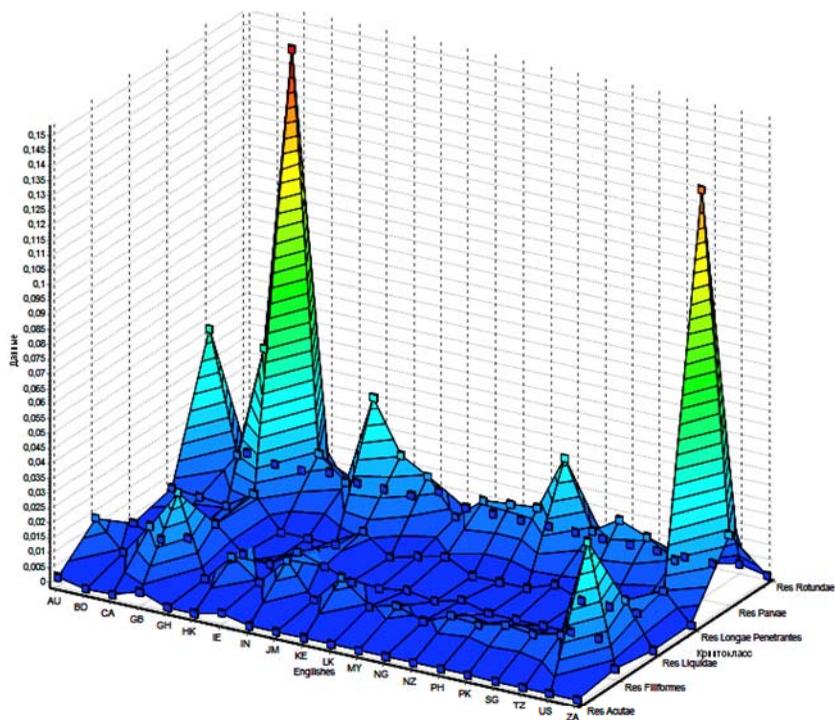


Рис. 5
**Многомерная диаграмма. Когнитивный континуум
 метафоризации политики по вариантам**

Изучение неологизмов политического дискурса корпусными методами

Макрокорпусные технологии позволяют наблюдать за редкими окказиональными языковыми явлениями, например неологизмами, которые возникают в периоды активных политических действий как естественная реакция общества на происходящие события. Исследование новой лексики в политическом дискурсе позволяет сделать ряд выводов о том, как воспринимаются политические процессы в обществе. Такие данные полезны для анализа и сравнения тональности сообщений, создаваемых носителями разных географических

вариантов английского языка. В качестве примера рассмотрим ряд неологизмов, образованных от имен и фамилий двух основных кандидатов на пост президента США в 2016 г.

Предвыборная президентская кампания 2016 г. в США послужила стимулом для появления неологизмов, отражающих отношение общества к политическим программам и действиям кандидатов во время президентской гонки. Эти неологизмы получили широкое распространение не только в новостном дискурсе США, но и во всех 20 англоязычных странах, что позволяет проследить, как политическое событие в одной стране отражается на жизни людей по всему миру.

Фамилия Дональда Трампа использовалась в текстах СМИ в качестве словообразовательной базы гораздо чаще, чем фамилия его соперницы. Кроме того, количество новой лексики с корнем Trump- значительно превышает число новых слов, образованных от фамилии Clinton.

В таблице 3 представлены неологизмы, полученные в результате поискового запроса *trump*, абсолютная частота которых на начало января 2017 г. превышает 100 словоупотреблений.

Таблица 3

Д. Трамп: Неологизмы (по данным корпуса NOW)

№	Неологизм	Абсолютная частота словоупотребления	Словообразовательная модель	Возможный русский эквивалент
1	Anti-Trump	2880	Добавление приставки anti-	Анти-трамповский
2	Pro-Trump	891	Добавление приставки pro	Про-трамповский
3	Trumpism	866	Добавление суффикса -ism	Трампизм
4	Trumpian	752	Добавление суффикса -ian	Трамповский
5	Trump-like	282	Основосложение	Трампо-подобный
6	Trump-branded	226	Основосложение	Бренд, созданный Трампом
7	Trump-land	202	Основосложение	Трамплэнд
8	Post-Trump	198	Добавление приставки post-	Пост-трамповский
9	Trump-led	115	Основосложение	Ведомый Трампом
10	Non-Trump	114	Добавление приставки non-	Не-трамповский
11	Trumponomics	104	Контаминация Trump+economics	Трампономика

Для создания неологизмов носители разных вариантов английского языка используют одинаковые словообразовательные модели, что создает благоприятные условия для увеличения ком-

муникативного обращения слова и повышает вероятность его закрепления в узусе. Любопытно, что большинство контекстов, содержащих неологизмы, выражают негативную оценку Дональда Трампа, его персональных качеств и действий. Особенно ярко эта тенденция проявляется в тех случаях, когда авторы текста не просто создают новое слово, но используют языковую игру. Отметим, что по сравнению с неологизмами, приведенными в таблице 3, такие примеры встречаются реже (примеры 12–14):

(12) *Trumpocalypse* (40 вхождений)

KGB, RNC, FBI and Wikileaks feel like the four shameless horsemen of the Trumpocalypse.

КГБ, Национальный комитет Республиканской партии, ЦРУ и Викиликс – как будто четыре наглых всадника Трампокалипсиса.

(13) *Trumpflation* (34 вхождения)

But then along came Trump and now investors are being warned of «Trumpflation».

А потом появился Трамп, и теперь инвесторов предупреждают о «Трампфляции».

(14) *Trumpmania* (26 вхождений)

Trump hypnotizes the media; Trump goes global; Trumpmania is the new Obamania...

Трамп гипнотизирует СМИ, Трамп становится мировой фигурой, Трампомания – это новая Обамания.

Имя и фамилия Хиллари Клинтон тоже использовались для создания неологизмов, однако их общее количество и частотность в текстах СМИ гораздо ниже. Неологизмы с абсолютной частотой выше 100 словоупотреблений представлены в таблице 4.

Таблица 4

Х. Клинтон: Неологизмы (по данным корпуса NOW)

№	Неологизм	Абсолютная частота словоупотребления	Словообразовательная модель	Возможный русский эквивалент
1	Pro-Clinton	279	Добавление приставки pro-	За Клинтон, в пользу Клинтон
2	Anti-Clinton	255	Добавление приставки anti-	Против Клинтон
3	Clinton-era	184	Основосложение	Эра Клинтон
4	Anti-Hillary	123	Добавление приставки anti	Против Хиллари
5	Pro-Hillary	123	Добавление приставки pro-	За Хиллари

Обращает на себя внимание тот факт, что прилагательное anti-Trump используется в СМИ почти в 3 раза чаще, чем pro-Trump. В случае с Хиллари Клинтон ситуация иная: pro-Clinton встречается в целом чаще, чем anti-Clinton, а неологизмы, образованные аналогичным способом от ее имени, употребляются равное количество раз.

Неологизмы, в которых присутствует языковая игра, также употребляются в контекстах, выражающих критику в отношении действий Клинтон (пример 15), причем часто эта критика является прямым указанием на то, что Клинтон по своим политическим взглядам является полной приемницей действующего президента:

(15) *Hillarycare*

*Obamacare's failures are now becoming part of **Hillarycare**, and she has to defend skyrocketing premiums that affect individuals and families across the country.*

Провалы реформы здравоохранения Обамы теперь становятся частью реформы, которую предлагает Хиллари, и ей приходится защищать непомерную стоимость медицинских страховок, которые касаются и отдельных людей, и семей по все стране.

Одним из наиболее интересных примеров стало появление неологизма *Killary*, который является контаминацией глагола *kill* (убивать) и имени кандидата в президенты США. Негативная коннотация нового способа номинации политика в данном случае очевидна; контексты, в которых появляется слово *Killary*, как правило, усиливают отрицательное отношение авторов к деятельности Клинтон (пример 16):

*(16) I am very sure you will gladly enjoy the benefits when Trump turns this world into a better place and prevent it from turning it into an ash heap with the wars Obama and **Killary** are pushing.*

*Я абсолютно уверен, что вы будете наслаждаться всеми преимуществами, когда Трамп превратит этот мир в лучшее место и сделает так, чтобы он не стал горой пепла со всеми войнами, которые проталкивают Обама и **Киллари (Убиллари)**.*

Таким образом, появление неологизмов, образованных от имен участников политических процессов, как правило, является показателем отрицательного отношения к политику. В тексте такие слова могут служить сигналом общей негативной тональности новостного сообщения.

Исследование политического дискурса на материале микрокорпусов

Изучение политической коммуникации может проводиться не только на материале многомиллионных корпусов, но и на основе данных, которые предоставляют массивы специально отобранных текстов, которые можно назвать микрокорпусами. Примером таких микрокорпусов можно считать текстовые массивы транскриптов выступлений отдельных политиков или представителей определенных политических течений. Тексты публичных выступлений политиков, транскрипты предвыборных дебатов используются как такой источник данных об особенностях речи участников политического дискурса, который позволяет выявлять предпочитаемые ими способы трансляции идеологии и влияния на сознание аудитории.

В данной работе для исследования особенностей предвыборного дискурса кандидатов на пост президента в США в 2016 г. были использованы два микрокорпуса: Корпус-1 объемом 60 тыс. словоупотреблений, содержащий транскрипты выступлений Хиллари Клинтон в качестве кандидата на пост Президента США от Демократической партии; Корпус-2 объемом 55 тыс. словоупотреблений, содержащий транскрипты речей Дональда Трампа в качестве кандидата на пост Президента США от Республиканской партии.

Несмотря на небольшой объем, эти массивы текстов могут дать нам информацию о частотности использования слов, об их лексической сочетаемости, а также о словах, которые играют роль ключевых в речи (и, более широко, в идеологии) двух политиков. Эти данные позволяют далее переходить от исследования формальных показателей (например, частотности употребления лексики) к исследованию семантики (т.е. изучению смысловой стороны коммуникации). Переход к семантическому анализу – это необходимый шаг, который позволяет проводить дальнейшее описание идеологии и системы ценностей двух политиков.

Корпусное исследование ключевых слов предполагает использование специализированных программ-конкордансеров. В данном исследовании была использована программа AntConc, которая позволяет не только строить конкордансы (т.е. автоматически выделять все контексты, в которых употребляется то или иное слово), но и анализировать лексическую сочетаемость языковых единиц, а также сравнивать два корпуса и выделять слова, частота встречаемости которых

в одном корпусе значительно отличается в большую или меньшую сторону от частоты их использования во втором корпусе.

Частотность – один из показателей востребованности слова в дискурсе, на основе которого можно делать выводы о значимости определенной темы или сферы жизнедеятельности для социума. Однако при использовании языковых корпусов для составления списков частотных слов приходится учитывать преобладание в верхней части списка слов, которые относятся к несамостоятельным частям речи (предлоги, союзы) либо являются дейктическими единицами (личные и указательные местоимения, наречия с семантикой места и времени), либо выполняют грамматические функции (вспомогательные глаголы, артикли). Поэтому далее мы остановимся только на функционировании существительных, вошедших в первую сотню наиболее частотных слов в речах двух политиков.

Наиболее частотными существительными в предвыборных выступлениях Трампа и Клинтон оказались имена, представленные в таблицах 5 и 6.

Таблица 5

**Наиболее частотные существительные
в предвыборных выступлениях Д. Трампа**

№	Существительное	Позиция среди 100 самых частотных словоформ	Кол-во употреблений
1	people	29	280
2	country	32	251
3	Hillary	42	176
4	Clinton	44	174
5	jobs	75	111
6	years	82	102
7	Obama	96	85
8	States	100	82

Таблица 6

**Наиболее частотные существительные
в предвыборных выступлениях Х. Клинтон**

№	Существительное	Позиция среди 100 самых частотных словоформ	Кол-во употреблений
1	2	3	4
1	people	28	294
2	country	49	180
3	America	55	162
4	Donald	71	143

Продолжение таблицы 6

1	2	3	4
5	president	73	142
6	jobs	75	137
7	police	83	121
8	Americans	89	113
9	Trump	90	112
10	families	98	95

Ожидаемым является частое употребление личных имен политических противников, что объясняется постоянной критикой в адрес оппонента. Оба политика активно используют слово *jobs* (*рабочие места*), и это подтверждает актуальность проблемы создания новых рабочих мест в американском обществе. Другие лексемы, обладающие одинаково высокой частотностью в речи Трампа и Клинтон, – это *country* (*страна*) и *people* (*люди, народ*). Можно ли утверждать, что одинаковая частота встречаемости лексем *country* и *people* в речах обоих политиков неинформативна в отношении особенностей тематической и риторической организации их выступлений? Чтобы ответить на этот вопрос, необходимо обратиться к контекстам, в которых эти словоформы употребляются.

Среди 251 контекста, в которых Д. Трамп использует слово *country*, наиболее частым является словосочетание *our country* с абсолютной частотой корпусных вхождений 144, 40 раз в корпусе встречается сочетание *this country*, 19 раз – *the country*. Поскольку референтом всех этих сочетаний являются США, это позволяет говорить о том, что тема государства постоянно находится в фокусе внимания политика. Трамп использует прилагательные (*great, incredible*), которые описывают США как великую страну, основная проблема которой – власть, не способная обеспечивать людям достойный образ жизни (пример 17):

(17) *I began this campaign because I was so tired of seeing such foolish things happen to our country. **This is a great country. This is a great land.***

*Я начал эту кампанию, поскольку я устал смотреть на те глупости, которые творятся в нашей стране. **Это великая страна. Это великая земля.***

В выступлениях Х. Клинтон из 180 контекстов, в которых встречается слово *country*, 110 раз оно сочетается с местоимением *our*, 17 раз – с артиклем *the*, и девять раз – с указательным место-

имением *this*. Помимо этого Клинтон использует прилагательные в сравнительной степени, что позволяет ей неявным образом выразить идею сравнения настоящего с будущим и обозначить пути дальнейшего улучшения жизни в стране (пример 18):

(18) *I'm running for President with the belief that we need to face up to these challenges and fix them in order to become a stronger, fairer country.*

Я борюсь за пост президента, будучи уверенной в том, что нам необходимо открыто смотреть на вызовы и справляться с ними, чтобы стать более сильной, более справедливой страной.

В речах Д. Трампа лексема *people* наиболее часто используется в сочетании *the American people* – 17 вхождений, из них шесть – *you the American people* (вы, американский народ). Словосочетание *our people* имеет 12 вхождений, *one people* – три вхождения. Таким образом, Трамп активно использует стратегию объединения аудитории (пример 19):

(19) *Our movement is about replacing a failed, and corrupt, and when I say corrupt I'm talking about totally corrupt, political establishment with a new government controlled by you the American people.*

Наш курс направлен на замену провального и коррумпированного (и когда я говорю «коррумпированного», я говорю об абсолютно коррумпированном) политического истеблишмента новым правительством, подконтрольным вам, американскому народу.

Словосочетание *one people* (единый народ) используется для акцентирования внимания аудитории на идее единения американского народа, а не разделения его на многочисленные меньшинства (пример 20):

(20) *Just imagine what our country could accomplish if we started working together as One People, under One God, saluting One American Flag.*

Только представьте себе, чего могла бы добиться наша страна, если бы мы начали работать вместе, как Единый Народ, под властью Единого Бога, отдавая честь Единому американскому флагу.

Идея национального объединения проявляется еще и в том, каким образом в речи отражается идея социальных различий и осуществляется референция к отдельным социальным группам. Для Дональда Трампа характерна обобщенная референция, без указания на социально-экономические, этнические или иные признаки, без называния конкретных цифр (пример 21):

(21) *Some people warned me this campaign would be a journey to hell. But they are wrong, it will be a journey to heaven because we will help so many people.*

Некоторые предупреждали меня, что эта кампания превратится в дорогу в ад. Но они ошибаются, это будет дорога в рай, потому что мы поможем очень многим людям.

Хиллари Клинтон же использует лексему *people* для отсылки к конкретным социальным группам (пример 22), что может интерпретироваться как желание политика четко обозначить свой электорат:

(22) *We spent a year and a half bringing together millions of people from every corner of our country to say with one voice that we believe that the American dream is big enough for everyone. For people of all races, and religions, for men and women, for immigrants, for LGBT people, and people with disabilities.*

Мы потратили полтора года на то, чтобы объединить миллионы людей во всех уголках нашей страны, чтобы заявить в один голос, что американской мечты хватит на всех. На людей всех рас и религий, на мужчин и женщин, на иммигрантов, на людей с нетрадиционной ориентацией, на людей с ограниченными возможностями.

Характерные для Д. Трампа сочетания в речи Х. Клинтон встречаются гораздо реже: *our people* – восемь вхождений, *the American people* – пять вхождений, словосочетание *one people* в нашем корпусе не встретилось ни разу. Говоря о проблемах, Клинтон предпочитает оперировать конкретными цифрами (пример 23), тем самым задействуя не только эмоциональный, но рациональный способ восприятия информации аудиторией.

(23) *Nine million people lost their jobs. Five million people lost their homes.*

Девять миллионов людей, девять миллионов людей потеряли работу. Пять миллионов людей потеряли свои дома.

Как оказалось, идея многообразия в восприятии американского общества и рациональное начало, к которым апеллировала Х. Клинтон, оказались малоэффективными, в то время как идея национального единства Д. Трампа, «преподнесенная» электорату посредством эмоционального воздействия, позволила ему стать президентом.

Теперь посмотрим на слова, которые часто употребляются только одним из политиков.

Появление словоформы *years* в выступлениях Дональда Трампа связано с постоянным указанием на то, что его политические соперники находятся у власти в течение достаточно долгого промежутка времени (пример 24), однако за время правления им не удалось решить ключевые социальные проблемы:

(24) *It's been defective for a long time, many years, but the politicians haven't done anything about it.*

Это соглашение остается неправильным в течение долгого времени, многие годы, но политики ничего с этим не сделали.

Вхождение в первую сотню слова *family* (семья) означает, что для Х. Клинтон одной из ключевых оказывается тема семейного благополучия (пример 25):

(25) *And I want us to do more to support people who are struggling to balance family and work.*

И я хочу, чтобы мы оказывали больше поддержки людям, которые пытаются совмещать семью и работу.

Появление среди частотных существительного *president* (президент) и анализ контекстов, в которых оно встречается, указывает на две риторические стратегии, регулярно используемые Клинтон в своих выступлениях: первая стратегия связана с использованием комиссивных речевых актов, в которых кандидат на пост президента дает предвыборные обещания. Вторая риторическая стратегия связана с упоминанием других президентов США, членов Демократической партии, чьи действия, по мнению Клинтон, следует признать успешными:

(26) *When my husband was president, we went from a \$300 billion deficit to a \$200 billion surplus and we were actually on the path to eliminating the national debt. When President Obama came into office, he inherited the worst economic disaster since the Great Depression. He has cut the deficit by two-thirds.*

Когда мой муж был президентом, мы ушли от долга в 300 миллиардов долларов к профициту в 200 миллиардов, и мы уже были на пути к избавлению от национального долга. Когда президент Обама начал свою работу, он унаследовал наихудшую экономическую ситуацию со времен Великой депрессии. Он снизил дефицит на две трети.

Использование имен успешных политиков – одна из стратегий повышения собственного авторитета, основанная на зарабатывании «символического капитала» [Bourdieu, 1991].

Анализ частоты употребления слов – это способ выявления коммуникативно значимых языковых единиц внутри одного корпуса. Частотность – это, безусловно, важная характеристика, однако она описывает только то, что «лежит на поверхности». Более сложный способ анализа корпусов текстов связан с возможностью сопоставления двух текстовых массивов и выявления ключевых слов.

Ключевые слова как «сигналы» эпохи

Обращение к проблеме ключевых слов в современной лингвистике связано с признанием того факта, что в любом тексте значимость слов различна в том смысле, что каждая словоформа вносит свой вклад в общий смысл текста, но некоторые словоформы оказываются более важными в смысловой организации текста. Такие слова принято называть *ключевыми*. Анализ ключевых слов необходим для определения тематической (содержательной) направленности текста, для выявления основных идей или стилистических предпочтений автора.

В современной лингвистике понятие *ключевого слова* имеет несколько трактовок. Одно из наиболее популярных пониманий – это понимание ключевого слова как важного элемента культуры. В фокусе внимания исследователей оказываются слова, которые описывают характерные для данной культуры понятия и ценности; специфика семантики таких слов часто вызывает трудности при переводе [Вежбицкая, 2001]. Глобальная цель таких исследований – с помощью лингвистического анализа описать языковую картину мира.

Второй вариант трактовки понятия ключевого слова связан не с возможностью обозначения им специфического понятия, характерного для определенной культуры, а с ролью отдельных слов в понимании конкретного текста. Например, в тексте научной статьи ключевые слова играют роль метаданных, которые в самом общем виде представляют тематику исследования [Кретов, 2012; Кашкин, Шилихина, 2014].

В сфере автоматической обработки естественного языка трактовку понятия *ключевого слова* можно считать «гибридной»: ключевые слова рассматриваются и как инструмент понимания текста, и как способ выражения определенных идей.

В рамках данного исследования были выявлены ключевые слова в речах Дональда Трампа и Хиллари Клинтон. Получение списков ключевых слов проводилось по следующему алгоритму: при сопоставлении двух массивов текстов выделялись статистически значимые слова в Корпусе-1, при этом Корпус-2 использовался в качестве справочного. Затем был составлен список ключевых слов для Корпуса-2, при этом в качестве справочного использовался Корпус-1. Инструментом статистической оценки, на основе которой были выделены слова, встречаемость которых в речи одного политика выше, чем в речи другого, стала функция логарифмического правдоподобия. Данные о первых 30 ключевых словах представлены в таблице 7.

Таблица 7

Ключевые слова в речах Д. Трампа и Х. Клинтон

№	Дональд Трамп	Хиллари Клинтон
1	She	He
2	Hillary	Donald
3	Trump	Need
4	They	Police
5	Going	His
6	Illegal	Work
7	Will	Families
8	Are	To
9	Her	That
10	You	Us
11	Very	Together
12	Bad	Economy
13	Great	Officers
14	Story	Too
15	NAFTA	Black
16	Immigration	Still
17	Not	Americans
18	Administration	Young
19	Disaster	Everyone
20	Obamacare	Children
21	Corruption	Here
22	Media	In
23	Tremendous	Let
24	Borders	How
25	Emails	Each
26	Leaving	Kids
27	Countries	Do
28	Corrupt	Stand
29	Under	Can
30	False	Sure

Результат, представленный в таблице, позволяет увидеть, какие ценности и идеи являются для двух политиков приоритетными: для Хиллари Клинтон это семья и дети, молодежь, взаимоотношения рядовых граждан с полицией. Кроме того, в число ключевых попали слова, выражающие уверенность в собственной правоте, в будущем. Для Трампа важными являются темы коррупции, иммиграции, противостояния администрации Барака Обамы.

Более детальный анализ первых 100 ключевых слов позволяет увидеть тенденции в выборе способа воздействия на аудиторию. Обращает на себя внимание обилие прилагательных с отрицательной коннотацией: *illegal* (нелегальный), *bad* (плохой), *disaster* (катастрофа), *corruption* (коррупция), *corrupt* (коррумпированный), *horrible* (ужасный), *radical* (радикальный), *crime* (преступление), *depleted* (истощенный), *nothing* (ничего).

Х. Клинтон пользовалась противоположной стратегией: в отличие от своего оппонента она использовала лексику, в семантике которой преобладает положительная оценка: *together* (вместе), *young* (молодой), *sure* (уверенный), *clear* (ясный, понятный), *stronger* (сильнее), *trust* (доверие), *faith* (вера), *progress* (прогресс), *hope* (надежда), *affordable* (доступный), *opportunity* (возможность), *health* (здоровье), *investments* (инвестиции), *best* (лучший), *fairer* (более справедливый).

Заключение

Подводя итоги, отметим многообразие вариантов применения методов корпусной лингвистики для анализа политической коммуникации. Корпусные технологии позволяют выявлять метафорические модели, к которым обращаются как политики, так и наивные пользователи языка, отслеживать географию и выявлять ареальную специфику метафорических процессов в англоговорящем пространстве, наблюдать за редкими языковыми явлениями, например неологизмами, на малом временном отрезке, что в свою очередь позволяет делать выводы о способах осмысления политических процессов и явлений в современном обществе.

Анализ частотности словоформ и выделение ключевых слов в специализированном корпусе публичных выступлений политиков позволяет объективно оценить тематическое содержание тек-

стов и определить основные идеи, которые политические деятели стремятся донести до аудитории, культурные и идеологические импликации, которые содержатся в тексте и с помощью которых имплицитно осуществляется влияние на массовое сознание.

Список литературы

- Вежибицкая А.* Сопоставление культур через посредство лексики и прагматики / Пер. с англ. А.Д. Шмелева. – М.: Языки славян. культуры, 2001. – 272 с.
- Донина О.В., Борискина О.О.* Эмотивная лексика в аспекте ареальной вариативности // Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация. – Воронеж, 2016. – № 4. – С. 41–45.
- Кашкин В.Б., Шилихина К.М.* Концептуальный диссонанс при переводе российской лингвистической терминологии // Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация. – Воронеж, 2014. – № 1. – С. 72–77.
- Кретов А.А.* Маркемы и ключевые слова в научных текстах // Мир лингвистики и коммуникации. – Тверь, 2012. – Т. 1, № 27. – С. 1–13.
- Bourdieu P.* Language and symbolic power. – Cambridge: Harvard univ. press, 1991. – 302 p.
- Davies M.* Corpus of news on the web (NOW): 3+ billion words from 20 countries, updated every day. – Provo, 2013. – Mode of access: <http://corpus.byu.edu/now/> (Дата посещения: 12.01.2017.)
- Khokhlova M.V.* Large Corpora and Frequency Nouns // Computational linguistics and intellectual technologies: Proceedings of the international conference «Dialogue-2016». – Moscow, 2016. – June 1–4. – P. 237–250.
- Lakoff G.* Metaphor and war: The metaphor system used to justify war in the Gulf // Peace Research. – Winnipeg, 1991. – N 23(2/3). – P. 25–32.
- Wiedemann G.* Opening up to big data: Computer-assisted analysis of textual data in social sciences // Historical social research. – Mannheim, 2013. – Vol. 38, N 4 (146). – P. 332–357. – Mode of access: <http://www.jstor.org/stable/24142701> (Дата посещения: 12.01.2017.)