

А. В. Колмогорова, А. А. Калинин, Ю. А. Горностаева
Красноярск, Россия

РАЗРАБОТКА КОМПЬЮТЕРНОЙ ПРОГРАММЫ АВТОМАТИЧЕСКОГО АНАЛИЗА И КЛАССИФИКАЦИИ ПОЛЯРИЗОВАННЫХ ПОЛИТИЧЕСКИХ ТЕКСТОВ НА АНГЛИЙСКОМ ЯЗЫКЕ ПО УРОВНЮ ИХ МАНИПУЛЯТИВНОГО ВОЗДЕЙСТВИЯ: ПРАКТИЧЕСКИЕ РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

АННОТАЦИЯ. Статья посвящена обсуждению практических результатов, полученных в рамках проекта по разработке компьютерной программы — классификатора англоязычных политических текстов о России по уровню их манипулятивного воздействия. Актуальность проекта обусловлена тем, что в условиях информационной войны и проницаемости национальных массмедийных пространств англоязычные манипулятивные тексты могут служить не только подогреванию русофобии среди западных читателей, но и, в случае репоста или перепечатки, стать инструментом манипулирования сознанием владеющих английским языком представителей российского гражданского общества. Методология создания программы-классификатора основана на лингвистических методах дискурс-анализа и социолингвистического эксперимента, а также на алгоритме прецедентного машинного обучения «Дерева решений». В качестве параметров оценки используется относительное количество в тексте следующих маркеров манипуляции: военной лексики, нацистского лексикона, лексем с префиксами «pro-» и «anti-», морфемы «soviet», маркеров манипулятивности, полученных в социолингвистическом эксперименте, и имен политических лидеров, в частности Владимира Путина. Результатом проверки является атрибуция текста как принадлежащего к одному из 4 классов, сопровождаемая соответствующим сообщением: класс 1 — тексты, не содержащие манипуляции (сообщение: «This text doesn't seem to have any bias»), класс 2 — тексты с низким уровнем манипулятивности (сообщение: «Bias degree is low»), класс 3 — тексты со средним уровнем манипулятивности (сообщение: «Bias degree is medium»), класс 4 — тексты с высоким уровнем манипулятивности (сообщение: «Bias degree is high»). В настоящее время прототип программы успешно апробирован в Междисциплинарном центре исследований цифровой экономики СФУ.

КЛЮЧЕВЫЕ СЛОВА: манипулятивное воздействие; политический дискурс; поляризованный дискурс; компьютерные программы; медиадискурс; СМИ; средства массовой информации; политические тексты; английский язык.

СВЕДЕНИЯ ОБ АВТОРАХ: Колмогорова Анастасия Владимировна, доктор филологических наук, профессор, заведующий кафедрой романских языков и прикладной лингвистики, Сибирский федеральный университет; 660041, Россия, г. Красноярск, пр-т Свободный, 82, стр. 1, к. 234; e-mail: nastiakol@mail.ru.

Горностаева Юлия Андреевна, аспирант кафедры романских языков и прикладной лингвистики, Сибирский федеральный университет; 660041, Россия, г. Красноярск, пр-т Свободный, 82, стр. 1, к. 333б; e-mail: yulyatald@ya.ru.

Калинин Александр Андреевич, старший преподаватель кафедры романских языков и прикладной лингвистики, Сибирский федеральный университет; 660041, Россия, г. Красноярск, пр-т Свободный, 82, стр. 1, к. 333б; e-mail: verbalab@yandex.ru.

Введение

Данная статья посвящена обсуждению компьютерной программы — классификатора англоязычных политических текстов по уровню выраженности в них манипулятивного воздействия, созданной научным коллективом Лаборатории прикладной лингвистики и когнитивных исследований Сибирского федерального университета.

Разработка проекта начата в 2015 г. на фоне нарастающей политической конфронтации между Россией и США в контексте «крымского вопроса». Проект направлен на решение задач информационной безопасности российского массмедийного пространства и предусматривает разработку и внедрение компьютерной программы с удобным интерфейсом, способной автоматически обрабатывать большие массивы политических текстов на английском языке, посвященных России, и классифицировать их по степени выраженности в них техник и приемов манипуляции массовым сознанием читателя. Актуальность разработки связана с тем, что в условиях информационной войны западных СМИ против России манипуляция массовым сознанием используется как в отношении западного читателя для подогревания русо-

фобии (при этом манипулятивные тексты в случае репоста или перепечатки могут превратиться в орудие внутренней информационной войны в медиапространстве России), так и в целях намеренного искажения картины мира молодого поколения россиян, владеющих английским языком.

Компьютерная программа адресована рядовым интернет-пользователям и информационным агентствам для самостоятельного анализа и отбора качественного англоязычного массмедийного контента.

В настоящее время программа, созданная на основе статистической обработки лингвистических маркеров манипуляции в тексте и технологии машинного обучения по прецедентам, апробирована в Междисциплинарном центре исследований цифровой экономики СФУ. В дальнейшем планируется разработка мобильного приложения, а также аналогичного классификатора для русскоязычных текстов.

1. Теоретический и технологический базис дизайна программы

Под манипуляцией мы понимаем, вслед за Т. ван Дейком [Dijk 2006: 360], коммуникативную и интерактивную практику, которая предполагает, что манипулятор, занимаю-

© Колмогорова А. В., Калинин А. А., Горностаева Ю. А., 2017

щий определенную авторитетную социальную позицию, действуя в своих интересах и злоупотребляя своей властью, устанавливает контроль над другими людьми, как правило, против их воли. Манипулятор, преследуя собственные цели, заставляет людей совершать определенные поступки, которые лежат в сфере интересов манипулятора и противоречат интересам манипулируемых [Dijk 2006].

Разграничивая вслед за рядом авторов [Паршин 2007; Стернин 2012] собственно речевое воздействие и манипуляцию, при отборе материала исследования мы руководствовались таким принципом: если в политическом дискурсе *отсутствуют открытые проявления воздействия* (например, призывы, оскорбления, предикаты негативно-оценочного характера и др.), но при этом дискурс имеет целью *ущемление интересов* объекта воздействия в виде намеренного искажения его картины мира посредством активизации в его сознании негативных эмоциональных состояний страха, агрессии, отвращения и введения на данном фоне некоторой, зачастую фактологически верной, информации (о разграничении манипуляции информацией и манипуляции людьми см.: [Billing, Marinho 2014]), то такое воздействие мы считаем манипулятивным.

Материал исследования составили тексты, репрезентирующие так называемый поляризованный дискурс — дискурс, возникающий на фоне конфликта политических интересов двух крупных социально-политических субъектов (партий, политических лагерей, государств) и пропагандирующий стереотипные представления об оппоненте, очерняющий его репутацию, целью которого является деление глобальной социальной среды на две группы — своих (in-groups) и чужих (out-groups) [Eissa et al. 2014].

Таким образом, манипуляция в рамках поляризованного политического дискурса в контексте отношений в паре *Россия — Запад* должна обнаруживать следующие свойства: формирование политическими элитами гипертрофированного образа врага на фоне подогревания чувств страха, тревоги, агрессии по отношению к нему; при этом собственно фактическая информация, которая сама по себе не дает никаких поводов для подобных чувств, в таком контексте начинает осмысливаться искаженно, в определенном, нужном субъектам манипуляции ракурсе.

Первоначальная выборка составила 100 статей, посвященных отношениям США и России на фоне украинского кризиса, из американских периодических изданий *The*

New York Times, *The Washington Post* за период с 2014 по 2015 г. Мы исходили из методологической посылки о том, что в рамках англоязычного поляризованного дискурса субъектом манипуляции является американская политическая элита, ее объектом, как правило, — массовый американский читатель, а мишенью — Россия и российское гражданское общество, соответственно, именно представители последнего будут наиболее чувствительны к искажениям информации о себе, к маркерам манипуляции, мишенью которой являются они сами. Таким образом, в результате социолингвистического эксперимента (см. подр.: [Колмогорова, Калинин, Талдыкина 2016]) были выявлены первичные маркеры манипуляции массовым сознанием в англоязычном политическом дискурсе о России. В дальнейшем анализ выборки текстов, в которых встречаются данные «первичные» маркеры, позволил выявить еще 5 маркеров манипуляции: 1) военная терминология (на данный момент представлен список из 960 лексических единиц); 2) лексика по тематике нацизма (160 лексических единиц); 3) морфема *soviet*; 4) лексемы из составленного по итогам эксперимента списка маркеров манипуляции (тематика страха, агрессии, вторжения); 5) упоминание прецедентных для данной тематики личностей и феноменов (например, Владимир Путин); 6) прилагательные с антонимическими приставками *anti-* и *pro-*.

Затем, используя двухвыборочный t-критерий Стьюдента для независимых выборок, мы проанализировали статистическую значимость отличий в характеристических функциях выявленных 6 маркеров в контрольной (150 контрольных текстов, случайно взятых с сайтов тех же американских изданий) и экспериментальной (150 манипулятивных текстов, проанализированных по методике Т. ван Дейка экспертом-лингвистом) выборках. В качестве порога статистической значимости было выбрано значение в 5 % (0,05).

Для всех маркеров, кроме маркера «лексика по тематике нацизма», порог статистической значимости различий в частотности в контрольной и экспериментальной выборках был преодолен. Таким образом, пять маркеров оказались статистически значимыми именно для манипулятивных текстов. Шестой маркер, не показавший статистической значимости, после дополнительного анализа также не был отброшен: было принято решение принимать его во внимание тогда, когда статистически значимым оказывается еще хотя бы один маркер.

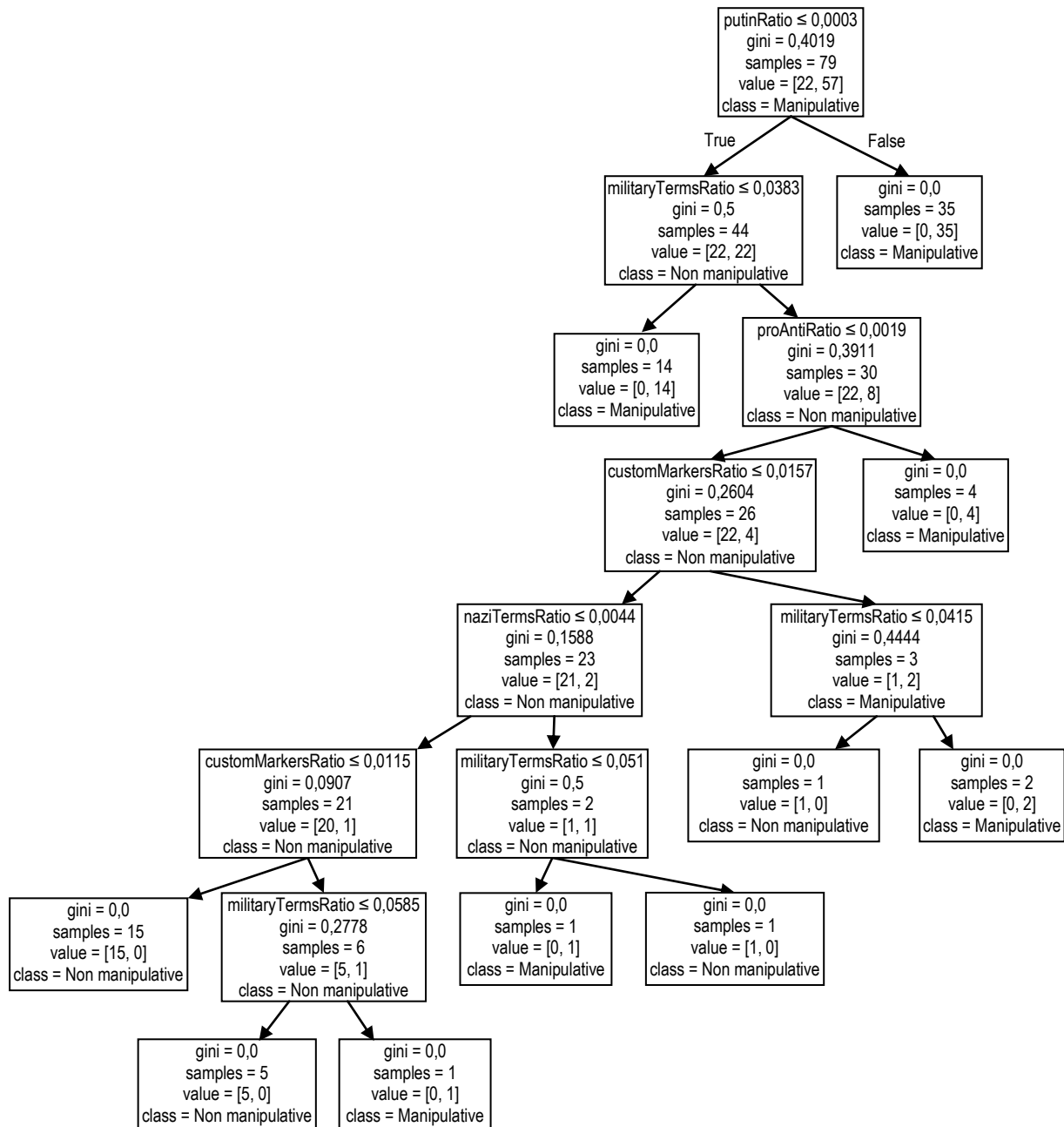


Рис. 1. Пример работы алгоритма «Деревья решений» при классификации одного из текстов

В качестве подхода к разработке классификатора мы выбрали технологию машинного обучения с учителем (обучение по прецедентам), основной принцип которой — по частным данным, представляющим набор пар «объект — ответ», выявить закономерности, присущие не только конкретной обучающей выборке, но и генеральной совокупности данных [Прикладная и компьютерная лингвистика 2016]. В нашем случае в качестве обучающей выборки рассматривалось множество объектов — массмедийных политических англоязычных текстов о России, преследующих цель манипуляции массовым сознанием читателя. Каждому объекту из обучающей выборки приписывался признак

«Является манипулятивным» на основании измерения шести маркеров манипуляции, которые в контексте машинного обучения получили статус параметров (feature). Параметр, или фича, в машинном обучении определяется как индивидуальное измеряемое свойство (признак) исследуемого объекта (*a feature is an individual measurable property of a phenomenon being observed* [Bishop 2006]). «Натренированная» на размеченных данных (пары «объект — ответ») модель затем позволяет распространить усвоенные при работе с ними закономерности на «свежую» выборку текстовых данных.

Что касается этапов работы, то независимо от сферы применения можно выделить

общие этапы решения задачи машинного обучения [Ножина, Луценко 2016], основные из которых: постановка задачи исследования, определение требований к данным, сбор первичных данных, их предварительный анализ, улучшение качества данных, выбор наилучшего алгоритма обучения, настройка параметров алгоритма для улучшения качества получаемой модели, интерпретация полученных результатов.

В рамках нашей работы в качестве алгоритма для обучения модели был выбран алгоритм «Деревья решений» [Люгер 2005]. Построение деревьев классификации — совокупность методов аналитической обработки больших массивов информации с целью выявить в них значимые закономерности и/или систематические связи между предикторными переменными, которые затем можно применить к новым совокупностям измерений. Деревья решений представляют собой последовательные иерархические структуры, состоящие из узлов, которые содержат правила, т. е. логические конструкции вида «если... то...». Конечными узлами дерева являются «листья», соответствующие найденным решениям и объединяющие некоторое количество объектов классифицируемой выборки.

Так, на рис. 1 представлено одно из первых «деревьев решений» для классификации множества из 79 текстов. Построение дерева решений идет сверху вниз. В первом узле компьютер констатирует, что из 79 загруженных предположительно манипулятивных текстов 35 содержат статистически большое количество использований имени собственного *Poutine*, что уже является основанием для их классификации как манипулятивных. Оставшиеся 44 текста проверяются на наличие военной лексики: в 14 ее количество статистически значимо, а в 30 — нет. Первые 14 попадают в категорию манипулятивных, а оставшиеся 30 проверяются на наличие других маркеров и т. д. В дальнейшем данный алгоритм был модифицирован по причинам, которые обсуждаются ниже, но тем не менее принцип «Деревья решений» был сохранен.

2. Интерфейс онлайн-версии программы и алгоритм работы с ней

При поддержке Центра цифровой экономики СФУ полученная модель была использована в онлайн-версии программы, доступной в режиме тестирования [Text manipula-

tive bias estimation]. На рис. 2 представлен скриншот начальной страницы программы.

Для проверки текста на манипулятивность необходимо вставить его в форму «Paste your text», затем набрать кодовое слово в поле «Enter your code», предварительно запросив его по ссылке xyz@verba.lab.ru (расположена после вопроса «Don't have code?»), и нажать на кнопку «Check text».

Рассмотрим результаты проверки трех произвольно взятых из англоязычных СМИ текстов.

Текст 1 [Nardelli, Rankin, Arnett 2015].

Vladimir Putin's approval rating at record levels

Almost nine out of 10 Russians approve of their president, according to survey that also highlights support for Ukraine strategy

Vladimir Putin's approval rating is at record levels, with nine out of 10 Russians saying they have a positive view of their president. Putin had an approval of 87% in July, and an all-time high of 89% in June, according to Levada Centre polling.

Following a drop in popularity in 2012 and 2013, when Putin's approval ratings dropped into the 60s, the Russian president's popularity picked up again last year on the back of events in Ukraine.

According to separate Levada figures from June, 66% believe that western sanctions are meant to humiliate and weaken Russia, and only 5% think they are about ending the conflict in Ukraine.

*Some 70% of Russians believe the country should stick to its current position on Ukraine, while 20% say it would be better to make concessions in order to avoid sanctions. 87% support **the annexation of Crimea**, and only 4% think that the eastern regions of Donetsk and Luhansk should return to their pre-conflict status.*

However, when it comes to the economy, only 13% describe Russia's current predicament as good, while 53% describe it as average and 31% as either bad or very poor.

*With nine out of 10 Russians approving of Putin, **the president's ratings are now better than they were in 2008, at the start of the Russian-Georgian war.***

They are also unusually high compared with the popularity of other world leaders. Only Angela Merkel comes anywhere near Putin, with nearly seven in 10 Germans approving of the chancellor.

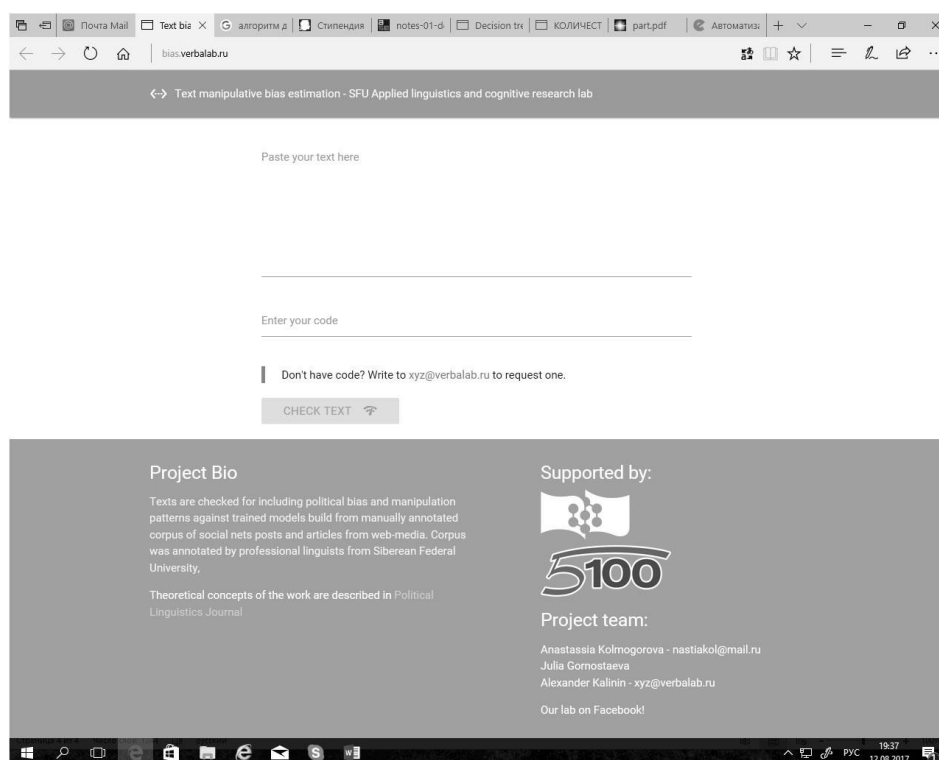


Рис. 2. Скриншот начальной страницы онлайн-версии программы

Bias degree is medium!	
Feature name	Feature ratio (%)
Custom markers (opionated adjectives and strong violent verbs)	0.017857142857142856
Military terms	0.017857142857142856
Terms related to Nazism	0
Political persons mentioned	0.017142857142857143
Words starting with 'pro-' and 'anti-' prefixes	0
Terms related to 'Soviet' lexeme	0
BACK TO FORM ←	

Рис. 3. Результаты проверки классификатором текста 1

При проверке классификатор помещает данный текст в категорию «среднего уровня манипулятивности» (Bias degree is medium) (рис. 3), поскольку выявлено присутствие трех параметров — первичные маркеры манипуляции, установленные в эксперимен-

тальной работе (*custom markers*), военная лексика (*military terms*) и политические персоналии (*political persons mentioned*). Однако статистические показатели данных параметров не достигают уровня, характерного для текстов с ярко выраженной манипуляцией.

Если мы проведем собственно дискурсивно-коммуникативный анализ текста 1, то увидим, что, хотя в нем нет открытых стратегий, направленных на дискредитацию России и российского президента, — в статье говорится о небывало высоком рейтинге В. Путина среди россиян по опросам 2015 г., — тем не менее всякое упоминание высокого уровня поддержки россиянами своего президента сопровождается ссылкой на такие события, как «украинский кризис», «аннексия Крыма», «военный конфликт между Россией и Грузией»: *рейтинг президента, после спада в 2012—2013 гг., вновь поднялся на фоне событий на Украине; 87 % россиян поддерживают аннексию Крыма; сейчас рейтинг президента даже выше, чем в 2008, когда начался военный конфликт между Россией и Грузией*. Таким образом, посредством намеренного соположения информации манипуляция все-таки осуществляется (формируется мнение о том, что позиция Кремля в отношении Украины — не борьба за национальные геополитические интересы, а всего лишь ловкий предвыборный ход), хотя и не имеет той силы воздействия, что, например, в тексте 2.

Текст 2 [Putting Russia's Crimean Intervention in Context 2014].

Текст объемом около 3000 знаков озаглавлен «Putting Russia's Crimean Intervention in Context». Статистическая оценка показывает, что в тексте содержатся пять маркеров манипуляции из шести: «первичные» маркеры, выделенные в экспериментальной работе; военная лексика; нацистская лекси-

ка; лексемы с префиксами *pro-* и *anti-*; лексемы, содержащие морфему *Soviet*. Текст классифицируется программой как имеющий высокую степень манипулятивности (Bias degree is high) (рис. 4).

Рассмотрим несколько примеров из текста 2:

Russia's military intervention in Crimea and subsequent buildup of forces throughout the region has raised speculation that Russia could choose to intervene elsewhere in the former Soviet space, including in mainland Ukraine, Moldova, Georgia or perhaps even the Baltic states (военное вторжение России в Крым и последующее наращивание военного присутствия в регионе наводит на размышления о том, что в принципе Россия может вторгнуться в любую страну постсоветского пространства: в центральные области Украины, Молдавию, Грузию или, возможно, даже в страны Балтии).

Из фактической информации в данном пассаже артикулируется только то, что территория Крыма вошла в состав Российской Федерации, и то, что в прошлом Украина, Молдова, Грузия и страны Балтии входили в состав Советского Союза. Однако использование лексического повтора с семой «вторжение» (*intervention, intervene*), военной лексики (*military, buildup of forces*), а также апелляция к образу «Советов» в эпоху холодной войны актуализируют в сознании читателя образ врага, угрозы, вызывающей чувство тревоги за будущее, на фоне которого факты воспринимаются искаженно и предсказуемо предвзято.

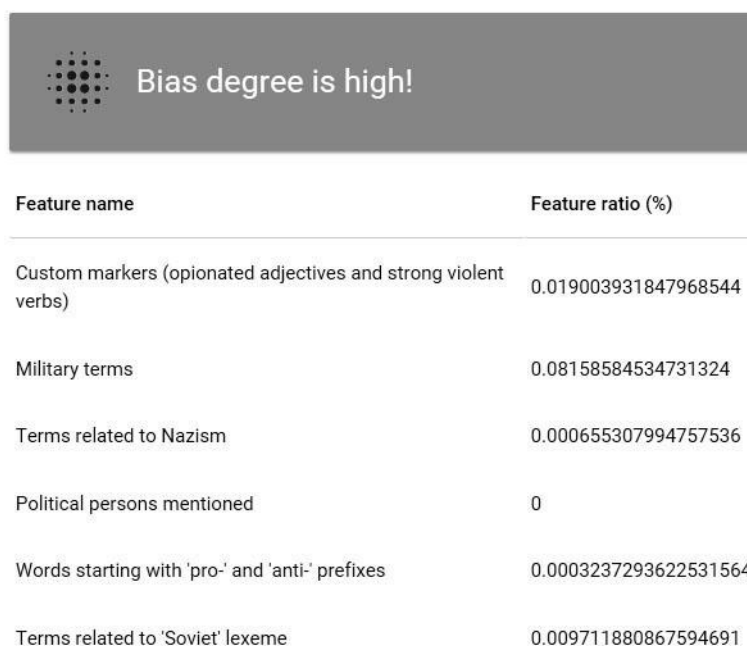


Рис. 4. Результаты проверки классификатором текста 2

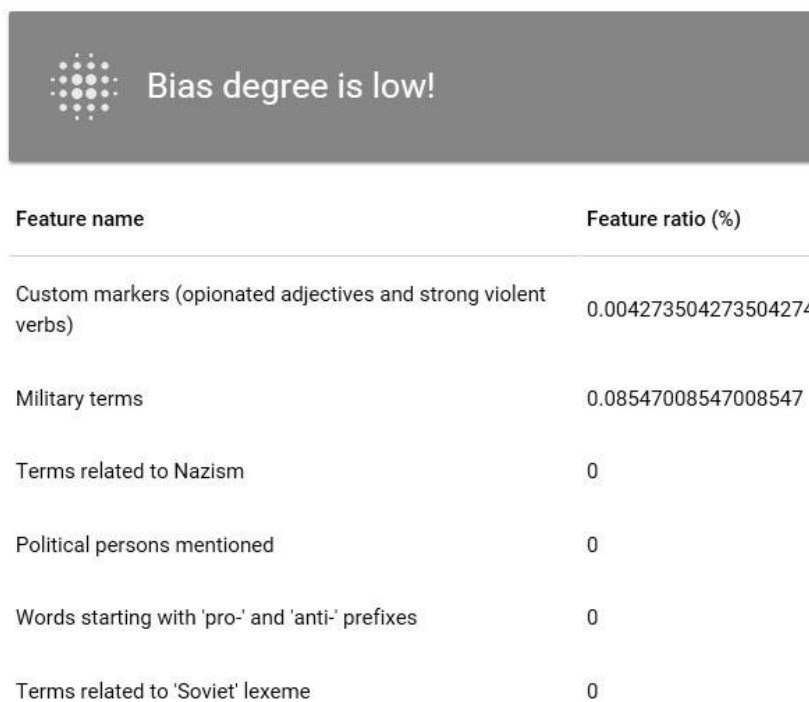


Рис. 5. Результаты проверки классификатором текста 3

Russia's most recent actions in Crimea follow a pattern similar to that of its interventions in other parts of the former Soviet space over the past 25 years. This pattern, though it varies in terms of specific application, broadly consists of three categories: the organization of ethnic Russian or pro-Russian social and political groups and movements; the deployment or support of informal or unofficial security forces in key areas; and finally the launching of formal military operations (недавние действия России в Крыму выстроены по привычному сценарию вторжения, многократно опробованному в течение последних 25 лет в других регионах бывшего СССР. Этот сценарий хотя и варьируется в зависимости от конкретных условий, но в общих чертах состоит из трех этапов: создание общественных движений, состоящих из этнических русских и сочувствующих русским представителей других этнических групп; формирование или поддержка уже существующих сил безопасности в ключевых точках региона; и, наконец, начало официальной военной операции).

В данном пассаже мы вновь наблюдаем набор военной лексики (*intervention, military operations, security forces*), упоминание СССР (*the former Soviet space*) и деление, с одной стороны, на русских и прорусски настроенных людей (*pro-Russian social and political groups*), с другой — всех остальных. Создав таким образом необходимое эмоциональное состояние у читателя, авторы тек-

ста переходят к изложению фактов (ввод советских войск на территорию бывшей в ту пору советской республикой Литвы в 1991 г. после объявления последней в одностороннем порядке своей независимости; советское, а затем российское военное присутствие в Приднестровье; ввод войск в Абхазию и Южную Осетию в 2008 г. для защиты геополитических интересов России), которые в уже заданном фокусе восприятия приобретают то смысловое содержание, которое и нужно было субъекту манипуляции.

Наконец, рассмотрим текст 3 [White 2017].

An unarmed Russian air force jet flew over multiple sensitive Washington, DC area locations on Wednesday, according to a report by CNN.

The reported flight was conducted in accordance with the Treaty on Open Skies, which allows participating members to conduct observational flights over other countries' military facilities. A source told CNN that American military personnel were on board.

The US Capitol, the Pentagon, Central Intelligence Agency and the presidential retreat at Camp David were among the sites that the aircraft reportedly flew over — in addition to the Trump National Golf Club in Arlington, Virginia. Sources told CNN the aircraft was also scheduled to visit the New Jersey town of Bedminster, where Donald Trump is currently on vacation.

Between the Treaty on Open Skies taking effect in 2002 and 2016, the United States flew more than twice as many flights over Russia than vice versa, according to a US State Department fact sheet.

"The Treaty's primary value is its role in building transparency and confidence, not intelligence gathering," the fact sheet says, "allowing members to "avoid surprises in a cooperative way."

While America already has some imaging capacity, the fact sheet adds, "there are significant parts of Russia best imaged by Treaty aircraft. The Treaty provides valuable information, especially for our Allies and partners that do not have the same imaging capabilities as the US."

Классификатор помещает данный текст в категорию текстов с низким уровнем манипуляции (рис. 5) — «Bias degree is low».

Текст британского издания посвящен факту полета российского самолета-разведчика над стратегическими объектами в Вашингтоне — Капитолием, Пентагоном, ЦРУ. Отмечается, что полет был осуществлен в рамках Договора о так называемом «открытом небе», согласно которому самолеты США также совершают наблюдательные полеты над российскими объектами.

Хотя в тексте содержится военная лексика и несколько маркеров, выделенных в ходе экспериментальной работы, однако данные лексемы и речевые клише не используются для нагнетания атмосферы страха и создания образа врага — употребление военной лексики обусловлено тематикой сообщаемой информации.

3. Проблемы, обнаруженные во время тестирования

Во время тестирования прототипа программы исследовательский коллектив столкнулся с некоторыми проблемами, для большинства которых решение было найдено.

Проблема выборки. Первоначально выборка, на которой тренировался алгоритм, была составлена только из текстов, в которых по результатам дискурсивного анализа лингвистами-экспертами был установлен факт наличия манипулятивных стратегий и техник. После обучения алгоритм опирался на статистические закономерности именно таких текстов, в результате, встретив имя *Poutine*, программа классифицировала текст как манипулятивный. Тогда обучающая выборка была расширена за счет неманипулятивных текстов, содержащих некоторые из маркеров манипулятивности. Например, в выборку были включены тексты, рассказывающие о полете В. Путина с журавлями,

или репортажи с мест военных событий в Сирии, где используется много военной терминологии, но текст не затрагивает отношения России и США и не имеет манипулятивного характера.

Проблема учета количества параметров. На рис. 1 было показано, что первоначально в результате применения алгоритма «Деревья решений» текст классифицировался как манипулятивный уже в том случае, если у него присутствовал один ярко выраженный параметр (и эта проблема связана с предыдущей — несбалансированностью обучающей выборки). Однако тестирование показало, что при таком подходе практически все политические поляризованные тексты классифицируются как манипулятивные. Тогда было принято решение о шкалировании по критерию количества ненулевых параметров. Так, если в тексте параметры с ненулевым значением составляют от 0 до 1, то текст неманипулятивный, от 2 до 3 — низкой степени манипулятивности, 4 — средней, 5—6 — высокоманипулятивный текст.

В перспективе валидность получаемых результатов должна улучшаться: чем большее количество текстов «пройдет» через классификатор, тем больше статистических закономерностей будет выявлено, а значит, тем точнее будет анализ.

Заключение

Созданный научным коллективом Лаборатории прикладной лингвистики и когнитивных исследований классификатор англоязычных политических текстов является научной разработкой, потенциальными потребителями которой могут стать интернет-пользователи, в том числе акторы *social media* (блогеры, ютуберы), стремящиеся к получению и обсуждению объективной качественной информации из англоязычных источников, а также государственные органы, проводящие мониторинг объективности информации, поступающей в российское массмедийное пространство. В качестве перспектив проекта рассматривается продвижение программы в форме мобильного приложения, доступного для скачивания, а также создание аналогичного классификатора для русскоязычных текстов.

ИСТОЧНИКИ

1. Nardelli A., Rankin J., Arnett G. Vladimir Putin's approval rating at record levels // The Guardian. 2015. 23 July. URL: <https://www.theguardian.com/world/datablog/2015/jul/23/vladimir-putins-approval-rating-at-record-levels> (date of access: 12.08.2017).
2. Putting Russia's Crimean Intervention in Context // Stratfor. 2014. 12 Apr. URL: <https://worldview.stratfor.com/article/putting-russias-crimean-intervention-context>.
3. Text manipulative bias estimation. URL: <http://bias.verbalab.ru/>.

4. White J. B. Russian airforce jet flies over Washington DC // Independent. 2017. 9 Aug. URL: <http://www.independent.co.uk/news/world/americas/russia-jet-flight-washington-dc-a7885331.html> (date of access: 14.08.2017).

ЛИТЕРАТУРА

5. Колмогорова А. В., Калинин А. А., Горностаева Ю. А. Языковые маркеры манипуляции в поляризованном политическом дискурсе: опыт параметризации // Политическая лингвистика. 2016. № 4 (58). С. 194—200.

6. Люгер Дж. Ф. Искусственный интеллект: стратегии и методы решения сложных проблем. — М.: Вильямс, 2005.

7. Ножина Е. П., Луценко А. В. Типовые этапы решения задачи обучения по прецедентам // Перспективы развития информационных технологий. 2016. № 29. С. 11—15.

8. Паршин П. Б. Речевое воздействие: основные сферы и разновидности. — М.: Эксмо, 2007.

9. Прикладная и компьютерная лингвистика / под ред. И. С. Николаева, О. В. Митрениной, Т. М. Ландо. — М.: Ленанд, 2016.

10. Стернин И. А. Основы речевого воздействия. — Воронеж, 2012.

11. Billig M., Marinho Ch. Manipulating information and manipulating people // Critical Discourse Studies. 2014. № 11: 2. P. 158—174. DOI: 10.1080/17405904.2013.852982.

12. Bishop C. Pattern recognition and machine learning. — Berlin: Springer, 2006.

13. Dijk. T. A. van. Discourse and manipulation // Discourse and Society. 2006. № 17 (2). P. 359—383.

14. Eissa M. M. Polarized discourse in the news // Procedia — Social and Behavioral Sciences. 2014. P. 70—91.

A. V. Kolmogorova, A. A. Kalinin, J. A. Gornostaeva
Krasnoyarsk, Russia

COMPUTER PROGRAM DESIGN FOR CLASSIFYING ENGLISH POLARIZED POLITICAL TEXTS BY THEIR MANIPULATIVE IMPACT: RESULTS AND DISCUSSION

ABSTRACT. The article discusses the results in terms of the project devoted to designing a computer classifier of English polarized political texts about Russia by their manipulative impact. The project relevance is due to the fact that English texts largely using manipulative technics and strategies could serve as an instrument of information war against the Russian civil society. In the situation of national mass media areas penetrance, such texts, after reposting and reprinting by Russian Internet users or Internet media, could have a serious impact on the worldview of young Russian people speaking English. The computer classifier is built on the linguistic methodology of discourse analysis and sociolinguistic experiment, while its technological basis is grounded on the supervised machine learning approach using decision trees algorithm. Our feature list includes the Soviet lexicon items, the Nazi lexicon items, military terms, discursive markers of manipulation selected by the respondents, the prefixes “pro-”, “anti”, the morpheme “Soviet” and precedent names or political personalities. After checking the text fragment, the classifier attributes it to one of the four classes and displays on the screen the appropriate message: texts without any manipulation (This text doesn't seem to have any bias), texts with low degree of manipulation (Bias degree is low), texts with medium degree of manipulation (Bias degree is medium), highly manipulative texts (Bias degree is high). The computer program's prototype has been successfully tested in the Centre of Digital Economics of Siberian Federal University.

KEYWORDS: manipulation; political discourse; polarized discourse; computer program; media discourse; mass media; media; political text; the English language.

ABOUT THE AUTHORS: Kolmogorova Anastasia Vladimirovna, Doctor of Philology, Full Professor, Head of Department of Romance Languages and Applied Linguistics, Siberian Federal University, Krasnoyarsk, Russia.

Gornostaeva Julia Andreevna, Post-graduate Student, Department of Romance Languages and Applied Linguistics, Siberian Federal University, Krasnoyarsk, Russia.

Kalinin Alexander Andreevich, Senior Lecturer of the Chair of Romance Languages and Applied Linguistics, Siberian Federal University, Krasnoyarsk, Russia.

REFERENCES

1. Nardelli A., Rankin J., Arnett G. Vladimir Putin's approval rating at record levels // The Guardian. 2015. 23 July. URL: <https://www.theguardian.com/world/datablog/2015/jul/23/vladimir-putins-approval-rating-at-record-levels> (date of access: 12.08.2017).

2. Putting Russia's Crimean Intervention in Context // Stratfor. 2014. 12 Apr. URL: <https://worldview.stratfor.com/article/putting-russias-crimean-intervention-context>.

3. Text manipulative bias estimation. URL: <http://bias.verba-lab.ru/>.

4. White J. B. Russian airforce jet flies over Washington DC // Independent. 2017. 9 Aug. URL: <http://www.independent.co.uk/news/world/americas/russia-jet-flight-washington-dc-a7885331.html> (date of access: 14.08.2017).

5. Kolmogorova A. V., Kalinin A. A., Gornostaeva Yu. A. Ya-zykovye markery manipulyatsii v polyarizovannom politicheskom diskurse: opyt parametrizatsii // Politicheskaya lingvistika. 2016. № 4 (58). S. 194—200.

6. Lyuger Dzh. F. Iskusstvennyy intellekt: strategii i metody resheniya slozhnykh problem. — M.: Vil'yams, 2005.

7. Nozhina E. P., Lutsenko A. V. Tipovye etapy resheniya zadachi obucheniya po pretsedentam // Perspektivy razvitiya informatsionnykh tekhnologiy. 2016. № 29. S. 11—15.

8. Parshin P. B. Rechevoe vozdeystvie: osnovnye sfery i raznovidnosti. — M.: Eksmo, 2007.

9. Prikladnaya i komp'yuternaya lingvistika / pod red. I. S. Nikolaeva, O. V. Mitreninoy, T. M. Lando. — M.: Lenand, 2016.

10. Sternin I. A. Osnovy rechevogo vozdeystviya. — Voronezh, 2012.

11. Billig M., Marinho Ch. Manipulating information and manipulating people // Critical Discourse Studies. 2014. № 11: 2. P. 158—174. DOI: 10.1080/17405904.2013.852982.

12. Bishop C. Pattern recognition and machine learning. — Berlin: Springer, 2006.

13. Dijk. T. A. van. Discourse and manipulation // Discourse and Society. 2006. № 17 (2). P. 359—383.

14. Eissa M. M. Polarized discourse in the news // Procedia — Social and Behavioral Sciences. 2014. P. 70—91.