

Определение лингвистических показателей сетевого социально-политического дискурса с помощью интеллектуального анализа текстов*

О. Г. Григорьев, А. А. Чуганская, М. А. Станкевич

Федеральный исследовательский центр «Информатика и управление» РАН, г.Москва, Россия

Аннотация. В статье проведено изучение социально-политического дискурса с использованием методов интеллектуального анализа текстов по социальной и политической психологии. Предложена методология применения инструментария в социально-политических исследованиях. Определены и проанализированы наиболее значимые психолингвистические показатели для комментариев на политических и неполитических русскоязычных YouTube-каналах.

Ключевые слова: интеллектуальный анализ текста, лингвистические маркеры, социальная группа, политическая психология, общественно-политический дискурс, YouTube.

DOI 10.14357/20718594220104

Введение

Анализ общественно-политического дискурса, который представлен комментариями к видеороликам политической и неполитической направленности в русскоязычных YouTube-каналах, привлекает всё большее внимание исследователей. Предполагается, что пропагандистские ролики вызывают более эмоциональную реакцию, чем ролики учебной и бытовой направленности; обсуждение своих и чужих эмоций и их причин идёт более активно в сетевых дискуссиях, разгорающихся под влиянием политической пропаганды. Сформулированные гипотезы были проверены на больших данных сетевых дискуссий с помощью метода реляционно-ситуационного анализа текстов. Используя инструмент интеллектуального анализа

текста TITANIS, были выявлены наиболее значимые психолингвистические показатели политического дискурса. Проведённое исследование имеет значимую методическую функцию для изучения больших социальных групп.

1. Методы социально-психологических и политологических исследований

Проведение социо-гуманитарных исследований на современном уровне требует разработки методологии выделения субъекта исследований для разных уровней взаимодействия как индивидов, так и групп. Это связано с необходимостью классифицировать общественное поле исследований более дифференцированно по сравнению с простым выделением человека и/или группы.

* Работа выполнена при частичной финансовой поддержке РФФИ (проект опн №21-011-31638)

✉ Чуганская Анфиса Анваровна. E-mail: anfisa.makh@gmail.com

В современных психологических исследованиях выделяют три уровня рассмотрения социально-психологических понятий: макроуровень – большие группы и сообщества, мезо-уровень – непосредственное окружение индивида (семья, места проживания и работы, вовлеченность в группы с «живым» общением), микроуровень – социальные особенности самой личности и ее взаимодействия с другими [1]. Изменение структуры общения, растущая роль дистанционных и виртуальных форм взаимодействия людей приводят к необходимости более детального рассмотрения больших групп, а именно сетевых сообществ, которые приобретают специфические черты за счет коммуникации в пространстве Internet. Особенно актуальными становятся социально-психологические и политологические исследования.

М. Ю. Урнов отмечает, что в рамках социально-политических исследований ведущими методами остаются социологический опрос и анализ текстовой информации, представленной в СМИ и в сетевом дискурсе [2]. Для анализа социально-политических тематик наблюдается тенденция увеличения значимости оценки эмоционального компонента, настроений и переживаний. Во-первых, эмоциональные состояния во многом влияют на восприятие, оценку и политическое поведение субъектов в группе. Так, Л. Берковиц отмечает, что «позитивные чувства часто продуцируют позитивные мысли, в то время как негативные настроения порождают негативные и даже враждебные мысли – если только не иницируется нацеленный на ослабление негативного аффекта процесс саморегуляции, мотивирующий индивида избегать “плохих” мыслей» [3, с. 157]. Именно поэтому в социально-политических исследованиях уделяется большое внимание как косвенным признакам оценки потребностей членов общества через оценку материального благополучия, так и степени удовлетворения базовых потребностей. Во-вторых, существует различие между влиянием длительных, «фоновых» настроений, которые могут стать почвой для кардинальных политических изменений и краткосрочными аффективными воздействиями связанными, как правило, с психологией толпы. С. Хантингтон отмечает важность институционализации политических процессов и существенное влияние эмоционального фактора в переходных общест-

вах: «Обществу со слабыми политическими институтами нелегко справляться с последствиями личных и групповых вождедений» [4, с. 42].

Исследования взаимосвязи эмоциональности и ее влияния на последующее поведение во многом основано на примитивизации индивидуального поведения и заражения коллективными эмоциями человека в толпе, что было описано еще в работах Г. Лебона [5]. В этом же направлении рассуждает и французский психолог С. Московичи: «Внушение или влияние ... предполагают: уход от логического мышления, даже его избегание, и предпочтение алогичного мышления; раскол рационального и иррационального в человеке, его внутренней и внешней жизни» [6, с. 42].

Понятие «эмоциональной атмосферы общества» основывается на работах Г. Тарда и К. Левина. В конце XIX века Г. Тард вводит понятие «социальной системы чувств» [7], которое описывает преобладающие в обществе эмоции и настроения. Социальная система чувств обеспечивает с одной стороны стабильность общества, единую систему ценностей и построение идентичности, определяемой принадлежностью к той или иной группе, а с другой – сама она подвержена изменениям под влиянием политических событий (волнения, протесты, выборы, избирательные кампании) и ее динамика является необходимой для развития общества. Синхронизацию настроений небольших групп и индивидов осуществляют «настройщики», к которым Г. Тард отнес как харизматичных лидеров (в современном обществе к ним относятся также и инфлюэнсеры в сетевом общении), так и общественные массовые мероприятия (например, крупные праздники, спортивные или культурные события).

В 30-х годах XX века К. Левин ввел новое понятие «социальной атмосферы», «группового климата» [8]. Он связал его со структурными характеристиками группы, а именно: типом лидерства, способом принятия решений, социальными статусами, эмоциональными характеристиками, которые во многом описывают доминирующее настроение в группе. При этом социальную группу К. Левин определял «как динамическое целое, основанное не столько на сходстве, сколько на взаимной зависимости» [8, с. 344]. Это выделяет значимость эмоционального компонента при образовании групп

(сообществ), что во многом характерно для сетевого взаимодействия. В последующем идеи К. Левина развил Ж. де Ривер, который отмечал, что эмоциональный климат влияет «на склонность людей поддерживать или отторгать инициативы политических лидеров, решать проблемы с помощью переговоров или, напротив, конфликтов и пр.» [2, с. 28].

Таким образом, актуальность и значимость анализа эмоциональности в социально-политических исследованиях продиктована необходимостью изучения больших социальных групп и их представленностью в Internet пространстве. Это связано и с возрастающей ролью соцсетей в общественно-политической жизни. Использование сетевого контента, представленного Twitter, Facebook, YouTube, LiveJournal и Instagram, становится одним из обязательных параметров оценки социальной активности субъекта, причем не только на уровне отдельной личности (политического лидера), но и на уровне больших социальных групп [9]. Развитие новых медиа и дистанционных технологий общения вносит большой вклад в систему получения новых знаний субъектом и процессы коммуникации, что подчеркивает актуальность исследований в области политологии, социологии и психологии групп, опирающихся на средства автоматизированного интеллектуального анализа. Перспективным направлением в сфере анализа сетевого контента общественно-политического дискурса является контент-анализ с учетом семантических показателей [10, 11].

Важную роль в оценке эмоционального компонента общественно-политического дискурса в сетевом контенте играют методы интеллектуального анализа текстов, которые позволяют учитывать различные лингвистические показатели и семантико-синтаксическое своеобразие национального языка. Одним из таких методов является предложенный Г.С. Осиповым и др. метод реляционно-ситуационного анализа текстов [12], основанный на коммуникативной грамматике русского языка Г. А. Золотовой [13]. Этот метод реализует идею описания семантико-синтаксических компонентов в контексте коммуникативно значимого целого (высказывания).

В исследованиях лингвистических параметров корпусов текста, в том числе и социально-политической направленности, используются, как правило, такие ресурсы, как WordNet,

WordNet Affect, SentiWordNet и их различные аналоги. Отдельно можно выделить разработанный J.W. Pennebaker компьютерный инструмент Linguistic Inquiry and Word Count (LIWC) [14], который позволяет определять частотность аффективных слов, маркеры тем, частоту встречаемости в корпусе текстов разных частей речи и их форм. Результаты выявления типизированных лингвистических параметров текста обычно основываются на лексических и/или частотных характеристиках текста. Однако существенным ограничением этого инструментария является трудность его адаптирования из-за необходимости учета особенностей национальных языков.

2. Инструментарий и методика исследования

Для анализа различий политических и неполитических комментариев в социальных сетях использовался инструмент интеллектуального анализа текста TITANIS, созданный в Федеральном исследовательском центре "Информатика и управление" РАН, который в первой версии именовался «Машина РСА» [15-17]. Инструмент предназначен для изучения психолингвистических показателей корпусов текста и описывает более 150 показателей. В инструменте TITANIS разработаны принципы хранения предикатных слов, базирующиеся на методе реляционно-ситуационного анализа, который позволяет проводить учет не только самого глагольного предиката, содержащегося в словаре предикатных слов, но и его именных зависимых (аргумента предиката) [14, 15]. Описание предикатной (преимущественно глагольной) лексики строится с учетом семантических и формальных свойств их аргументов (категориальный класс, падеж, семантическая роль) и получает воплощение в словаре предикатных слов (СПС). Каждый аргумент принадлежит к определенной специально разработанной тематической группе слов (ТГС).

Корпус текстов комментариев к YouTube-каналам формировался с использованием официального сервиса YouTube Data API, который предоставлял доступ к различным материалам YouTube. Комментарии пользователей к видео на YouTube имели двухуровневую структуру, объединяя, с одной стороны, комментарии непосредственно к видео, а с другой стороны –

комментарии на комментарий. Был проведен анализ следующей информации о комментариях обоих уровней: текст комментария, пользовательское имя автора комментария, дата и время публикации комментария, количество “лайков”, идентификатор комментария первого уровня (для комментариев второго уровня). Было также выполнено статистическое сравнение двух выборок - экспериментального корпуса (комментарии к видео общественно-политической тематики) и контрольного корпуса (комментарии к видео неполитической тематики), а также проведена проверка статистических гипотез о различии данных текстовых характеристик между корпусами.

Корпус текстов комментариев составили комментарии за период с 30 апреля 2020 г. по 30 апреля 2021 г. к роликам следующих каналов:

- политические государственнической направленности – Куликов, Соловьёв, Киселёв, Примаков, Багдасаров, Сатановский;

- политические оппозиционно-либеральной направленности – Соболь, Навальный, Яшин, Волков, Явлинский, Дудь;

- неполитические (фоновые) – Кулинарная стихия Ларисы Рубальской, KoLena, Шеф-повар Василий Емельяненко, Секреты Кулинарии, MARFFA, Хобби Тракториста, Будни Тракториста, Домашняя Кулинария, Gotovim s Asmoу, КУЛИНАРИЯ Любовь Ким, Лум Планет, Центр Архэ, Итальянцы by Kuzno Productions, Anna Chi, Кулинарим с Таней, Денис Косяков.

Каналы были выбраны на основе экспертной оценки с учетом содержания видео, численных показателей просмотра, скачивания, активности комментаторов к ним. В контрольную выборку были включены видео, которые не имели пересе-

чений с политической тематикой, однако по численным показателям просмотров и реакций комментаторов были сравнимы с политическими. В итоговый корпус текстов были собраны в общей сложности 6 млн. 288 тыс. 558 комментариев к 4807 видеороликам. Численная характеристика исследованных корпусов представлена в Табл. 1.

Чтобы выявить значимые лингвистические параметры общественно-политического корпуса текстов, с помощью методов машинного обучения была получена экспериментальная оценка предсказательной эффективности психолингвистических маркеров и тематических групп слов в качестве текстовых признаков. Проведены 4 отдельных эксперимента по бинарной классификации комментариев к видеороликам YouTube. В первом эксперименте была построена модель для дифференциации политических и неполитических видео, во втором, третьем и четвертом – для пар “государственные-оппозиционные”, “государственные-неполитические” и “оппозиционные-неполитические”. Материалом для извлечения признаков служили все относящиеся к данному видео комментарии, склеенные в один текст. Качество классификации оценивалось в каждой задаче, используя (1) функцию для выявления важности признаков CatBoostClassifier из библиотеки CatBoost [18]; (2) элементы библиотеки scikit-learn [19], в том числе метод опорных векторов (SVM), алгоритм построения классификатора случайного леса (RandomForestClassifier), оценка на кросс-валидации (cross_val_score), функция оценки f1-метрики (f1_score). Результаты бинарной классификации комментариев представлены в Табл. 2, жирным шрифтом отмечены классы, взятые как положительные при расчёте f1-метрики.

Табл.1. Количественное описание корпусов текстов

	Политические каналы			Неполитические каналы
	Государственные	Оппозиционно-либеральные	Вместе	
Число видео	49%	6%	55%	45%
Общее число сообщений	20%	60%	80%	20%
Медиана числа сообщений на одно видео	150	1243	187	165

Табл. 2. Результаты классификации комментариев

Каналы	SVM	RandomForest	CatBoost
Политические / Неполитические	96.8%	97.1%	98.0%
Государственные / Оппозиционные	89.6%	78.4%	89.6%
Государственные / Неполитические	96.9%	97.2%	97.6%
Оппозиционные / Неполитические	92.4%	94.8%	95.6%

Наилучшее качество классификации во всех экспериментах было получено при использовании классификатора CatBoost. При этом задача разделения политических и неполитических дискуссий решается заметно лучше, чем задача классификации государственных и оппозиционных каналов. Лучший классификатор CatBoost для каждого из 4 экспериментов был использован для извлечения наиболее информативных признаков.

3. Результаты и их обсуждение

Рассмотрим вначале результаты, полученные при выделении 10 параметров, наиболее значимых в процедуре классификации двух корпусов текстов – комментариев к каналам политической направленности и к фоновым, неполитическим каналам, имеющим развлекательно-бытовую тематику (Табл. 3).

Как следует из Табл. 3, комментарии к каналам политической направленности характери-

зуются преобладанием слов из тематических групп лексики отрицательной социальности, власти, обценной лексики. В комментариях к политическим каналам больше доля прописных символов, больше средняя длина слова, чаще встречаются глаголы 3 лица. В комментариях фоновых каналов, имеющих неполитическую тематику преобладают слова из тематической группы слов позитивных эмоций, а также местоимения и глаголы 1 лица.

Такой результат является вполне ожидаемым, исходя из преобладания негативных эмоций в политических обсуждениях, повышенной эмоциональности (доля прописных символов), направленности высказываний на описание поведения политиков (доля глаголов 3 лица). Как отмечает Т. Парсонс, политическое поведение и общение в сфере политических представлений во многом определяется ролевыми установками. Он выделял пять основных характеристик роли: эмоциональность, способ получения

Табл. 3. Рейтинг 10 наиболее значимых лингвистических показателей комментариев

Рейтинг 10 наиболее значимых лингвистических показателей комментариев к политическим и неполитическим (фоновым) каналам	Признак	Медиана больше в классе
1	Тональность текста на основе словаря linis-crowd	Фоновые
2	ТГС Социальность отрицательная	Политические
3	Доля прописных символов	Политические
4	Доля местоимений 1 лица единственного числа	Фоновые
5	ТГС Тематическая Власти	Политические
6	Доля глаголов 3 лица	Политические
7	Средняя длина слов (в числе символов)	Политические
8	Доля глаголов 1 лица	Фоновые
9	ТГС Эмоции позитивные	Фоновые
10	ТГС Обценная лексика	Политические

роли, масштаб, который определяется полем взаимодействия людей, формализация правил поведения роли, мотивация [20, с. 75].

Надо отметить, что политическая тематика подразумевает выраженность, в том числе в ролевом плане, эмоциональной составляющей [21], которая образует, как отмечает М.Ю. Урнов, и процесс идентификации с той или иной политической группой [2]. М.К. Соколон [22] отмечает, что еще в трактатах Аристотеля обозначаются четырнадцать политических эмоций.

Рассмотрим результаты, полученные при выделении 10 параметров, наиболее значимых при классификации комментариев на каналах с государственнической пропагандой и фоновых, неполитических каналов, с развлекательно-бытовым содержанием (Табл. 4).

Для комментариев на государственнических каналах отмечается высокая доля прописных символов, что связано как с высокой значимостью эмоционального компонента, так и упоминанием имён политиков и организаций, также преобладают тематические группы слов отрицательной социальности и власти, больше число употреблений уникальных слов. Комментарии на фоновых каналах развлекательно-бытового содержания характеризуются преобладанием местоимений и глаголов 1 лица. Здесь люди чаще описывают свой опыт, а не апеллируют к опыту той или иной социальной группы.

Полученные результаты хорошо интерпретируются в рамках концепции эффекта «мы и они». В социальной психологии этот эффект был введен В.Ф. Поршневым и понимается как

чувство принадлежности к определенной группе людей (эффект «мы») и одновременно чувство отстраненности от других, размежевания с другими группами (эффект «они») [23]. Эффект «мы-они» возникает на основе общности целей и деятельности членов группы, чувства единства, что становится источником формирования группы и ее развития. Для возникновения чувства «мы» требуется обособление с каким-то «они», с другой социальной группой [24].

В социальной психологии принадлежность к группе, идентификация с ней, включает в себя два дополнительных эффекта: сопричастности и эмоциональной поддержки. Во многом они направлены на поддержание внутригрупповой сплоченности и выступают мощным эмоциональным стимулом усиления процессов идентификации с группой. Эффект сопричастности выражается во включенности члена группы в проблемы, достижения, в деятельность группы, к которой он себя причисляет. Эффект эмоциональной поддержки проявляется в ожидании членом группы эмоциональной поддержки, понимания, сопереживания, помощи со стороны других членов группы [23].

Можно сказать, что комментаторы каналов с государственнической направленностью больше ориентированы на переживание групповой идентичности и на получение эмоциональной поддержки от единомышленников. При комментировании же на каналах развлекательно-бытовой направленности люди скорее выступают как индивидуалисты, не склонные к групповой поддержке или сегрегации.

Табл. 4. Рейтинг 10 наиболее значимых лингвистических показателей, дифференцирующих комментарии на каналах государственнической пропаганды и неполитических (фоновых) каналах

Рейтинг	Признак	Медиана больше в классе
1	Тональность текста на основе словаря <i>linis-crowd</i>	Фоновые
2	Доля прописных символов	Государственные
3	ТГС Социальность отрицательная	Государственные
4	Доля местоимений 1 лица единственного числа	Фоновые
5	Доля глаголов 3 лица	Государственные
6	Средняя длина слов (в количестве символов)	Государственные
7	Доля глаголов 1 лица	Фоновые
8	Число уникальных слов / Число слов	Государственные
9	ТГС Тематическая Власти	Государственные
10	ТГС Социальность предметная Здравоохранение, демография и экология	Фоновые

Рассмотрим 10 наиболее значимых лингвистических показателей, разделяющих комментарии на каналах либерально-оппозиционной пропаганды и фоновой, неполитической направленности (Табл. 5).

В комментариях на каналах либерально-оппозиционной направленности также преобладает тематика власти, отрицательной социальности, силовых структур (как косвенный признак обращения к государственным группам), прописные символы, негативные эмоции и местоимения 3 лица. Комментарии на фоновых каналах отличаются большим количеством уникальных слов, позитивных эмоций, положительной социальности, местоимений 1 лица.

Как можно видеть, в отличие от государственнической пропаганды, либерально-оппозиционная пропаганда приводит к снижению переживаний позитивного спектра – снижаются

показатели позитивных эмоций и положительной социальности. Надо отметить, что в целом для каналов либерально-оппозиционной направленности наблюдается большее число сообщений в комментариях. Рассмотрим Табл. 1. Доля комментариев именно этой группы составляет 60% от общей выборки, медиана числа комментариев 1243. При этом число проанализированных видео составляет всего 6% от общего числа. Таким образом, имеется большая активность комментаторов на каналах либерально-оппозиционной направленности.

Рассмотрим, наконец, разделение внутри самого политического дискурса и укажем на признаки, дифференцирующие комментарии на каналах с двумя видами пропаганды: государственнической и либерально-оппозиционной (Табл. 6).

Табл. 5. Рейтинг 10 наиболее значимых лингвистических показателей, дифференцирующих комментарии на каналах либерально-оппозиционной пропаганды и неполитических (фоновых)

Рейтинг	Признак	Медиана больше в классе
1	ТГС Тематическая Власти	Оппозиционные
2	ТГС Социальность отрицательная	Оппозиционные
3	Число уникальных слов / Число слов	Фоновые
4	ТГС Эмоции позитивные	Фоновые
5	ТГС Тематическая Силовые структуры	Оппозиционные
6	Доля прописных символов	Оппозиционные
7	Доля местоимений 1 лица единственного числа	Фоновые
8	ТГС Эмоции негативные, стеничные	Оппозиционные
9	Доля местоимений 3 лица	Оппозиционные
10	ТГС Социальность положительная	Фоновые

Табл. 6. Рейтинг 10 наиболее значимых лингвистических показателей, дифференцирующих комментарии на каналах государственнической и либерально-оппозиционной пропаганды

Рейтинг	Признак	Медиана больше в классе
1	Число уникальных слов / Число слов	Государственные
2	ТГС Тематическая Власти	Оппозиционные
3	Доля прописных символов	Государственные
4	Тональность текста на основе словаря linis-crowd	Оппозиционные
5	ТГС Молодежный жаргон	Оппозиционные
6	Средняя длина предложения (в числе слов)	Государственные
7	Доля глаголов 1 лица	Оппозиционные
8	ТГС Рациональная оценка положительная	Оппозиционные
9	ТГС Социальность предметная Здравоохранение, демография и экология	Государственные
10	Доля местоимений 1 лица единственного числа	Оппозиционные

Сравнение показателей внутри политического дискурса дает общую картину, характерную для обеих групп. Однако в наибольшей степени параметры касаются оппозиционных комментариев: тематика власти, молодежный жаргон (косвенный признак отнесенности высказывающихся к определенной возрастной группе). Появляется доля местоимений и глаголов 1 лица, положительная рациональная оценка, что в позиции субъективного описания характеризует сближение комментариев оппозиционных и фоновых каналов развлекательно-бытовой направленности, как инструмента самовыражения субъекта.

Заключение

Применение инструмента интеллектуального анализа текста TITANIS при описании лингвистических показателей социально-политического дискурса в русскоязычном сегменте YouTube показало следующее.

Инструмент TITANIS позволил выделить значимые параметры для разделения корпусов текстов комментариев на каналах политической и развлекательно-бытовой направленности. Для каналов с политическим содержанием отмечается наличие следующих лингвистических параметров: преобладание слов из тематических групп лексики отрицательной социальности, власти, обценной лексики, большая доля прописных символов, большая средняя длина слова, высокая частота встречаемости глаголов 3 лица.

При сравнении текстов комментариев в каналах государственнической и либерально-оппозиционной направленности с комментариями к фоновым, неполитическим каналам отмечается увеличение употребления слов из тематической группы негативных и стеничных эмоций в либерально-оппозиционной группе.

В комментариях на политических каналах значительно преобладают местоимения и глаголы 3 лица, в отличие от фоновых (развлекательно-бытовых каналов), где чаще упоминаются местоимения и глаголы 1 лица. Для политических каналов это отражает социально-политический эффект «мы-они» и большую направленность на групповую принадлежность и обозначение позиции своей группы относительно других.

Исследование социально-политического дискурса, проведенное с помощью метода интеллектуального анализа, свидетельствует о перспективности использования предложенных методов для исследования поведения больших Интернет-сообществ. Однако остается ряд ограничений и вопросов относительно оценки внешней валидности метода. Существенным ограничением является отсутствие выявления истинной позиции авторов комментариев и определения комментариев ботов, что значимо для построения содержательных выводов. Полученные результаты требуют аккуратной интерпретации и во многом зависят от тщательности разработки экспериментальной схемы социо-гуманитарного исследования, включающей критерии выбора контрольной и экспериментальной групп.

Интеллектуальный анализ текстов и сетевого контента, в том числе в области социально-гуманитарных наук, стремительно развивается и остро нуждается в новых методах, один из которых описан в данной статье.

Литература

1. Микляева А.В. Психология межвозрастных отношений. МО.: СВИВТ, М.: Перо, 2014. 159 с.
2. Урнов М.Ю. Эмоции в политическом поведении. М.: Аспект Пресс, 2008. 540с.
3. Берковиц Л. Агрессия: причины, последствия и контроль / Пер. с англ. СПб.: ПРАЙМ-ЕВРОЗНАК, 2002. 510 с.
4. Хантингтон С. Политический порядок в меняющихся обществах / Пер. с англ. М.: Прогресс-Традиция, 2004. 480 с.
5. Лебон Г. Психология народов и масс / Пер. с фр. СПб.: Макет, 1995. 311 с.
6. Московичи С. Век толп. Исторический трактат по психологии масс / Пер. с фр. М.: Центр психологии и психотерапии, 1998.
7. Тард Г. Социальная логика / Пер. с фр. СПб.: Социально-психологический центр. 1996. 554 с.
8. Левин К. Разрешение социальных конфликтов / Пер. с англ. СПб.: Речь, 2000. 407 с.
9. Arnaboldi V., Passarella A., Conti M., Dunbar R. Structure of ego-alter relationships of politicians in twitter. // J. Comput. Mediated Commun. 2017, 22 (5): 231–247.
10. Bayrakdar S. Yucedag I., Şimşek M., Doğru I. Semantic analysis on social networks: A survey. // International Journal of Communication Systems, 2020, 33. e4424. 10.1002/dac.4424.
11. Semenkovich S., Tsukanova O.A. On the Algorithms of Identifying Opinion Leaders in Social Networks. Procedia Computer Science. 7th International Conference on Information Technology and Quantitative Management (ITQM 2019): Information technology and quantitative

- management based on Artificial Intelligence. Elsevier, 2019: 778-785.
12. Осипов Г.С., Смирнов И.В., Тихомиров И.А. Реляционно-ситуационный метод поиска и анализа текстов и его приложения // Искусственный интеллект и принятие решений. 2008. № 2. С. 3–10.
 13. Золотова Г.А. Синтаксический словарь. Репертуар элементарных единиц русского синтаксиса. М.: Наука, 1998.
 14. Pennebaker J. W., Francis M.E., Booth R.J. Linguistic Inquiry and Word Count (LIWC): LIWC2001. Mahwah: Lawrence Erlbaum Associates, 2001.
 15. Ениколопов С.Н., Кузнецова Ю.М., Осипов Г.С., Смирнов И.В., Чудова Н.В. Метод реляционно-ситуационного анализа текста в психологических исследованиях // Психология. Журнал Высшей школы экономики. 2021. Т. 18. № 4. С. 748–769.
 16. Ениколопов С.Н., Кузнецова Ю.М., Смирнов И.В., Станкевич М.А., Чудова Н.В. Создание инструмента автоматического анализа текста в интересах социогуманитарных исследований. Часть 1. Методические и методологические аспекты // Искусственный интеллект и принятие решений. 2019. № 2. С. 28–38.
 17. Кузнецова Ю.М., Смирнов И.В., Станкевич М.А., Чудова Н.В. Создание инструмента автоматического анализа текста в интересах социогуманитарных исследований. Часть 2. Машина PCA и опыт ее использования // Искусственный интеллект и принятие решений. 2019. № 3. С. 40–51.
 18. Prokhorenkova L. et al. CatBoost: unbiased boosting with categorical features //arXiv preprint arXiv:1706.09516. 2017.
 19. Pedregosa F. et al. Scikit-learn: Machine learning in Python // Journal of Machine Learning Research. 2011. Т. 12. С. 2825-2830.
 20. Смелзер Н. Социология / Пер. с англ. М.: Феникс, 1994. 688 с.
 21. Koziak B. Retrieving Political Emotion: Gender, Thumos and Aristotle. New York. 2000.
 22. Ward A. Political Emotions: Aristotle and the Symphony of Reason and Emotion. // Canadian Journal of Political Science. 2007. 40. 543 – 544.
 23. Поршнев Б. Ф. Социальная психология и история. 2-е изд., доп. и испр. М.: Наука, 1979.
 24. Приходько А.И., Гулевич О.А. Межгрупповые эмоции в условиях политической конкуренции // Психологические исследования: электрон. науч. журн. 2010. N 1(9). URL: <http://psystudy.ru> (дата обращения: 30.12.2021).

Григорьев Олег Георгиевич. Доктор технических наук. Главный научный сотрудник, Федеральный исследовательский центр «Информатика и управление» Российской академии наук. Области исследований: информационные технологии, обработка естественного языка, извлечение знаний. E-mail: olegpolikvart@yandex.ru

Чуганская Анфиса Анваровна. Кандидат психологических наук. Научный сотрудник, Федеральный исследовательский центр «Информатика и управление» Российской академии наук. Области исследований: искусственный интеллект, социальная психология, когнитивные исследования. E-mail: anfisa.makh@gmail.com

Станкевич Максим Алексеевич. Инженер-исследователь, Федеральный исследовательский центр «Информатика и управление» Российской академии наук. Области исследований: обработка естественного языка, информационные технологии, интеллектуальный анализ текста. E-mail: maxastan95@gmail.com

Identification of Linguistic Indicators of Network Socio-Political Discourse Using Text Mining

O. G. Grigoriev, A. A. Chuganskaya, M. A. Stankevich

Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia

Abstract. The paper studies the socio-political discourse using methods of intellectual analysis of texts in social and political psychology. A methodology for applying tools in socio-political research is proposed. The most significant psycholinguistic indicators for comments on political and non-political Russian-language YouTube channels have been identified and analyzed.

Keywords: text mining, linguistic markers, social group, political psychology, socio-political discourse, YouTube.

DOI 10.14357/20718594220104

References

1. Miklyaeva, A.V. 2014. Psihologiya mezhvozzrastnyh otноshenij [Psychology of inter-age relations]. MO.: SVIVT, M.: Pero. 159 p.
2. Urnov, M.YU. 2008. Emocii v politicheskom povedenii [Emotions in political behavior]. М.: Aspekt Press. 240 p.
3. Berkowitz, L.J. (1993) Aggression: Its Causes, Consequences, and Control. New York: MacGraw Hill. 512 p.
4. Huntington, S. P. (1996) Political order in Changing Societies New Haven; London: Yale University Press. 488 p.
5. Lebon, G. (1982) Psychologie der Massen. 15. Aufl. Stuttgart: Kröner. 469 p

6. Moskovici, S. (1981) *L'Age des foules: un traité historique de psychologie des masses*. Paris: Fayard. 503 p.
7. Tarde, G. (1895) *Logique sociale*. Paris: Félix Alcan. 464 p.
8. Lewin, K. (1997) *Resolving social conflicts & Field theory in social science*. Washington, D.C: American Psychological Association. 243 p.
9. Arnaboldi, V., Passarella A., Conti M., Dunbar R. Structure of ego-alter relationships of politicians in twitter. *J. Comput. Mediated Commun.* 2017, 22 (5): 231–247.
10. Bayrakdar, S., Yucedag I., Şimşek M., Dođru I. Semantic analysis on social networks: A survey. // *International Journal of Communication Systems*, 2020, 33. e4424. 10.1002/dac.4424.
11. Semenkovich S., Tsukanova O.A. On the Algorithms of Identifying Opinion Leaders in Social Networks. *Procedia Computer Science. 7th International Conference on Information Technology and Quantitative Management (ITQM 2019): Information technology and quantitative management based on Artificial Intelligence*. Elsevier, 2019: 778-785.
12. Osipov, G.S., Smirnov I.V., Tihomirov I.A. Relyacionno-situacionnyj metod poiska i analiza tekstov i ego prilozheniya [Relational-situational method of search and analysis of texts and its applications] // *Iskusstvennyj intellekt i prinyatie reshenij [Artificial intelligence and decision-making]*. 2008. № 2. S. 3–10.
13. Zolotova, G.A. *Sintaksicheskij slovar'*. Repertuar elementarnyh edinic russkogo sintaksisa [The repertoire of elementary units of Russian syntax]. M.: Nauka, 1998.
14. Pennebaker, J. W., Francis M.E., Booth R.J. *Linguistic Inquiry and Word Count (LIWC): LIWC2001*. Mahwah: Lawrence Erlbaum Associates, 2001.
15. Enikolopov, S.N., Kuznecova Yu.M., Osipov G.S., Smirnov I.V., Chudova N.V. Metod relyacionno-situacionnogo analiza teksta v psihologicheskikh issledovaniyah [Method of relational-situational analysis of the text in psychological research] // *Psihologiya. Zhurnal Vysshej shkoly ekonomiki [Psychology. Journal of the Higher School of Economics]*. 2021. T. 18. № 4. S. 748–769. DOI: 10.17323/1813-8918-2021-4-748-769
16. Enikolopov, S.N., Kuznecova YU.M., Smirnov I.V., Stankevich M.A., Chudova N.V. Sozdanie instrumenta avtomaticheskogo analiza teksta v interesah socio-gumanitarnyh issledovanij. Chast' 1. Metodicheskie i metodologicheskie aspekty [Creation of an automatic text analysis tool in the interests of socio-humanitarian research. Part 1. Methodological and methodological aspects] // *Iskusstvennyj intellekt i prinyatie reshenij [Artificial intelligence and decision-making]*. 2019. № 2. C. 28–38.
17. Kuznecova YU.M., Smirnov I.V., Stankevich M.A., Chudova N.V. Sozdanie instrumenta avtomaticheskogo analiza teksta v interesah socio-gumanitarnyh issledovanij. Chast' 2. Mashina RSA i opyt ee ispol'zovaniya [Creation of an automatic text analysis tool in the interests of socio-humanitarian research. Part 2. The PC machine and the experience of its use] // *Iskusstvennyj intellekt i prinyatie reshenij [Artificial intelligence and decision-making]*. 2019. № 3. C. 40–51.
18. Prokhorenkova L. et al. CatBoost: unbiased boosting with categorical features // *arXiv preprint arXiv:1706.09516*. 2017.
19. Pedregosa F. et al. Scikit-learn: Machine learning in Python // *Journal of machine Learning research*. 2011. T. 12. S. 2825-2830.
20. Smelser, N. (ed.). (1988). *Handbook of sociology*. Newbury Park, Calif.: Sage Publications. 824 p.
21. Koziak B. *Retrieving Political Emotion: Gender, Thumos and Aristotle*. New York. 2000.
22. Ward A. Political Emotions: Aristotle and the Symphony of Reason and Emotion. *Canadian Journal of Political Science*. 40. 543 – 544, 2007. DOI: 10.1017/S0008423907070527.
23. Porshnev B. F. *Social'naya psihologiya i istoriya [Social psychology and history]*. 2-e izd., dop. i ispr. M.: Nauka, 1979.
24. Prihod'ko A.I., Gulevich O.A. Mezhruppovye emocii v usloviyah politicheskoy konkurencii [Intergroup emotions in conditions of political competition] // *Psihologicheskie issledovaniya: elektron. nauch. Zhurn [Psychological research: electron. scientific journal]*. 2010. N 1(9). URL: <http://psystudy.ru> (data obrashcheniya: 30.12.2021)

Grigoriev Oleg G. Doctor of Technical Sciences. Chief Researcher, Federal Research Center “Computer Science and Control”, the Russian Academy of Sciences. Research areas: information technologies, natural language processing, knowledge extraction. E-mail: oleggpolikvart@yandex.ru

Chuganskaya Anfisa A. PhD in Psychology. Researcher, Federal Research Center “Computer Science and Control”, the Russian Academy of Sciences. Research areas: artificial intelligence, social psychology, cognitive research. E-mail: anfisa.makh@gmail.com

Stankevich Maxim A. Research engineer, Federal Research Center “Computer Science and Control”, the Russian Academy of Sciences. Research areas: natural language processing, information technology, intelligent text analysis. E-mail: maxastan95@gmail.com