

Е.В. КРУЧИНСКАЯ*

**ЯЗЫК НЕНАВИСТИ КАК ИНДИКАТОР
АФФЕКТИВНОЙ ПОЛИТИЧЕСКОЙ ПОЛЯРИЗАЦИИ
В УСЛОВИЯХ МОБИЛИЗАЦИИ: ОТ ИЗМЕРЕНИЯ
К ПРОГНОЗИРОВАНИЮ**

Аннотация. Статья фокусируется на установлении эмпирической взаимосвязи между аффективной политической поляризацией и протестной мобилизацией. Проверяются два предположения, опирающиеся на анализ предыдущих исследований. Согласно первому, протест является фактором усиления проявления поляризации в обществе, но не является причиной проявления поляризации. Согласно второму, протест значительно влияет на расколы и идентификацию: в период протестной мобилизации большую значимость приобретает выражение политической идентификации, нежели социальной.

Для проверки предположений были собраны данные в русскоязычной сети «ВКонтакте» в два периода: экспериментальный, июль – сентябрь 2019 г. (московские протесты), и контрольный, март – май 2019 г. (непротестный период), $N_{total} = 141517$. Были использованы методы автоматической разметки на основании моделей *ruBERT* и последующей разметки кодировщиков, анализа временных рядов (кросс-корреляции, модель сезонного прогноза *SARIMAX*). Результаты двух моделей кросс-корреляционного анализа, основанных на сравнении проявления языка ненависти (на размеченных данных) в контрольный и экспериментальный период, подтверждают оба выдвинутых предположения статьи. Таким образом, подтверждено, что методология сбора и разметки данных, использованная в статье, позволяет строить прогнозные модели для оценки распространения языка ненависти в социальной сети. Это может получить дальнейшее развитие в исследованиях о прогнозе аффективной поляризации в ответ на конкретные протестные события.

* **Кручинская Екатерина Владиславовна**, старший преподаватель кафедры высшей математики, преподаватель Департамента политики и управления, Национальный исследовательский университет «Высшая школа экономики» (Москва, Россия), e-mail: ekruchinskaya@hse.ru

Ключевые слова: аффективная политическая поляризация; язык ненависти; политическая онлайн-коммуникация; расколы; *ruBERT*; прогнозирование языка ненависти.

Для цитирования: Кручинская Е.В. Язык ненависти как индикатор аффективной политической поляризации в условиях мобилизации: от измерения к прогнозированию // Политическая наука. – 2025. – № 1. – С. 156–180. <http://www.doi.org/10.31249/poln/2025.01.07>

Аффективная поляризация: между социальной и политической идентичностью

Поляризация имеет богатую историю изучения в социальных науках. Работы, посвященные этой проблеме, – самостоятельная область политологии, возникшая еще в предыдущем столетии [Esteban, Ray, 1994; DiMaggio, Evans, Bryson, 1996]. Понятие поляризации как политического явления успешно поддается измерению [Алескеров, Голубенко, 2003]. Тем не менее усиление поляризации в мире в XXI в. (обострение арабско-израильского конфликта, движение *Black Lives Matter* в США и «желтых жилетов» во Франции, «цветные» революции) сопровождается новыми методологическими и теоретическими дискуссиями относительно причин проявления поляризации, а также возможности прогнозирования ее всплесков [Kirkizh et al., 2024].

В связи с тем, что эпизоды политической мобилизации поставили множество новых вопросов перед исследователями поляризации, в настоящей работе изучается значимость различий в проявлении аффективной поляризации в условиях политической мобилизации и в отсутствии таких условий. В том числе изучается структура связанных с поляризацией политических конфликтов [Collins, 2012; Vartanova, Gladkova, Dunas, 2022], что составляет отдельный раздел исследований, относящихся к аффективной поляризации.

Относительно чего рассматривают проявления политического феномена поляризации? У этого концепта есть несколько контекстов проявления. Поляризация рассматривается как в контексте усиления разногласий и чувства вражды между политическими оппонентами и партиями [Gerber, Green, 1999; Zingher, Flynn, 2018], так и на фоне обострения ценностных расколов [Johnston, Manley, Jones, 2016], усиления чувства социальной идентичности или принадлежности к социальной группе [Tajfel, Turner, 1979; Iyengar, Sood, Lelkes, 2012; Costa, 2020].

Тип поляризации, в рамках которого формируется отношение политических масс или элит друг к другу как к аутгруппам и относящийся к типу отношений «свой – чужой», называют аффективной поляризацией. Аутгруппа – термин, в литературе понимаемый как группа, к которой не принадлежит индивид и с ней себя не ассоциирует [Tajfel, Turner, 1979; Агадуллина, Гулевич, 2013]. В этом случае (и далее в статье) считается, что аутгруппа формируется на основании политической или социальной принадлежности оппонента.

Согласно ряду современных исследований [Druckman et al., 2019], аффективная поляризация тесно связана с феноменом идентификации (или самоидентификации) индивида. Следует сделать оговорку, что в настоящей работе мы используем именно понятие идентификации как процесса, возникающего, в том числе, в ответ на внешние шоки. В англоязычной литературе, на которую сделаны ссылки, это понятие в основном также формулируется через слово *identification* [Greene, 1999], хотя в отдельных случаях используется и понятие *identity*.

В настоящей статье аффективная поляризация рассматривается как центральный тип поляризации, поскольку его изучение, во-первых, междисциплинарное, во-вторых – содержит в себе наибольшее количество лакун и «белых пятен», что будет показано далее.

Сегодня возникает все больше дискуссий о том, какую роль в проявлении аффективной поляризации играет самоидентификация, что может усиливать проявления аффективной поляризации, какие факторы делают ее проявления более интенсивными [Iyengar et al., 2019; Costa, 2020].

Аффективную поляризацию во всем ее многообразии обычно рассматривают через две призмы идентичностей: политическую, т.е. через идентификацию с политической или партийной группой [Коргунюк, 2012; Малинова, 2013; Webster, Abramowitz, 2017], или как идентификацию (или самоидентификацию) с социальными характеристиками ингруппы и непринятием социальной принадлежности аутгруппы, что обычно изучается в рамках социально-психологического подхода с позиций теории социальной идентичности, так называемой *social identity approach* [Rosenberg et al., 1995; Tajfel, Turner, 1979; Iyengar et al., 2019]. При этом понятие социальной идентичности объясняется идентификацией индивида с одной или несколькими группами по признакам пола, расы, национальности или иных социально-демографических характеристик [Druckman, Green, Iyengar, 2023; Orr, Flower, Huber, 2023; Nezi, 2024].

Можно предположить, что такое разделение на политическое и социально-психологическое довольно условно даже в теории. Какие теоретические положения могут быть подтверждением этого?

Представление о том, что политическая принадлежность есть отдельная форма социальной идентичности, было подтверждено еще в прошлом столетии [Campbell et al., 1960; Greene, 1999; Green, Palmquist, Schickler, 2002]. Тогда же исследователи установили, что политические предпочтения тесно связаны с понятиями идентичности (и идентификации) и самоидентификации в социальных группах [Rosenberg et al., 1995; Tajfel, Turner, 1979].

Некоторые авторы современной школы эмпирически подтверждают, что в проявлениях аффективной поляризации сочетаются как социальная, так и политическая идентичности [Mason, Wronski, 2018; Iyengar et al., 2019; Costa, 2020]. Эти же авторы приходят к выводу о том, что растущее проявление политической идентичности прямо сопряжено с ростом идентичности по половой, этнической или гендерной принадлежностям – т.е. с ростом того, что мы в работе называем социальной идентификацией.

В отдельном недавнем исследовании [Smith et al., 2024] с опорой на теорию групповой поляризации [Isenberg, 1986] и теорию Московичи и Заваллони [Moscovici, Zavalloni, 1969] явно подчеркнута, что социальная идентификация лежит в основе поляризации, проявляемой из-за ощущения политической идентичности. Похожие выводы делаются и в отечественных исследованиях [Bodrunova et al., 2019].

Вывод о связи социальной и политической компоненты идентификации также подтверждается исследованием, в котором проводится анализ связи идентификации с силой враждебности по отношению к аутгруппе [Mason, Notes, 2016]. Считается, что формирование негативного отношения к сторонникам партий-противников формируется не только на идее политического противостояния, но и из-за социально-идентичностного противопоставления себя с ними, например, на основании характеристик пола, возраста, расы [Druckman, Green, Iyengar, 2023; Orr, Flower, Huber, 2023; Nezi, 2024].

Вышеприведенный анализ в целом формирует одно из основных теоретических положений статьи, которое впоследствии находит подтверждение в эмпирике: идентичность индивида формируется на основании двух факторов: идентификации последнего как с политическим явлением или фактором, так и с социально-демографическими характеристиками. Тем не менее мы предпола-

гаем, что баланс идентичности может меняться в ответ на внешние события, что далее покажем с помощью эмпирического анализа.

А что это за внешние события? Открывая еще один «клик» поляризации как явления, стоит отметить, что поляризация связана и с понятием политического конфликта. Основы теории социального конфликта, сформулированные еще в прошлом столетии [Simmel, 1908; Coser, 1956], сегодня применяются в изучении поляризации в рамках анализа массовых конфликтов (переворотов и революций) с целью объяснения причины последних [Collins, 2008; 2012].

Другой подход – объяснение поляризации как таковой через понятие политического конфликта: так, Шедлер [Schedler, 2023] определяет политическую поляризацию как форму общественного конфликта, основанную на столкновении идеологических установок и связанной с ними партийных (политических) идентичностей. Это определение выводится автором на основании сопоставления двух принципиально разных подходов к поляризации. Во-первых, это теория поляризованного плюрализма Сартори [Sartori, 1976], развитая в ряде работ в дальнейшем [Graham, Svobik, 2020] и рассматривающая поляризацию как выражение конфликтов на основании социальных или политических расколов и увеличение идеологической дистанции между политическими оппонентами. Во-вторых, автор опирается на теорию социальной идентичности [Tajfel, Turner, 1979], согласно которой поляризация возникает на основании факта столкновения идентичностей – для этого не нужен конфликт как причина.

Выводы некоторых работ [Bauer et al., 2016; Theelen, Bohm, 2021; Harel et al., 2024], опирающихся на вышеназванные теории, выглядят противоречиво: с одной стороны, есть конкретные эмпирические подтверждения того, что в ситуации политического конфликта аффективная поляризация может расти, при этом проявления идентичности – не меняться [Rosler, 2016]. С другой стороны, в отдельных исследованиях делаются выводы о том, что конфликт может снижать аффективную поляризацию, являясь «внешней угрозой» и причиной сплочения [Bauer et al., 2016; Theelen, Bohm, 2020], хотя при этом и политическая, и социальная идентичность индивида могут увеличиваться, а отношение к аутгруппе эскалироваться [Harel et al., 2024].

Как объяснить этот исследовательский пазл? Для разрешения этого противоречия имеет смысл вернуться к тому, что ранее мы называли внешним фактором, влияющим на баланс идентичности.

Согласно основным положениям теории конфликтологии, таким внешним фактором может быть политическая мобилизация [Collins, 2008; 2012].

Связан ли рост поляризации и интенсификация эпизодов протестной мобилизации? Некоторые работы указывают на такую связь: изучается пример протестов и распространения петиций, возникших на фоне миграционного кризиса 2015 г. после распространения фотографии Айлана Курди (*Aylan Kurdi*)¹ [Thomas et al., 2016]. Авторы работы на основании проведенного социологического опроса утверждают, что в активную фазу протестов и развития миграционного кризиса поляризация относительно снимка была значимо выше, чем через год после. Это позволяет утверждать, что исследование фактора политической мобилизации может рассматриваться как внешнее событие, влияющее на идентификацию индивида в рамках его общей идентичности.

Как установить связь аффективной поляризации и протестной мобилизации: роль онлайн-коммуникации

С точки зрения теории социальной идентичности аффективная политическая поляризация может выражаться в прямых отсылках к групповой (политической) принадлежности оппонента, упомянутых во враждебном ключе, а не конкретно к оппоненту как к личности [Olteanu et al., 2018], т.е. с помощью языка ненависти (*hate speech*). Язык ненависти в этом случае понимается как негативно окрашенное эмоциональное высказывание, проявленное к политическим оппонентам по причине их принадлежности к аутгруппе.

Язык ненависти сегодня все чаще изучается как явление, проявляющееся в онлайн-среде [Castaño-Pulgarín et al., 2021], хотя мы признаем, что это не единственная форма его проявления.

Считается, что язык ненависти является самостоятельным индикатором аффективной политической поляризации при усло-

¹ Айлан Курди, мальчик трех лет сирийского происхождения, трагически погиб вместе с несколькими родственниками 2 сентября 2015 г. при попытке семьи пересечь Средиземное море ввиду развития проблемы европейского миграционного кризиса. Фотографии тела Айлана, вынесенного на побережье Турции вблизи города Бодрум, были опубликованы во многих СМИ и привлекли большое внимание социальных медиа, стали причиной митингов и петиций.

вии, что оба концепта используются в рамках теории социальной идентичности [Harel, Jameson, Maoz, 2020]. Наиболее важно то, что результаты дискуссии между политическими оппонентами на платформах социальных медиа – самостоятельный эмпирический материал, который можно собирать и анализировать с помощью современных методов машинного обучения [Cantini et al., 2022].

Таким образом, политическая онлайн-коммуникация – явление, которое отдельно не изучается в работе, а лишь позволяет увидеть и зафиксировать связь между аффективной поляризацией и протестной мобилизацией. Такой подход к оценке поляризации является основным ограничением настоящего исследования, у которого тем не менее есть предшественники, использующие похожий тип эмпирического анализа, в том числе сбор данных [Gomez et al., 2020; Smetanin, 2020; Pronoza et al., 2021; Jahan, Oussalah, 2023].

Выводы отдельных исследований подтверждают, что онлайн-сообщения между оппонентами, в которых используется язык ненависти, могут считаться индикаторами аффективной поляризации [Del Vicario et al., 2016]. В некоторых работах делается предположение о том, что язык ненависти в онлайн-коммуникации может быть причиной усиления расколов и размежеваний в обществе, в отдельных случаях – проявления насилия [Siegel, 2009].

В этой связи эмпирическая часть исследования базируется на предположении о том, что аффективную поляризацию можно операционализировать через факт наличия сообщения в социальной сети, содержащей язык ненависти [Кручинская, 2023].

Выше мы рассмотрели некоторые инструментальные характеристики того, почему принятый подход к операционализации может быть обоснован. Тем не менее необходимо пояснить, почему такой подход обоснован содержательно.

Сегодня активно изучается связь протестной мобилизации и онлайн-коммуникации [Филиппов, 2020; Petrov et al., 2021; Ахременко, 2024] или эффекты выражения аффективной поляризации в онлайн-коммуникации [Bakshy, Messing, Adamic, 2015; Lelkes, Westwood, 2017]. Существует зафиксированная эмпирическая зависимость между аффективной поляризацией и распространением онлайн-коммуникаций, что в последнее десятилетие находится в авангарде исследовательской дискуссии в социальных науках [Bakshy, Messing, Adamic, 2015; Törnberg et al., 2021; Tucker et al., 2018].

Наряду с этим существует ряд исследований, подчеркивающих, что протестная мобилизация значимо сказывается на распро-

странении сообщений в онлайн-коммуникации – это подтверждено как для русскоязычных социальных медиа [Филиппов, 2020; Ахременко, 2024], так и для зарубежных [Breuer, Landman, Farquhar, 2015; Clark, Bland, Livingston, 2017; Estrada, Juarez, Pina-Garcia, 2022].

Наконец, эмпирически доказано, что онлайн-коммуникация является фактором связи протеста и аффективной поляризации [Del Vicario et al., 2016; O'Reilly et al., 2023], т.е. онлайн-контекст способствует росту аффективной поляризации. В одном из исследований показано, что онлайн-пространство является средой распространения этнически направленной аффективной поляризации [Prnopa, 2021]. Также утверждается, что онлайн-среда является благоприятной для распространения эмоционально окрашенных сообщений, которые могут быть индикаторами аффективной поляризации [Smetanin, 2020].

Вышеописанные аргументы являются содержательным фактором того, что в исследовании принят обоснованный ранее эмпирический дизайн и подход к операционализации.

Цифровые следы языка ненависти и подходы к их обработке

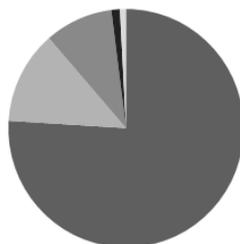
Для анализа связи аффективной поляризации и протестной мобилизации автором были собраны данные через *get* запросы с использованием *API VK* (метод *newsfeed.search*) по ключевым словам: «протест», «навальнист», «прокремлевский», «пропагандист», «диктатура». Эти слова можно считать как индикаторами политической онлайн-коммуникации [Kirkizh et al., 2024], так и эмоционально-окрашенными словами [Aken et al., 2018].

Период сбора сообщений: 14.07.2019 г. – 14.09.2019 г., (экспериментальный период, московские летние протесты 2019 г. во время выборов в Мосгордуму), 14.03.2019 г. – 14.05.2019 (период, предшествующий протестам). Собирались все сообщения, которые содержат упоминания указанных слов в заданные периоды.

Обратимся к некоторым визуальным описательным статистикам.

Диаграммы (рис. 1–2) показывают, что в экспериментальном (протестном) и контрольном периодах наблюдается одинаковая структура частоты использования ключевых слов, но в период мобилизации интенсивность их использования значительно возрастает.

Общее количество сообщений с использованием ключевых слов
(за контрольный и экспериментальный период)



Протест ■ Диктатура ■ Пропагандист ■ Навальнист ■ Прокремлевский

Рис 1.

**Распределение слов сообщений с использованием
ключевых слов (за контрольный и экспериментальный период)**

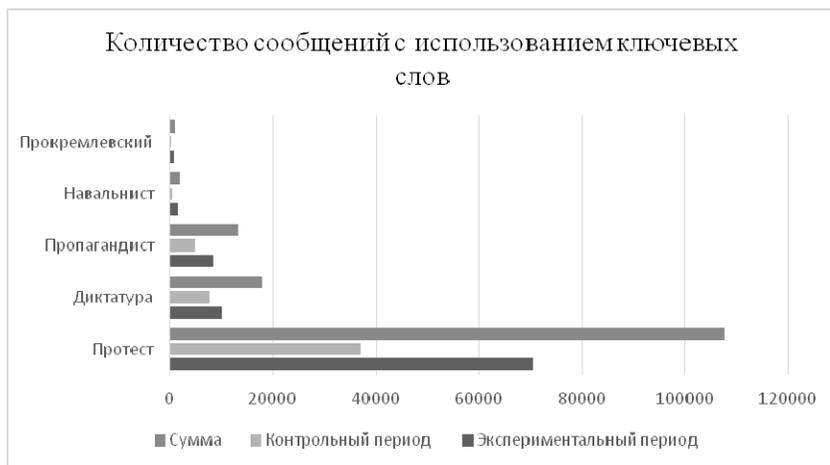


Рис. 2.

**Распределение слов сообщений с использованием
ключевых слов (за контрольный и экспериментальный период)**

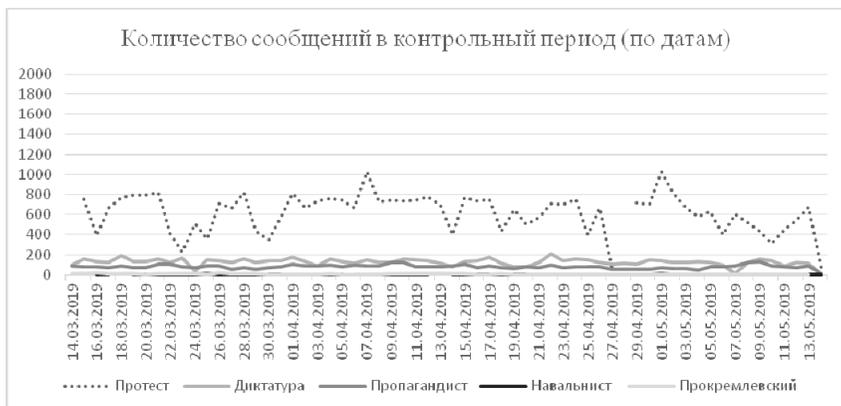


Рис. 3.
Динамика сообщений с использованием ключевых слов
(за контрольный период)



Рис. 4.
Динамика сообщений с использованием ключевых слов
(за экспериментальный период)

Из рис. 3–4 видно, что наиболее популярным словом как в экспериментальном, так и в контрольном периоде является «протест». Это может косвенно означать, что такое политическое событие в целом (как в период протеста, так и в период отсутствия мобилизации) является наиболее популярной причиной дискуссии

в социальных сетях (при прочих равных, т.е. в сравнении с другими используемыми словами).

Сбор данных по ключевым словам – не финальный этап работы с эмпирическим материалом, поскольку обязательным для анализа условием является факт наличия языка ненависти в сообщении.

Зарубежные исследователи уже научились применять подходы по автоматизированному поиску языка ненависти в социальных сетях [Schmidt, Wiegand, 2017; Tucker et al., 2018; Jahan, Oussalah, 2023; Gandhi et al., 2023]. Для русскоязычных текстов и сообщений такой подход также активно разрабатывается [Стукал, Беленков, Филиппов, 2023; Платонов, Руденко, 2022]. Сбор данных обычно ведется либо через поиск по ключевым словам [Филиппов, 2020], как и в этой работе, либо через сбор всех опубликованных сообщений по конкретным сообществам и группам [Курганская, Степанова, 2024].

Как в этом случае зафиксировать наличие языка ненависти в сообщении? Подобная задача требует лингвистического анализа, но современные методы работы с текстовыми массивами предлагают использование автоматической разметки текстов с помощью методов машинного обучения [Husunbeyi, Akar, Ozgur, 2022; Gandhi et al., 2023; Kirkizh et al., 2024] или больших языковых моделей [Smetanin, 2020].

С учетом этого опыта в настоящем исследовании для разметки текстов использовались два классификатора, созданные (добученные) на основе больших моделей: *ruBERT* и *Russian Toxicity Classifier*, доступные в открытом репозитории *HuggingFace*¹. Дубли сообщений и спам исключались с помощью автоматических методов.

Таблица 1

Результаты автоматической разметки

Период	<i>ruBERT</i>			<i>Toxicity Classifier</i>	
	<i>Negative</i>	<i>Neutral</i>	<i>Positive</i>	<i>Toxic</i>	<i>Neutral</i>
Контрольный	4813	1454	61	3758	2566
Экспериментальный	9406	2711	97	7455	4757

Источник: составлено автором.

¹ Hugging Face. – Mode of access: <https://huggingface.co/> (accessed: 30.08.2024).

Из табл. 1 видно, что собранный массив текстов в основном содержит негативные и токсичные сообщения, но для контрольного периода количество нейтральных сообщений больше. Кроме того, разные классификаторы показывают, что токсичные сообщения, скорее, всегда негативные, но негативные сообщения не всегда токсичные.

Тем не менее предположение о том, что эта разметка с безошибочной точностью определяет, есть ли язык ненависти, не имеет устойчивых оснований. Сегодня ни один классификатор пока не может установить, есть ли в сообщении выражение чувства ненависти к аутгруппе, т.е. имеется ли объект ненависти, поскольку язык ненависти – более сложное явление, хоть и связанное с токсичностью и негативом. Для выявления этого фактора данные должны быть размечены вручную независимым кодировщиком, что и было предпринято в настоящей работе. Более подробное описание дизайна см. в следующем разделе.

Составляющие языка ненависти и идентификации: как на них влияет политическая мобилизация и поддается ли это прогнозированию

Есть ли у языка ненависти конкретные лингвистические проявления? В литературе предлагается подход, основанный на выявлении его трех составляющих: оскорбления, призыв к насилию и дискриминация [Стукал, Ахременко, Петров, 2022; Berntzen, Kelsall, Hartevelde, 2024]. Считается, что проявление хотя бы одного из этих компонентов в онлайн-коммуникации может свидетельствовать об аффективной поляризации. Это допущение было использовано для инструкций независимым разметчикам.

Три составляющие языка ненависти не единственные показатели для разметки. В связи с теоретическими предпосылками, отраженными в первой части статьи, считается, что политическая мобилизация может являться фактором для усиления политической идентичности индивида по сравнению с социальной. В этой связи кодировщики в рамках разметки фиксировали и тему сообщения: межнациональный / гражданский конфликт, проявление оппозиционного настроения, проявление патриотического настроения, этнический вопрос. Предположение (допущение) настоящей работы состоит в том, что первые из трех тем могут

свидетельствовать о проявлении политической идентичности [Nezi, 2024], а последняя – о социальной идентичности.

Из общего сета для ручной разметки случайным образом были выбраны 9 тыс. сообщений, размеченных как *toxic, negative* и 1 тыс. нейтральных сообщений с сохранением баланса изначальной выборки (эксперимент / контроль). Работу выполняли 10 независимых кодировщиков, которые не видели критериев автоматической разметки и не имели информации о периоде публикации сообщения (контрольный или экспериментальный). Таким образом, разметчикам был доступен только сам текст сообщения.

Ручная разметка была выполнена по следующим критериям.

1. **Offtopic_spam (1 / 0)**¹: Является ли комментарий спамом или нерелевантным для указанной темы? 1 – да, 0 – нет.

2. **Violence (1 / 0)**: Выражает ли комментарий призыв к насилию или угрозу применения насилия в отношении человека или группы лиц из-за их принадлежности к той или иной группе? 1 – да, 0 – нет.

3. **Insult (1 / 0)**: Содержится ли в комментарии оскорбление в отношении человека или группы лиц из-за их принадлежности к той или иной группе (аутгруппе)? 1 – да, 0 – нет.

4. **Discrimination (1 / 0)**: Содержится ли в комментарии призыв к дискриминации или угроза дискриминации человека или группы лиц из-за их принадлежности к той или иной группе? 1 – да, 0 – нет

5. Кто / что является темой сообщения², выберите из опций:

1. **Theme_civil (1 / 0)**: Посвящено ли сообщение теме внешне-политического или гражданского конфликта, где 1 – да, 0 – нет;

2. **Theme_opposition (1 / 0)**: Посвящено ли сообщение оппозиционной теме, где 1 – да, 0 – нет;

3. **Theme_patriotic (1 / 0)**: Посвящено ли сообщение патриотической теме, где 1 – да, 0 – нет;

4. **Theme_ethnic (1 / 0)**: Посвящено ли сообщение теме этнических разногласий, где 1 – да, 0 – нет.

¹ Если да, то сообщение исключалось из анализа.

² Главный критерий выбора той или иной темы – упоминание ключевых слов, явно свидетельствующих о теме. Для исключения смещений конкретный перечень слов кодировщикам не приводился. Кодировщик мог выбрать неограниченное количество тем.

Результаты разметки показали выдвинутые исследовательские предположения: действительно, протестный период способствует интенсификации проявлений аффективной поляризации. Также мобилизация является значимым фактором для того, чтобы фокус идентичности смещался на политический – в сравнении с социальным.

Поскольку собранные данные представляют собой временные ряды, для оценки значимости фактора политической мобилизации предлагается воспользоваться стандартным методом работы с временными рядами – кросс-корреляционной функцией.

Для каждого периода – контрольного и экспериментального – предлагается оценить взаимную корреляцию каждого из составляющих языка ненависти. Знак корреляции означает, как во времени связаны между собой проявления двух составляющих языка ненависти, т.е. сигналы. Считается, что если коэффициент корреляции выше 0,4, то два независимых сигнала похожи друг на друга.

Таблица 2

**Кросс-корреляция разных проявлений языка ненависти
для нулевого лага при первых разностях**

<i>Types of hatespeech</i>	<i>Crosscorr (Violence&Insult)</i>		<i>Crosscorr (Insult&Discrimination)</i>		<i>Crosscorr (Discrimination& Violence)</i>	
	<i>Experimental</i>	<i>Control</i>	<i>Experimental</i>	<i>Control</i>	<i>Experimental</i>	<i>Control</i>
<i>Lag0</i>	0,40	0,30	0,46	0,32	0,42	0,09

Таблица 2 показывает, что коэффициент кросс-корреляции значительно выше для всех проявлений языка ненависти именно в протестный период. Самый значимый отрыв у кросс-корреляции проявления призывов к насилию и дискриминации, самый незначительный у проявления призывов к насилию и оскорблений. Тем не менее эти различия имеют среднюю значимость.

Таким образом, мы подтвердили первое исследовательское положение: протест сам по себе не является причиной появления языка ненависти, который есть и в непротестные периоды, но является причиной для значимой интенсификации проявления аффективной поляризации.

Как различается структура идентичности и связанные с этим проявления идентификации?

Таблица 3 показывает, что кросс-корреляция суммарного показателя различных проявлений языка ненависти (гнев, призыв

к насилию, дискриминация) с темами сообщений в протестном периоде всегда имеет более высокое значение для протестного периода, но только за исключением этнической темы. Таким образом, темы сообщений, которые свидетельствуют о проявлении политической идентичности (*Civil, Opposition, Patriotic*), более значимы в период протестной мобилизации, а тема сообщений, свидетельствующая о социальной идентичности (*Ethnic*), сильнее проявляется в контрольном периоде: подтверждение нашло и второе исследовательское предположение.

Таблица 3

Кросс-корреляция суммы проявлений языка ненависти к темам сообщений для нулевого лага при первых разностях

Period	<i>Hatespeech& Civil</i>		<i>Hatespeech& Opposition</i>		<i>Hatespeech& Patriotic</i>		<i>Hatespeech& Ethnic</i>	
	<i>Experimental</i>	<i>Control</i>	<i>Experimental</i>	<i>Control</i>	<i>Experimental</i>	<i>Control</i>	<i>Experimental</i>	<i>Control</i>
<i>Lag0</i>	0,64	0,42	0,84	0,69	0,80	0,74	0,37	0,50

Для дополнительного подтверждения первого исследовательского вывода прогнозируемость данных проверена на модели *SARIMAX*. Эта модель представляет собой подвид семейства моделей *ARIMA*, часто используемых для прогнозирования будущих значений зависимой переменной.

Поскольку качество прогноза такой модели обозначает общий уровень предсказуемости феномена, использование *SARIMAX* может считаться некоторой проверкой первого вывода работы на робастность.

Были построены две модели *SARIMAX* – для контрольного и экспериментального периодов. Стационарность временных рядов (для контрольного и экспериментального периода) проверена с помощью теста Дики – Фуллера и подтверждена его результатами.

Общий вид моделей описывается следующей спецификацией:

$$ARIMA(p, d, q)(P, D, Q)_s \quad (1),$$

где: p – порядок модели $AR(p)$; d – порядок интегрирования исходных данных; q – порядок модели $MA(q)$; P – порядок сезонной составляющей модели $SAR(P)$; D – порядок интегрирования сезонной

составляющей; Q – порядок сезонной составляющей $SMA(Q)$; s – размерность сезонности.

Параметры модели для контрольного периода:

$$ARIMA(1, 1, 1)(1, 1, 1, 12) \quad (2).$$

Параметры модели для экспериментального периода:

$$ARIMA(0, 1, 1)(1, 1, 1, 12) \quad (3).$$

Для удобства восприятия оценим точность прогноза на графиках.



Рис. 5.
Результаты прогнозной модели *SARIMAX*
для контрольного периода

Рисунок 5 показывает достаточно низкую предсказательную способность модели для контрольного периода: пиковые точки не совпадают, схожесть паттерна зафиксирована лишь частично.

Рисунок 6 показывает, что для экспериментального периода мы имеем значительно более четкие результаты: пиковые точки частично совпадают, зафиксирована схожесть паттерна.



Рис. 6.
Результаты прогнозной модели *SARIMAX*
для экспериментального периода

Эти выводы могут быть подтверждены не только визуально, но и статистически.

Таблица 4

Model 1 – модель *SARIMAX* для контрольного периода

<i>Dependent variable</i>	
<i>Hatespeech</i>	
<i>ma.L1</i>	-0,9
<i>Observations</i>	62

Таблица 5

Model 2 – модель *SARIMAX* для экспериментального периода

<i>Dependent variable</i>	
<i>Hatespeech</i>	
<i>ma.L1</i>	-0,6***
<i>Observations</i>	63

Из таблиц 4–5 видно, что в экспериментальном периоде коэффициент при предикторе модели значим, тогда как в контрольном – нет. У данных значений нет содержательной интерпретации кроме той, что значимость коэффициента подтверждает прогнозную способность модели.

Таким образом, при общей сопоставимости временных рядов по длине результаты прогноза оказываются различными. Публикации, сделанные в период политической мобилизации, легче поддаются прогнозированию. Это может означать, что протест является значимым фактором для распространения языка ненависти в социальных сетях. Это еще раз подтверждает один из основных выводов исследования – протест не является причиной аффективной поляризации, но выступает ее катализатором в социальных медиа.

Дискуссия и выводы

Настоящее исследование, направленное на изучение связи аффективной поляризации и политической мобилизации, было проведено с помощью эмпирического анализа онлайн-коммуникации в социальной сети «ВКонтакте». Были проверены и подтверждены два предположения: протест является фактором усиления проявления поляризации в обществе, но не является причиной проявления поляризации; протест значимо влияет на расколы и идентификацию: в период протестной мобилизации большую значимость приобретает выражение политической идентификации, нежели социальной. Оба предположения подтвердились.

Основное ограничение исследования состоит в том, что аффективная поляризация изучается в онлайн-контексте. С одной стороны, такой подход позволяет избежать рисков социальной желательности, которые связаны с использованием социологического инструментария при изучении аффективной поляризации. С другой стороны, это повышает риск субъективности выводов, основанных только на одном из проявлений основного понятия работы (в онлайн-контексте).

Эмпирический анализ на основе кросс-корреляционного анализа размеченных данных показал, что протест является значимым фактором для усиления, но не появления языка ненависти (аффективной поляризации). Прогнозные модели, построенные на данных контрольного и экспериментального периодов, также показали, что фактор протеста делает данные более подходящими для прогноза, а значит – более однородными и предсказуемыми.

Также удалось подтвердить значимость различий, связанных с конкретными типами идентификации: социальной и политической. Так, для протестного периода большее значение приобретает идентификация, связанная с гражданственностью, патриотизмом

или оппозиционным настроем, то есть в настоящем дизайне – политическая. Для непротестного периода большую статистическую значимость имеет социальная идентификация, связанная с национальным признаком и этническими разногласиями на его почве.

Результаты настоящего исследования могут стать заделом как для изучения паттернов проявления аффективной поляризации в русскоязычном пространстве, так и для поиска новых решений по прогнозированию распространения языка ненависти – возможно, в связи с конкретными событиями, проходящими в рамках протестных движений и с учетом характеристик этих событий.

E.V. Kruchinskaia*

Hate speech as an indicator of affective political polarization during mobilization: from measurement to forecasting

Abstract. The article examines the phenomenon of affective political polarization, which is operationalized through the fact of detected hate speech. In this study we have made two assumptions. Firstly, it is believed that during political mobilization, affective polarization (hate speech) will be significantly stronger than during the non-protest period. Although, hate speech will be detected for both periods. Secondly, we believe protest significantly affects cleavages. During the period of protest mobilization, political cleavages become more important than social ones. To verify the assumptions, we examined data collected on the VK social network in two periods: July–September 2019 (Moscow protests) and March–May 2019 (without protests). The results of the analysis, based on the automatic mapping (ruBERT) and on the human mapping of independent encoders, confirm both assumptions. We prove both assumptions via cross correlation analysis. Moreover, present approach to data collection and mapping allows us to create predictive models (SARIMAX) to predict hate speech in a social network.

Keywords: affective political polarization; hate speech; online political communication; cleavages; ruBERT; prediction of hate speech.

For citation: Kruchinskaia E.V. Hate speech as an indicator of affective political polarization during mobilization: from measurement to forecasting. *Political science (RU)*. 2025, N 1, P. 156–180. DOI: <http://www.doi.org/10.31249/poln/2025.01.07>

* **Kruchinskaia Ekaterina**, HSE University (Moscow, Russia), e-mail: ekruchinskaya@hse.ru

References

- Agadullina E.R., Gulevich O.A. “We” and “they”: assistance in intergroup relations. *Psychological research: electronic scientific journal*. 2013, Vol. 6, N 28. DOI: <https://doi.org/10.54359/ps.v6i28.703> (In Russ.)
- Aken B. van, Risch J., Krestel R., Löser A. Challenges for Toxic Comment Classification: An In-Depth Error Analysis. In: Fišer D., Huang R., Prabhakaran V., Voigt R., Waseem Z., Wernimont J. (eds). *Proceedings of the 2nd Workshop on Abusive Language Online (ALW2)*. Brussels, Belgium: Association for computational linguistics, 2018, P. 33–42. DOI: <https://doi.org/10.18653/v1/W18-5105>
- Aleskerov F.T., Golubenko M.A. On the assessment of the symmetry of political views and the polarization of society. *Working paper WP7/2003/04 series “Mathematical methods of decision analysis in economics, business and politics”*. Moscow: Higher School of Economics, 2003, 24 p. (In Russ.)
- Akhremenko A.S. Internet shutdown as a theoretical problem of political science, or what we (don't) understand about online protest mobilization. *Polis. Political research*. 2024, N 2, P. 118–134. DOI: <https://doi.org/10.17976/jpps/2024.02.09> (In Russ.)
- Bakshy E., Messing S., Adamic L.A. Exposure to ideologically diverse news and opinion on Facebook. *Science*. 2015, N 348, P.1130–1132.
- Bauer M., Christopher B., Chytilova J., Joseph H., Miguel E., Mitts T. Can war foster cooperation? *Journal of economic perspectives*. 2016, N 30 (3), P. 249–274.
- Berntzen L.E., Kelsall H., Harteveld E. Consequences of affective polarization: Avoidance, intolerance and support for violence in the United Kingdom and Norway. *European journal of political research*. 2024, N 63, P. 927–949. DOI: <https://doi.org/10.1111/1475-6765.12623>
- Bodrunova S., Blekanov I., Smoliarova A., Litvinenko A. Beyond left and right: real-world political polarization in Twitter discussions on inter-ethnic conflicts. *Media and communication*. 2019, N 7 (3), P. 119–132. DOI: <https://doi.org/10.17645/mac.v7i3.1934>
- Breuer A., Landman T., Farquhar D. Social media and protest mobilization: Evidence from the Tunisian revolution. *Democratization*. 2015, N 22 (4), P. 764–792. DOI: <https://doi.org/10.1080/13510347.2014.885505>
- Campbell A., Converse P., Miller W., Stokes D. *The American voter*. New York: John Wiley & Sons, 1960, 576 p.
- Cantini R., Marozzo F., Talia D., Trunfio P. Analyzing political polarization on social media by deleting bot spamming. *Big data and cognitive computing*. 2022, N 6 (1), P. 1–16. DOI: <http://doi.org/10.3390/bdcc6010003>
- Castaño-Pulgarín S.A., Suárez-Betancur N., Vega L.M., López H.M. Internet, social media and online hate speech. Systematic review. *Aggression and violent behavior*. 2021, N 58 (6), 101608. DOI: <http://doi.org/10.1016/j.avb.2021.101608>
- Clark M.D., Bland D., Livingston J.A. Lessons from# McKinney: Social media and the interactive construction of police brutality. *The Journal of social media in society*. 2017, N 6 (1), P. 284–313.
- Collins R. C-Escalation and D-Escalation: A theory of the time-dynamics of conflict. *American sociological review*. 2012, N 77 (1), P. 1–20. DOI: <https://doi.org/10.1177/0003122411428221>
- Collins R. *Violence: A micro-sociological theory*. Princeton: Princeton University Press, 2008, 584 p. DOI: <http://doi.org/10.1515/9781400831753>

- Coser L. *Functions of social conflict*. New York: Free Press, 1956, 188 p.
- Costa M. Ideology, not affect: what Americans want from political representation. *American journal of political science*. 2020, Vol. 65, N 2, P. 342–358 DOI: <https://doi.org/10.1111/ajps.12571>
- Del Vicario M., Vivaldo G., Bessi A., Zollo F., Scala A., Caldarelli G., Quattrociocchi W. Echo chambers: emotional contagion and group polarization on Facebook. *Scientific reports*. 2016, N 6, P. 1–14. DOI: <https://doi.org/10.1038/srep37825>
- DiMaggio P., Evans J., Bryson B. Have American's social attitudes become more polarized? *American journal of sociology*. 1996, N 102 (3), P. 690–755.
- Druckman J.N., Green D.P., Iyengar S. Does affective polarization contribute to democratic backsliding in America? *The ANNALS of the American Academy of political and social science*. 2023, Vol. 708, N 1, P. 137–163. DOI: <https://doi.org/10.1177/00027162241228952>
- Druckman J.N., Gubitza S.R., Levendusky M.S., Lloyd A.M. How incivility on partisan media (De)Polarizes the electorate. *The Journal of politics*. 2019, N 81 (1), P. 291–295. DOI: <https://doi.org/10.1086/699912>
- Esteban J.-M., Ray D. On the measurement of polarization. *Econometrica*. 1994, N 62 (4), P. 819–851. DOI: <https://doi.org/10.2307/2951734>
- Estrada M.S., Juarez Y., Pina-Garcia C.A. Toxic social media: affective polarization after feminist protests. *Social media and society*. 2022, Vol. 8, N 2, P. 1–12. DOI: <https://doi.org/10.1177/20563051221098343>
- Gandhi A., Ahir P., Adhvaryu K., Shah P., Lohiya R., Cambria E., Poria S., Hussain A. Hate speech detection: a comprehensive review of recent works. *Expert systems*. 2023, Vol. 41, N 8, P. 1–24. DOI: <https://doi.org/10.1111/exsy.13562>
- Gerber A., Green D. Misperceptions about perceptual bias. *Annual review of political science*. 1999, N 2 (1), P. 189–210. DOI: <https://doi.org/10.1146/annurev.polisci.2.1.189>
- Gomez R., Gibert J., Gomez L., Karatzas D. Exploring hate speech detection in multimodal publications. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2020, P. 1470–1478.
- Graham M.H., Svoboda M.W. Democracy in America? Partisanship, polarization, and the robustness of support for democracy in the United States. *American political science review*. 2020, N 114 (2), P. 392–409. DOI: <https://doi.org/10.1017/S0003055420000052>
- Green D., Palmquist B., Schickler E. *Partisan hearts and minds: political parties and social identities of voters*. Yale: Yale University Press, 2002, 266 p.
- Greene S. Understanding party identification: a social identity approach. *Political psychology*. 1999, N 20, P. 393–403. DOI: <https://doi.org/10.1111/0162-895X.00150>
- Harel T.O., Jameson J.K., Maoz I. The normalization of hatred: identity, affective polarization, and dehumanization on Facebook in the context of intractable political conflict. *Social media and society*. 2020, N 6 (2), P. 1–10. <https://doi.org/10.1177/2056305120913983>
- Harel T.O., Nir N., Vandermeulen D., Maoz I., Halperin E. A threat to cohesion: intragroup affective polarization in the context of intractable intergroup conflict. *Journal of conflict resolution*. 2024, Vol. 0 (0), P. 1–27. <https://doi.org/10.1177/00220027241247033>
- Husunbeyi Z.M., Akar D., Ozgur A. Identifying hate speech using neural networks and discourse analysis techniques. In: Adebayo K. (ed.). *Proceedings of the first workshop on language technology and resources for a fair, inclusive, and safe society within the 13-th language resources and evaluation conference*. Paris: European language resources associate, 2022, P. 32–41.

- Isenberg D.J. Group polarization: A critical review and meta-analysis. *Journal of personality and social psychology*. 1986, N 50, P. 1141–1151. DOI: <https://doi.org/10.1037/0022-3514.50.6.1141>
- Iyengar S., Sood G., Lelkes Y. Affect, not ideology: a social identity perspective on polarization. *Public opinion quarterly*. 2012, Vol. 76, N 3, P. 405–431. DOI: <https://doi.org/10.1093/poq/nfs038>
- Iyengar S., Lelkes Y., Levendusky M., Malhotra N., Westwood S.J. The origins and consequences of affective polarization in the United States. *Annual review of political science*. 2019, N 22 (1), P. 129–146. DOI: <https://doi.org/10.1146/annurev-polisci-051117-073034>
- Jahan M.S., Oussalah M. A systematic review of hate speech automatic detection using natural language processing. *Neurocomputing*. 2023, Vol. 546, P. 1–30. DOI: <https://doi.org/10.1016/j.neucom.2023.126232>
- Johnston R., Manley D., Jones K. Spatial polarization of presidential voting in the United States, 1992–2012: The “Big Sort” revisited. *Annals of the American association of geographers*. 2016, N 106 (5), P. 1047–1062. DOI: <https://doi.org/10.1080/24694452.2016.1191991>
- Kirkizh N., Ulloa R., Stier S., Pfeffer J. Predicting political attitudes from web tracking data: a machine learning approach. *Journal of information technology and politics*. 2024, N 21 (4), P. 564–577. DOI: <https://doi.org/10.1080/19331681.2024.2316679>
- Korgunyuk Y.G. The structure of electoral divisions in the 2011–2012 election cycle and possible scenarios for the development of the situation. *Politiya*. 2012, N 3, P. 84–99. DOI: <https://doi.org/10.30570/2078-5089-2012-66-3-84-99> (In Russ.)
- Kruchinskaya E.V. Affective political polarization: proposals for the operationalization of the concept through the assessment of the language of hostility in social media. *Information society*. 2023, N 3, P. 97–107. (In Russ.)
- Kurganskaia E., Stepanova N. Identification of “toxicity” in social networks based on the semantic proximity criterion. *Philology. Theory and practice*. 2024, N 17, P. 1596–1603. DOI: [10.30853/phil20240231](https://doi.org/10.30853/phil20240231) (In Russ.)
- Lelkes Y., Westwood S.J. The Limits of partisan prejudice. *Journal of politics*. 2017, N 79 (2), P. 485–501. DOI: <https://doi.org/10.1086/688223>
- Malinova O.Y. Changing institutional conditions of production and competition of political ideas in Russia: from the 1990s to the 2010s. In Yarskaya-Smirnova E.R., Romanov P.V. (eds.). *Public sphere: theory, methodology, case study*. Moscow: OOO “Variant”: TSPGI, 2013, P. 55–74. (In Russ.)
- Mason L., Notes A.A. Cross-cutting calm: how social sorting drives affective polarization. *Public opinion quarterly*. 2016, Vol. 80, N S1, P. 351–377. DOI: <https://doi.org/10.1093/poq/nfw001>
- Mason L., Wronski J. One tribe to bind them all: how our social group attachments strengthen partisanship. *Political psychology*. 2018, Vol. 39, N S1, P. 257–277. DOI: <https://doi.org/10.1111/pops.12485>
- Moscovici S., Zavalloni M. The group as a polarizer of attitudes. *Journal of personality and social psychology*. 1969, N 12 (2), P. 125–135.
- Nezi P. A tale of two crises: affective polarization in Greece. *Frontiers in political science*. 2024, N 6, Article 1256199. DOI: <https://doi.org/10.3389/fpos.2024.1256199>
- O’Reilly C., Maher P.J., Smith E.M., MacCarron P., Quayle M. Social identity emergence in attitude interactions and the identity strengthening effects of cumulative

- attitude agreement. *European journal of social psychology*. 2023, N 54, P. 97–117. DOI: <https://doi.org/10.1002/ejsp.3000>
- Olteanu A., Castillo C., Boy J., Varshney K. The Effect of Extremist Violence on Hateful Speech Online. In *Proceedings of the Twelfth International AAAI conference on web and social media (ICWSM 2018)*. 2018, Vol. 12, N 1, P. 221–230. DOI: <https://doi.org/10.1609/icwsm.v12i1.15040>
- Orr L.V., Flower A., Huber G.A. Is affective polarization driven by identity, loyalty, or substance? *American journal of political science*. 2023, Vol. 67, N 4, P. 948–962. DOI: <https://doi.org/10.1111/ajps.12796>
- Petrov A., Akhremenko A.S., Zheglov S.A., Kruchinskaia E.V. Is network structure important for protest mobilization? Findings from agent-based modeling. *Monitoring obshchestvennogo mneniya: ekonomicheskie i sotsial'nye peremeny*. 2021, N 6, P. 226–253. DOI: <https://doi.org/10.14515/monitoring.2021.6.2021> (In Russ.)
- Philippov I.B. Quantitative textual characteristics as a factor of success in promoting publications within the framework of protest communication on social networks. *Outlines of global transformations: politics, economics, law*. 2020, Vol. 13, N 2, P. 249–268. DOI: <https://doi.org/10.23932/2542-0240-2020-13-2-13> (In Russ.)
- Philippov I.B. The “Moscow case” as a factor of protest communication activity in the social network “VK”. *RUDN Journal of political science*. 2022, Vol. 24, N 3, P. 530–544. DOI: <https://doi.org/10.22363/2313-1438-2022-24-3-530-544> (In Russ.)
- Platonov E.N., Rudenko V.Yu. Identification and classification of toxic statements using machine learning methods. *Modeling and data analysis*. 2022, N 12 (1), P. 27–48. DOI: <https://doi.org/10.17759/mda.2022120103> (In Russ.)
- Pronoza E., Panicheva P., Koltsova O., Rosso P. Detecting ethnicity-targeted hate speech in Russian social media texts. *Information processing and management*. 2021, N 58 (6), Article 102674.
- Rosenberg M., Schooler C., Schoenbach C., Rosenberg F. Global self-esteem and specific self-esteem: different concepts, different outcomes. *American sociological review*. 1995, Vol. 60, N 1, P. 141–156. DOI: <https://doi.org/10.2307/2096350>
- Rosler N. Not as simple as that: how leaders faced the challenges of pursuing the peace process in Northern Ireland. *Peace and conflict: Journal of peace psychology*. 2016, N 22 (2), P. 177–180. DOI: <https://doi.org/10.1037/pac0000122>
- Sartori G. *Parties and party systems: a framework for analysis*. Cambridge: Cambridge University Press, 1976, Vol. 1, 390 p.
- Schedler A. Rethinking political polarization. *Political science quarterly*. 2023, N 138, P. 335–359. DOI: <https://doi.org/10.1093/psquar/qqad038>
- Schmidt A., Wiegand M. A survey on hate speech detection using natural language processing. In: Ku L.-W., Li Ch.-T. (eds). *Proceedings of the fifth international workshop on natural language processing for social media*. Valencia: Association for computational linguistics, 2017, P. 1–10. DOI: <https://doi.org/10.18653/v1/W17-1101>
- Siegel D.A. Social networks and collective action. *American journal of political science*. 2009, N 53 (1), P. 122–138.
- Simmel G. *Conflict and the web of group affiliations*. New York: The Free Press, 1956, 208 p.
- Smetanin S. Toxic comments detection in Russian. In: Selegey V. (ed.). *Computational linguistics and intellectual technologies: proceedings of the international conference “Dialogue 2020”*. Moscow: Publishing House of RSSU, 2020, P. 1149–1159.

- Smith L.G.E., Thomas E.F., Bliuc AM. Polarization is the psychological foundation of collective engagement. *Communications psychology*. 2024, Vol. 2, N 1, Article 41. DOI: <https://doi.org/10.1038/s44271-024-00089-2>
- Stukal D.K., Akhremenko A.S., Petrov A.P. Affective political polarization and hate speech: are they made for each other? *RUDN Journal of political science*. 2022, Vol. 24, N 3, P. 480–498. (In Russ.)
- Stukal D.K., Belenkov V.E., Philippov I.B. Methods of data science in political research: analysis of protest activity in social networks. *Political Science*. 2021, N 1, P. 46–75. DOI: <https://doi.org/10.31249/poln/2021.01.02> (In Russ.)
- Tajfel H., Turner J.C. An integrative theory of intergroup conflict. In Austin W.G., Worchel S. (eds.). *The social psychology of intergroup relations*, Monterey, CA: Brooks / Cole, 1979, P. 33–37.
- Theelen M., Bohm R. The conflict-cooperation effect persists under intragroup payoff asymmetry. *Group processes and intergroup relations*. 2021, N 24 (5), P. 815–835.
- Thomas E.F., McGarty C., Mavor K. Group interaction as the crucible of social identity formation: a glimpse at the foundations of social identities for collective action. *Group process and intergroup relations*. 2016, N 19, 137–151.
- Törnberg P., Andersson C., Lindgren K., Banisch S. Modeling the emergence of affective polarization in the social media society. *PLoS ONE*. 2021, N 16 (10), Article e0258259. DOI: <https://doi.org/10.1371/journal.pone.0258259>
- Tucker J.A., Guess A., Barberá P., Vaccari C., Siegel A., Sanovich S., Stukal D., Nyha B. Social media, political polarization, and political disinformation: a review of the scientific literature. *SSRN electronic journal*. 2018, P. 1–66. DOI: <https://doi.org/10.2139/ssrn.3144139>
- Vartanova E., Gladkov A., Dunas D. On representation of conflict in modern academic discourse. *Online media and global communication*. 2022, N 1 (4), P. 794–814. DOI: <https://doi.org/10.1515/omgc-2022-0063>
- Webster S., Abramowitz A. The Ideological Foundations of Affective Polarization in the US Electorate. *American Politics Research*. 2017, N 45, P. 621–647. DOI: <https://doi.org/10.1177/1532673X17703132>
- Zingher J.N., Flynn M.E. From on high: the effect of elite polarization on mass level attitudes and behaviors, 1972-2012. *British journal of political science*. 2018, Vol. 48, N 1, P. 23–45. DOI: <https://doi.org/10.1017/S0007123415000514>

Литература на русском языке

- Агадуллина Е.Р., Гулевич О.А. «Мы» и «они»: помощь в межгрупповых отношениях // Психологические исследования: электронный научный журнал. – 2013. – Т. 6, № 28. – С. 1–9.
- Алескеров Ф.Т., Голубенко М.А. Об оценке симметричности политических взглядов и поляризованности общества. Препринт WP7/2003/04 «Математические методы анализа решений в экономике, бизнесе и политике». – М.: ГУ ВШЭ. – 2003. – 24 с.
- Ахременко А.С. Отключение интернета как теоретическая проблема политической науки, или Что мы (не) понимаем в сетевой протестной мобилизации // Полис.

- Политические исследования. – 2024. – № 2. – С. 118–134. – DOI: <https://doi.org/10.17976/jpps/2024.02.09>
- Коргунок Ю.Г.* Структура электоральных размежеваний в избирательном цикле 2011–2012 годов и возможные сценарии развития ситуации // *Полития*. – 2012. – № 3. – С. 84–99. – DOI: <https://doi.org/10.30570/2078-5089-2012-66-3-84-99>
- Кручинская Е. В.* Аффективная политическая поляризация: предложения к операционализации понятия через оценку языка вражды в социальных медиа // *Информационное общество*. – 2023. – № 3. – С. 97–107.
- Курганская Е.В., Степанова Н.В.* Выявление «токсичности» в социальных сетях на основании критерия семантической близости // *Филологические науки. Вопросы теории и практики*. – 2024. – Т. 17, № 5. – С. 1597–1603. – DOI: <https://doi.org/10.30853/phil20240231>
- Малинова О.Ю.* Изменение институциональных условий производства и конкуренции политических идей в России: от 1990-х к 2010-м // *Публичная сфера: теория, методология, кейс стади* / под ред. Е.Р. Ярской-Смирновой, П.В. Романова. – Москва: ООО «Вариант»: ЦСПГИ, 2013. – С. 55–74.
- Платонов Е.Н., Руденко В.Ю.* Выявление и классификация токсичных высказываний методами машинного обучения // *Моделирование и анализ данных*. – 2022. – Т. 12, № 1. – С. 27–48. – DOI: <https://doi.org/10.17759/mda.2022120103>
- Стукал Д.К., Ахременко А.С., Петров А.П.* Аффективная политическая поляризация и язык ненависти: созданы друг для друга? // *Вестник Российского университета дружбы народов. Серия: Политология*. – 2022. – Т. 24, № 3. – С. 480–498.
- Стукал Д.К., Беленков В.Е., Филиппов И.Б.* Методы наук о данных в политических исследованиях: анализ протестной активности в социальных сетях // *Политическая наука*. – 2021. – № 1. – С. 46–75. – DOI: <https://doi.org/10.31249/poln/2021.01.02>
- Филиппов И.Б.* Количественные текстовые характеристики как фактор успешности продвижения публикаций в рамках протестной коммуникации в социальных сетях // *Контурь глобальных трансформаций: политика, экономика, право*. – 2020. – Т. 13, № 2. – С. 249–268. – DOI: <https://doi.org/10.23932/2542-0240-2020-13-2-13>