«Демобилизация» дагестанских национальностей: анализ дискурса «Дагестанской правды» методами обработки естественного языка

Рекомендация для цитирования:
Хамазаев Р. М. (2026) «Демобилизация» дагестанских национальностей: анализ дискурса «Дагестанской правды» методами обработки естественного языка. *Социология власти*, 38 (3): 166-201. EDN: JLFWCS

For citation:
Khamazaev R. M. (2026) The "Demobilization" of Dagestan Nationalities: A Discourse Analysis of Dagestanskaya Pravda Using Natural Language Processing. *Sociology of Power*, 38 (3): 166-201.

Поступила в редакцию: 05.05.2026; прошла рецензирование: 11.07.2026; принята в печать: 15.07.2026
Received: 05.05.2026; Revised: 11.07.2026; Accepted: 15.07.2026

© Author, 2026
This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Руслан М. Хамазаев

Национальный исследовательский университет «Высшая школа экономики», Москва, Российская Федерация

https://orcid.org/0009-0000-5713-6759

Резюме: В статье на материалах газеты «Дагестанская правда» исследуется дискурсивная мобилизация этнических категорий в региональной государственной прессе полиэтничной республики. Теоретической рамкой исследования выступает когнитивная теория этничности Р. Брубейкера, согласно которой этничность представляет собой не свойство групп, а ситуативно активируемую когнитивную схему, реализующуюся в конкретных коммуникативных практиках. Эмпирическую основу работы составляют два корпуса газеты, разделенные более чем семью десятилетиями: советский (1939–1945 гг., 1914 выпусков) и современный (2018–2024 гг., 1349 выпусков) — суммарным объемом около 26 миллионов токенов. Применяется сочетание дистрибутивно-семантического анализа на основе векторных представлений слов (Word2Vec), частотного анализа коллокаций, теста на ассоциативные смещения (WEAT) и качественного анализа извлеченных контекстов. Установлено, что в советском подкорпусе этнонимы дагестанских национальностей и обобщающая категория «дагестанец» функционировали как взаимозаменяемые мобилизационные единицы, активируемые в военных, трудовых и культурных нарративах. В постсоветском подкорпусе зафиксирована функциональная дифференциация категорий: этнонимы статистически значимо смещаются в сферу культурно-тради-

ционной репрезентации («демобилизация» национальностей), тогда как мобилизационная функция концентрируется на региональной категории «дагестанец», частотность которой возросла приблизительно в восемь раз. Полученные результаты эмпирически иллюстрируют тезис Брубейкера о переменной активации этничности и фиксируют дискурсивный коррелят институциональной рецентрализации Дагестана: культурализация этнонимов сопровождается конструированием регионально-гражданской идентичности как доминирующей рамки лояльности региональной и федеральной власти. Эмпирические данные вносят вклад в изучение медиатизации этничности в постсоветских контекстах и предлагают воспроизводимую методологию диахронического анализа для других регионально-этнических кейсов России и сопредельных государств.

Ключевые слова: этничность, Р. Брубейкер, конструктивизм, медиадискурс, Дагестан, диахронический анализ, векторные представления слов, WEAT, мобилизация, надэтническая идентичность

The "Demobilization" of Dagestan Nationalities: A Discourse Analysis of Dagestanskaya Pravda Using Natural Language Processing

Ruslan M. Khamazaev

National Research University «Higher School of Economics», Moscow, Russian Federation
https://orcid.org/0009-0000-5713-6759

Abstract: This article examines the discursive mobilization of ethnic categories in the regional state press of a multi-ethnic republic, drawing on materials from the newspaper *Dagestanskaya Pravda*. The theoretical framework is Rogers Brubaker's cognitive theory of ethnicity, according to which ethnicity constitutes not a property of groups but a variably activated cognitive schema realized through concrete communicative practices. The empirical basis consists of two corpora of the newspaper separated by more than seven decades: a Soviet corpus (1939–1945, 1,914 issues) and a contemporary corpus (2018–2024, 1,349 issues), totaling approximately 26 million tokens. The methodology combines distributional semantic analysis based on word vector representations (Word2Vec), collocational frequency analysis, the Word Embedding Association Test (WEAT), and qualitative analysis of extracted contexts. The study finds that in the Soviet subcorpus, ethnonyms of Dagestan nationalities and the generalizing category *dagestanets* ('Dagestani') functioned as interchangeable units of mobilization, activated across military, labor, and cultural narratives. In the post-Soviet subcorpus, a functional differentiation of categories is observed: ethnonyms shift in a statistically significant manner toward the domain of cultural-traditional representation (the "demobilization" of nationalities), while the mobilization function becomes concentrated

in the regional category *dagestanets*, whose frequency increased approximately eightfold. These findings empirically corroborate Brubaker's thesis of the variable activation of ethnicity and identify a discursive correlate of Dagestan's institutional recentralization: the culturalization of ethnonyms proceeds alongside the construction of a regional civic identity as the dominant frame of loyalty.

Keywords: ethnicity, Brubaker, constructivism, media discourse, Dagestan, diachronic analysis, word embeddings, mobilization, supra-ethnic identity

Введение

Вопрос о том, каким образом этнические категории приобретают социальную значимость и политическое измерение, остается одним из центральных в современной социологии. Конструктивистский подход, развитый в работах Роджерса Брубейкера, предлагает ответ, согласно которому этничность не является изначально данным групповым свойством, а воспроизводится через когнитивные схемы и коммуникативные практики, в том числе через институциональные механизмы производства публичного дискурса (Brubaker 2004). Этот ответ принципиально важен и для медиаисследований: именно средства массовой информации выступают одной из ключевых арен, на которой этнические категории обретают эмоциональную насыщенность, нормативный статус и политическую значимость — превращаясь из абстрактных классификаций в практические схемы восприятия социальной реальности. Есть свидетельства того, что регулярное взаимодействие с этнически ориентированными медиа не только укрепляет субъективную значимость этнической идентичности, но и оказывает воздействие на электоральное поведение (Lemole 2024).

Особое аналитическое значение в этом контексте приобретает пресса, функционирующая в условиях государственного учредительства и зависимости от бюджетного финансирования. В России подобная модель является системообразующей: государство одновременно выступает регулятором, законодателем, собственником и финансовым донором этнических СМИ, что принципиально определяет характер производимого дискурса (Gladkova et al. 2019). Государственные медиа в этих условиях становятся инструментом легитимации властных практик и управления публичным пространством (Wijermars 2022), тогда как редакторы и журналисты оказываются в состоянии структурной зависимости от субсидий и институционального надзора (Lowrey, Erzikova 2013). Понимание механизмов, посредством которых такая пресса конструирует этнические категории, является необходимым условием для ана-

лиза более широких процессов этнополитической мобилизации и управления разнообразием в многонациональном государстве.

Этнополитическая мобилизация в связке с государственными СМИ — тема особенно релевантная для постсоветского пространства: для репрезентации национальностей советская медиасистема заложила устойчивые дискурсивные рамки, которые не исчезли после распада СССР, но все же значительно изменились. Советская пресса функционировала как инструмент конструирования национальных идентичностей в рамках проекта «этнографической колонизации» — административного закрепления этнических категорий, создания национальных элит и институционализации этнических различий через государственную бюрократию (Hirsch 2005). Важное место в советском медиадискурсе об этничности занимала так называемая «экономика подарков»: пресса конструировала образ национальных республик как участников общесоюзного обмена символическими «дарами» — трудовыми достижениями, лояльностью, вкладом в общее дело — в ответ на которые центр обеспечивал признание, покровительство и ресурсы (Brooks 2000). Газеты в этой связи становились не просто каналом трансляции этих нарративов, но активными производителями определенной этнической таксономии (Lenoe 2004).

Постсоветский период ознаменовался не только институциональной реструктуризацией медиапространства, но и существенным переопределением самих дискурсивных рамок, через которые осмысляется этничность. Если советский дискурс выстраивал этнические категории преимущественно в логике «дружбы народов» и совместного созидания, то постсоветская репрезентация тяготеет к иным семантическим полям: миграционным нарративам, разговорам о порядке и безопасности, а нередко — к криминализирующему фреймингу отдельных этнических групп (Нагорный 2017). Вопрос о том, насколько радикален этот сдвиг, насколько он устойчив и как проявляется на уровне конкретного регионального издания, остается открытым.

В качестве объекта данного исследования выбрана газета «Дагестанская правда» — старейшее издание Республики Дагестан, непрерывно выходящее на русском языке с 1918 года: это делает ее уникальным материалом для диахронического анализа. Выбор Дагестана в качестве кейса обусловлен рядом обстоятельств, делающих его исключительно значимым объектом для анализа. Во-первых, республика является одним из наиболее этнически разнообразных регионов не только России, но и мира — в ее границах насчитывается более тридцати различных языковых групп (Dobrushina 2023). Во-вторых, в отличие от большинства национальных республик Российской Федерации, Дагестан лишен единой титульной националь-

ности, что исторически обусловило формирование специфической модели консенсусного этнополитического управления, основанной на принципах пропорционального представительства различных народов в органах власти (Ware, Kisriev 2001). В-третьих, Дагестан располагает документированной историей неравного включения этнических групп в государственные институты — от призывной политики периода Великой Отечественной войны до современных ограничений доступа к военной и правоохранительной службе, — что создает уникальное поле для исследования соотношения между медиальной репрезентацией этнических категорий и институциональными практиками их использования.

Новые методы обработки естественного языка расширили методологический инструментарий для социальных исследований, позволяющий фиксировать семантические сдвиги в масштабных текстовых корпусах с ранее недоступной точностью. Распространение дистрибутивно-семантических моделей, основанных на векторных представлениях слов (word embeddings), и сопутствующих им инструментов (в том числе теста на ассоциативные смещения, WEAT) открыло возможности для систематической реконструкции стереотипов, имплицитных ассоциаций и их изменений во времени. Гарг с соавторами показали, что векторные эмбеддинги способны количественно зафиксировать сто лет эволюции гендерных и этнических стереотипов в англоязычном корпусе (Garg et al. 2018); Чарлзуорт с коллегами распространили этот подход на двухсотлетний период, охватив четырнадцать социальных групп в американском английском (Charlesworth et al. 2022); Сорато с соавторами применили диахроническую логику к анализу репрезентации мигрантов в многоязычном политическом корпусе (Sorato et al. 2024). Для русскоязычных региональных советских и постсоветских текстов сопоставимого исследования до настоящего момента не проводилось.

Настоящая статья восполняет эту лакуну. На материале двух корпусов газеты «Дагестанская правда» — советского (1939–1945) и современного (2018–2024), суммарным объемом около 26 миллионов токенов — мы анализируем, как изменилась дискурсивная репрезентация советских дагестанских национальностей в региональных государственных медиа за восемь десятилетий. Центральный исследовательский вопрос формулируется следующим образом: каковы направление и масштаб семантических сдвигов в репрезентации этнических категорий в «Дагестанской правде» при переходе от советского к постсоветскому периоду, и в какой мере эти сдвиги свидетельствуют о переопределении самой рамки, через которую пресса конструирует этничность? Для ответа на этот вопрос применяется сочетание дистрибутивно-семантического ана-

лиза (Word2Vec), выявления частотных коллокационных паттернов и теста на ассоциативные смещения (WEAT), дополненное качественным анализом извлеченных контекстов. Работа вносит вклад в три области: социологию этничности (в части изучения того, как этнические категории конструируются в медиа), российские медиаисследования (в части диахронической динамики государственной региональной прессы) и вычислительную социальную науку (в части применения NLP-методов к русскоязычным историческим корпусам).

Обзор литературы

Медиафреймирование и производство этнических категорий

Отправной точкой для понимания роли средств массовой информации в конструировании этничности служит концепция фрейминга. Роберт Энтман определяет фрейминг как процесс отбора отдельных аспектов воспринимаемой реальности и наделения их особой значимостью в тексте таким образом, чтобы продвигать определенную интерпретацию проблемы, ее причин, моральной оценки и возможных решений (Entman 1993). Приложенная к этническому измерению, эта концепция указывает на то, что СМИ не просто описывают существующие этнические различия, но активно производят их: выбирая, каким образом называть группы, с какими событиями и качествами их ассоциировать, медиаинституты формируют когнитивные схемы, через которые аудитория воспринимает межгрупповые отношения. В этом смысле медийный дискурс является не нейтральным каналом передачи информации, а одним из ключевых пространств, где этнические категории обретают социальную весомость и эмоциональную насыщенность — именно то, что Брубейкер описывает как превращение категорий в практические схемы восприятия (Brubaker 2004). Ван Дейк показал, что медийные практики производства и воспроизводства текстов систематически отражают и укрепляют отношения власти, в том числе цементируют этническую категоризацию (Van Dijk 2008).

Эмпирически эта теоретическая интуиция находит подтверждение в нескольких направлениях исследований. Абрахано, Хайнал и Хасселл убедительно показали, что устойчивое негативное освещение иммиграции в новостях связано не только с ужесточением взглядов на саму иммиграцию, но и с более глубокими сдвигами в партийной самоидентификации белых американцев — вплоть до смены партийных предпочтений (Abrajano, Hajnal, Hassell 2017). Гауз, Мур и Остфельд дополнили эту картину: систематический

контент-анализ показал, что протесты с участием людей смешанного расового происхождения значительно чаще описываются в новостях через призму гнева и страха по сравнению с аналогичными протестами, организованными преимущественно белыми участниками (Gause, Moore, Ostfeld 2023). Нигерийский кейс, проанализированный Адисой, Мохаммедом и Ахмадом, указывает на наиболее острые последствия такого подхода: подстрекательский язык и двойные стандарты в газетах, детерминированные интересами медиавладельцев, активно усугубляют этнические разногласия и способствуют эскалации межэтнического конфликта (Adisa et al. 2024).

Медиа, этническая идентичность и политическая мобилизация

Воздействие медиа на этническую идентичность и ее политическое измерение формирует самостоятельную исследовательскую область. Лемоли показывает, что регулярное взаимодействие с этнически ориентированными СМИ в контексте радиовещания на баскском языке не только укрепляет субъективную значимость этнической самоидентификации, но и ведет к большей поддержке этнических политических партий: медиаинституты функционируют как механизмы «продуцирования групповости», активно конституирующие этнические категории в качестве политически значимых (Lemole 2024). Чуба демонстрирует, что кандидаты стратегически приспосабливают свои мобилизационные стратегии к этническому составу избирателей, что подтверждает тезис Брубейкера об этничности как активно конструируемой коммуникативными действиями категории (Czuba 2023).

Важным дополнением к этому корпусу служит историческая перспектива Пенгла, Весслера и Руэды. Анализируя долгосрочное воздействие распространения печатных технологий на политизацию этничности в странах Африки к югу от Сахары, они установили, что общение с печатными СМИ на африканских языках укрепляло «воображаемые сообщества», делая этническую идентичность более заметной когнитивно при относительно прозрачных групповых границах (Pengl, Roessler, Rueda 2022). Эта работа перекликается с классическим тезисом Андерсона о нации как воображаемом сообществе, чье производство неотделимо от технологий печатного капитализма (Anderson 2006).

Отдельного внимания в контексте настоящего исследования заслуживает институциональный пласт этнической мобилизации — использование этнических категорий государством при конструировании вооруженных сил. Седерман, Гледич и Бухауг показывают,

что неравенство между этническими группами создает напряжения для этнонационалистической мобилизации, а исключение из властных институтов систематически повышает риск вооруженного конфликта (Cederman, Gleditsch, Buhaug 2013). Харкнесс в работе об «этническом стакинге» в африканских армиях показывает, что лидеры, создающие военизированные формирования из представителей собственной этнической группы, удерживают власть существенно дольше (Harkness 2021). Джонсон и Турбер на данных проекта Security-Force Ethnicity показывают, что этнический состав вооруженных сил является значимым предиктором как риска переворота, так и государственных репрессий, причем аналитически критично различать представительство этнических групп на уровне офицерства и рядового состава (Johnson, Thurber 2017). Советская практика комплектования армии демонстрирует особую логику этнической селекции: советское государство открывало институт армии лишь для «лояльных» национальностей (Rousseau, Blauvelt 1998). Безугольный на архивных материалах реконструирует динамику этнических ограничений призыва в годы Великой Отечественной войны, включая приказы НКО № 0585 от 30 июля 1942 года и № 0974 от 21 декабря 1942 года, выводившие горские народы Северного Кавказа из общего призыва (Безугольный 2020). Современные данные показывают, что этническая селекция при комплектовании силовых структур в России воспроизводится и сегодня, а частичная мобилизация 2022 года актуализировала эти паттерны заново (Мохов, Простаков 2014; Арутюнова 2023).

Советская пресса и производство этнических категорий

Взаимосвязь между государственными медиа и этническими категориями в советском контексте имеет отчетливую институциональную специфику. Советская пресса функционировала в рамках централизованной системы производства публичного дискурса, в которой партийные лидеры устанавливали рамки для всех способов самовыражения. Брукс показывает, что советский публичный дискурс организовывал этнические взаимоотношения посредством специфической «экономики подарков»: поддерживалась иерархия, в которой русские представлялись как локомотив модернизации, а национальные республики — как создатели особых культурных и материальных ценностей (Brooks 2000). Хирш реконструирует этот процесс на институциональном уровне: советский проект «этнографической колонизации» предполагал активное государственное вмешательство в классификацию населения, создание и закрепление национальных категорий через паспортную систему (Hirsch

2005). Ленои показывает, что советская пресса сегментировалась по аудиториям — для крестьянства, для молодежи, для национальностей, — что само по себе воспроизводило и закрепляло определенные категориальные различия (Lenoe 2004). Фицпатрик описывает более широкий контекст советской повседневности, в которой медиадискурс оправдывал правовое насилие и конструировал образы «нового человека», разделяя общество по этапам формации в соответствии с идеями марксизма-ленинизма (Fitzpatrick 1999).

Российские региональные медиа: структурный контекст

В постсоветский период структурная логика отношений между государством и прессой формально изменилась, но сохранила многие содержательные черты. Гладкова с соавторами показывают, что в России государство одновременно выступает регулятором, законодателем, собственником и финансовым донором этнических СМИ (Gladkova et al. 2019). Виермарс свидетельствует о том, что государственные СМИ играют важную роль в легитимации властных практик и что государство все активнее проникает в сферу информационных систем для перераспределения медиарынков (Wijermars 2022). Лоури и Эрзикова в исследовании российских региональных журналистов фиксируют состояние структурной зависимости от государственных субсидий, олигархических интересов и институционального надзора, что создает специфические ограничения для производства дискурса (Lowrey, Erzikova 2013).

Вычислительные методы в исследованиях этнического дискурса: от частотного анализа к диахроническим эмбеддингам

За последние полтора десятилетия методологический арсенал вычислительных исследований этничности претерпел значительную эволюцию. Первые исследования концентрировались на дискурсе в социальных медиа и опирались на стандартные частотные метрики и простые классификаторы (Bruns, Liang 2012), которые позволяли изучать связь настроений с различными событиями (Burnap et al. 2015). Переход к методам машинного обучения открыл принципиально иные аналитические горизонты: стало возможным фиксировать сдвиги общественных настроений в реальном времени, а многолетний анализ данных на нескольких платформах позволил связывать изменения в тональности контента с конкретными политическими событиями (Sorato et al. 2024).

В российском исследовательском контексте аналогичная эволюция происходила с некоторым запозданием. Значимым этапом стал проект по выявлению хейт-спича в этническом дискурсе на данных из социальных сетей, в рамках которого были разработаны первые методологии для выявления этнорелевантных дискурсов (Pronoza et al. 2021). Проект EthnoHate продвинулся дальше: для детекции этнически мотивированной ненависти были апробированы нейросетевые архитектуры, включая тонкую настройку BERT (Pronoza et al. 2021).

Наиболее релевантным для настоящей работы направлением является применение диахронических векторных эмбеддингов к анализу исторических изменений в репрезентации социальных групп. Гамильтон, Лесковец и Юрафски установили регулярные закономерности семантического изменения и разработали воспроизводимую методологию для сопоставления эмбеддингов через исторически разделенные корпуса (Hamilton, Leskovec, Jurafsky 2016). Как уже отмечалось во введении, эту методологию последовательно развили Гарг с соавторами, Чарлзуорт с коллегами и Сорато с соавторами — от столетней ретроспективы гендерных и этнических стереотипов в англоязычном корпусе до двухсотлетнего охвата и мультиязычного анализа репрезентации мигрантов (Garg et al. 2018; Charlesworth et al. 2022; Sorato et al. 2024). Тийс, Триллинг и Кроон применили контекстуализированные эмбеддинги к анализу этнических предубеждений в новостных текстах, обнаружив систематические ассоциативные паттерны в репрезентации этнических аутгрупп (Thijs, Trilling, Kroon 2024).

Принципиальным аргументом в пользу использования статических эмбеддингов (Word2Vec) в настоящей работе является специфика задачи. BERT-подобные модели обучены на современных корпусах и не адаптированы к языку советской прессы 1939–1945 годов; тонкая настройка на историческом домене потребовала бы разметки, которой для данного корпуса не существует. Word2Vec, напротив, позволяет обучить полностью доменно-специфичные векторы непосредственно на целевых корпусах. WEAT, валидированный применительно к статическим эмбеддингам (Caliskan et al. 2017), обеспечивает строгую количественную оценку ассоциативных смещений и дает возможность непосредственно сравнивать структуры ассоциаций между двумя корпусами.

Исследовательская лакуна и значимость исследования

Настоящее исследование работает с двумя хронологически разделенными срезами репрезентации этнических категорий — советским военным периодом (1939–1945) и современным российским

(2018–2024) — и принципиально включает в анализ современный публичный дискурс как «естественный язык». Это методологическое решение вытекает из самой постановки проблемы: если этничность, согласно Брубейкеру, существует не как вещь, но как процесс категоризации, разворачивающийся в конкретных коммуникативных практиках (Brubaker 2004), то ее исследование требует обращения к тем регистрам языка, в которых эта категоризация совершается здесь и сейчас. Сопоставление двух срезов позволяет выявить, как одни и те же номинативные единицы — названия народов, регионов, конфессиональных общностей — меняют свои ассоциативные поля в зависимости от институционального и политического контекста.

Теоретическое обоснование такого методологического сдвига предлагает Варшавер в недавней работе, систематизирующей конструктивистские подходы к исследованию этничности в контексте когнитивного поворота (Варшавер 2024а). Автор показывает, что задача конструктивистского исследования состоит в реконструкции механизмов производства этничности в коммуникации, и предлагает дескриптивную модель, фиксирующую, какие аспекты этничности подлежат эмпирической операционализации. Диахроническое сопоставление эмбеддингов, таким образом, становится инструментом, позволяющим видеть этничность не как статичную систему обозначений, а как подвижный конструкт, переопределяемый во времени вместе с изменениями в социальной, институциональной и политической среде.

На этом теоретическом фоне область пересечения исследований медиафреймирования этничности и вычислительных диахронических методов применительно к русскоязычной региональной прессе остается практически неразработанной. Существующие вычислительные исследования этнического дискурса в России сосредоточены преимущественно на социальных сетях и неформальных текстах (Pronoza et al. 2021), тогда как систематический диахронический анализ советских и постсоветских газетных корпусов с применением векторных эмбеддингов до сих пор не проводился. Параллельно сформировавшийся корпус исследований об этничности как операциональном критерии военной мобилизации (Cederman, Gleditsch, Buhaug 2013; Harkness 2021; Johnson, Thurber 2017; Безугольный 2020) фиксирует институциональный контекст, в котором этнические категории функционировали как маркеры лояльности, — однако систематической связки этого контекста с количественным анализом репрезентаций тех же категорий в медиадискурсе до сих пор предложено не было.

Значимость проведенного исследования обусловлена тремя ключевыми факторами. Во-первых, оно впервые предлагает диа-

хронический вычислительный анализ репрезентации этнических категорий в региональном государственном издании на материале корпусов, разделенных хронологически и институционально, что позволяет эмпирически проверить тезис Брубейкера о переменной активации этничности применительно к конкретной многоэтничной общности с документированной историей неравного включения в государственные институты. Во-вторых, включение современного дискурса как одного из срезов анализа позволяет соотнести исторические паттерны с актуальными практиками категоризации и сделать видимой преемственность или разрыв между советской и постсоветской логиками производства этнических различий. В-третьих, дагестанский кейс обладает специфической аналитической ценностью: как показывают архивные материалы о призывной политике 1942–1943 годов (Безугольный 2020) и современные данные о доступе дагестанцев к военной и правоохранительной службе, этничность здесь действовала и продолжает действовать как институциональный фильтр. Таким образом, работа располагается на пересечении трех исследовательских траекторий — когнитивной теории этничности, вычислительных методов диахронической семантики и институциональной социологии этнической мобилизации — и восполняет лакуну, образовавшуюся между ними.

Контекст

Газетный дискурс опирается на уже существующие категориальные системы, и анализ репрезентации этнических категорий в дагестанской прессе требует обращения к исторически сложившимся формам социальной организации, поверх которых работала и продолжает работать журналистская категоризация. До советской модернизации Дагестан отличался локальной изолированностью при исключительном языковом разнообразии — более тридцати языковых групп, многие из которых не обладают взаимопонятностью (Dobrushina 2023). Основной социальной единицей выступала сельская община-джамаат, структурированная вокруг тухумов — патрилинейных родственных групп (Wixman 1980). Этническая принадлежность в этот период не была политизирована и административно фиксирована: ситуативность самоопределений, многоязычие и пересекающиеся локальные лояльности не создавали условий для устойчивых «групповых» идентичностей в современном понимании.

Переломным моментом стал советский проект институционализации этничности. Описанный Хирш проект «этнографической колонизации» предполагал активное государственное вмешательство в классификацию населения: создание и закрепление нацио-

нальных категорий, формирование национальных элит, введение паспортной системы с обязательным указанием национальности, развитие национальных языков и культурных институтов (Hirsch 2005). Советская модель предполагала сосуществование множества «национальных квартир» в общем политическом пространстве (Slezkine 1994). В Дагестане этот процесс приобрел особый масштаб: список из сорока национальностей переписи 1926 года был редуцирован к 1939 году примерно до пятнадцати позиций. Варшавер, Орлова и Шульга показывают, что следствием реклассификаций стало распространение в Дагестане особой «народной социологии этничности» — документарного конструктивизма, в рамках которого этничность воспринимается прежде всего как административная характеристика (Варшавер, Орлова, Шульга 2023). Этничность стала ресурсом политического распределения, а пресса — пространством его публичной легитимации.

Распад СССР ознаменовался не угасанием советских категорий, а их острой политизацией. В Дагестане — республике без единой титульной нации — был сформирован Государственный совет на принципах пропорционального этнического представительства; Уэр и Кисриев описывают эту модель как консоциональную демократию, основанную на договорных балансах между крупнейшими национальными группами и предотвращающую перерастание этнической конкуренции в открытый конфликт (Ware, Kisriev 2001). Параллельно возникали новые типы напряжений, связанные с миграцией горцев на равнины и конкуренцией за земельные и городские ресурсы (Карпов, Капустина 2011). Упразднение Государственного совета в 2006 году в ходе федеральной рецентрализации совпало по времени с интенсивными процессами урбанизации и пространственной мобильности 2000–2010-х годов: центром внутренней миграции стала Махачкала, усилилась внешняя миграция в другие регионы России (Карпов, Капустина 2011). В городской среде и за пределами республики начала складываться надэтническая «общедагестанская» идентичность, функционирующая на уровне «дагестанец», а не «аварец» или «лезгин» (Варшавер и др. 2025), а исламское возрождение и вытеснение местного многоязычия русским как lingua franca повседневности дополнительно снижали социальную релевантность национальных категорий (Dobrushina 2023; Варшавер и др. 2024).

В этих условиях постсоветский медиаландшафт Дагестана формировался при сохранении государственного учредительства основных изданий. Гладкова с соавторами показывают, что этнические СМИ в России функционируют между государственной моделью и альтернативными голосами, причем государство сохраняет роль доминирующего собственника и спонсора (Gladkova et

al. 2019), а Виермарс демонстрирует ключевую роль постсоветских государственных медиа в легитимации властных практик (Wijermars 2022). «Дагестанская правда» выступает площадкой публичного закрепления официально санкционированных версий этнических отношений, артикулируемых уже в рамках миграционных нарративов, религиозных маркеров и логики безопасности. Из этого описания следует логика хронологических рамок исследования: советский корпус (1939–1945) охватывает период Великой Отечественной войны, когда мобилизующая функция этнических категорий достигала максимальной интенсивности (Brooks 2000; Hirsch 2005); современный корпус (2018–2024) охватывает период после демонтажа консоциональной логики 1990-х, трансформации социального состава аудитории под действием урбанизации и миграции и возвращения военной риторики в связи с событиями 2022 года. Сопоставление корпусов позволяет зафиксировать долгосрочные семантические сдвиги — от советской мобилизационной рамки к постсоветским дискурсивным конфигурациям.

Данные и методология

Данное исследование основано на двух корпусах газетных текстов издания «Дагестанская правда»: корпуса советского периода (1939–1945 гг.) и корпуса современного периода (2018–2024 гг.). В советском корпусе представлено 1914 выпусков, а в постсоветском — 1349.

Параллельно с вычислительным анализом проводилось качественное исследование репрезентации советских и дагестанских этнических категорий в публичном дискурсе. Статьи и цитаты, содержащие эксплицитные отсылки к советским национальностям, подвергались ручному отбору, анализу и тематической классификации. В общей сложности было извлечено около 1100 цитат из более чем 500 публикаций. На основе данного массива были выявлены и описаны основные дискурсивные стратегии использования этнических категорий в региональных средствах массовой информации Республики Дагестан.

Предобработка данных, обучение моделей и последующий анализ осуществлялись на языке Python с использованием реализации Word2Vec из библиотеки gensim (Mikolov et al. 2013a; Mikolov et al. 2013b; Řehůřek, Sojka 2010). Ввиду специфики качества оцифрованных текстов советского периода была применена автоматическая орфографическая коррекция с использованием библиотеки SymSpell для русского языка. Данная процедура сопровождалась рядом ограничений, которые подробно обсуждаются в соответствующем разделе работы. Лемматизация выполнялась с помощью библиотеки pymorphy3 (Korobov 2015).

После удаления стоп-слов (включая служебные части речи, такие как частицы, местоимения и союзы) объем корпуса советского периода составил 11 966 114 токенов, современного корпуса — 13 889 370 токенов. Последовательности, содержащие менее двух токенов, исключались из анализа. Процесс предобработки текста продемонстрирован на Рис. 1–3. Параметры моделей и программный код представлены в приложении.

… знатный чабан аварец Омар Пакалов на мгновенье…

… Зайнутдин, аварец из Гунибского района, вступив в неравный бой…

… Гаджиев аварец, до войны жил и работал в селении Гергибель…

Рис. 1. Исходный текст
Fig. 1. Original text

… знатный чабан аварец Омар Пакалов на мгновенье…

… Зайнутдин, аварец из Гунибский район, вступить в неравный бой…

… Гаджиев аварец, до война жить и работать в селение Гергибель…

Рис. 2. Текст после приведения к стандартной морфологической форме
Fig. 2. Text after normalization to the standard morphological form

… знатный чабан аварец Омар Пакалов мгновенье…

… Зайнутдин аварец Гунибский район вступить неравный бой…

… Гаджиев аварец война жить работать селение Гергибиль…

Рис. 3. Текст после удаления стоп-слов (англ. stopwords) и корректировки ошибок
Fig. 3. Text after removing stopwords and correcting errors

Для систематического сопоставления семантических различий формы слов, обозначающие советские дагестанские национальности, а также соответствующие прилагательные формы были заменены на унифицированные токены «dag_nation_NOUN» и «dag_nation_ADJ». Это позволило сопоставить их семантические роли в двух корпусах и сравнить их с семантическим окружением лексемы «дагестанец». Дополнительно была проведена оценка эксплицитной предвзятости путем извлечения и анализа наиболее частотных би- и триграмм в обоих временных периодах (Kurita et al. 2019; Basta et al. 2019). Унифицированные токены использовались и на этом этапе анализа для обеспечения сопоставимости результатов.

Анализ семантических изменений в исследовании базируется на методе векторных представлений (англ. vector embeddings) — подходе к кодированию дискретных объектов (слов, предложений, изображений, графов и др.) в виде непрерывных векторов в многомерном пространстве (см. Рис. 4). Основная идея метода заключается в том, чтобы сохранить семантические или структурные отношения между объектами: близкие по смыслу или свойствам объекты имеют близкие векторные представления — малое расстояние в пространстве (см. Рис. 5).

На основе этих представлений для количественной оценки семантических изменений во времени применялся Word Embedding Association Test (WEAT), направленный на выявление ассоциаций между целевыми этническими категориями и двумя семантическими наборами — традиционным (культура, язык, музей) и модернистским (будущее, гражданин, модернизация). Статистическая значимость выявленных ассоциаций определялась с использованием пермутационного теста (Caliskan et al. 2017).

Корпусы текстов были предварительно обработаны, и для каждого корпуса были получены векторы представлений слов (см. Рис. 6 и 7). На основе того, как близко относительно друг друга находятся векторы слов категорий советских национальностей и релевантных для настоящего исследования слов (например, СВО, боец), реконструируется семантическое значение советских национальностей за соответствующие временные отрезки.

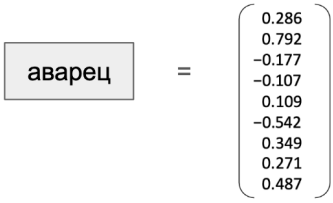


Рис. 4. Представление слова в виде вектора чисел
Fig. 4. Representation of a word as a vector

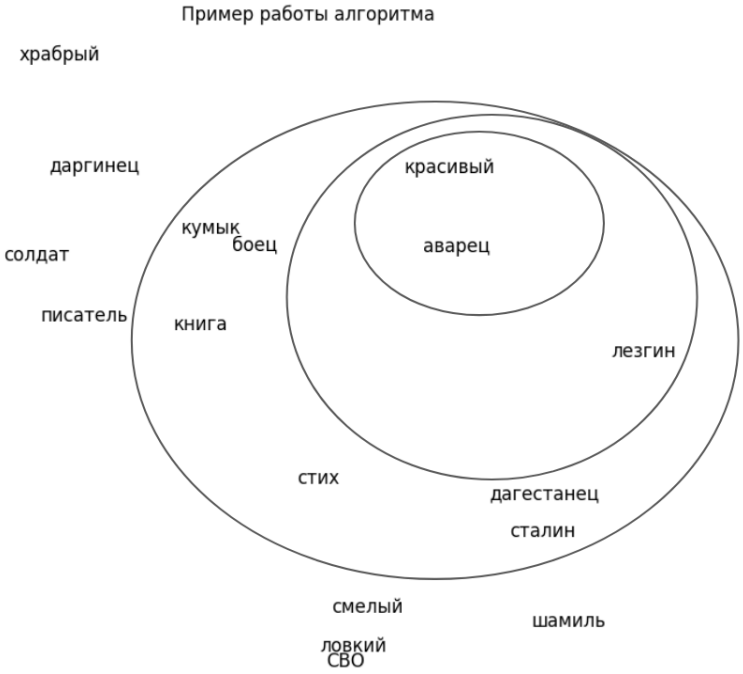


Рис. 5. Семантические отношения на основе векторного расстояния
Fig. 5. Semantic relationships based on vector distance

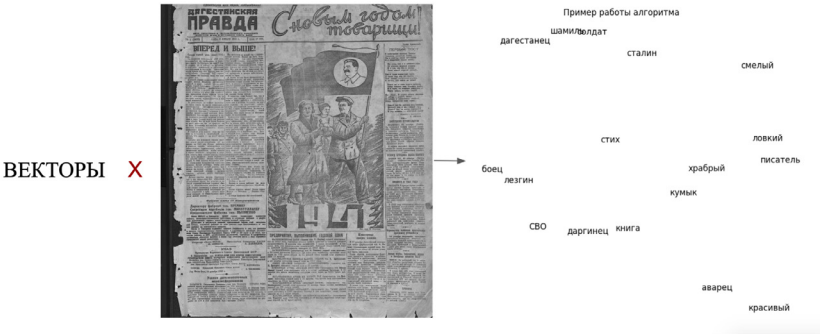


Рис. 6. Алгоритм работы с советским корпусом
Fig. 6. Workflow for processing the Soviet corpus



Рис. 7. Алгоритм работы с постсоветским корпусом
Fig. 7. Workflow for processing the post-Soviet corpus

В основном мы следовали протоколу проверки гипотез WEAT, представленному Калискан и соавторами (Caliskan et al. 2017). Входными данными служила предполагаемая категория слов с этническим предубеждением, представленная двумя списками: **X** и **Y** — списками **целевых слов**, то есть слов, которые предположительно смещены в сторону одной или другой группы категорий (например: X = {воин, боец, война}, Y = {мир, благосостояние, спокойствие}). Проверялось, относится ли X или Y в большей степени к тому или иному списку категорий, или же между этими двумя списками нет разницы в предвзятости. Осуществлялся компаративный анализ предвзятости в отношении двух лексических списков, репрезентирующих семантические поля с доминирующей ролью этнических категорий в газетном дискурсе. Эти списки были подобраны на основе качественного анализа и состояли из семантических полей «война» — «мир», «традиция» — «современность/прогресс». Предвзятость сравнивалась по отношению к двум спискам слов, которые однозначно представляют две группы категорий: советские национальности и дагестанскость (например, A = {аварец, аварский, лезгин}, B = {дагестанка, дагестанец, дагестанский}).

Нулевая гипотеза H_0 заключалась в том, что нет различий между X и Y с точки зрения их относительного (косинусоидального) сходства с A и B. При условии, что существует вектор вложения слов w (обученный на некотором корпусе из некоторой предметной области) для каждого слова w в X, Y, A и B, вычислялась следующая тестовая статистика:

$s(X, Y, A, B) = \sum s(x, A, B) - \sum s(y, A, B)$, где $s(w, A, B)$ является мерой ассоциации между целевым словом w и атрибутивными словами:

$$s(w, A, B) = mean_{a \in A} cos(\underline{w}, \underline{a}) - mean_{b \in B} cos(\underline{w}, \underline{b})$$

В работе Caliskan et al. (2017) H_0 тестируется с помощью теста перестановок, в котором $A \cup B$ разбивается на альтернативные целевые списки A' и B' в экстремальных условиях и вычисляется одностороннее значение $p[s(X, Y, \hat{A}, \hat{B}) > s(X, Y, A, B)]$ — т. е. доля перестановок разделов A и B, в которых тестовая статистика $s(X, Y, \hat{A}, \hat{B})$ больше, чем наблюдаемая тестовая статистика $s(X, Y, A, B)$. Это p-value — вероятность получить наблюдаемое или более экстремальное значение тестовой статистики при условии истинности H_0. Другими словами, это вероятность того, что между A и B нет разницы (по отношению к X и Y) и, следовательно, что категория слова не является предвзятой. Чем выше это значение p-value, тем меньше оснований отвергнуть нулевую гипотезу.

Целевые группы включали:

- X (дагестанская группа): дагестанец, дагестанка, дагестанский;
- Y (группа советских национальностей): лакец, лачка, лакский, кумык, кумычка, кумыкский, аварец и др.

Атрибутивные списки включали:

- A: музей, язык, картина, культура, представление, выступление, традиция, театр, праздник, литература, произведение, книга, царский, угнетение, герой, эпос, диалект, деревня, аул;
- B: будущее, город, страна, гражданин, современность, житель.

Выбор данных лексических единиц обусловлен их устойчивой ассоциацией с соответствующими семантическими полями в русскоязычных корпусах и предшествующих качественных наблюдениях над советским и постсоветским медиадискурсом о Дагестане: список A объединяет лексику, регулярно сопровождающую репрезентацию этнонимов в культурно-историческом и фольклорном ключе, тогда как список B — лексику, маркирующую современность, гражданственность и территориально-государственную принадлежность.

Описанная методология применялась для анализа разницы между ассоциациями советских национальностей и категорий для описания дагестанцев с: а) войной и б) советским дискурсом о национальностях.

Результаты

Для начала были определены леммы (базовые формы слова), чаще всего соседствующие с основными дагестанскими национально-

стями и категорией «дагестанец». Оказалось, что — наряду с леммой «язык» — наиболее частотными «соседями» национальностей являются другие национальности. Так, например, в советском массиве лемматизированная категория «аварец» соседствует с леммами (в порядке убывания частоты) «кумык», «лезгин», «даргинец», «лакец», «русский». Еще более выраженная картина характерна для постсоветского массива: все четыре наиболее частотные леммы-соседи оказываются другими дагестанскими национальностями. Категория же «дагестанец» «держится автономно» от национальностей или любых других этнических категорий. Самые частотные ее леммы-соседи в советском массиве — «год», «тысяча», «советский», «война», «герой», а в постсоветском — «республика», «Дагестан», «свой», «год», «наш». Иначе говоря, национальности в основном выступают «в связке» (Табл. 1).

Таблица 1. Статистика лемм-соседей по подкорпусам
Table 1. Statistics of lemma neighbors across subcorpora

Слово	Токен 1	Частота	Токен 2	Частота	Токен 3	Частота	Токен 4	Частота	Токен 5	Частота
Советский подкорпус										
лезгин	язык	98	кумык	92	аварец	61	русский	53	царский	2
кумык	язык	153	царский	96	лезгин	92	русский	85	год	81
аварец	кумык	78	лезгин	61	год	55	даргинец	46	дагестан	45
даргинец	кумык	56	аварец	46	лезгин	41	лакец	27	дагестан	26
лакец	тов	99	год	86	район	78	работа	73	колхоз	72
дагестанец	год	124	советский	94	дагестан	78	тысяча	70	война	67
лезгин	аварец	41	лакец	41	даргинец	39	кумык	37	дагестан	29

кумык	аварец	53	даргинец	46	лакец	45	лезгин	37	дагестан	35
аварец	лакец	56	кумык	53	даргинец	51	дагестан	41	лезгин	41
даргинец	аварец	51	кумык	46	лакец	46	лезгин	39	дагестан	36
лакец	аварец	56	даргинец	46	кумык	45	лезгин	41	дагестан	38
дагестанец	республика	1035	дагестан	850	свой	780	год	756	наш	654

Чем, однако, эти «связки» (равно как и категория «дагестанец») семантизированы? Параллельный качественный анализ массива позволил сформулировать предположение о наиболее частотных семантиках. Так, в советском подкорпусе национальности и категория «дагестанец» регулярно упоминались в контексте военной или трудовой мобилизации:

> *«Табасаранцы, агульцы, рутульцы! Мстите проклятому Гитлеру за страдания и смерть наших людей! Грудью вставайте на защиту завоеваний октября!» («Дагестанская правда», 1942, 30 августа, № 218).*
> *«Основной, сгонный цех завода трудится с боевым подъемом. Сгонщик коммунист Гасанов — аварец. Он плохо знает русский язык и мало говорит с товарищами по работе, но всю свою большую силу беспредельной любви к родине он вкладывает в самоотверженный труд» («Дагестанская правда», 1941, 8 ноября, № 284).*
> *«В окопах под Ленинградом их встретили дагестанцы, а в предгорьях Кавказа русские люди с Невы и Волги вместе с казаками Кубани» («Дагестанская правда», 1942, 17 декабря, № 309).*

Другим частотным контекстом для советского подкорпуса являлся культурный:

> *«Культурная революция в Дагестане сказывается на обогащении языка его народов. Кумык, аварец, лезгин, даргинец теперь хотят читать на грамотном родном языке классиков марксизма-ленинизма, новости социалистической науки и техники, художественную литературу» («Дагестанская правда», 1939, 5 марта, № 52).*

Культурный и военный контексты упоминания национальностей сохранились и в постсоветском подкорпусе:

> *«Даргинцы-цудахарцы с. Куппа клялись ликом («Берла дяхтлера») и лучами солнца («Берла нурлера»). Клятва солнцем бытовала и у даргинцев-сюргинцев. До сих пор у лакцев бытуют фразы: «Клянусь солнцем над нами!» («КIа ялувсса бургъин хъана»). Клялись солнцем и луной и другие дагестанские народы: лезгины — «Клянусь солнцем» («Ракъунал кьин хьуй»), («Вацрал кьин хьуй»), табасаранцы — «Клянусь солнцем» («Ригъ'инди ху апIурза»), («Ригъдин гьякьниинди»), («Вазлиинди ху апгурза»). У кумыков сохранились клятвы («Гюннюярыгъы гьакъына» — «Клянусь светом солнца!») и благопожелания, связанные с солнцем («Гюн иржайсын ата юртунга» — «Пусть солнце улыбнется на отцовский дом»)» («Дагестанская правда», 2018, 4 мая, № 124).*

Военный регистр в постсоветском материале, напротив, оказывается сосредоточен на категории «дагестанец», а не на отдельных национальностях:

> *«Для каждого россиянина, для каждого дагестанца спецоперация — это возможность задушить фашизм. Боевой дух наших ребят не просто преобладает, он становится все сильнее и тверже...» («Дагестанская правда», 2023, 12 января, № 3).*

По результатам качественного анализа была сформулирована гипотеза, что соотношение указанных контекстов (мобилизация — культура) изменилось и что национальности во многом потеряли свой мобилизационный функционал. Эту гипотезу было решено проверить количественно с помощью WEAT-анализа. Для его осуществления были сконструированы два семантических пространства, заданных дихотомиями «война — мир» и «традиционная культура — современность». Затем были определены леммы соответствующих контекстов (например, для военного контекста — «воин», «боец», «война» и др., для мирного — «мир», «благосостояние», «спокойствие» и др.) и проведены тесты, позволявшие оценить, насколько семантика национальностей и категории «дагестанец» изменилась относительно друг друга и в каком направлении. Результаты тестов приведены в Таблице 2.

Тесты показали, что если в советском подкорпусе национальности и дагестанцы в равной степени семантизированы парами «война/мир» и «традиционная культура/современность», то в постсоветском корпусе национальности продолжают быть семантизированы культурой, но перестают использоваться в военных контекстах, тогда как категория «дагестанец» продолжает в них использоваться. Образно говоря, дагестанские национальности были «демобилизованы», «оставив на войне» дагестанцев.

Таблица 2. Результаты WEAT-анализа: ассоциации целевых групп с атрибутивными полями
Примечание: p-value — значение, вычисленное с помощью пермутационного теста (N = 10 000 итераций).
Table 2. Results of the WEAT analysis: associations of target groups with attribute fields
Note: The p-value is computed using a permutation test (N = 10 000 iterations).

Тест	Целевые группы (X / Y)	Атрибуты (A / B)	Корпус	p-value
WEAT-1	Дагестанцы/ Советские национальности	Традиция/ Современность	Советский	0,0899
WEAT-2	Дагестанцы/ Советские национальности	Традиция/ Современность	Постсоветский	0,0084
WEAT-3	Дагестанцы/ Советские национальности	Война/Мир	Советский	0,106
WEAT-4	Дагестанцы/ Советские национальности	Война/Мир	Постсоветский	0,0338

Последним действием стал подсчет простой встречаемости национальностей и категории «дагестанец» в связи с предположением, что последняя стала употребляться чаще (см. Рис. 8.). Действительно, если в советском подкорпусе категория «дагестанец» встречается примерно с той же частотой (844 упоминания), что и национальные категории (например, «лакец» — 864), то в постсоветском подкорпусе ситуация меняется радикальным образом: категория «дагестанец» оказывается частотнее (7288) каждой отдельной национальности (например, «лакец» — 176) более чем в 40 раз, а всех вместе взятых — примерно в 8 раз.



Рис. 8. Статистика появления этнических категорий в диахронии
Fig. 8. Diachronic statistics of ethnic category occurrences

Проведенный анализ газеты «Дагестанская правда» за периоды 1939–1944 и 2018–2024 гг. позволил зафиксировать существенную трансформацию семантики этнических категорий, транслируемую региональным изданием. Если в советском подкорпусе и национальности, и «дагестанцы» в равной мере семантизировались дихотомиями «война/мир» и «традиция/современность», то в постсоветском материале национальности сохранили преимущественно культурный регистр, утратив мобилизационную функцию, которая оказалась перенесена на региональную категорию «дагестанец». Подсчет частотности подтвердил эту тенденцию: встречаемость категории «дагестанец» возросла кратно, превысив частотность отдельных национальностей более чем в сорок раз. С позиций конструктивистской когнитивной теории, полученные данные свидетельствуют об изменениях в транслируемых властью этнических когнитивных схемах: национальные категории семантически редуцируются до культурно-фольклорного измерения, а консолидирующая функция закрепляется за региональной категорией.

Дискуссия

Переменная активация этничности и функциональная дифференциация категорий

Общая теоретическая значимость полученных результатов состоит в эмпирической демонстрации тезиса Брубейкера о том, что этничность не является свойством групп, а представляет собой переменно активируемую когнитивную схему, реализующуюся в конкретных коммуникативных практиках (Brubaker 2004). Диахроническое со-

поставление двух корпусов позволяет увидеть не просто изменение содержания этнических категорий, но перестройку конструкции категориального аппарата, которым пресса оперирует для описания дагестанского общества. В советский период этнонимы и «дагестанец» были взаимозаменяемы: обе категории с равной регулярностью активировались в мобилизационных контекстах — военных и трудовых — и в культурных нарративах об «обогащении языков народов». В постсоветский период произошло то, что можно определить как функциональную дифференциацию категорий: они перестали быть взаимозаменяемыми и закрепились за качественно различными регистрами публичного дискурса: этнонимы — за культурно-фольклорным регистром, а категория «дагестанец» — за мобилизационно-гражданским.

Эта дифференциация имеет принципиальное теоретическое значение. Брубейкер настаивает на смещении исследовательского фокуса с вопроса «что представляют собой группы?» на вопрос «когда и как категории приобретают групповую значимость?» (Brubaker 2004); наши данные показывают, что применительно к дагестанскому кейсу этот вопрос должен быть переформулирован еще раз: для какой задачи активируется та или иная категория? Наблюдаемый сдвиг не может быть объяснен исключительно демографическими процессами — урбанизацией, миграцией, ослаблением локальных джамаатских структур (Карпов, Капустина 2011; Варшавер и др. 2024b). Он представляет собой сдвиг в структуре самого дискурса, отражающий трансформацию модели «желаемых» этнических схем, которую транслирует государственно учрежденное издание (Gladkova et al. 2019; Wijermars 2022).

Преемственность и разрыв: мобилизационная функция этнических категорий

Сопоставление двух корпусов позволяет уточнить характер преемственности между советской и постсоветской моделями производства этнических категорий. Советский дискурс 1939–1945 годов функционировал в логике, которую Брукс описывает как «экономику подарков»: отдельные национальности репрезентировались как активные субъекты, вносящие вклад в общее дело модернизации и военной победы (Brooks 2000). Характерные для советского подкорпуса контексты — «табасаранцы, агульцы, рутульцы, мстите Гитлеру», «сгонщик коммунист Гасанов — аварец», «в окопах под Ленинградом их встретили дагестанцы» — демонстрируют, что этнонимы функционировали как полноценные мобилизационные категории: они адресовались к конкретным группам и призывали их к действию. Это согласуется с выводами Хирш о том, что советский про-

ект «этнографической колонизации» предполагал не подавление, а институционализацию национальных категорий (Hirsch 2005; Slezkine 1994).

Сопоставление двух корпусов позволяет сформулировать более общую гипотезу о логике, по которой государство в многоэтничных обществах с одной доминирующей этнической группой мобилизует периферийные этнические сообщества для участия в войне. Исследовательская литература по этническому составу силовых структур последовательно фиксирует, что такие государства выстраивают иерархию доступа к институту армии, в которой ядро составляют представители доминирующей группы, а периферийные этнические сообщества оказываются на разных — и исторически подвижных — позициях этой иерархии (Cederman, Gleditsch, Buhaug 2013; Harkness 2021; Johnson, Thurber 2017). Логика иерархии определяется соображениями лояльности: доминирующая группа рассматривается как политически надежный ресурс, тогда как периферийные группы проблематизируются в терминах предполагаемой ненадежности — из-за памяти о недавнем сопротивлении центру, культурной и религиозной отличности или трансграничных связей. Данные Безугольного о приказах НКО 1942–1943 годов, исключавших горские народы Северного Кавказа из общего призыва (Безугольный 2020), представляют собой характерный пример такой логики.

Принципиально важно, что эта иерархия не является статичной: она перестраивается в зависимости от хода войны и состояния ресурсов доминирующей группы. Когда ресурс лояльной группы оказывается исчерпан — из-за потерь, длительности конфликта или одновременных требований нескольких фронтов, — государство вынужденно обращается к периферийным этническим сообществам как к резервному источнику человеческих ресурсов. Именно в этот момент запрет на призыв снимается или смягчается, а официальный дискурс перестраивается таким образом, чтобы сделать мобилизацию ранее исключенных групп легитимной и желательной. Упомянутые выше материалы о добровольческой вербовке в Дагестане 1943 года точно соответствуют этой логике. Мобилизация исторически маргинализированных групп выступает механизмом одновременного управления дефицитом ресурсов и легитимации иерархии, при которой ранее исключенные сообщества получают «возможность доказать лояльность» через непропорциональное участие в опасных военных задачах.

Дискурсивная работа, осуществляемая прессой, является неотъемлемым элементом этой мобилизационной перестройки. Чтобы периферийные этнические группы были готовы отвечать на призыв, необходимо произвести соответствующий категориальный аппарат: этнонимы, прежде проблематизированные в терминах

ненадежности, должны быть переопределены в регистре героизма и преданности общему делу. Ван Дейк показал, что медийный дискурс систематически воспроизводит отношения доминирования не только через эксклюзию, но и через управляемую инклюзию — дозированное предоставление публичной видимости тем группам, чье участие становится необходимым в конкретный исторический момент (Van Dijk 2008). Советский случай демонстрирует, что в условиях экзистенциального военного давления дискурсивная инклюзия может опережать институциональную: этнонимы, которые в практике призыва 1942–1943 годов функционировали как маркеры исключения, на страницах «Дагестанской правды» одновременно активно встраивались в общесоюзный мобилизационный нарратив, формируя мобилизационную готовность этнических сообществ быстрее, чем государство было готово их реально призвать.

Постсоветская конфигурация воспроизводит ту же базовую логику, но в иной институциональной и дискурсивной форме. В контексте растянувшейся на годы СВО государство вновь столкнулось с задачей мобилизации периферийных этнических сообществ, включая народы Северного Кавказа. Государство в подобных конфигурациях не просто реагирует на демографические или военные обстоятельства, но активно перенастраивает дискурсивный режим этничности под свои текущие задачи: когда участие периферийных сообществ в военной службе политически нежелательно — этнические категории закрепляются за ними как маркеры исключения и ненадежности, как это было зафиксировано в приказе НКО 1942 года (Безугольный 2020); когда же мобилизация этих сообществ становится необходимой, те же категории переопределяются в регистре лояльности и общего дела. Одновременно сохраняются и ограничения: практики доступа к военной службе в Дагестане, документированные в публицистике[1] и в социологических исследованиях (Мохов, Простаков 2014), показывают, что доступ к элитным родам войск и офицерским позициям остается ограниченным. Можно предположить, что длящаяся с 2022 года военная кампания способна временно ослаблять и эти ограничения — по логике военного времени, описанной выше для советского случая, — однако проверка этой гипотезы требует данных, выходящих за хронологические рамки анализируемого постсоветского корпуса (до 2024 года).

Именно здесь обнаруживается принципиальное различие в дискурсивной механике двух периодов. Советская пресса осуществляла мобилизацию периферийных сообществ через их эксплицит-

1 Непризывные национальности (2012). Черновик. URL: https://chernovik.net/content/novosti/neprizyvnye-nacionalnosti-2

ное называние: она перечисляла народы, призывала их поименно, встраивала конкретные этнонимы в героические нарративы. Постсоветская пресса оперирует иначе: ее мобилизационная риторика разворачивается вокруг обобщающей категории «дагестанец», вытесняющей отдельные этнонимы из активного политического регистра. Формула «для каждого россиянина, для каждого дагестанца спецоперация — это возможность задушить фашизм» фиксирует этот сдвиг с эмблематической точностью: двухуровневая идентификация «россиянин — дагестанец» выполняет работу мобилизации, которая в советский период распределялась между этнонимами и «дагестанцем» совместно. Агрегированная категория «дагестанец» позволяет мобилизовать периферийное население для войны, не восстанавливая ни советской риторики «дружбы народов», ни публичной артикуляции внутренних этнических различий и неравенств.

«Демобилизация» национальностей и культурализация этнонимов

Особого аналитического внимания заслуживает паттерн «демобилизации» дагестанских национальностей — сдвиг их ассоциативного поля из военно-мобилизационного в сугубо культурно-традиционный регистр. Статистически значимое смещение этнонимов в сторону семантического поля «традиция, музей, язык, эпос, аул» при их одновременном исключении из военных контекстов представляет собой то, что можно определить как культурализацию этнических категорий: они сохраняются в дискурсе, но оказываются функционально ограничены сферой репрезентации культурного наследия. Характерные для постсоветского корпуса контексты — клятвы солнцем у даргинцев-цудахарцев, фразеологизмы лакцев и табасаранцев, кумыкские благопожелания — активируют этнонимы в этнографическом, фольклорном, музейном нарративе, но не в обсуждении актуальных политических, экономических или мобилизационных задач.

Этот сдвиг согласуется с наблюдениями Добрушиной о трансформации языковых идеологий в современном Дагестане, где языковая витальность продиктована языковыми идеологиями населения (Dobrushina 2023). Одновременно он соотносится с более общим паттерном, фиксируемым Пронозой с соавторами в современном российском публичном дискурсе: закреплением отдельных этнических категорий за специфическими семантическими полями — миграции, криминализации, культурной экзотизации — при одновременном ослаблении их политической артикуляции (Pronoza et al. 2021). В случае дагестанских национальностей такой семантикой оказывается не криминализация (характерная для дискурса о «кавказцах» в федеральных российских медиа), а культуризация и традиционализация.

Теоретически значимо и то, что культурализация этнонимов сопровождается их частотным вытеснением из дискурса: в постсоветском подкорпусе отдельные национальности упоминаются на порядок реже обобщающей категории. Это позволяет говорить не просто о семантическом сдвиге, но о систематическом отходе от этнонимической логики репрезентации общества. Энтман, формулируя концепцию фрейминга, подчеркивал, что выбор одних категорий при игнорировании других является одним из ключевых механизмов продвижения определенной интерпретации социальной реальности (Entman 1993); перераспределение частотности между этнонимами и «дагестанцем» представляет собой именно такой механизм.

Конструирование надэтнической категории и ее политические функции

Резкий рост частотности категории «дагестанец» в постсоветском подкорпусе — более чем в восемь раз по сравнению с суммарной частотностью всех отдельных этнонимов — нельзя трактовать как простое отражение демографических или урбанизационных процессов, поскольку рост носит резкий, а не постепенный характер и хронологически совпадает с периодом институциональной рецентрализации Дагестана, а не с существенно более медленными темпами урбанизации или миграции: он представляет собой результат систематической, институционально организованной дискурсивной работы регионального государственного издания по конструированию надэтнической категории как доминирующей рамки коллективной самоидентификации. Семантическое окружение этой категории в постсоветском корпусе — «республика», «Дагестан», «свой», «наш» — указывает на ее функционирование прежде всего как политико-территориальной, а не этнической в узком смысле. «Дагестанец» в современном дискурсе «Дагестанской правды» — это в первую очередь житель Республики Дагестан, участник ее политического пространства, субъект лояльности региональной и федеральной власти, и только во вторую очередь — носитель какой-либо специфической этнической принадлежности.

Этот результат соотносится с выводами о формировании надэтнической общедагестанской идентичности как функционального ответа на условия миграции, урбанизации и контакта с внешними группами за пределами республики (Варшавер 2026). Наши данные указывают на то, что региональная государственная пресса не просто отражает этот процесс, но активно участвует в его производстве: систематическое закрепление категории «дагестанец» в мобилизационных и политических контекстах превращает ее из ситуативного маркера в институционализированную рамку публичной иден-

тификации. «Дагестанская правда» выполняет функцию, которую Пенгл, Весслер и Руэда описывают применительно к распространению печатных технологий в Африке: работает как инструмент формирования «воображаемого сообщества» в андерсоновском смысле (Anderson 2006; Pengl, Roessler, Rueda 2022), но на уровне не этнического, а регионально-гражданского сообщества.

Политическая функция этого сдвига становится видимой при обращении к институциональному контексту постсоветского Дагестана. Упразднение Государственного совета в 2006 году и переход к системе назначаемого главы республики означали демонтаж консоциональной модели, основанной на пропорциональном этническом представительстве (Ware, Kisriev 2001), и его замену вертикалью, встроенной в федеральную политическую архитектуру. В этих условиях артикулированная этническая дифференциация утратила легитимное место в публичном пространстве, а ее место заняла обобщающая категория, согласующаяся с логикой общероссийской политической идентификации. Зафиксированный в наших данных частотный и семантический перевес «дагестанца» над этнонимами может быть интерпретирован как дискурсивный коррелят институциональной рецентрализации.

Методологический вклад и ограничения

Методологический вклад настоящего исследования состоит в применении к русскоязычному региональному медиакорпусу инструментария диахронического анализа распределенных семантических представлений, разработанного в работах Гарга с соавторами (Garg et al. 2018), Чарлзуорт, Калискан и Банаджи (Charlesworth et al. 2022) и Сорато с соавторами (Sorato et al. 2024) преимущественно на англоязычных и западноевропейских корпусах. Продемонстрированная возможность количественно фиксировать сдвиги в семантике этнических категорий на материале региональной прессы, разделенной хронологически более чем на восемьдесят лет, расширяет методологическое поле вычислительных исследований этничности и показывает, что сочетание статических эмбеддингов и WEAT-тестирования продуктивно для русскоязычного исторического материала.

Вместе с тем следует зафиксировать несколько методологических ограничений. Во-первых, корпусы различаются не только хронологически, но и по ряду структурных параметров — объему, качеству оцифровки, типичному составу материалов; автоматическая орфографическая коррекция на основе SymSpell, необходимая для работы с советскими текстами, неизбежно вносит некоторый шум. Во-вторых, статические эмбеддинги Word2Vec дают единое векторное представление для всех контекстов употребления слова, что может скры-

вать локальные вариации значения. В-третьих, WEAT-тестирование фиксирует ассоциативные сдвиги, но не устанавливает причинные механизмы их производства — для этого необходимо сочетание с качественным и историко-институциональным анализом, что и было реализовано в настоящей работе. Наконец, принципиальное ограничение касается статуса источника: выводы относятся не к этническим категориям в восприятии дагестанского общества, а к моделям этнических схем, транслируемым государственно учрежденным изданием. В рамках конструктивистской рамки Брубейкера, впрочем, это уточнение не снижает аналитической ценности результатов: именно институциональные механизмы производства публичного дискурса являются одним из ключевых объектов исследования (Brubaker 2004), а «Дагестанская правда» с ее непрерывной историей с 1918 года представляет собой один из немногих доступных длинных временных рядов институционализированного этнического дискурса.

В концептуальном плане представленное исследование показывает, что этнические категории в публичном дискурсе не исчезают и не сохраняются в неизменном виде — они перестраиваются. Характер этой перестройки становится видимым только при систематическом количественном анализе ассоциативных структур, разворачиваемом на диахронической оси. Перспективные направления дальнейшей работы включают расширение анализа за счет контекстуализированных архитектур с доменной адаптацией, включение дополнительных временных срезов — позднесоветского периода 1960–1980-х годов и «горячего» постсоветского десятилетия 1990-х, — а также сопоставление «Дагестанской правды» с региональной прессой других полиэтничных республик (Татарстана, Башкортостана, республик Северного Кавказа) и с негосударственными изданиями. Дагестанский кейс — с его документированной историей институционализации национальностей, уникальной непрерывностью регионального государственного издания и современной конфигурацией, в которой надэтническая идентичность сосуществует с сохраняющимися этническими категориями, — представляет собой привилегированный объект для изучения механизмов, через которые этничность производится, переопределяется и функционально специализируется в современных публичных пространствах.

Использование ИИ

При подготовке настоящей работы автор использовал три генеративные языковые модели для вспомогательных задач. Языковая модель YandexGPT (Яндекс) применялась для поиска русскоязычной научной литературы. ChatGPT (OpenAI) использовался для генера-

ции и отладки программного кода на языке Python, реализующего предобработку корпусов, обучение векторных моделей и WEAT-тестирование. Claude (Anthropic) применялся для поиска англоязычной литературы, составления первичных аннотаций к источникам, стилистической правки русскоязычного текста и проверки логической связности аргументации. Все библиографические ссылки, выявленные с помощью указанных инструментов, были самостоятельно проверены автором по первоисточникам.

Благодарности / Acknowledgments

Автор благодарит Е.А. Варшавера за научное руководство данным проектом.
The author thanks E.A. Varshaver for his academic supervision of this project.

Список источников / References

Арутюнова Е. М. (2023) Социальное самочувствие и ключевые проблемы: кейс Республики Саха (Якутия) в условиях внешнего давления на Россию. *Социологическая наука и социальная практика*, 11(3), с. 76–94. EDN: IOSYEQ. https://doi.org/10.19181/INAB.2023.1.6

— Arutyunova E. M. (2023) Social well-being and key problems: The case of the Republic of Sakha (Yakutia) under external pressure on Russia. *Sociological Science and Social Practice*, 11(3), pp. 76–94. (in Russ.) https://doi.org/10.19181/INAB.2023.1.6

Безугольный А. Ю. (2020) Отрицательный отбор: запреты и ограничения на комплектование Красной армии по этническому признаку в годы Великой Отечественной войны. *Вестник Российского университета дружбы народов. Серия: История России*, 19(4), с. 869–888. https://doi.org/10.22363/2312-8674-2020-19-4-869-888

— Bezugolny A. Y. (2020) Negative selection: Bans and restrictions on the recruitment of the Red Army on ethnic grounds during the Great Patriotic War. *RUDN Journal of Russian History*, 19(4), pp. 869–888. (in Russ.) https://doi.org/10.22363/2312-8674-2020-19-4-869-888

Варшавер Е. А. (2026) Роль национальностей в Дагестане: взгляд через призму когнитивного поворота в конструктивистских исследованиях этничности. *Журнал фронтирных исследований*, 11(2), с. 72–108. https://doi.org/10.46539/jfs.v11i2.685

— Varshaver E. A. (2026) The Role of Nationalities in Dagestan: A Perspective from the Cognitive Turn in Constructivist Ethnicity Studies. *Journal of Frontier Studies*, 11(2), pp. 72–108. (in Russ.) https://doi.org/10.46539/jfs.v11i2.685

Варшавер Е. А. (2024а) Что именно исследуется, когда исследуется этничность? Дескриптивная модель для конструктивистских исследований этничности в контексте когнитивного поворота. *Социологический журнал*, 30(1), с. 8–32. https://doi.org/10.17323/1728-192x-2024-3-94-126

— Varshaver E. A. (2024a) What exactly is being studied when ethnicity is studied? A descriptive model for constructivist studies of ethnicity in the context of the cognitive turn. *Sociological Journal*, 30(1), pp. 8–32. (in Russ.) https://doi.org/10.17323/1728-192x-2024-3-94-126

Варшавер Е. А., Орлова А. В., Шульга Н. А. (2023) Документарный конструктивизм: этничность как административная категория в восприятии дагестанцев. *Журнал социологии и социальной антропологии*, 26(3), с. 54–81. EDN: MOURPT. https://doi.org/10.31857/S0869541523060106

— Varshaver E. A., Orlova A. V., Shulga N. A. (2023) Documentary constructivism: Ethnicity as an administrative category in the perception of Dagestanis. *The Journal of Sociology and Social Anthropology*, 26(3), pp. 54–81. (in Russ.) https://doi.org/10.31857/S0869541523060106

Варшавер Е. А., Иванова Н. С., Гупалова Ю. В., Кузнецова С. В. (2024b) Какую роль играют национальности в регуляции брачного поведения в Дагестане? *Журнал фронтирных исследований*, 9(2), с. 98–133. EDN: NJKCMO. https://doi.org/10.46539/jfs.v9i2.566

— Varshaver E. A., Ivanova N. S., Gupalova J. V., Kuznetsova S. V. (2024b) What Role do Soviet Nationalities Play in Regulating the Marital Behavior in Dagestan? *Journal of Frontier Studies*, 9(2), pp. 98–133. (in Russ.) https://doi.org/10.46539/jfs.v9i2.566

Карпов Ю. Ю., Капустина Е. Л. (2011) *Горцы после гор: Миграционные процессы в Дагестане в XX — начале XXI века, их социальные и этнокультурные последствия и перспективы.* СПб.: Петербургское востоковедение. EDN: QJBBMZ.

— Karpov Yu. Yu., Kapustina E. L. (2011) *Highlanders beyond the highlands: Migration processes in Dagestan in the 20th — early 21st centuries, their social and ethnocultural consequences and prospects*. St. Petersburg: Peterburgskoe Vostokovedenie. (in Russ.)

Мохов С. В., Простаков С. П. (2014) Межэтнические конфликты в армии и оценка опыта службы в опросах демобилизовавшихся солдат и офицеров. *Мониторинг общественного мнения: экономические и социальные перемены*, (5), с. 96–110. EDN: QALNZJ. https://mirros.hse.ru/article/view/4999

— Mokhov S. V., Prostakov S. P. (2014) Inter-ethnic conflicts in the army and the evaluation of service experience in surveys of demobilized soldiers and officers. *Monitoring of Public Opinion: Economic and Social Changes*, (5), pp. 96–110. (in Russ.)

Нагорный О. С. (2017) Репрезентация этничностей в русскоязычных социальных медиа. *Мониторинг общественного мнения: Экономические и социальные перемены*, (4), с. 165–184. EDN: YLYNVI. https://doi.org/10.14515/monitoring.2017.4.11

— Nagornyy O. S. (2017) Representation of ethnicity in the Russian-speaking social media. *Monitoring of Public Opinion: Economic and Social Changes*, (4), pp. 165–184. (in Russ.) https://doi.org/10.14515/monitoring.2017.4.11

Abrajano M. A., Hajnal Z., Hassell H. J. (2017) Media framing and partisan identity: The case of immigration coverage and white macropartisanship. *Journal of Race, Ethnicity, and Politics*, 2(1), pp. 5–34. https://doi.org/10.1017/rep.2016.25

Adisa R. M., Segbefia S. K., Mohammed S., Trofimova G. N. (2024) Misinformation and disinformation in ethno-religious conflicts: a comparative study of media in Ghana and

Nigeria. *Вестник Российского университета дружбы народов. Серия: Литературоведение, журналистика*, 29(2), pp. 346–357. https://doi.org/10.22363/2312-9220-2024-29-2-346-357

Anderson B. (2006) *Imagined communities: Reflections on the origin and spread of nationalism (Rev. ed.).* London: Verso.

Basta C., Costa-jussà M. R., Casas N. (2019) Evaluating the underlying gender bias in contextualized word embeddings. *arXiv.* https://doi.org/10.48550/arXiv.1904.08783

Brooks J. (2000) *Thank you, comrade Stalin! Soviet public culture from revolution to Cold War.* Princeton, NJ: Princeton University Press.

Brubaker R. (2004) *Ethnicity without groups.* Cambridge, MA: Harvard University Press.

Bruns A., Liang Y. E. (2012) Tools and methods for capturing Twitter data during natural disasters. *First Monday*, 17(4), pp. 1–8. https://doi.org/10.5210/fm.v17i4.3937

Burnap P., Colombo W., Scourfield J. (2015) Machine classification and analysis of suicide-related communication on twitter. In: Proceedings of the 26th ACM Conference on Hypertext & Social Media. ACM, pp. 75–84. https://doi.org/10.1145/2700171.2791023

Caliskan A., Bryson J. J., Narayanan A. (2017) Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), pp. 183–186. https://doi.org/10.1126/science.aal4230

Card D., Chang S., Becker C., Mendelsohn J., Voigt R., Boustan L., Abramitzky R., Jurafsky D. (2022) Computational analysis of 140 years of US political speeches reveals more positive but increasingly polarized framing of immigration. *Proceedings of the National Academy of Sciences*, 119(31), Article e2120510119. https://doi.org/10.1073/pnas.2120510119

Cederman L. E., Gleditsch K. S., Buhaug H. (2013) *Inequality, grievances, and civil war.* Cambridge: Cambridge University Press.

Charlesworth T. E., Caliskan A., Banaji M. R. (2022) Historical representations of social groups across 200 years of word embeddings from Google Books. *Proceedings of the National Academy of Sciences*, 119(28), e2121798119. https://doi.org/10.1073/pnas.2121798119

Czuba K. (2023) Political mobilization of layered ethnic identities. *Ethnic and Racial Studies*, 46(10), pp. 2138–2162. https://doi.org/10.1080/01419870.2022.2154614

Dobrushina N. (2023) Language ideology in an endogamous society: The case of Daghestan. *Journal of Sociolinguistics*, 27(2), pp. 159–176. https://doi.org/10.1111/josl.12590

Entman R. M. (1993) Framing: Towards clarification of a fractured paradigm. *McQuail's reader in mass communication theory*, pp. 390, 397. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Fitzpatrick S. (1999) *Everyday Stalinism: Ordinary life in extraordinary times — Soviet Russia in the 1930s.* New York: Oxford University Press.

Garg N., Schiebinger L., Jurafsky D., Zou J. (2018) Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16), E3635–E3644. https://doi.org/10.1073/pnas.1720347115

Gause L., Moore S. T., Ostfeld M. C. (2023) Anger, fear, and the racialization of news media coverage of protest activity. *Journal of Race, Ethnicity, and Politics*, 8(3), pp. 446–468. https://doi.org/10.1017/rep.2023.27

Gladkova A., Aslanov I., Danilov A., Danilov A., Garifullin V., Magadeeva R. (2019) Ethnic media in Russia: between state model and alternative voices. *Russian Journal of Communication*, 11(1), pp. 53–70. https://doi.org/10.1080/19409419.2018.1564355

Hamilton W. L., Leskovec J., Jurafsky D. (2016) Diachronic word embeddings reveal statistical laws of semantic change. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, pp. 1489–1501. https://doi.org/10.18653/v1/P16-1141

Harkness K. A. (2021) The Ethnic Stacking in Africa Dataset: When leaders use ascriptive identity to build military loyalty. *Conflict Management and Peace Science*, 39(1), pp. 3–21. https://doi.org/10.1177/0738894220972996

Hirsch F. (2005) *Empire of nations: Ethnographic knowledge and the making of the Soviet Union.* Ithaca, NY: Cornell University Press.

Johnson P., Thurber C. (2017) The Security-Force Ethnicity (SFE) Project: Introducing a new dataset. *Conflict Management and Peace Science*, 37(1), pp. 106–129. https://doi.org/10.1177/0738894217709012

Korobov M. (2015) Morphological analyzer and generator for Russian and Ukrainian languages. In: Analysis of Images, Social Networks and Texts. Springer, pp. 320–332. https://doi.org/10.1007/978-3-319-26123-2_31

Kurita K., Vyas N., Pareek A., Black A. W., Tsvetkov Y. (2019) Measuring bias in contextualized word representations. In: Proceedings of the First Workshop on Gender Bias in Natural Language Processing. Association for Computational Linguistics, pp. 166–172. https://doi.org/10.18653/v1/W19-3823

Lemole G. (2024) Ethnic Media and the Mobilization of Identity. *Comparative Political Studies*, 58(5), pp. 890–923. https://doi.org/10.1177/00104140241252076

Lenoe M. (2004) *Closer to the masses: Stalinist culture, social revolution, and Soviet newspapers.* Cambridge, MA: Harvard University Press.

Lowrey W., Erzikova E. (2013) One profession, multiple identities: Russian regional reporters' perceptions of the professional community. *Mass Communication and Society*, 16(5), pp. 639–660. https://doi.org/10.1080/15205436.2013.770031

Mikolov T., Chen K., Corrado G., Dean J. (2013a) Efficient estimation of word representations in vector space. *arXiv.* https://doi.org/10.48550/arXiv.1301.3781

Mikolov T., Yih W., Zweig G. (2013b) Linguistic regularities in continuous space word representations. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, pp. 746–751. https://aclanthology.org/N13-1090

Sorato D., Lundsteen M., Ventura C. C., Zavala-Rojas D. (2024) Using word embeddings for immigrant and refugee stereotype quantification in a diachronic and multilingual setting. *Journal of Computational Social Science*, 7(1), pp. 469–521. https://doi.org/10.1007/s42001-023-00243-6

Pengl Y. I., Roessler P., Rueda V. (2022) Cash crops, print technologies, and the politicization of ethnicity in Africa. *American Political Science Review*, 116(1), pp. 181–199. https://doi.org/10.1017/S0003055421000782

Pronoza E., Panicheva P., Koltsova O., Rosso P. (2021) Detecting ethnicity-targeted hate speech in Russian social media texts. *Information Processing & Management*, 58(6), 102674. https://doi.org/10.1016/j.ipm.2021.102674

Řehůřek R., Sojka P. (2010) Software framework for topic modelling with large corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. ELRA, pp. 45–50. https://doi.org/10.13140/2.1.2393.1847

Rousseau D. L., Blauvelt T. (1998) War, mobilization, and democratization: The experience of minority groups. Paper presented at the annual meeting of the International Studies Association, Minneapolis, MN.

Slezkine Y. (1994) The USSR as a communal apartment, or how a socialist state promoted ethnic particularism. *Slavic review*, 53(2), pp. 414–452. https://doi.org/10.2307/2501300

Thijs G., Trilling D., Kroon A. C. (2024) Contextualized word embeddings expose ethnic biases in news. In: Proceedings of the 16th ACM Web Science Conference. ACM, pp. 290–295. https://doi.org/10.1145/3614419.3643994

Van Dijk T. A. (2008) *Discourse and context: A sociocognitive approach.* Cambridge: Cambridge University Press.

Ware R. B., Kisriev E. (2001) Ethnic parity and democratic pluralism in Dagestan: A consociational approach. *Europe-Asia Studies*, 53(1), pp. 105–131. https://doi.org/10.1080/09668130124419

Wijermars M. (2022) Selling internet control: The framing of the Russian ban of messaging app Telegram. *Information, Communication & Society*, 25(15), pp. 2190–2206. https://doi.org/10.1080/1369118X.2021.1933562

Wixman R. (1980) *Language aspects of ethnic patterns and processes in the North Caucasus.* Chicago: University of Chicago.

Об авторе / About the author

Хамазаев Руслан Максимович — магистрант Национального исследовательского университета «Высшая школа экономики», Москва, Россия. Научные интересы: социология этничности, когнитивная социология, методы обработки естественного языка, AI Safety.

https://orcid.org/0009-0000-5713-6759. E-mail: ruslan.khamazaev1@gmail.com

Ruslan M. Khamazaev — master's student at the National Research University Higher School of Economics, Moscow, Russian Federation. Research interests: sociology of ethnicity, cognitive sociology, Natural Language Processing, AI Safety.

https://orcid.org/0009-0000-5713-6759. E-mail: ruslan.khamazaev1@gmail.com