

оригинальная статья

Кросс-региональный анализ настроений пользователей сетевых сообществ ВКонтакте «Типичный Краснодар» и «Типичный Кемерово»

Рябченко Наталья Анатольевна

Кубанский государственный университет, Краснодар, Россия
<https://orcid.org/0000-0001-6980-2894>
rrrnatali@mail.ru

Малышева Ольга Петровна

Кубанский государственный университет, Краснодар, Россия
<https://orcid.org/0000-0001-8285-0508>

Поступила в редакцию 02.09.2022. Принята после рецензирования 10.11.2022. Принята в печать 02.12.2022.

Аннотация: Исследование посвящено описанию комплексного метода анализа сетевых лингвистических данных, основанного на семантическом моделировании дискурсивных полей и сетевом анализе. Анализ глобальных дискурсивных полей как социальных графов посредством визуализации и анализа семантических ядер позволяет оперативно выявлять тематики, вызывающие широкий общественный резонанс, и проводить предиктивную аналитику вектора культурных и социально-политических тенденций. Результаты позволили предложить органам власти ряд рекомендаций по организации эффективного информационного управления и нивелирования социальной напряженности в регионах за счет дискурсивного регулирования коммуникативных процессов в онлайн-пространстве.

Ключевые слова: дискурсивные поля, онлайн-сообщества, семантическое ядро, сетевая лингвистика, социальные медиа, ВКонтакте, сетевые лингвистические данные

Цитирование: Рябченко Н. А., Малышева О. П. Кросс-региональный анализ настроений пользователей сетевых сообществ ВКонтакте «Типичный Краснодар» и «Типичный Кемерово». *Виртуальная коммуникация и социальные сети*. 2022. Т. 1. № 4. С. 215–220. <https://doi.org/10.21603/2782-4799-2022-1-4-215-220>

full article

Cross-Regional Sentiment Analysis of VKontakte Communities: Typical Krasnodar vs. Typical Kemerovo

Natalia A. Ryabchenko

Kuban State University, Krasnodar, Russia
<https://orcid.org/0000-0001-6980-2894>
rrrnatali@mail.ru

Olga P. Malysheva

Kuban State University, Krasnodar, Russia
<https://orcid.org/0000-0001-8285-0508>

Received 2 Sep 2022. Accepted after peer review 10 Nov 2022. Accepted for publication 2 Dec 2022.

Abstract: This research tested a complex analytical method based on semantic modeling of discursive fields and network analysis that can be applied to networked linguistic data. The copyright software package Monitoring and Analysis of Social Networks, Internet Communities, and Users provided automatic data processing to restore the semantic core of the discursive fields generated by Typical Krasnodar and Typical Kemerovo online communities in 2021–2022. The analysis was visualized as social graphs and semantic core models. It identified a number of sensitive topics that generated a lot of feedback. The method also proved efficient in predicting cultural and socio-political trends. The complex methodology provided rapid identification of amplified discursive activity, conditions for social unrest, and bifurcation points for destructive social practices. The results may contribute to effective information management and social strain relief in the regions through discursive control of online communication.

Keywords: discursive fields, online communities, semantic core, network linguistics, social media, VKontakte, network linguistic data

Citation: Ryabchenko N. A., Malysheva O. P. Cross-Regional Sentiment Analysis of VKontakte Communities: Typical Krasnodar vs. Typical Kemerovo. *Virtual Communication and Social Networks*, 2022, 1(4): 215–220. (In Russ.) <https://doi.org/10.21603/2782-4799-2022-1-4-215-220>

Введение

Социальные сети и социальные медиа как платформы реализации сетевого коммуникативного взаимодействия в онлайн-пространстве привлекают внимание философов, социологов, политологов, лингвистов, специалистов по большим данным (data scientists), бренд-менеджеров, маркетологов, PR-менеджеров, SMM- и SEO-специалистов, ученых, напрямую не связанных с анализом социальных сетей и больших данных [Aaditi 2020]. Исследователи в области социальных наук рассматривают платформы социальных сетей как фактор, детерминирующий формирование когнитивных и поведенческих моделей пользователей за счет дискурсивного управления. Дискурсивные поля, формируемые в результате дискурсивного взаимодействия пользователей онлайн-сообществ, предопределяют принятие решений на различных уровнях.

Исследования социальных медиа имеют значимые теоретические и прикладные результаты. Изучение особенностей коммуникации пользователей онлайн-сообществ на платформах социальных сетей [McDonald, Forte 2021] открывает перспективы исследования сути (мотивов, факторов, паттернов) социального взаимодействия [Wagner, González-Howard 2018], способствует выработке междисциплинарных подходов к анализу дискурсивных полей, формирующихся на платформах социальных медиа, которые используются для исследования закономерностей дискурсивного управления настроениями пользователей в онлайн- и офлайн-пространстве. Применение методологии изучения больших данных (data science) и когнитивных, корпусных исследований позволяет строить семантические графы («графы знаний» и семантические сети) для машинного обучения [Chen et al. 2022; Ji et al. 2021], разрабатывать программные продукты для автоматической обработки текста и таким образом расширять объемы исследуемого материала, повышая репрезентативность выборки и качество результатов исследования.

Современные исследования социальных сетей и онлайн-сообществ связаны с изучением различных типов или жанров дискурса (медицинского, политического, дидактического, административного, религиозного, дипломатического, рекламного, спортивного) [Wu, Li 2018]; тематик (расистский дискурс,

язык ненависти и неравенства [Papatzani 2021; Pasta 2022]; вопросы вакцинации [Manganello et al. 2022]); дискурсивного поведения политических лидеров [Kalantari et al. 2020]; фейков и дезинформации [Zhang et al. 2021]; аудиовизуальной коммуникации и аудиовизуальных экосистем [Рябченко и др. 2021б]; гендерных вопросов [Гнедаш 2022]; социальных последствий цифрового активизма как явления целенаправленного распространения дискурсов и нарративов в онлайн-пространстве [Leong et al. 2019].

Таким образом, прикладные исследования социальных сетей и дискурсивных взаимодействий набирают популярность в силу высокой практической значимости и эффективности в решении социально-политических проблем, таких как выявление точек пересечения и линий разломов в ожиданиях населения и действиях правительства [Ryabchenko et al. 2019] или формирования и корректировки имиджа субъекта (индивида, организации, региона или страны).

Методы и материалы

Материалом для кросс-регионального исследования настроений пользователей российского сегмента сети Интернет послужили сетевые лингвистические данные, извлеченные методом сплошной выборки из двух крупных пабликов социальной сети ВКонтакте: «Типичный Краснодар» и «Типичный Кемерово». Сообщества относятся к тематике Lifestyle (стиль жизни) и непрерывно генерируют большое количество контента – постов и длинных веток комментариев, свидетельствующих об актуальности публикуемого контента и заинтересованности в нем пользователей.

Мы определяем сетевые лингвистические данные [Рябченко и др. 2021а], которые генерируются дискурсивными полями, формирующимися на платформах социальных сетей, как массивы неструктурированных или неаннотированных вербальных и невербальных данных, которые представлены в произвольной форме, не соответствуют определенной модели организации данных и не зафиксированы в базах данных, что отличает их от текстовых корпусов.

Объекты массивов сетевых данных могут быть организованы по сетевому принципу и визуализированы

в виде социального графа. Визуализация сетевых лингвистических данных в глобальном масштабе позволяет выявлять имплицитные смыслы дискуссий на платформах социальных сетей, определять доминирующие нарративы, актуальные тенденции социально-культурного развития и вектор формирования культурной идентичности. Таким образом, подобные данные позволяют прогнозировать социальное действие, результат реализации которого может иметь как конструктивные, так и деструктивные последствия для устойчивости социально-политических систем.

Работа с сетевыми лингвистическими данными, функционирующими в рамках дискурсивных полей в онлайн-пространстве, предполагает сплошную выгрузку сетевых данных с платформ социальных сетей и их предобработку: систематизацию и формирование датасетов, которые представляют собой специализированные корпуса текстов – веб-корпусы сообщений и веток комментариев. Автоматическая и полуавтоматическая обработка веб-корпусов сообщений и веток комментариев позволяет визуализировать семантическое ядро дискурсивного поля и оперативно выявлять актуальные проблемы социально-политической жизни.

Конструирование семантических ядер основано на теории структурного и семантического синтаксиса, грамматике зависимостей [Tesnière 2015] и ядерных синтаксических структур [Chomsky 2002]. Исходя из того, что любая грамматика проецирует конечную совокупность наблюдаемых высказываний на множество грамматически правильных структур, ядерные модели являются основой речевой деятельности. Они могут подвергаться разнообразным трансформациям в зависимости от требований контекста, дополняться второстепенными элементами, однако это не влияет на пропозиционный смысл высказывания. В данном исследовании для конструирования семантического ядра была выбрана лаконичная и показательная с точки зрения характера социального действия структура SVO, в которой S (Subject) и O (Object) – именованные сущности, актуализированные номинативом существительного или личного местоимения, называющего или указывающего на одушевленного или неодушевленного производителя действия, а V (Verb) – глагол, отражающий характер предикативной связи между элементами S и O. Таким образом, семантическое ядро дискурсивного поля,

представленное в виде направленного графа, делает возможным определение не только актуальных тематик, но и семантических ролей акторов (узлов графа).

Построение семантического ядра дискурсивных полей, сформированных сообществами «Типичный Кемерово» и «Типичный Краснодар», проводилось в несколько этапов.

1. Сбор данных. Критериями для определения точки входа для сбора данных стали: тип платформы (ВКонтакте), сетевое сообщество («Типичный Кемерово» и «Типичный Краснодар», считающиеся наиболее популярными и распространенными сообществами с высоким рейтингом доверия ввиду качества новостей и оперативности их публикации), временной интервал (август 2021 – август 2022 гг.), тип и объем выборки. Использование программного интерфейса ВКонтакте (API ВКонтакте) позволило нам отсортировать все посты по популярности, выбрать 1000 наиболее комментируемых постов (89045 и 531462 комментария в выборке постов по сообществам «Типичный Кемерово» и «Типичный Краснодар» соответственно), сформировав датасеты «Топ-1000 Кемерово» и «Топ-1000 Краснодар».
2. Предварительная обработка данных. Датасеты были подвергнуты автоматической обработке с помощью авторского программного комплекса «Мониторинг и анализ социальных сетей, сообществ и пользователей в интернет-пространстве»¹. После автоматической очистки выборки от стоп-слов были проведены токенизация текста², парсинг с целью анализа синтаксической структуры и выделение связки лексических элементов, соответствующих модели SVO, благодаря чему были сформированы датасеты «Топ-1000 Кемерово-SVO» и «Топ-1000 Краснодар-SVO».
3. Визуализация и интерпретация. Посредством программы визуализации данных Gephi³, обработанные данные датасетов были представлены в виде ориентированных графов, анализ которых позволил выделить в анализируемых дискурсивных полях семантическое ядро, определить доминирующие тематики, характер предикативных связей и семантические роли акторов (элементов S и O в описанной модели).

¹ Свидетельства о государственной регистрации программы для ЭВМ № 2018665564; № 2018663499; № 2019666985; № 2020667724; № 2020667728; № 2021610210; № 2020667357.

² Natural Language Toolkit. *NLTK Project*. URL: <http://www.nltk.org/> (accessed 17 Aug 2022).

³ The Open Graph Viz Platform. *Gephi.org*. URL: <https://www.gephi.org> (accessed 17 Aug 2022).

Результаты

Кросс-региональный анализ дискурсивных полей (рис. 1) показал, что центром семантического ядра дискурсивного поля для обоих регионов является название города – *Кемерово, Краснодар*; президент РФ – *Путин* и жители – *Кемеровчане, Краснодарцы*.

Однако для дискурсивного поля, сформированного сообществом «Типичный Кемерово», центральной тематикой стало обсуждение событий, происходящих во всем регионе – узел *Кузбасс* имеет наибольший диаметр. В дискурсивном поле, сформированном сообществом «Типичный Краснодар», максимальное количество комментариев получили сообщения с единицей *Краснодарцы*. В дискурсивном поле «Типичный Кемерово» предпочтение отдается темам *Кузбасс* и *Россия* в отличие от «Типичного Краснодара», в котором наиболее популярным является обсуждение самих жителей, при этом узел *Краснодарцы* по диаметру в два раза превышает *Кемеровчане*.

В дискурсивном поле, формируемом сообществом «Типичный Краснодар», наблюдается более активное реагирование пользователей на гораздо более широкий круг тематик, чем в дискурсивном поле «Типичный Кемерово». Крупный диаметр узлов *КТТУ, Вакцинация, Автомобилисты, ДТП, Таксист, Видео*, а также *работу России* и *России* свидетельствует о том, что тематики, соответствующие данным объектам, вызывают широкий отклик у пользователей сообщества, а значит являются актуальными или проблемными для данного региона. В дискурсивном поле «Типичный Кемерово», помимо упомянутых выше тем, актуальным стало обсуждение

действий полиции – *Полиция–установила–личность мужчины; Полиция–выясняет–обстоятельства происшествия*. Таким образом, социально-политическая повестка дня, создаваемая публикациями в обоих сообществах, направлена на всестороннее освещение действительности.

Сообщество «Типичный Краснодар» направлено на привлечение максимального количества участников за счет широкого разброса тем, в отличие от «Типичного Кемерово», демонстрирующего сфокусированность на определенном круге тематик.

Анализ предикативных связей в эго-сетях узлов *Кемеровчане* и *Краснодарцы* (рис. 2) показывает, что основными семантическими ролями элементов *Кемеровчане* и *Краснодарцы* являются роль агенса: *Кемеровчане–познают, Краснодарцы–сообщают, интересуются*; бенефактива: *Кемеровчане–смогут* и экпериментера: *Кемеровчане–обеспокоены; Краснодарцы–обеспокоены, расперживались*.

Лингводискурсивный анализ текстов сообщений в датасетах «Топ-100 Кемерово-SVO» и «Топ-1000 Краснодар-SVO» подтвердил, что актуальными социальными проблемами, вызывающими максимальный отклик у Краснодарцев, стали проблемы с вывозом мусора и необъяснимые явления, ассоциирующиеся с повышением уровня опасности в регионе, у Кемеровчан – отсутствие фонтанов. Кемеровчане и Краснодарцы позиционируются как акторы, проявляющие активно-пассивную позицию: они имеют возможность влиять на принятие решений, в частности в определении названий строящихся микрорайонов;

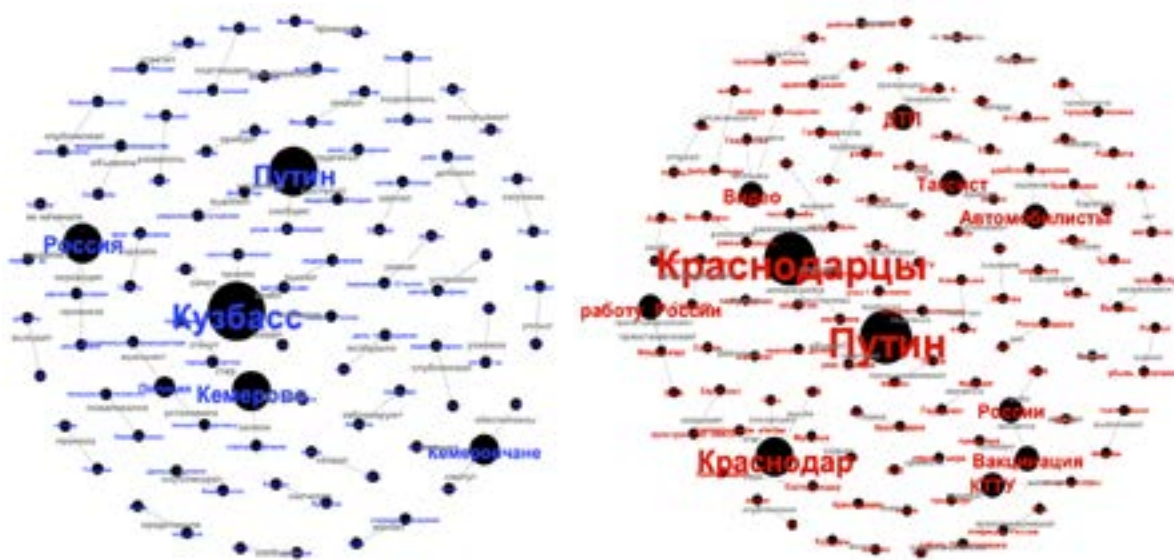


Рис. 1. Семантическое ядро дискурсивных полей, сформированных сообществами «Типичный Кемерово» и «Типичный Краснодар»

Fig. 1. Semantic core of the discursive fields: *Typical Kemerovo* vs. *Typical Krasnodar*

проявляют беспокойность в отношении технического обслуживания города и культуры пользования электрическими самокатами, уровень которой повышается за счет активного участия самих жителей; сообщают о проблемах с вывозом мусора; обращаются за разъяснениями по поводу волнующих их ситуаций и явлений.

Таким образом, внимание местных властей должно быть в первую очередь обращено на решение выявленных проблем: необходимо уделить внимание техническому обслуживанию города (запустить фонтаны; решить проблему с кондиционерами в общественном транспорте). Рекомендованной информационной стратегией для нивелирования напряженности, возникшей в результате распространения тревожных новостей, слухов об экологическом состоянии и острых социально-экономических проблемах в Кузбассе и Краснодаре, может стать публикация новостей о предпринимаемых руководством управленческих решениях и гражданских инициативах или опровержение данных слухов.

Заключение

Составление веб-корпусов сетевых лингвистических данных и визуализация семантических ядер дискурсивных полей, постоянно формирующихся онлайн-сообществами на платформах социальных сетей, позволяет составлять базы данных, которые в дальнейшем могут быть использованы для обработки естественного языка; разработки новых алгоритмов машинного обучения и выявления областей повышенной дискурсивной активности (информационных каскадов), свидетельствующей о критической фазе социального конфликта, актуализированного в коммуникативном взаимодействии; для совершенствования обучения нейросетей, что позволит оперативно выявлять линии разлома между ожиданиями населения

в отношении актуальных социально-политических проблем и реальными действиями в условиях глобальных кризисов. Помимо этого, прикладная значимость подобных исследований заключается в возможности нивелирования областей социальной напряженности за счет дискурсивного регулирования коммуникативных практик пользователей в онлайн-пространстве.

Конфликт интересов: Авторы заявили об отсутствии потенциальных конфликтов интересов в отношении исследования, авторства и / или публикации данной статьи.

Conflict of interests: The authors declared no potential conflicts of interests regarding the research, authorship, and / or publication of this article.

Критерии авторства: Авторы в равной степени участвовали в подготовке и написании статьи.

Contribution: All the authors contributed equally to the study and bear equal responsibility for information published in this article.

Финансирование: Исследование выполнено при финансовой поддержке РФФИ (Отделение гуманитарных и общественных наук) в рамках научного проекта № 20-012-00033 «Лингвистические модели социально-политической коммуникации в online-пространстве: дискурсивные поля, паттерны и гибридная методология анализа сетевых данных» (2020–2022 гг., рук. В. В. Катермина).

Funding: The research was supported by The Russian Foundation for Basic Research (Department of Humanitarian and Social Science), research project no. 20-012-00033: Linguistic models of sociopolitical online communication: discursive fields, patterns, and hybrid methodology of network data analysis (2020–2022; project leader – V. V. Katermina).

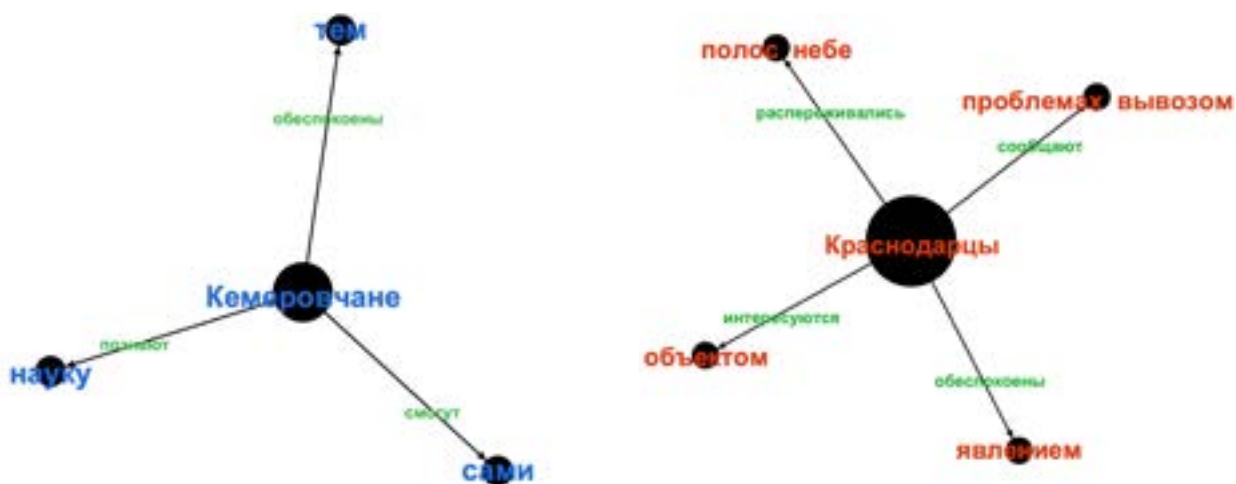


Рис. 2. Визуализация его-сетей узлов Кемеровчане и Краснодарцы
Fig. 2. Visualization of ego-network nodes: Kemerovo and Krasnodar

Литература / References

- Гнедаш А. А. Четвертая волна феминизма: политический дискурс и лидеры мнений в социальной сети Твиттер. *Вестник Российского университета дружбы народов. Серия: Политология*. 2022. № 24. С. 64–89. [Gnedash A. A. The fourth wave of feminism: political discourse and opinion leaders in Twitter. *RUDN Journal of Political Science*, 2022, (24): 64–89. (In Russ.)] <https://doi.org/10.22363/2313-1438-2022-24-1-64-89>
- Рябченко Н. А., Гнедаш А. А., Мальшева О. П., Катермина В. В. Управление политическим контентом в онлайн-пространстве современных государств: как Twitter не позволил Д. Трампу выиграть президентские выборы в 2020 г.? *Политическая наука*. 2021а. № 4. С. 135–160. [Ryabchenko N. A., Gnedash A. A., Malysheva O. P., Katermina V. V. Managing political content in the online space of modern states: how Twitter prevented D. Trump from winning the 2020 presidential election? *Political science*, 2021, (4): 135–160. (In Russ.)] <https://doi.org/10.31249/poln/2021.04.06>
- Рябченко Н. А., Мальшева О. П., Катермина В. В., Гнедаш А. А. Модель коммуникации «говорящий – слушающий» в условиях цифровизации лингвистического поворота: опыт сетевого и лингводискурсивного анализа контента видеохостинга «YouTube». *Политическая лингвистика*. 2021б. № 1. С. 81–94. [Ryabchenko N. A., Malysheva O. P., Katermina V. V., Gnedash A. A. Communication model "Speaker–Listener" in the context of digitalization of the linguistic turn: experience of network and linguodiscursive analysis of YouTube content. *Political Linguistics*, 2021, (1): 81–94. (In Russ.)] https://doi.org/10.26170/1999-2629_2021_01_07
- Aaditi K. A Study about Influence of Social Media in Agriculture Marketing with reference to India. *Advances in Management*, 2020, 13(4): 50–65.
- Chen B., Zhu X., Shui H. Socio-semantic network motifs framework for discourse analysis. *LAK22: 12th International Learning Analytics and Knowledge Conference*, 2022, 500–506. <https://doi.org/10.1145/3506860.3506893>
- Chomsky N. *Syntactic Structures*. Berlin, N. Y.: Mouton de Gruyter, 2002. <https://doi.org/10.1515/9783110218329>
- Ji S., Pan S., Cambria E., Marttinen P., Philip S. Y. A survey on knowledge graphs: representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 33(2): 494–514. <https://doi.org/10.1109/TNNLS.2021.3070843>
- Kalantari K. E., Frrokhi B. M. H., Ghonchepour M. A Critical discourse analysis of Trump's tweets based on Van Dijk model. *Journal of Zabanpazhubi*, 2020, 12(34): 155–156.
- Leong C., Pan S. L., Bahri S., Fauzi A. Social media empowerment in social movements: power activation and power accrual in digital activism. *European Journal of Information Systems*, 2019, 28(2): 173–204. <https://doi.org/10.1080/0960085X.2018.1512944>
- Manganello J. A., Chiang S. C., Cowlin H., Kearney M. D., Massey P. M. HPV and COVID-19 vaccines: social media use, confidence, and intentions among parents living in different community types in the United States. *Journal of Behavioral Medicine*, 2022, 17(15): 1–17. <https://doi.org/10.1007/s10865-022-00316-3>
- McDonald N., Forte A. Powerful privacy norms in social network discourse. *Proceedings of the ACM on Human-Computer Interaction*, 2021, 5(CSCW2): 1–27. <https://doi.org/10.1145/3479565>
- Papatzani E. Encountering everyday racist practices: sociospatial negotiations of immigrant settlement in Athens, Greece. *International journal of urban and regional research*, 2021, 45(1): 61–79. <https://doi.org/10.1111/1468-2427.12984>
- Pasta S. Social network conversations with young authors of online hate speech against migrants. *Cyberhate in the Context of Migrations*. Palgrave Macmillan: Cham, 2022, 187–214. https://doi.org/10.1007/978-3-030-92103-3_8
- Ryabchenko N. A., Gnedash A. A., Malysheva O. P., Shestakova A. A., Nikolaeva M. V. Socio-political content and regional discourse of modern Russia: the issues discussed by citizens in the online space and the solutions offered by the candidates for governor in their election manifestos (intersection points and fault lines). *South Russian Journal of Social Sciences*, 2019, 20(4): 27–48.
- Tesnière L. *Elements of structural syntax*. Amsterdam & Philadelphia, PA: John Benjamins, 2015, 780.
- Wagner C. J., González-Howard M. Studying discourse as social interaction: the potential of social network analysis for discourse studies. *Educational Researcher*, 2018, 47(6): 375–383. <https://doi.org/10.3102/0013189X1877741>
- Wu D. D., Li C. Emotional branding on social media: a cross-cultural discourse analysis of global brands on Twitter and Weibo. *Intercultural Communication in Asia. Education, language and values*, eds. Curtis A., Sussex R. Springer, 2018, 225–240. https://doi.org/10.1007/978-3-319-69995-0_11
- Zhang J., Featherstone J. D., Calabrese C., Wojcieszak M. Effects of fact-checking social media vaccine misinformation on attitudes toward vaccines. *Preventive Medicine*, 2021, 145. <https://doi.org/10.1016/j.ypmed.2020.106408>