

Научная статья

УДК 81'33

doi: 10.17223/19986645/92/5

Оценочный фон риск-коммуникации в сфере здоровья: новости vs Твиттер (на материале текстов о пандемии COVID-19)

Зоя Ивановна Резанова¹, Юлия Евгеньевна Сыпченкова²

^{1,2} *Национальный исследовательский Томский государственный университет,
Томск, Россия*

¹ *rezanovazi@mail.ru*

² *korovina.juliaa@gmail.com*

Аннотация. Представлены результаты проведенного на основе нейронных сетей анализа оценочного фона социально значимого варианта риск-коммуникации – отражения пандемии COVID-19 в российских СМИ и русскоязычном сегменте Твиттера. Охарактеризованы варианты тональности текстов в соотношении с динамикой развития пандемии, фоновым событийным рядом в центре и регионах: новости в целом имели негативную тональность, но новости о коронавирусе были в большей степени нейтральны; твиты характеризовались равным соотношением негативного и нейтрального оценочного фона.

Ключевые слова: пандемия COVID-19, инфодемия, тональность текста, новости, Твиттер, трансформеры

Благодарности: статья подготовлена при финансовой поддержке Российского научного фонда, проект №23-28-01001.

Для цитирования: Резанова З.И., Сыпченкова Ю.Е. Оценочный фон риск-коммуникации в сфере здоровья: новости vs Твиттер (на материале текстов о пандемии COVID-19) // Вестник Томского государственного университета. Филология. 2024. № 92. С. 94–111. doi: 10.17223/19986645/92/5

Original article

doi: 10.17223/19986645/92/5

Sentiment analysis on risk communication about health: News VS Twitter (based on texts about the COVID-19 pandemic)

Zoya I. Rezanova¹, Yulia E. Sypchenkova²

^{1,2} *National Research Tomsk State University, Tomsk, Russian Federation*

¹ *rezanovazi@mail.ru*

² *korovina.juliaa@gmail.com*

Abstract. The article presents the results of a comparative analysis of sentiment presented in Russian-language news discourse and Twitter discourse while reflecting the dynamics of the pandemic and the accompanying event series. The analysis was

performed using NLP methods. The material for the analysis was as follows: texts that in any way reflect the theme of the coronavirus and were published by RIA and TASS agencies in 2020. The volume of materials about Moscow was 3135 texts, and the volume about regions was 1404 texts. To immerse the texts about the coronavirus in the general context of the event radar of this period, the analysis included texts with "incidents", "science", and "culture" topics in the volume of 66,096 texts (31,890 texts reflect events in the regions and 34,206 in Moscow). The sentiment of Twitter texts was analyzed using 178,673 texts. The RuBERT-based blanchefort/rubert-base-cased-sentiment model was used to analyze the sentiment of news texts, and the sismetanin/rubert-rusentitweet model (<https://huggingface.co/sismetanin/rubert-rusentitweet>) was used to analyze Twitter texts. The results of the sentiment analysis were connected with official statistics on the development of the pandemic during this period (<https://yandex.cloud/ru/marketplace/products/yandex/coronavirus-dashboard-and-data>) and information on the chronicle of events (<https://ria.ru/20210305/koronavirus-1599707836.html>). The analysis showed that news about the coronavirus had a mostly neutral tone, while the tone of the general news flow was negative. No changes in the sentiment were found in Twitter texts: tweets of negative and neutral sentiment were equally distributed almost throughout the whole period under study. What unites the two discourses is the tendency to minimize the number of texts carrying positive sentiments. When analyzing news texts about the coronavirus, we identified several dates on which the negative sentiment was out of the general distribution. The increase in the negative sentiment on certain dates in June and November is connected with the peak of disease incidence in the regions – July and the beginning of the second wave of disease growth in 2020 in November and a new wave of restrictive measures. As the analysis showed, the sentiment of the news flow was influenced by the statistics on the spread of the coronavirus and the government's response to the situation by introducing regulatory measures that affected almost all aspects of social and private life in the country. We also documented days when the level of positive sentiment in news content increased in response to a decrease in disease incidence and reports of vaccine development. In general, the findings are consistent with research conducted on English-language news articles from COVID-19-affected countries around the world.

Keywords: COVID-19 pandemic, infodemic, sentiment analysis, news, Twitter, transformers

Acknowledgements: The study is supported by the Russian Science Foundation. Project No. №23-28-01001.

For citation: Rezanova, Z.I. & Sypchenkova, Yu.E. (2024) Sentiment analysis on risk communication about health: News VS Twitter (based on texts about the COVID-19 pandemic). *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology*. 92. pp. 94–111. (In Russian). doi: 10.17223/19986645/92/5

Пандемия новой коронавирусной инфекции COVID-19, как уже было неоднократно отмечено, явилась причиной значимых трансформаций в личной и социальной жизни людей во всех уголках земного шара. Отражением пандемии в сфере коммуникаций стала инфодемия – «массированное распространение (в том числе по каналам СМИ и соцмедиа) во время некоторой эпидемии избыточного объема информации о ней (в том числе искаженной и недостоверной)» [1]. Вместе с тем этот тип формирования и распространения информационных потоков может быть охарактеризован как один из

вариантов риск-коммуникации, предметом которой являются ситуации, связанные со значительными рисками в различных сферах общественной и частной жизни человека. Подобная тематика неизбежно включает в смысловое пространство дискурсов риск-коммуникации широкий спектр эмоциональных оценок происходящего разными субъектами, включенными в данные процессы.

Аналитики инфодемии подчеркивают наличие разных акторов формирования информационных потоков, отражающих кризисные явления в сфере здоровья, и разнонаправленность интересов субъектов коммуникации. В систему информационного сопровождения – информирования и интерпретации – вовлекаются субъекты с неравным уровнем компетенций в различных сферах медицины и общественной жизни, связанных с пандемией. Отметим, что пандемия новой коронавирусной инфекции COVID-19 интерпретируется и как медицинский феномен, и как глобальное социальное явление, в котором неизученный тип вируса обусловил не только сложное протекание заболевания, чрезвычайно быстро распространяющегося, но и как следствие – множественные социальные трансформации. Сложность данного феномена, его многоакторность, разноаспектность отражаются в гетерогенной природе информационных потоков инфодемии.

Привлечение внимания научного сообщества к инфодемии уже на ранних стадиях ее формирования было обусловлено осознанием острой социальной значимости негативных аспектов инфодемии: распространения недостоверных слухов, смещения эмоционального отношения к феномену в целом в сферу негативной оценки, что в конечном итоге как неоднократно отмечалось, способствовало порождению панических настроений. В результате уже в начале 2020 года были проведены исследования тематической фокусировки информационных потоков пандемии и их эмоционально-оценочного фона, представленные в значительном количестве публикаций, отражающих аспекты проявления инфодемии в разных странах мира. При этом в фокусе внимания были социальные сети, прежде всего Твиттер, исследования текстов которого значительно превышают количество обращений к другим социальным сетям, о чем свидетельствуют, например, данные сайта <https://arxiv.org>, на котором оперативно располагаются публикации по актуальным темам, а также результаты метаанализа статей, посвященных медийному отражению COVID-19 и других инфекционных заболеваний [2].

Однако, отмечая отсутствие четких границ между социальными и основными СМИ в современном медийном пространстве, исследователи подчеркивают, что дискурс социальных сетей во многом отражает то, что происходит в новостях [3]. Именно СМИ запускают информационные потоки, которые становятся триггерами внутрисетевого обмена информацией, а также являются объектами оценочных суждений. Высокий уровень значимости СМИ в поддержании глобальных общественных настроений и психического здоровья в условиях продолжающегося кризиса, связанного с COVID-19, подчеркивают К. Yakunin и соавторы [4]. Информирование о пандемии в

новостном контенте в разных странах было предметом анализа в ряде статей, в некоторых из которых приводятся данные о медийном отражении COVID-19 в Российской Федерации в сравнении с другими странами (например, [4, 5]).

При этом исследователями как социальных сетей, так и новостного дискурса получены неоднозначные данные. Так, например, А. Faheem с соавторами [6] на материале англоязычных СМИ нескольких стран утверждают, что около 52% заголовков новостей выражали негативные оценки и только 30% – положительные, тогда как 18% были нейтральными. А. Chakraborty и S. Bose на основе анализа полных текстов новостей в нескольких странах также фиксируют преобладание негативного оценочного фона в новостном контенте [4], выявляя, что наиболее нейтральный тон характерен для новостей в России, преимущественно негативный – во Франции, позитивный – в Италии. Эти наблюдения были подтверждены в работе А.С. Панфиловой и Д.В. Ушакова [7]. А. Chakraborty и S. Bose отмечают влияние смежных социальных факторов (потеря работы, проблемы рабочих мигрантов и под.) на смещение эмоционального фона новостей к негативному полюсу, а также прослеживают наличие корреляций между статистикой заболеваний и оценочным фоном новостей [4]. А. Ashima Yadav выявил тематическую обусловленность варьирования эмоционального фона обсуждения пандемии в англоязычном сегменте Твиттера: положительные эмоции были выражены в твитах, связанных с работниками на переднем крае, развлечениями, мотивацией и качественным проведением времени с семьей. Негативные настроения – с социально-экономическими факторами, такими как расовая несправедливость, уровень безработицы, фейковые новости и смертность [8].

При наличии большого количества статей, в которых представлены результаты исследований тематического, оценочного отражения пандемии новой коронавирусной инфекции COVID-19 в Твиттере и в новостном контенте в разных странах, сравнительные исследования единичны, исследования сегмента социальных сетей и СМИ в РФ в данном аспекте отсутствуют. Отметим в этом случае глубокий анализ соотношения новостного и сетевого контента, построенного на материале 172 основных интернет-источников в 11 странах, отражающих пандемию в работе [9].

На основании изучения данных работ и российских социологических исследований ([1, 10, 11] и др.) нами была выдвинута гипотеза о возможном разрыве в фокусировках событийного ряда и в сопровождающем эмоциональном фоне отражения пандемии в дискурсах социальных сетей и новостей, что, по нашему предположению, могло стать потенциальным источником негативных оценок в социальных сетях, стимулировать распространение страхов, слухов, основанных на недоверии к официально распространяемой информации. Для проверки гипотезы на первом этапе нами было проведено исследование тематического и фреймового моделирования ситуации COVID-19 в русскоязычном сегменте Твиттера и на материалах новостей информационных агентств РИА и ТАСС [12, 13], в том числе в соотношении

с динамикой развития пандемии и сопровождающих ее событий [14] с использованием методов автоматического анализа текстов. Однако проведенный тематический и фреймовый анализ новостей и твитов, связанных с возникновением, распространением, социальными регулирующими мероприятиями со стороны государства и их последствиями, не выявил значительных разрывов в отражении пандемии в двух типах дискурсов. Также анализ показал, что повышение активности в отражении тем в новостях соотносилось с динамикой развития пандемии.

На основе результатов первого этапа исследования, был проведен сентимент-анализ текстов новостей и Твиттера.

Цель данной статьи – представление сравнительного анализа эмоционального фона русскоязычного новостного дискурса и дискурса социальной сети Твиттер в отражении динамики пандемии и сопровождающего ее событийного ряда с использованием методов автоматического анализа текстов.

Актуальность такой фокусировки исследования определяется необходимостью выявления того, есть ли «разрывы» в эмоционально-оценочном представлении одного и того же событийного ряда в разных потоках риск-коммуникации, обусловленных разнонаправленностью интересов ее субъектов. Представляется значимым сравнение транслируемого эмоционально-оценочного отношения к пандемии в новостном дискурсе, институциональном, представляющем интересы государства, и личностном, персональном, дающем некий усредненный срез интерпретаций непосредственно событий и новостного потока участниками сетевого сообщества в соответствии с направленностью личности и индивидуальным опытом. Важно также подчеркнуть, что применяемый метод позволяет выявить не только явно выраженные оценки, но и те, которые обозначаются соотношенными в тексте лексемами, оцениваемыми носителями русского языка как потенциально несущими положительную, нейтральную или отрицательную информацию. В настоящее время, когда цикл развития инфодемии завершен, проведенный анализ с применением методов автоматического анализа к значительному объему текстов позволяет получить объективные данные о роли СМИ в формировании эмоционально-оценочного отношения аудитории к кризисным ситуациям в РФ с учетом локально-административной дифференциации, что может быть использовано при выработке медийных стратегий в отражении кризисных ситуаций в сфере здоровья в будущем.

Материал и методы

Для выявления тональности текста в настоящее время используются методы, основанные на трех подходах (см., например, обзор применяемых методов анализа тональности текстов в работе [4]). Кратко охарактеризуем наиболее часто реализуемые методы. В рамках лингвистико-инженерного подхода сформированы *методы, основанные на применении словарей и правил*. Для анализа русскоязычных текстов созданы словари RuSentiLex [15],

RuSentiFrames [16], КартаСловСент [17], LINIS Crowd [18]. Однако в результате апробации этих методов обнаружилось и проблемные места в их использовании – прежде всего, невозможность описать лексическую и грамматическую структуру естественного языка на основе ограниченного набора словарей и правил. *Методы классического машинного обучения* (метод опорных векторов, случайный лес, регрессия, наивный байесовский классификатор) преодолевают эти ограничения, автоматически выделяя признаки, которые влияют на тональность. На этапе представления анализируемого текста набором признаков, например «мешком слов», дополнительно может учитываться грамматическая и лексическая характеристика лексем, пунктуация, наличие/отсутствие эмодзи. Однако создание репрезентативного набора таких признаков для обучения и составление датасета является трудоемкой задачей, так же как и создание словаря и алгоритмов. Методы машинного обучения с учителем весьма эффективны, но только при наличии качественных размеченных данных и обширного словаря предметной области. Проблема существующих датасетов заключается в том, что они размечены не единообразно: построенная модель плохо переносится на другую предметную область. Тем не менее для анализа корпуса текстов на русском языке на материале Твиттера существует датасет общего назначения (об этом см.: [19]).

Ограничения в применении методов классического машинного обучения могут быть нивелированы использованием *нейросетевого подхода, предобученных нейронных сетей*. В настоящее время SOTA (state-of-the-art) моделями, показывающими лучший результат при решении задач NLP 9 Автоматического анализа языка, являются модели на основе архитектуры трансформер [20].

Для русского языка разработана модель RuBERT [21], для которой с применением трансфера была адаптирована мультиязыковая модель BERT. Для анализа тональности созданы модели на основе RuBERT и RuBERT Conversational (<https://huggingface.co/DeepPavlov/rubert-base-cased-conversationa>). Для модели RuBERT использовались статьи Википедии и новостные тексты (какие именно, в источнике не уточняется), для RuBERT Conversational применялись корпуса «OpenSubtitles» [22] и «Тайга» [23], тексты блогов Dirty и Пикабу. Модели и описание к ним можно найти на сайте платформы Hugging-Face (<https://huggingface.co/>).

При анализе тональности новостных текстов нами была использована модель blanchefort/rubert-base-cased-sentiment, которая была построена на основе RuBERT Conversational с применением следующих датасетов: датасет, содержащий отзывы о медучреждениях (https://github.com/blanchefort/datasets/tree/master/medical_comments), RuTweetCorp [24], RuReviews [25], RuSentiment [26].

Анализ тональности новостных текстов проведен на текстах, которые каким-либо образом отражают тему коронавируса и были опубликованы агентствами РИА и ТАСС в 2020 г. Корпус включает метаинформацию из

текста новости – дату и место публикации (МОСКВА, 1 янв – РИА Новости). Объем материалов о Москве составил 3 135 текстов, о регионах – 1 404 текста. Распределение текстов о коронавирусе по датам и по отнесенности к Москве и регионам показано на рис. 1. На рис. 1–3 ось Y – количество текстов, опубликованных за день, ось X – временная шкала.

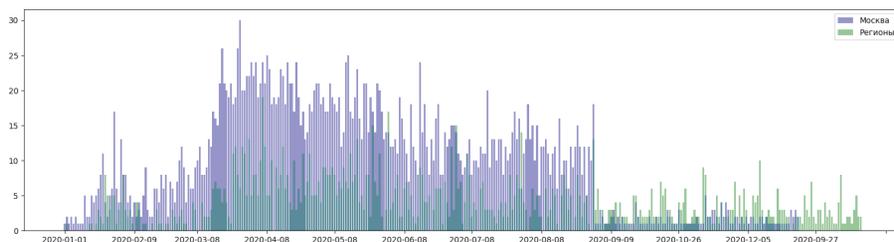


Рис. 1. Распределение новостных текстов о коронавирусе

Как можно видеть на рисунке, датасет не сбалансирован, что связано со спецификой работы программы для скачивания данных. В связи с этим на этапе анализа тональности для выявления дат, в которые количество положительных, негативных или же нейтральных текстов выделялось из общей выборки, нами была применена стандартизированная оценка (z-score).

Для погружения текстов о коронавирусе в общий контекст событийного ряда этого периода и последующего сравнения нами был использован расширенный материал: в анализ были включены тексты агентств РИА и ТАСС из рубрик «происшествия», «наука» и «культура» в объеме 66 096 текстов, из которых 31 890 текстов отражают события в регионах, 34 206 – в Москве. Распределение текстов расширенного датасета по датам и отнесению к Москве и регионам представлено на рис. 2.

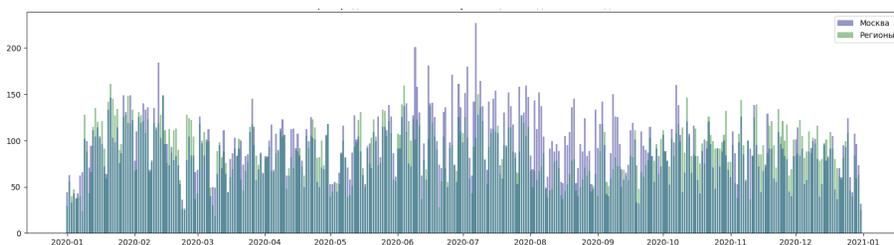


Рис. 2. Распределение новостных текстов расширенной тематики за 2020 г.

Как видно на графике, расширенный датасет более сбалансирован и характеризуется меньшей вариативностью, что позволяет опираться на него с большей уверенностью и продуктивно использовать его в дальнейшем.

При анализе тональности текстов Твиттера мы использовали модель **sismetanin/rubert-rusentitweet** (<https://huggingface.co/sismetanin/rubert->

rusentitweet). Она была получена путем обучения модели RuBERT на датасете RuSentiTweet [27], который содержит тексты, опубликованные с января по декабрь 2020 г.¹

Анализ тональности текстов Твиттера проходил на материале объемом 178 673 текста, который так же, как и датасет RuSentiTweet, охватывает 2020 г., что отражено на рис. 3.

В отличие от новостных текстов тексты Твиттера не содержат привязку к месту, поэтому их дифференциация относительно смысловой связанности с Москвой и регионами не была проведена.

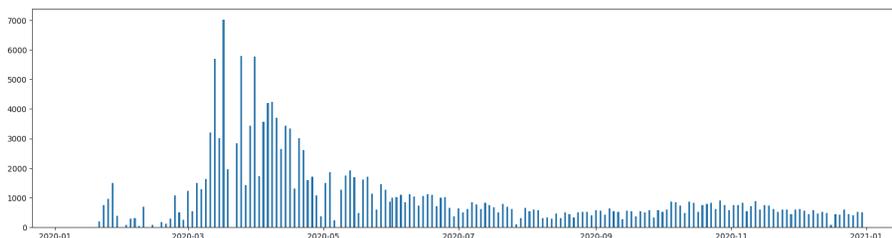


Рис. 3. Распределение текстов Твиттера о коронавирусе

В материале на основе текстов Твиттера также наблюдается несбалансированность: явный пик публикаций с марта по май 2020 г. Некоторые дни остались недоступными для анализа, что опять же связано со спецификой работы программы для скачивания данных в современных условиях. При анализе тональности текстов Твиттера нами также использовалась стандартизированная оценка.

Для решения поставленных задач мы соотнесли полученные с применением обозначенных методов sentiment-анализа данные о преобладающей тональности текстов новостей и Твиттера за каждый день 2020 г. (где это было доступно). Затем сравнили полученные результаты sentiment-анализа с данными официальной статистики развития пандемии за этот период на сервисе Yandex DataLens. Коронавирус. Дашборд (<https://yandex.cloud/ru/marketplace/products/yandex/coronavirus-dashboard-and-data>) и информацией о хронике развития на портале РИА (<https://ria.ru/20210305/koronavirus-1599707836.html>).

Результаты и обсуждение

С применением стандартизированной оценки были выявлены даты, когда количество текстов определенной тональности было выше среднего значения. Результаты анализа визуализированы на рис. 4.

¹ 4 марта 2022 г. Роскомнадзор заблокировал «Твиттер» на территории РФ.

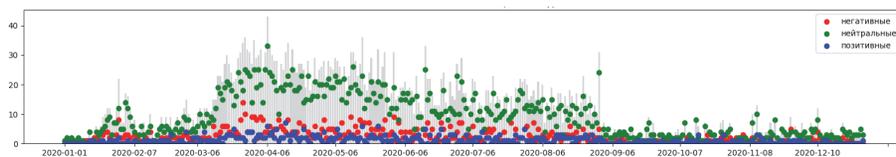


Рис. 4. Тональность новостей о коронавирусе за 2020 г.

Как можно видеть на рис. 4, эмоциональный фон новостей о коронавирусе был в большей степени нейтральным. Количество текстов позитивной тональности оказалось минимальным (отметим, что непосредственное количественное преобладание, отраженное на графике, не является информативным, так как отражает преобладание проанализированных текстов. Значимым является соотношение тональности новостей в каждый фиксируемый день анализируемого периода – взаимное расположение маркированных цветом точек).

При дифференцированном рассмотрении новостей о Москве и регионах нейтральность их эмоционально-оценочного фона новостей сохранялась, что можно видеть на рис. 5, 6, на которых также отчетливо визуализируется минимальность текстов с позитивной направленностью.

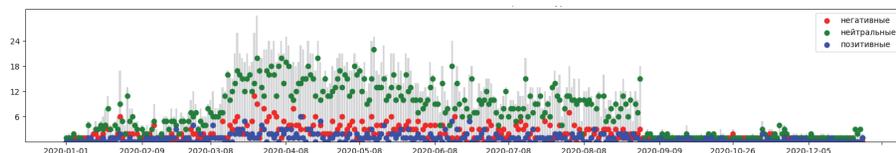


Рис. 5. Тональность новостей о коронавирусе, касающихся Москвы

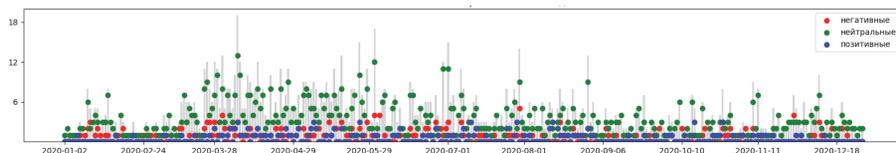


Рис. 6. Тональность новостей о коронавирусе, касающихся регионов

Полученные нами данные о доминирующем sentimentе соотносятся с опубликованными ранее выводами о преимущественном нейтральном фоне новостей о коронавирусе в России (А. Chakraborty и S. Bose [4] и А.С. Панфилова и Д.В. Ушакова [7]). По мнению авторов, это отличает sentiment новостного контента в РФ от эмоционального фона новостей о COVID-19 в других странах, сравнимыми с Россией по уровню заболеваемости.

Далее мы соотнесли полученные результаты sentiment-анализа текстов новостей с данными о развитии пандемии в стране в этот период. Приведем график, отражающий динамику заражений и выздоровлений на территории России за 2020 г. (рис. 7).

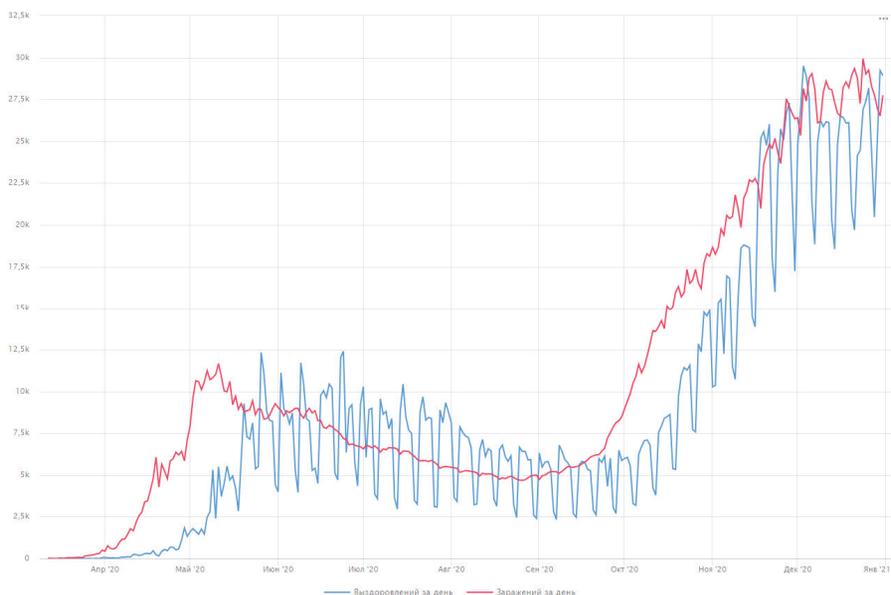


Рис. 7. Динамика заражений и выздоровлений за 2020 г. Yandex DataLens. Коронавирус. Дашборд и данные (дата обращения: 21.05.2024)

На графике виден пик заболеваемости в мае, постепенное снижение ее уровня в летний период и новая, вторая волна, начинающаяся в сентябре. Сравнение двух графиков свидетельствует об отсутствии прямого соотношения роста заболеваемости и изменения эмоционального тона отражения пандемии в текстах новостей на протяжении всего периода.

Также не находит непосредственного отражения динамика развития пандемии и в смене оценочного фона твитов, о чем свидетельствуют результаты проведенного сентимент-анализа текстов Твитера за 2020 г., в которых также наблюдается равное соотношение твитов с нейтральной и негативной эмоциональной оценкой на протяжении всего рассматриваемого периода (данные визуализированы на рис. 8). Такое соотношение преобладающего сентимента отличает тексты русскоязычного сегмента Твиттера, от, например, англоязычного американского. Авторы исследования (Hung M. и др. [28]) отмечают, что в целом настроения, связанные с пандемией, были в большей степени позитивными, что отражало надежду людей на решение беспрецедентного общественного кризиса в сфере здоровья; позитивный фон формировался прежде всего за счет выражения благодарности врачам и административным органам, обеспечивавшим поддержку уязвимым слоям населения. Такая же тенденция отражена и в статье Yin H. с соавторами [29], которые на основе анализа 13 миллионов англоязычных твитов, связанных с COVID-19, обнаружили, что положительных настроений было выражено больше, чем отрицательных. Авторы также отметили тематическую поляризацию сентимента твитов.

Очевидной причиной расхождения полученных нами результатов и представленных в приведенных исследованиях могут быть, наряду с социально-политическими факторами, различия в методике анализа: авторы использовали бинарную классификацию, мы – тернарную. Второе отличие нашего анализа – соотнесение данных о доминирующем сентименте не с тематическим отражением в контенте двух дискурсов, а с динамикой развития пандемии и значимыми событиями государственного реагирования на пандемию.

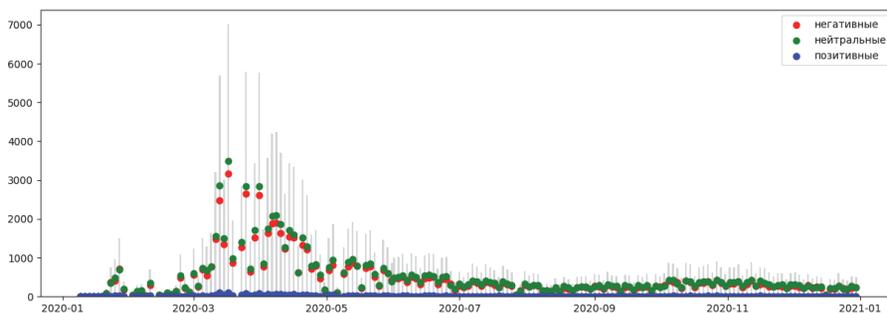


Рис. 8. Тональность текстов о коронавирусе Твиттера за 2020 г.

Отличительной особенностью текстов твитов в исследуемый период является и отсутствие выделенности по «всплескам» какого-либо оценочного фона в отдельные даты, которые наблюдаются в текстах новостей. Наличие такой оценочной дифференциации позволило нам в новостном контенте РИА и ТАСС определить дни, в которых новости выделялись доминирующим оценочным фоном в соотношении с общим новостным потоком.

В общем эмоциональном фоне повышением отрицательных оценок были отмечены следующие даты: в новостном потоке, отражающем события, связанные Москвой, это 26 марта, 27 марта, 30 марта, 11 апреля; в новостях о регионах – 31 марта, 17 мая, 1 июня, 3 июня, 30 июля, 27 ноября.

Мы сравнили эти данные с хроникой событий, связанных с пандемией (данные портала (<https://ria.ru/20210305/koronavirus-1599707836.html>)).

Повышение отрицательного эмоционального фона новостей, отражающих события в марте и апреле в Москве и в марте в регионах, может быть связано с тем, что в этот период усиливаются ограничения на передвижения, прежде всего международные, 19 марта был введен режим самоизоляции, с 23 марта Россия приостановила авиасообщение со всеми странами мира, 30 марта было ограничено пересечение государственной границы. Весь апрель и первая половина мая были объявлены нерабочими на фоне активного роста количества заболеваний.

Если летом в Москве наблюдалось уменьшение количества зараженных, то в регионах оно только росло – вторая волна пандемии в регионах впоследствии оказалась более суровой. Приведем данные о развитии пандемии в регионах на рис. 9.

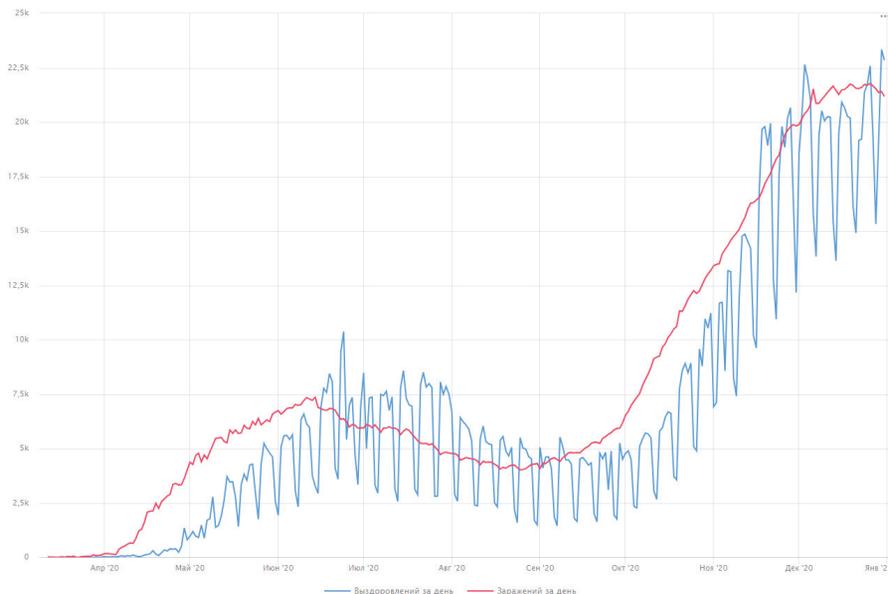


Рис. 9. Динамика заражений и выздоровлений в регионах за 2020 г. Yandex DataLens. Коронавирус. Дашборд и данные (дата обращения: 21.05.2024)

Повышение негативного фона в отдельные даты в июне и ноябре в новостях соотносится с пиком заболеваемости в регионах – в июле и в начале второй волны роста заболеваний в 2020 г. в ноябре и новой волной ограничительных мер. Как видим, на тональность новостного потока влияли статистика распространения коронавируса и реакция государства на ситуацию введением регулирующих мер, которые касались практически всех аспектов социальной и частной жизни человека в стране.

Если новости, связанные с коронавирусом, в основном имеют нейтральный фон, то, как показано на рис. 10, 11, общий фон новостного потока негативен, что может быть связано с отрицательным воздействием коронавируса на все сферы жизни. Мы предполагаем, что в текстах федеральных российских СМИ отразилась общая тенденция, отмеченная исследователями при анализе англоязычного новостного контента.

Нами также были зафиксированы дни, когда рос уровень положительной тональности: в новостном контенте о Москве – 22 марта, 14 апреля, 22 апреля, 15 июня, 22 июня; в новостях, касающихся регионов, – 6 апреля, 7 мая, 22 мая, 10 июля, 29 июля, 24 августа, 16 октября и 27 октября.

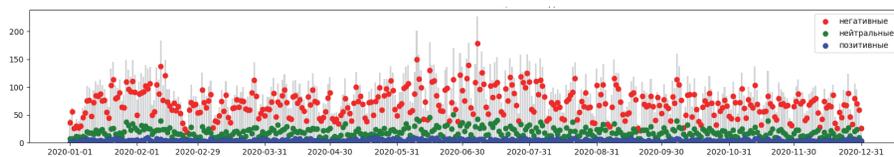


Рис. 10. Тональность новостей за 2020 г., касающихся Москвы

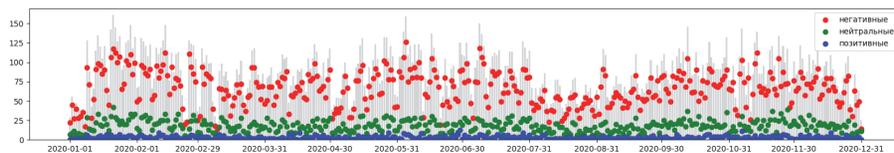


Рис. 11. Тональность новостей за 2020 г., касающихся регионов

С одной стороны, июнь и июль в Москве характеризовались резким спадом числа заболевших, в то время как в регионах росла заболеваемость, что нашло отражение, как было показано, в росте сообщений с отрицательной тональностью. В то же время значимыми для 2020 г. событиями стали регистрация вакцин «Спутник V» в августе и «ЭпиВакКорона» в октябре; к 30 сентября завершилась третья стадия клинических испытаний вакцины «Спутник V».

Выводы и обсуждение

Исходя из признания важной роли новостного потока в формировании информационных и интерпретационных потоков, отражающих кризисные социальные явления, мы поставили в центр обсуждения новости, дифференцированно рассмотрев информационный поток, формируемый федеральными информационными агентствами по двум основаниям: новости о Москве vs новости о регионах; новости о течении пандемии vs новости о пандемии в контексте более широкого тематического ряда и соотнесли полученные результаты sentiment-анализа с объективными данными о динамике заболеваемости и выделенными значимыми социальными событиями; новости vs твиты.

Новости о коронавирусе имели в большей степени нейтральную тональность, в то же время тональность общего новостного потока была негативной. В текстах Твиттера не было обнаружено изменений в оценочном фоне: твитов негативной и нейтральной тональности было поровну практически на протяжении всего исследуемого периода. Объединяет два дискурса и то, что количество текстов, несущих положительную оценку, было минимальным.

При анализе новостных текстов о коронавирусе нами были выделены несколько дат, в которые негативный фон выбивался из общего распределе-

ния. Полагаем, что это было связано с транслируемой статистикой заболеваемости и реакцией властей на текущую ситуацию, как, например, введение и ужесточение ограничительных мер.

В целом полученные данные согласуются с результатами метаисследования, проведенного на материале новостных статей всех стран, затронутых COVID-19 в мире, авторы которого приходят к выводу, что на глобальном уровне новостные настроения в значительной степени зависят от количества новых случаев или смертей, при этом эффект варьируется для разных стран, а также зависит от региональных социально-политических факторов [3]. В нашем анализе последнее положение подтверждается обнаруженными различиями в интерпретации течения пандемии в федеральном центре и регионах в одни и те же временные периоды.

В представленных англоязычных исследованиях были выявлены темы, которые обеспечивают сдвиг эмоционально-оценочного фона новостного потока и текстов Твиттера в позитивную сторону: наряду с темой снижения заболеваемости это темы разработки вакцины, личный героизм врачей, меры поддержки малоимущих [3, 8], а также темы в общем новостном потоке, которые получают преимущественно негативную оценку: потеря работы, проблемы рабочих-мигрантов и политические факторы (например, в Америке – продолжающаяся борьба между президентами США и Китая [3]). В данном исследовании мы, сосредоточившись на других аспектах сентимента информационных потоков, отметили единичные факты соотношения тем и эмоционального фона текстов. Проведенное нами ранее тематическое моделирование новостей и твитов о пандемии [12–14] позволяет продолжить анализ отражения COVID-19 в текстах двух дискурсов через призму эмоционально-оценочного фона наиболее частотных тем. Этот аспект моделирования информационного поля пандемии новой коронавирусной инфекции COVID-19 мы рассматриваем как перспективу исследования.

Список источников

1. *Инфодемия: существующие подходы к анализу паник, фобий, слухов, фейков во время эпидемий и предложения по борьбе с ними*. URL: <https://www.ranepa.ru/documents/monitoring/120-infodemiya.pdf> (дата обращения: 02.05.2024).
2. *Alamood A., Zaidan B.B., Zaidan A.A., Albahri O.S., Mohammed K.I., Malik R.Q., Almahdi E.M., Chyad M.A., Tareq Z., Albahri A.S., Hameed H., Alaa M.* Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review // *Expert Systems With Applications*. 2021. № 167. P. 114–155.
3. *Chen Lyu J., Le H.E., Luli C.* COVID-19 Vaccine-Related Discussion on Twitter: Topic Modeling and Sentiment Analysis // *J. Med Internet Res*. 2021 Jun 29. № 23 (6): e24435. doi: 10.2196/24435
4. *Chakraborty A., Bose S.* Around the world in 60 days: an exploratory study of impact of COVID-19 on online global news sentiment // *Journal of computational social science*. 2020. Vol. 3, № 2. P. 367–400.
5. *Yakunin K., Mukhamediev R.I., Zaitseva E., Levashenko V., Yelis M., Symagulov A., Kuchin Ya., Muhamedijeva E., Aubakirov M., Gopejkenko V.* Mass Media as a Mirror of the COVID-19 Pandemic // *Computation*. 2021. № 9 (12). P. 140. doi: 10.3390/computation9120140

6. *Aslam F., Mumtaz Awan T., Syed J.H., Kashif A., Parveen M.* Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak // Humanities and social sciences communications. 2020. № 7. P. 23. doi: 10.1057/s41599-020-0523-3
7. *Panfilova A.S., Ushakov D.V.* Sentiment Analysis of Russian, Italian, German and French Internet News Content during the Spread of the Coronavirus Pandemic // Psychology-Journal of the Higher School of Economics. 2022. № 19 (3). P. 562–586. URL: <https://pesquisa.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/resource/pt/covidwho-2244035>
8. *Ashima Yadav A.* Language-independent Network to Analyze the Impact of COVID-19 on the World via Sentiment Analysis // ACM Transactions on Internet Technology. 2021. Vol. 22, Is. 1. Article no: 28. P. 1–30. doi: 10.1145/3475867
9. *Krawczyk K., Chelkowski T., Laydon D.J., Mishra S., Xifara D., Gibert B., Flaxman S., Mellan T., Schwämmle V., Röttger R., Hadsund J.T., Bhatt S.* Quantifying Online News Media Coverage of the COVID-19 Pandemic: Text Mining Study and Resource // J. Med. Internet Res. 2021. № 23 (6): e28253. doi: 10.2196/28253
10. *Баринов Д.Н.* Медиавирус страха: особенности репрезентации российскими СМИ пандемии коронавирусной инфекции (COVID-19) в период первой волны (январь–июнь 2020 года) // Социодинамика. 2021. № 2. С. 73–86. doi: 10.25136/2409-7144.2021.2.35066
11. *Серегина Т.Н., Сухова С.К.* Информационные риски в условиях пандемии // Манускрипт. 2021. № 5. С. 940–944.
12. *Резанова З.И., Степаненко А.А.* Фреймирование кризисной ситуации в институциональных и персональных медиа (на материале текстов о COVID-19) // Вестник Томского государственного университета. 2023. № 495. С. 5–13. doi: 10.17223/15617793/495/1
13. *Резанова З.И., Степаненко А.А.* Дискурсивные варианты тематического моделирования пандемии COVID-19 (новостной медиадискурс vs социальные сети) // Вестник Томского государственного университета. Филология. 2023. № 86. С. 84–101. doi: 10.17223/19986645/86/6
14. *Резанова З.И., Коровина Ю.Е.* Риск-коммуникация в сфере здоровья: тематическая фокусировка событийного потока COVID-19 в новостном дискурсе // Вестник Томского государственного университета. Философия. Социология. Политология. 2023. № 76. С. 217–228. doi: 10.17223/1998863X/76/20
15. *Лукашевич Н.В., Левчик А.В.* Создание лексикона оценочных слов русского языка РуСентиЛекс // Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems (OSTIS-2016) : материалы VI междунар. науч.-техн. конф. Минск, 2016. С. 377–382.
16. *Loukachevitch N., Rusnachenko N.* Extracting sentiment attitudes from analytical texts // Computational Linguistics and Intellectual Technologies : proceedings of the International Conference “Dialogue 2018”. Moscow, 2018.
17. Кулагин Д.И. Открытый тональный словарь русского языка КартаСловСент. М., 2021. URL: <https://github.com/dkulagin/kartaslov/tree/master/dataset/kartaslovsent>; <https://www.dialog-21.ru/media/5570/kulagindi026.pdf>
18. LINIS Crowd. URL: <https://linis.hse.ru/news/176051028.html>
19. *Smetanin S.* RuSentiTweet: a sentiment analysis dataset of general domain tweets in Russian // PeerJ Computer Science. 2022. Vol. 8. P. e1039. URL: <https://peerj.com/articles/cs-1039/>
20. *Vaswani A. et al.* Attention is all you need // Advances in neural information processing systems. 2017. Vol. 30. P. 1–11.
21. *Kurатов Y., Arkhipov M.* Adaptation of deep bidirectional multilingual transformers for Russian language // Computational Linguistics and Intellectual Technologies : proceedings of the International Conference “Dialogue 2019”. Moscow, 2019.

22. Lison P., Tiedemann J. OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles // Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016). 2016. P. 923–929.

23. Shavrina T., Shapovalova O. To the methodology of corpus construction for machine learning: «TAIGA» syntax tree corpus and parser // CORPORA 2017 : International conference. Saint-Petersbourg, 2017. P. 78–84.

24. Рубцова Ю. Автоматическое построение и анализ корпуса коротких текстов (постов микроблогов) для задачи разработки и тренировки тонового классификатора // Инженерия знаний и технологии семантического веба. 2012. Т. 1. С. 109–116.

25. RuReviews: An Automatically Annotated SentimentAnalysis Dataset for Product Reviews in Russian. URL: <https://github.com/sismetanin/rureviews> (дата обращения: 02.05.2024).

26. Rogers A. et al. RuSentiment: An enriched sentiment analysis dataset for social media in Russian // Proceedings of the 27th international conference on computational linguistics. 2018. P. 755–763.

27. Smetanin S., Komarov M. Sentiment analysis of product reviews in Russian using convolutional neural networks // 2019 IEEE 21st conference on business informatics (CBI). IEEE, 2019. Vol. 1. P. 482–486. URL: <https://github.com/sismetanin/rureviews>

28. Hung M., Lauren E., Hon E.S., Birmingham W.C., Xu J., Su S., Hon S.D., Park J., Dang P., Lipsky M.S. Social network analysis of COVID-19 sentiments: Application of artificial intelligence // J. Med Internet Res. 2020. Vol. 18, № 22 (8). P. e22590. doi: 10.2196/22590

29. Yin H., Yang S., Li J. Detecting topic and sentiment dynamics due to COVID-19 pandemic using social media // Proceedings of the International Conference on Advanced Data Mining and Applications; International Conference on Advanced Data Mining and Applications; November 12–15, 2020. China, 2020. P. 610–623. URL: <https://arxiv.org/pdf/2007.02304>

References

1. RANEPА. (2024) *Infodemiya: sushchestvuyushchie podkhody k analizu panik, fobiy, slukhov, feykov vo vremya epidemiy i predlozheniya po bor'be s nimi* [Infodemics: existing approaches to the analysis of panics, phobias, rumors, fakes during epidemics and proposals for combating them]. [Online] Available from: <https://www.ranepa.ru/documents/monitoring/120-infodemiya.pdf> (Accessed: 02.05.2024).

2. Alamood, A. et al. (2021) Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. *Expert Systems With Applications*. 167. pp. 114–155.

3. Chen Lyu, J., Le, H.E. & Luli, C. (2021) COVID-19 Vaccine-Related Discussion on Twitter: Topic Modeling and Sentiment Analysis. *J. Med Internet Res*. Jun 29. 23 (6). e24435. doi: 10.2196/24435

4. Chakraborty, A. & Bose, S. (2020) Around the world in 60 days: an exploratory study of impact of COVID-19 on online global news sentiment. *Journal of Computational Social Science*. 3 (2). pp. 367–400.

5. Yakunin, K. Et al. (2021) Mass Media as a Mirror of the COVID-19 Pandemic. *Computation*. 9 (12). p. 140. doi: 10.3390/computation9120140

6. Aslam, F. et al. (2020) Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak. *Humanities and Social Sciences Communications*. 7. p. 23. doi: 10.1057/s41599-020-0523-3

7. Panfilova, A.S. & Ushakov, D.V. (2022) Sentiment Analysis of Russian, Italian, German and French Internet News Content during the Spread of the Coronavirus Pandemic. *Psychology-Journal of the Higher School of Economics*. 19 (3). pp. 562–586. [Online] Available from:

<https://pesquisa.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/resource/pt/covidwho-2244035>

8. Ashima Yadav, A. (2021) Language-independent Network to Analyze the Impact of COVID-19 on the World via Sentiment Analysis. *ACM Transactions on Internet Technology*. 22 (1). Article no: 28. pp. 1–30. doi: 10.1145/3475867

9. Krawczyk, K. Et al. (2021) Quantifying Online News Media Coverage of the COVID-19 Pandemic: Text Mining Study and Resource. *J. Med. Internet Res.* 23 (6). e28253. doi: 10.2196/28253

10. Barinov, D.N. (2021) Media virus of fear: the peculiarities of representation of COVID-19 pandemic by the Russian media during the first wave (January – June 2020). *Sotsiodinamika*. 2. pp. 73–86. (In Russian). doi: 10.25136/2409-7144.2021.2.35066

11. Seregina, T.N. & Sukhova, S.K. (2021) Informatsionnye riski v usloviyakh pandemii [Information risks in the context of a pandemic]. Manuscript. 5. pp. 940–944.

12. Rezanova, Z.I. & Stepanenko, A.A. (2023) Framing a crisis situation in institutional and personal media (based on COVID-19 texts). *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 495. pp. 5–13. (In Russian). doi: 10.17223/15617793/495/1

13. Rezanova, Z.I. & Stepanenko, A.A. (2023) Discursive variants of thematic modeling of COVID-19 (news media discourse VS social networks). *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology*. 86. pp. 84–101. (In Russian). doi: 10.17223/19986645/86/6

14. Rezanova, Z.I. & Sypchenkova, Yu.E. (2023) Risk communication about health: thematic focus of the COVID-19 event flow in news discourse. *Vestnik Tomskogo gosudarstvennogo universiteta. Filosofiya. Sotsiologiya. Politologiya – Tomsk State University Journal of Philosophy, Sociology and Political Science*. 76. pp. 217–228. (In Russian). doi: 10.17223/1998863X/76/20

15. Lukashevich, N.V. & Levchik, A.V. (2016) [Sozdanie leksikona otsenochnykh slov russkogo yazyka RuSentiLex]. *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh sistem = Open Semantic Technologies for Intelligent Systems (OSTIS-2016)*. Proceedings of the International Conference. Minsk. pp. 377–382. (In Russian).

16. Loukachevitch, N. & Rusnachenko, N. (2018) Extracting sentiment attitudes from analytical texts. *Computational Linguistics and Intellectual Technologies: proceedings of the International Conference "Dialogue 2018"*. Moscow.

17. Kulagin, D.I. (2021) *Otkrytyy tonal'nyy slovar' russkogo yazyka KartaSlovSent* [Open tonal dictionary of the Russian language KartaSlovSent]. Moscow. [Online] Available from: <https://github.com/dkulagin/kartaslov/tree/master/dataset/kartaslovsent>; <https://www.dialog-21.ru/media/5570/kulagindi026.pdf>

18. *LINIS Crowd*. [Online] Available from: <https://linis.hse.ru/news/176051028.html>

19. Smetanin, S. (2022) RuSentiTweet: a sentiment analysis dataset of general domain tweets in Russian. *PeerJ Computer Science*. 8. e1039. [Online] Available from: <https://peerj.com/articles/cs-1039/>

20. Vaswani, A. et al. (2017) Attention is all you need. *Advances in Neural Information Processing Systems*. 30. pp. 1–11.

21. Kuratov, Y. & Arkhipov, M. (2019) Adaptation of deep bidirectional multilingual transformers for Russian language. *Computational Linguistics and Intellectual Technologies: proceedings of the International Conference "Dialogue 2019"*. Moscow.

22. Lison, P. & Tiedemann, J. (2016) OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*. pp. 923–929.

23. Shavrina, T. & Shapovalova, O. (2017) To the methodology of corpus construction for machine learning: "TAIGA" syntax tree corpus and parser. *CORPORA 2017: International conference*. Saint Petersburg. pp. 78–84.

24. Rubtsova, Yu. (2012) [Automatic construction and analysis of a corpus of short texts (microblog posts) for the task of developing and training a tone classifier]. *Inzheneriya znaniy i tekhnologii semanticheskogo veba [All-Russian Youth Conference "Knowledge Engineering and Semantic Web Technologies" (KESW-2011)]*. Vol. 1. ITMO. pp. 109–116. (In Russian).

25. *RuReviews: An Automatically Annotated Sentiment Analysis Dataset for Product Reviews in Russian*. [Online] Available from: <https://github.com/sismetanin/rureviews> (Accessed: 02.05.2024).

26. Rogers, A. et al. (2018) RuSentiment: An enriched sentiment analysis dataset for social media in Russian. *Proceedings of the 27th International Conference on Computational Linguistics*. Association for Computational Linguistics. pp. 755–763.

27. Smetanin, S. & Komarov, M. (2019) Sentiment analysis of product reviews in Russian using convolutional neural networks. *2019 IEEE 21st conference on business informatics (CBI)*. IEEE, 2019. Vol. 1. pp. 482–486. [Online] Available from: <https://github.com/sismetanin/rureviews>

28. Hung, M. et al. (2020) Social network analysis of COVID-19 sentiments: Application of artificial intelligence. *J. Med Internet Res*. 18:22 (8). e22590. doi: 10.2196/22590

29. Yin, H., Yang, S. & Li, J. (2020) Detecting topic and sentiment dynamics due to COVID-19 pandemic using social media. *Proceedings of the International Conference on Advanced Data Mining and Applications; International Conference on Advanced Data Mining and Applications*. 12–15 November 2020. China. pp. 610–623. [Online] Available from: <https://arxiv.org/pdf/2007.02304>

Информация об авторах:

Резанова З.И. – д-р филол. наук, проф. кафедры общей, компьютерной и когнитивной лингвистики филологического факультета Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: rezanovazi@mail.ru

Сыпченкова Ю.Е. – лаборант лаборатории лингвистической антропологии филологического факультета Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: korovina.juliaa@gmail.com

Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

Z.I. Rezanova, Dr. Sci (Philology), professor, National Research Tomsk State University (Tomsk, Russian Federation). E-mail: rezanovazi@mail.ru

Yu.E. Syptchenkova, laboratory assistant, Laboratory of Linguistic Anthropology, National Research Tomsk State University (Tomsk, Russian Federation). E-mail: korovina.juliaa@gmail.com

The authors declare no conflicts of interests.

*Статья поступила в редакцию 29.05.2024;
одобрена после рецензирования 15.06.2024; принята к публикации 18.11.2024.*

*The article was submitted 29.05.2024;
approved after reviewing 15.06.2024; accepted for publication 18.11.2024.*