

## КОГНИТИВНАЯ МОДЕЛЬ АНАЛИЗА КОНТЕНТА ЦИФРОВЫХ КОММУНИКАЦИЙ

Пильгун Мария Александровна

Российский государственный социальный университет,  
Москва, Россия

### *Аннотация*

Статья посвящена разработке и апробации когнитивной модели для анализа больших объемов речевых данных и цифровых следов с позиций теории речевой деятельности при интеграции психолингвистических методик и нейросетевых технологий. Представлены результаты исследования восприятия речи акторов социальных медиа для анализа социального напряжения вокруг строительства Рублево-Архангельской линии метро (от ст. м. «Звенигородская» до ст. м. «Липовая аллея»). Материалом для исследования послужили данные микроблогов, блогов, мессенджеров, социальных сетей, видео-хостингов, интернет СМИ, тематических порталов, картографических сервисов и ТВ. Время сбора материала: 01.07.2020 23:59:59 – 24.03.2021 23:59:59. Для анализа контента применялся мультимодальный подход с использованием нейросетевых технологий, текстовый анализ, семантический анализ, контент-анализ, сентимент-анализ, анализ лексических ассоциаций. Результаты исследования показали, что интеграция психолингвистических методик и нейросетевых технологий с позиций теории речевой деятельности позволила построить когнитивную модель анализа социального напряжения в медиaprостранстве, провести исследование анализа мнений, восприятия речи акторов социальных медиа. Анализ данных показал отсутствие социальной напряженности вокруг строительства Рублево-Архангельской линии метро (от ст. м. «Звенигородская» до ст. м. «Липовая аллея»). Контент-анализ, выделение тематической структуры и саммаризация позволили уточнить ключевые темы, определить рейтинг контента; анализ цифровой агрессии, сентимент-анализ, анализ семантической и ассоциативной сетей подтвердили по большей части нейтральное восприятие акторов хода реализации проекта. Анализ лексических ассоциаций дал возможность выявить и интерпретировать имплицитные оценки и мнения акторов, подтверждающие результаты исследования.

**Ключевые слова:** теория речевой деятельности, восприятие речи, психолингвистика, социальные медиа, нейросетевые технологии

### **Введение**

Исследования с использованием данных социальных медиа, адаптация лингвистических и психолингвистических методик (в частности, разработанных в рамках московской психолингвистической школы) к новой эмпирической реальности для анализа Больших данных позволяют анализировать различные типы социальных ситуаций, например, выявлять наличие/отсутствие социальной напряженности, делать прогнозную аналитику и под.

Согласно концепции А.Н. Леонтьева, деятельность – это специфическая форма активности человека по познанию и преобразованию окружающей действительности. Речевая деятельность, как и другие виды деятельности, характеризуется, с одной стороны, предметным мотивом, целенаправленностью, эвристическим характером, а с другой стороны, состоит из нескольких последовательных фаз (ориентировка, планирование, реализация плана, контроль). Адекватное восприятие языковых знаков, также как их усвоение, производство и воспроизводство возможно только при корректном взаимосвязанном функционировании психофизиологических механизмов носителей языка [Леонтьев 1974].

Развитие цифровых коммуникаций позволило адаптировать традиционные психолингвистические методики, разработанные в рамках теории речевой деятельности, для анализа больших объемов цифровых данных.

В психолингвистической традиции представлены многочисленные исследования различных аспектов восприятия речи. Так, особенности субъективного восприятия проанализированы в [Ransom, Dale, et al. 2019], индивидуальные различия и взаимодействие восприятия и речепроизводства [Demuth 2018], восприятие лжи билингвами [Kreyßig, Krautz 2019], восприятие и обработка речи слепых билингвов и монолингвов [Phani, Arulmozi, Shiva 2020].

Использование потенциала нейронных систем для анализа восприятия и производства речи представлено в научных исследованиях уже достаточно давно. См., например, [Hickok, Poeppel 2000; Hickok 2001]. Интеграция поведенческих и нейронных функций в области восприятия речи, результаты современных исследований коркового и подкоркового представления речи, особенности восприятия речи взрослыми и младенцами описаны в работе [Holt, Peelle, et.al. 2021].

Очевидно, что изучение функционирования речевых структур в медиaprостранстве требует междисциплинарного подхода. Показательно, что в последнее время все больше работ ориентированы на интеграцию различных методов исследования (multi-method), так в исследовании Stefan Uhrig использован процессно-ориентированный подход к оценке качества речи, с применением субъективного, поведенческого и нейрофизиологического уровней анализа, что позволило выявить влияние качества передачи речи на обработку информации человеком [Uhrig 2021].

Между тем нейросетевой подход к изучению речевых структур акторов социальных медиа, в частности, для анализа восприятия речи, выявления социального напряжения применяется еще недостаточно широко. В то же время, очевидно, что большие объемы генерируемого пользователями мультимодального контента в виртуальном пространстве требуют автоматизированных методов обработки массивов данных. Рассматривая речевые данные социальных медиа и цифровые следы акторов как отражение реакции онлайн аудитории на происходящие события, психолингвистические методики и нейросетевые технологии могут быть использованы для оценки конфликтных ситуаций, выявления причин конфликта, основных аргументов сторон, а также для прогнозной аналитики социальной напряженности.

*Цель* данной работы: разработка и апробация варианта интеграции психолингвистических методик и нейросетевых технологий для обработки речевых данных и цифровых следов с позиций теории речевой деятельности. В работе представлены результаты исследования контента, генерированного пользователями,

и их цифровых следов для выявления наличия или отсутствия социального напряжения вокруг строительства Рублево-Архангельской линии метро (от ст. м. «Звенигордская» до ст. м. «Липовая аллея») на начальной стадии реализации проекта и определения возможных конфликтов, рисков потенциальной эскалации неактивного развития ситуации.

### Метод

Для анализа контента социальных медиа применялся мультимодальный подход с использованием нейросетевых технологий, текстовый анализ, семантический анализ, контент-анализ, сентимент-анализ, анализ лексических ассоциаций. Подробное описание методики представлено в работах [Kharlamov, Pilgun 2020; Pilgun, Raskhodchikov, Koreneva, Antonova 2022 и др.].

Автоматический анализ коннотации ситуации, описанной речевыми средствами, в настоящий момент, как правило, реализуется с помощью системы правил, методами дистрибутивной семантики и с использованием искусственных нейронных сетей (машинное обучение). Разработка системы правил под конкретную задачу представляет собой достаточно трудоемкий процесс. Методы дистрибутивной семантики, разработанные на материале английского языка, широко известны, в их основе лежит представление о текстах как о множестве несвязанных между собой слов (модель bag-of-words), что при анализе материала языков синтетического типологического класса (например, русского), в котором преобладают синтетические формы выражения грамматических значений, создает существенные ограничения для анализа. Применение большинства искусственных нейронных сетей требует для своего обучения большой по объему обучающей выборки, формирование которой не всегда может быть осуществлено автоматически (например, в области медицины) и потому стоит очень дорого.

Альтернативой существующим подходам к анализу коннотации языковых единиц может служить подход, основанный на нейросетевой парадигме использования нейроподобных элементов с временной суммацией сигналов, который, с одной стороны, работает с речевой моделью, отличной от «bag-of-words», с другой стороны, не требует большой обучающей выборки для формирования этой модели. Дополнительным преимуществом подхода является включение в цикл анализа эксперта (группы экспертов), которые привносят в анализ квалифицированные (хоть и субъективные) знания (в отличие от аутсорсеров, которых обычно привлекают для формирования обучающих выборок) [Kharlamov, Pilgun 2020].

Использование данного подхода позволяет сформировать гибкую методику анализа коннотации речевых структур, основанную на мнении экспертного сообщества. Мультилингвальные модули позволят анализировать большие объемы речевых данных на различных европейских языках (русском, английском, немецком, французском, испанском, итальянском).

Указанный подход позволяет решать разнообразные практические задачи. Так, анализ коннотации речевых данных основан на использовании лексических меток, характеризующих коннотацию текста на шкале «(-)(+)», или на шкале «(0)(+)», позволяющих характеризовать не только тональность речевых данных, но и любую коннотационную принадлежность речевых структур на заданной шкале. Выделенные экспертом маркеры, взвешенные им на принятой шкале, автоматически (в зависимости

от семантической наполненности) ранжируются на шкале «0-100%». Произведение автоматически полученного ранга на вес, заданный экспертом, характеризует вклад определенного маркера в коннотацию анализируемой речевой ситуации, что позволяет оценивать речевую структуру с точки зрения конкретных маркеров или оценивать суммарный вклад всех маркеров в коннотацию этого речевой структуры, то есть решить задачу автоматического коннотационного мониторинга речевых данных.

Автоматическое семантическое ранжирование речевых данных осуществляется с использованием двух алгоритмов: (1) алгоритма формирования однородной частотной сети текста с помощью искусственной нейронной сети на основе нейрореподобных элементов с временной суммацией сигналов и (2) итеративного хопфилдоподобного алгоритма ранжирования вершин сети на шкале «0-100%» [Kharlamov, Pilgun 2020].

Частотная сеть текста строится как множество пар слов, встречающихся в предложениях текста. Вершины сети взвешиваются их частотой встречаемости в тексте. Вес связи пары вершин в сети соответствует частоте встречаемости пар слов в предложениях текста.

В отличие от дистрибутивного представления текста, где вес вершины определяется суммарным нормированным весом соседних вершин, а вес связей пар слов в сети вычисляется на основе расстояний им соответствующих векторов в векторном пространстве, *n*-граммное представление формируется итеративным перевзвешиванием на заданном числе шагов или на основе критерия сходимости процесса ранжирования. То есть лексемы рассматриваются в рамках синтагм заданной (*n*) длины на семантической сети, сформированной на основе анализа текста. Первичная обработка текста, удаляющая часть слов из процесса анализа, не дискредитирует *n*-граммности подхода, так как удаление не является принципиальной частью алгоритма, который показал хорошие результаты и на полных данных, но первичная обработка уменьшает информационный шум, увеличивая робастность подхода [Etikan, Radev 2004].

Таким образом, междисциплинарный подход с использованием нейросетевых технологий представляется наиболее адекватным для решения задач, поставленных в исследовании.

Методика была апробирована в ходе выполнения более 30-ти практических проектов: Анализ социального напряжения вокруг реализации проектов Юго-восточная хорда – Южная рокада (Реконструкция ул. Верхние Поля на участке от ул. Марьинский парк до МКАД). Улично-дорожная сеть АДЦ «Коммунарка». Дорога «Варшавское шоссе – деревня Яковлево». Трасса «Солнцево – Бутово – Варшавское шоссе»; Исследования ситуации, возникшей вокруг строительства участка Коммунарской линии метро на улице Дмитрия Ульянова; Анализ социального напряжения вокруг строительства ст. м. «Ул. 800-летия Москвы» и «Лианозово»; Анализ социального напряжения вокруг проекта Северо-Восточная хорда в информационном поле сети Интернет»; Семантический анализ данных информационной ситуации (пользовательского контента сети Интернет) вокруг строительства первого участка Троицкой линии метро и др.

В данной статье предлагается когнитивная модель анализа социального напряжения в медиaprостранстве, представленная ниже на блок-схеме.



Когнитивная модель анализа социального напряжения в медиапространстве

Модель включает взаимосвязанные процедуры, составляющие несколько блоков:

1. Отбор и фильтрация контента.
  - 1.1. Выделение акторов с различным генезисом.
  - 1.2. Идентификация различных типов акторов для уточнения коммуникативной ситуации.
    - 1.3. Анализ цифровых платформ и формирование их рейтинга.
    - 1.4. Анализ цифровых следов акторов.
  2. Анализ данных.
    - 2.1. Проведение контент-анализа.
    - 2.2. Формирование тематической структуры.
    - 2.3. Проведение саммаризации.
    - 2.4. Выделении ключевых тем на основе результатов процедур 2.1.; 2.2.; 2.3.
    - 2.5. Проведение сентимент-анализа.
    - 2.6. Кластеризация контента.
    - 2.7. Анализ цифровой агрессии.
  3. Интерпретация данных.
    - 3.1. Анализ семантической сети.
      - 3.1.1. Формирование семантической сети.
        - 3.1.1.1. Извлечение ядра семантической сети (номинации с весом связей 98-100).
        - 3.1.1.1.1. Нейросетевой текстовый анализ семантического ядра.
      - 3.2. Анализ ассоциативной сети.
        - 3.2.1. Выполнение ассоциативного поиска.
        - 3.2.2. Построение ассоциативной сети.
        - 3.2.3. Анализ лексических ассоциаций.
    4. Вычисление индексов социального стресса и благополучия.

Эффективность данной модели апробируется на датасете, посвященном реализации проекта Рублево-Архангельской линии метро (от ст. м. «Звенигородская» до ст. м. «Липовая аллея»). Рублево-Архангельская линия метро – будущая линия Московского метрополитена. Она в перспективе должна соединить ММДЦ «Москва-Сити» и международный финансовый центр «Рублево-Архангельское», обеспечить транспортное обслуживание территории Рублево-Архангельское и максимального количества проживающих и работающих на территории районов Митино, Строгино и Хорошёво-Мнёвники, сокращая время в пути в 1,5 раза.

#### **Данные**

Материалом для исследования послужили данные социальных сетей, микроблогов, блогов, мессенджеров, видео-хостингов, интернет-СМИ, тематических порталов, картографических сервисов и ТВ, полученные в результате сплошной выборки. Время сбора материала: 01.07.2020 23:59:59 – 24.03.2021 23:59:59 (см. табл. 1).

*Таблица №1*

#### **Количественные характеристики данных**

Количество сообщений:	2709
Количество активных акторов:	771
Количество знаков:	4 384 470

## Результаты

### Общая характеристика контента

Контент, посвященный реализации проекта Рублево-Архангельской линии метро, представлен по большей части в социальных сетях (2 205 599), интернет СМИ (807 957), микроблогах (217 454) и мессенджерах (159 188) (Рис. 1).

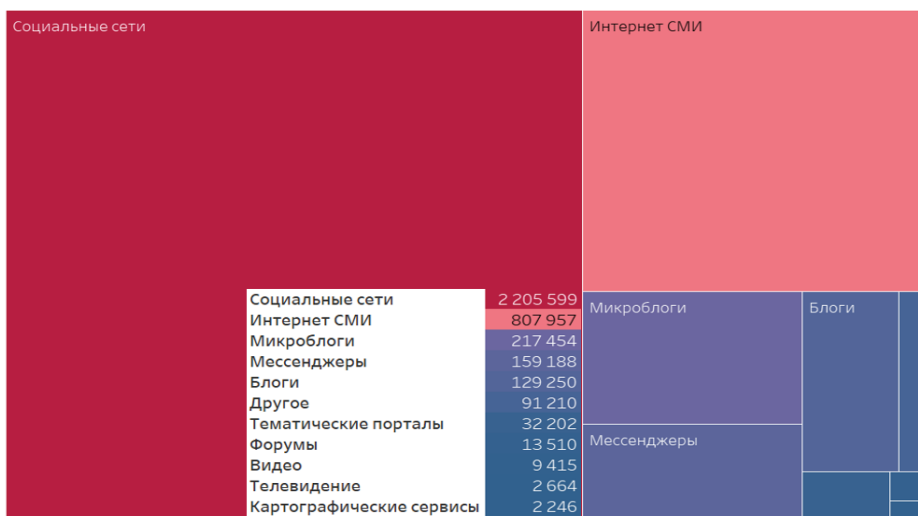


Рис. 1. Рейтинг типов источников.

Лидером по количеству генерированных сообщений среди цифровых источников была социальная сеть ВКонтакте (1 926 075), Твиттер<sup>1</sup> занимал второе место (217 454), третье место – Телеграм (153 188) (Рис. 2).

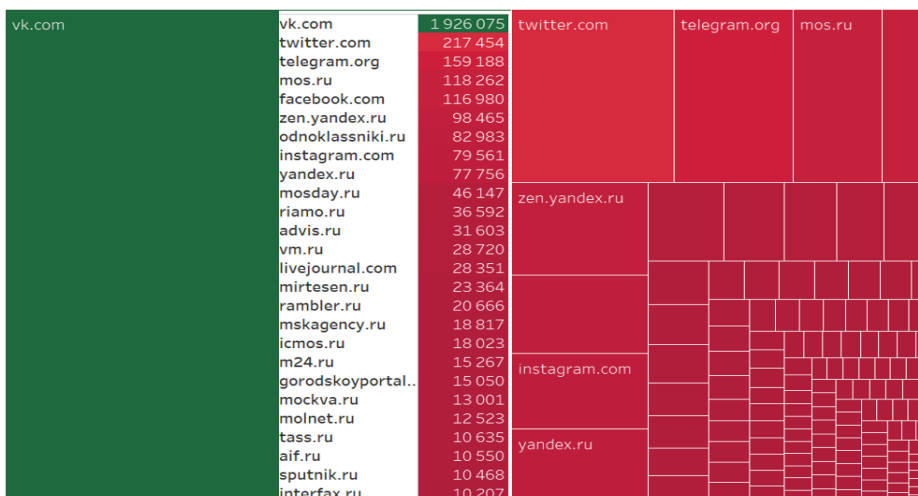


Рис. 2. Рейтинг цифровых источников.

<sup>1</sup> Twitter заблокирован в России с 4 марта 2022 г. на основании требования Генпрокуратуры.

Пик динамики генерирования сообщений приходится на 21.09.2020 (общее количество сообщений – 197, уникальных сообщений – 134), связан с официальной информацией о начале строительства Рублево-Архангельской линии метро. Также фиксируется пик 19.11.2020 (общее количество сообщений – 191, уникальных сообщений – 118), который вызван аналогичным сообщением о том, что строительство Рублево-Архангельской линии войдет в активную фазу после 2022 года (Рис. 3).



Рис. 3. Динамика общего числа и уникальных сообщений.

Динамика активности акторов также показывает топовые значения 21.09.2020 (108) и 19.11.2020 (86) (Рис. 4).

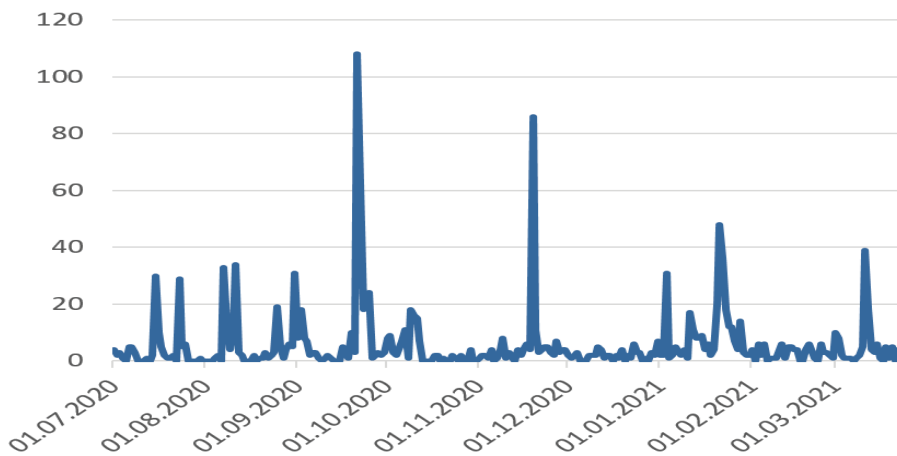


Рис. 4. Динамика активности пользователей.



Динамика просмотров контента, посвященного строительству Рублево-Архангельской линии метро (от ст. м. «Звенигородская» до ст. м. «Липовая аллея») имеет только один пик, который также приходится на 21.09.2020 (1 210 336) и связан с появлением информации в официальных источниках о начале строительства Рублево-Архангельской линии метро (Рис. 5).

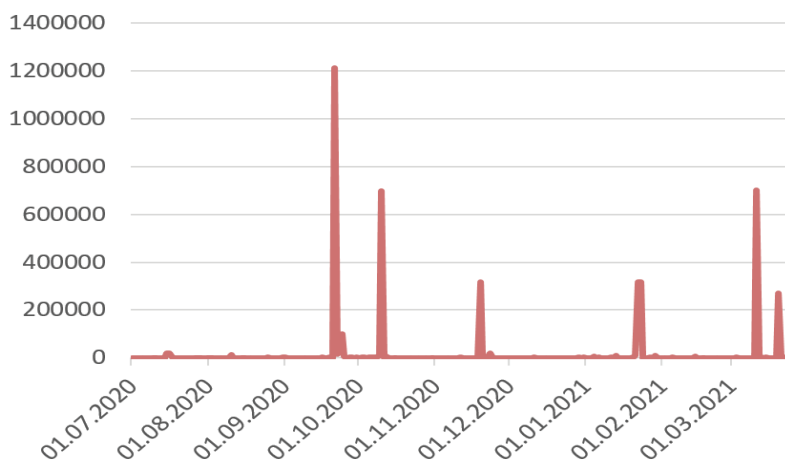


Рис. 5. Динамика просмотров

### Ключевые темы

В результате проведения контент-анализа, исследования тематической структуры, анализа результатов саммаризации и семантической сети были выделены ключевые темы контента, посвященного строительству Рублево-Архангельской линии метро.

Тематическая структура общей базы данных отражает набор ключевых понятий, позволяет выделить тематику, актуальную для жителей при обсуждении реализации проекта Рублево-Архангельской линии метро (Рис. 6).

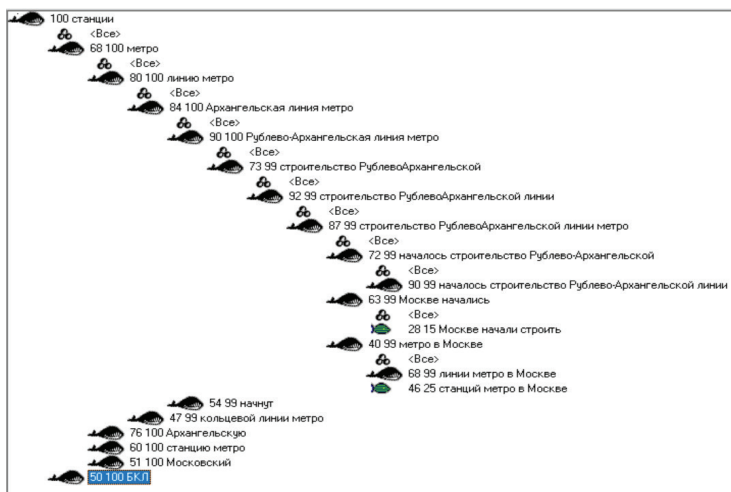


Рис. 6. Тематическая структура общей базы данных.

Анализ тематической структуры негативного кластера показывает, что максимальные веса связей получили номинации, связанные с вопросом о наименовании станций, что свидетельствует о негативной оценке жителей (Рис. 7.)

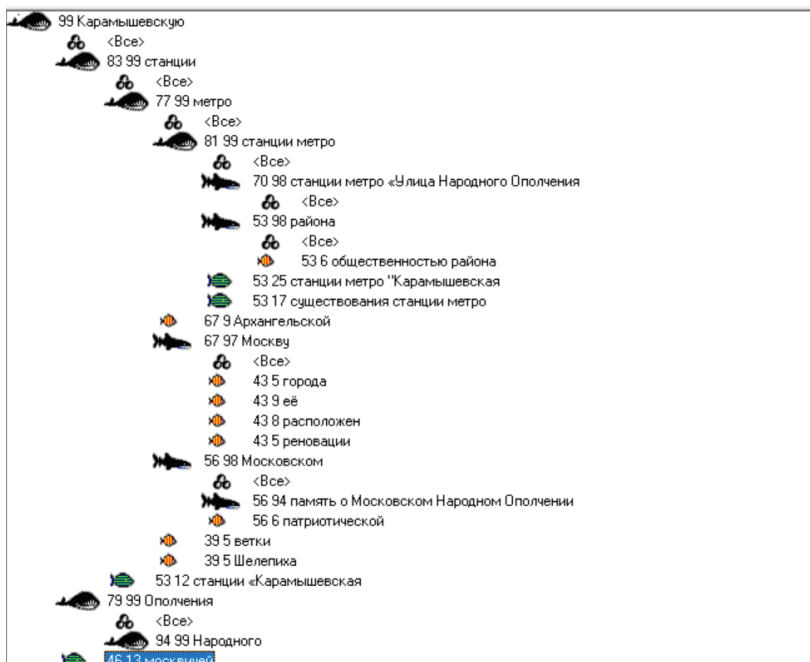


Рис. 7. Тематическая структура негативного кластера.

Контекст:

*Мы не против существования станции метро «Карамышевская», мы против ее появления в этом месте, где пересекаются улица Народного ополчения и проспект Маршала Жукова, где расположен Памятник и сквер Народного ополчения Москвы, где для жителей района и москвичей свята память о Московском Народном Ополчении.*

Для повышения эффективности результатов исследования сводная база контента была разделена на три кластера в соответствии с тональностью: негативный, позитивный, нейтральный. В каждом кластере были выделены и проанализированы тематическая структура, семантическая сеть и ядро семантической сети, проведен ассоциативный поиск, сформирована и исследована ассоциативная сеть.

*Негативный кластер*

• Споры при обсуждении названий станций:

*(...) Вот уже три месяца, с декабря 2020 года, идет противостояние между потомками московских ополченцев 1941 года, ветеранами района Хорошево-Мневники, депутатами района, равнодушной и патриотической общественностью района, с одной стороны, и «заинтересованной силой», которая руками «межведомственной городской комиссии» свершила свое деяние – с другой. Не слушая мнения общественности района, Москвы и граждан России, умные головы поменяли или, если хотите, дали новое название станции метро «Улица Народного Ополчения» и она*

стала «Карамышевской», типа, старое название было рабочим. Эти «бои местного значения» не видят или не хотят видеть московские и федеральные власти, а если видят, то никак не реагируют на возмущение москвичей и россиян-патриотов. (...)

Горожане активно участвовали в общественном обсуждении названий. Результаты обсуждения на «Активном гражданине» показали, что большинство москвичей (55%) проголосовало за название «Народное ополчение» (см. Рис.8):

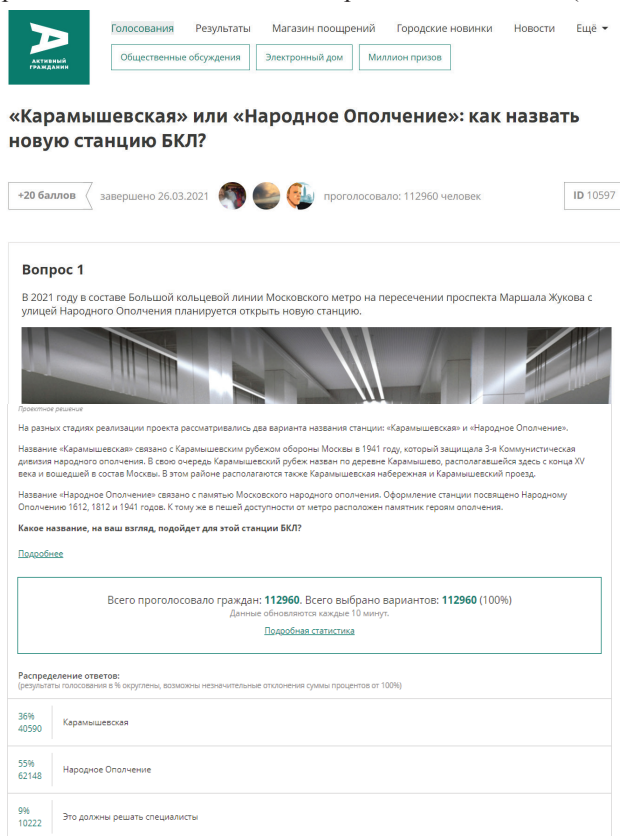


Рис. 8. Активный гражданин: опрос «“Карамышевская” или “Народное Ополчение”: как назвать новую станцию БКЛ?» Электронный ресурс <https://ag.mos.ru/poll/10597>.

- Недовольство москвичей вызвали также сроки общественного обсуждения названий станций (см. Рис. 9, 10).
- Строительство проекта объясняют корыстными мотивами, коррупцией в строительной отрасли (см. Рис. 11, 12).
- Недовольство жителей выбором месторасположения новой линии метро (см. Рис. 13, 14).
- Уничтожение зеленой зоны Москвы при строительстве (см. Рис. 15).

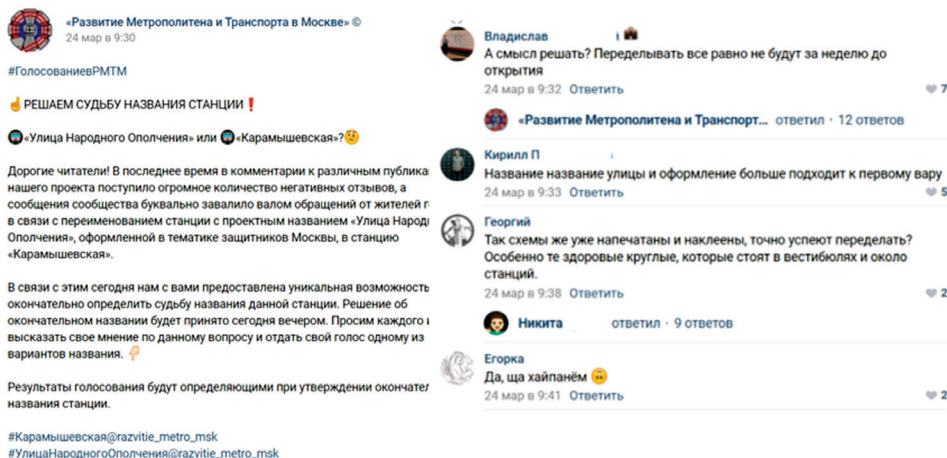


Рис. 9 и 10. Контент, генерированный пользователями. Электронный ресурс <https://vk.com/>.

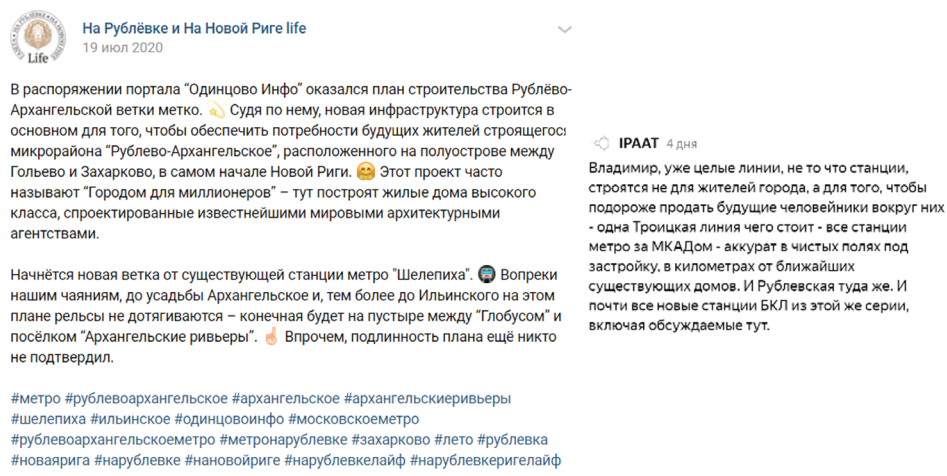


Рис. 11 и 12. Контент, генерированный пользователями. Электронный ресурс <https://vk.com/>.

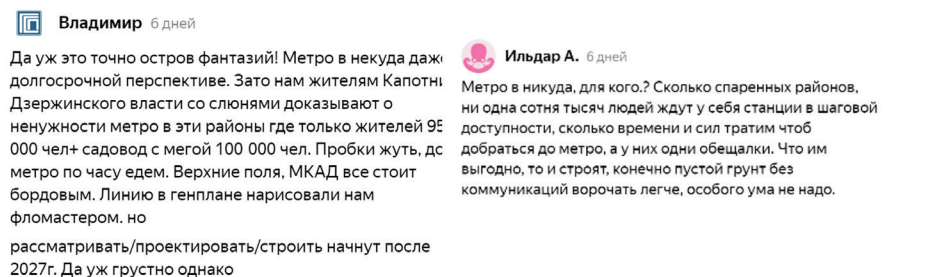


Рис. 13 и 14. Контент, генерированный пользователями. Электронный ресурс <https://vk.com/>.

*(...) Жителям Филевской поймы опять метро не видать, что на БКЛ, что РАЛ. И будут они ходить через Москву реку на Карамышевскую, что в мороз, что в дождь, что в пекло (...).*



Dima

5 дней

Зачем уничтожать зелёные лёгкие города. И так зелени с каждым годом меньше и меньше.



Олег:

3 дня

Хорошее место было для прогулок. Было...Зачем нужно было строить метро в этом месте. Тем более застраивать. Я не понимаю. Можно было сделать парк. Вышло как обычно через одно место. Это еще хорошо, что не стали строить парламентский центр. Провалился бы. Там раньше были колодца, откуда качали воду, для своих нужд тогдашние жители деревни Мневники, Терехово. Воду всю выкачали, остались пустоты. Выдержит ли такое количество построек, посмотрим.

Рис. 15. Контент, генерированный пользователями. Электронный ресурс <https://vk.com/>.

### Позитивный кластер

- Улучшение транспортной ситуации в Москве:

*(...) Открытие данного участка позволит улучшить транспортное обслуживание близлежащих районов с населением свыше 530 тыс (...)*

*(...) Новый маршрут с девятью станциями соединит деловой центр «Москва – Сити» и Рублево-Архангельское (район Кунцево (...)*

*(...) На линии длиной почти 20 км расположатся станции «Пресня», «Улица Народного Ополчения», «Бульвар Генерала Карбышева», «Живописная», «Строгино», «Троице-Лыково», «Рублево-Архангельское» и «Ильинская» (...).*

- Расширение системы московского метрополитена:

*(...) В будущем в Строгино появится еще одна станция метро – на новой Рублево-Архангельской линии (...)*

*(...) Кроме того, строится Троицкая линия метро, разрабатываются проекты Рублево-Архангельской и Бирюлевской веток (...).*

- Сроки строительства:

*(...) Ранее сообщалось, что Рублево-Архангельская линия метро может быть запущена до 2027 года (...).*

- Транспортная доступность Новой Москвы:

*(...) Проектируются Рублево-Архангельская линия метро, а также Бирюлевская, которая придет в Новую Москву от станции ЗИЛ на МЦК (...)*

*(...) Московское метро уходит за город: озвучены сроки начала строительства. О сроках начала строительства Рублево-Архангельской ветки метро рассказал замэра Москвы по вопросам градостроительной политики и строительства Андрей Бочкарев (...).*

- Информация о ходе строительных работ:

*(...) Продолжаются работы на пока что единственной площадке Рублево-Архангельской линии: ведется заливка временного основания монтажной щитовой камеры, из которой щиты начнут проходку тоннелей до станции «Пресня» («Звенигородская») (...).*

*Нейтральный кластер*

- Обсуждение хода реализации проекта:

(...) Рублево-Архангельскую линию метро начнут активно строить после 2022 года. Она соединит центр столицы с Рублево-Архангельским (...).

- Планы транспортного развития Москвы:

(...) Будущие маршруты метро в Москве: Первая станция на Рублево-Архангельской линии и продление в аэропорт Внуково (...)

- Продуманность встроенности Рублево-Архангельской линии в транспортную систему города:

(...) В ходе строительства БКЛ закладываются технические решения, которые позволят присоединить к ней новые радиусы метро: Рублево-Архангельскую линию – от станции «Улица Народного Ополчения» (...).

**Тональность контента**

Показательным представляется анализ рейтинга источников с различными типами тональности контента. Позитивный контент генерируется преимущественно на платформах: ВКонтакте, mos.ru, Телеграм; нейтральный: ВКонтакте, Телеграм, Твиттер<sup>2</sup>, а негативный – Фейсбук<sup>3</sup>, Zen.Yandex, ВКонтакте (см. Рис. 16–18).

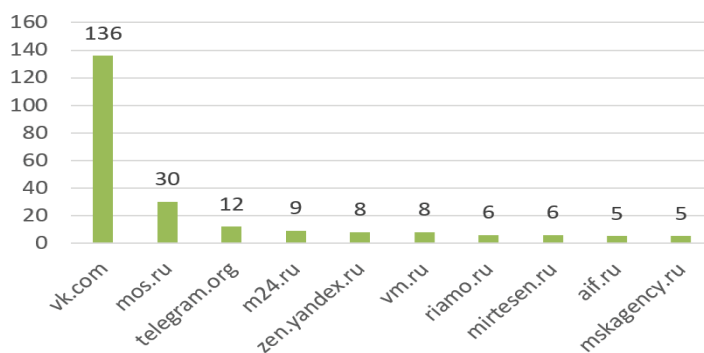


Рис. 16. Рейтинг источников с позитивным контентом.

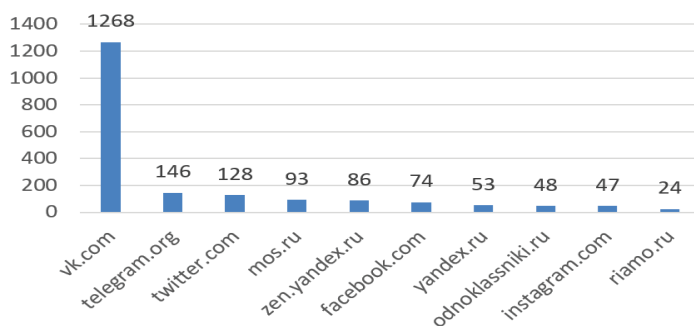


Рис. 17. Рейтинг источников с нейтральным контентом.

<sup>2</sup> Twitter заблокирован в России с 4 марта 2022 г. на основании требования Генпрокуратуры.

<sup>3</sup> Facebook заблокирован в России с 4 марта 2022 г. Роскомнадзор исключил Facebook (экстремистская организация, запрещенная на территории РФ) из реестра соцсетей.

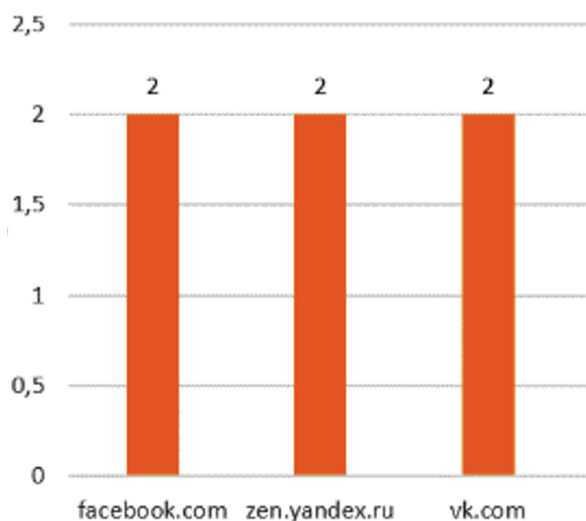


Рис. 18. Рейтинг источников с негативным контентом.

Геолокация акторов, генерирующих релевантный контент, показывает, что наибольшим разнообразием и широтой представленных стран отличается нейтральный контент (Россия, Латвия, Литва, США, Германия, Франция, Сингапур, Нидерланды, Украина, Индия, Казахстан, Италия, Великобритания, Кипр, Беларусь, Турция, Австралия, Катар, Эстония, Испания, Чехия и др.), позитивные сообщения создают акторы с более ограниченной геолокацией, негативный контент связан только с российскими пользователями, которые находятся преимущественно в Центральном округе, в то время как нейтральный и позитивный контент генерируют пользователи в различных регионах России.

Анализ контента позволил сделать вывод о том, что максимальный охват и внимание аудитории получило сообщение о начале строительства (690 173) (см. Рис. 19).

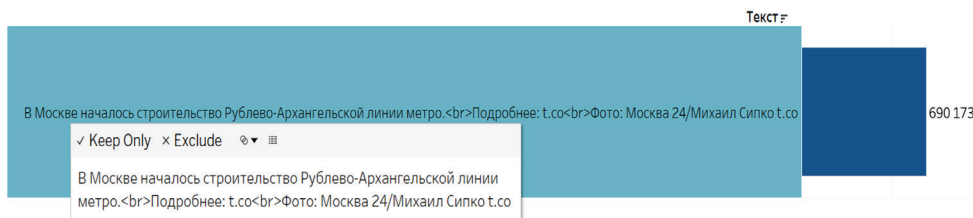


Рис. 19. Первое место в рейтинге текстов по охвату аудитории.

Сентимент-анализ позволил определить, что основная часть контента носит нейтральную тональность, небольшая часть – позитивную, негативные сообщения составляют незначительную часть базы данных (см. Рис. 20).

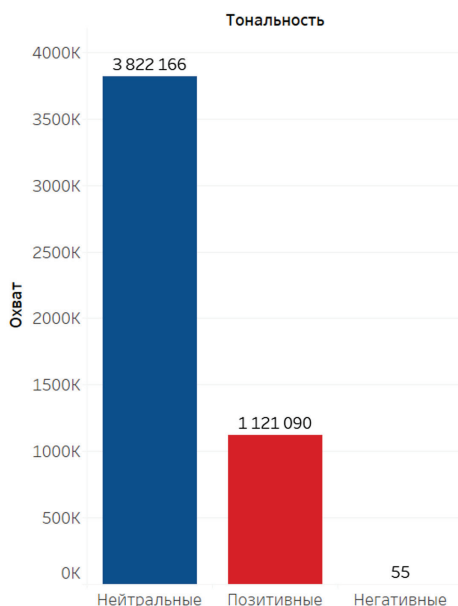


Рис. 20. Тональность контента по охвату аудитории.

Нейтральные и позитивные сообщения генерируются в микроблогах, мессенджерах и социальных сетях, негативные представлены только в социальных сетях (см. Рис. 21).

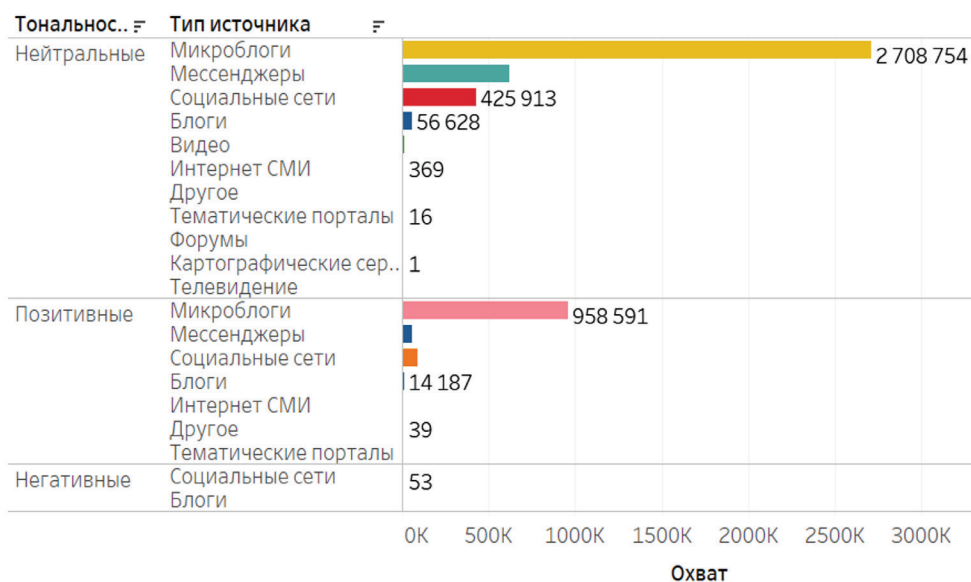


Рис. 21. Тональность контента в соответствии с типом источника.



Наибольшую коммуникативную активность акторы проявляли в социальных сетях, о чем свидетельствуют цифровые следы (Рис. 22).

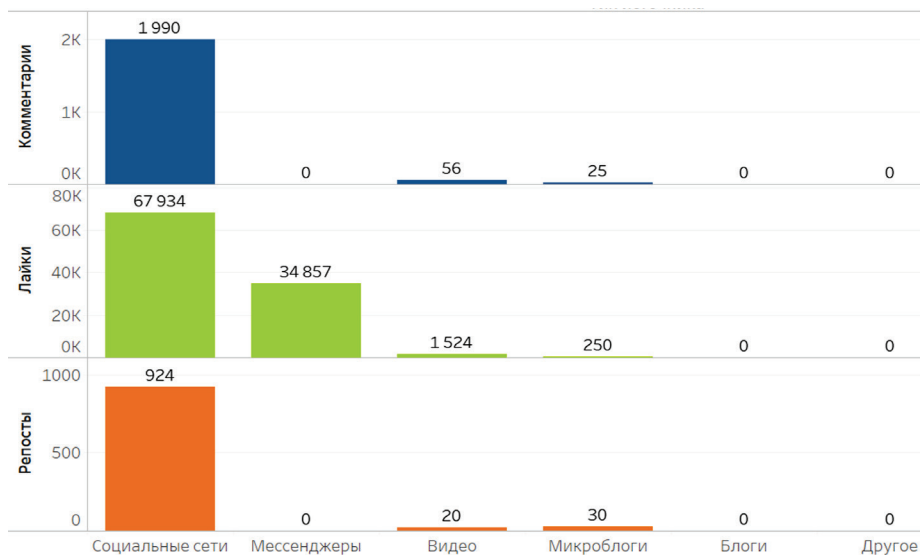


Рис. 22. Цифровые следы на разных типах платформ.

Основная часть цифровых следов пользователей также носит нейтральную тональность, небольшая часть – позитивную, негативные сообщения составляют незначительную часть от общего массива данных (см. Рис. 23).

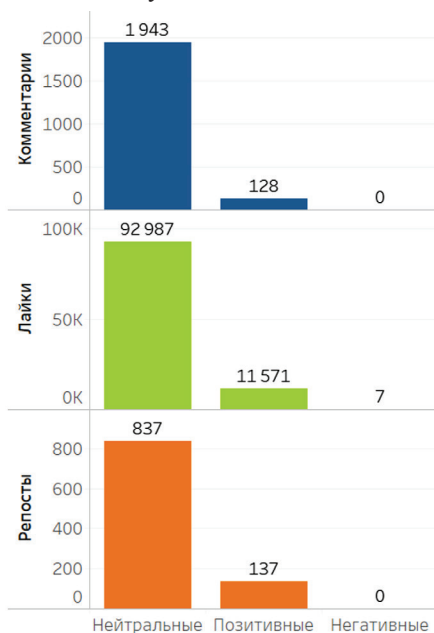


Рис. 23. Тональность цифровых следов акторов.

Рейтинг контента по вовлеченности возглавляют сообщения о начале строительства (см. Рис. 24).

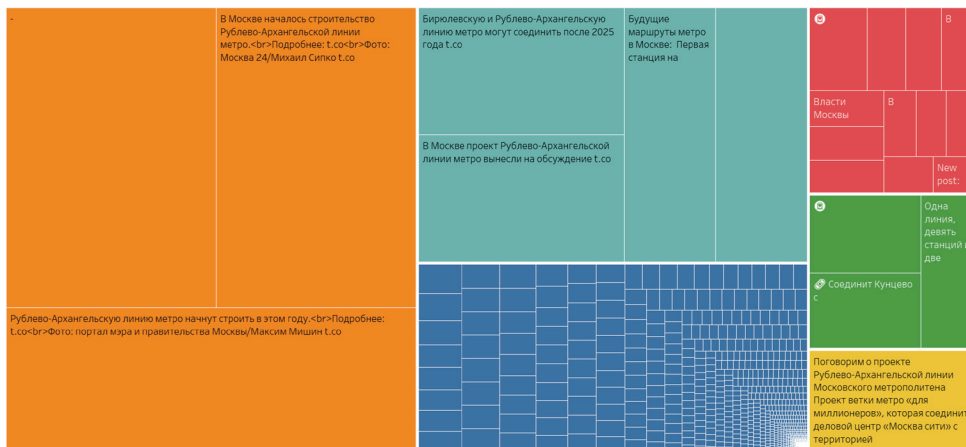


Рис. 24. Значимость контента по вовлеченности.

Лидером по уровню вовлеченности в обсуждения стал твит о начале строительства Рублево-Архангельской линии метро (от ст. м. «Звенигородская» до ст. м. «Липовая аллея»), который получил индекс 690 157 (см. Рис. 25):



Рис. 25. Твит о начале строительства Рублево-Архангельской линии метро.

### Ассоциативные сети

Ассоциативный поиск на больших объемах данных, построение и исследование ассоциативных сетей, анализ лексических ассоциаций позволяют выявлять

имплицитные оценки и мнения акторов виртуальных коммуникаций, а также, наряду с динамикой цифровой агрессии, служат основой для построения прогнозной аналитики (Рис. 26–29).

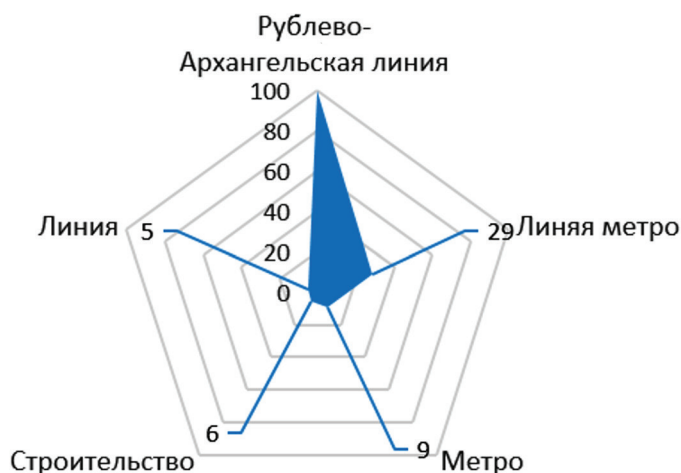


Рис. 26. Ядро ассоциативной сети стимула *Рублево-Архангельская линия метро* (10/6295) сводного датасета.

Контексты с реакциями:

*После ее открытия транспортная доступность улучшится для 350 тысяч жителей районов Обручевский, Черемушки и Коньково.* Ранее сообщалось, что строительство Рублево-Архангельской линии метро начнется в 2021 году.

*Лично пообщаться со специалистами центра можно с понедельника по пятницу с 9:00 до 20:00.* Ранее сообщалось, что строительство Рублево-Архангельской линии метро начнется в 2021 году.

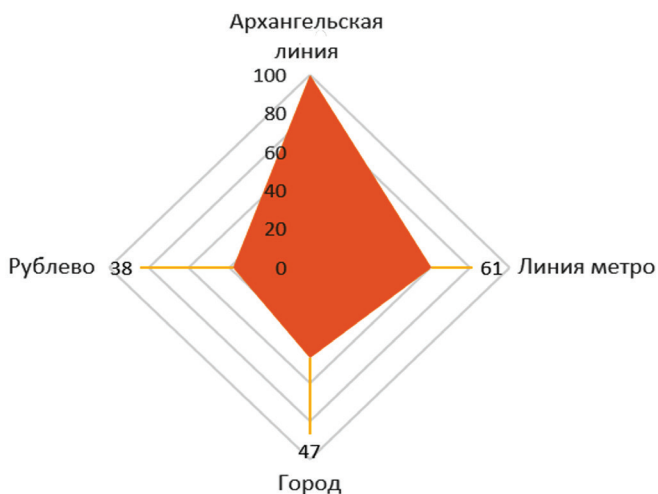


Рис. 27. Ядро ассоциативной сети стимула *Рублево-Архангельская линия метро* (6/6) негативного кластера.

Контексты с реакциями:

*Рублево-Архангельская линия метро, нужна ли она.*

*<br>На кону освоения в городе Москве встали, как солнце на рассвете, огромные, фантастические деньги реновации и вектор их особенно хорошо ложится в направлении на Рублево, чуть не сказал на Рублевку.*

*Судя по нему, новая инфраструктура строится в основном для того, чтобы обеспечить потребности будущих жителей строящегося <br>Показать полностью... микрорайона "Рублево-Архангельское", расположенного на полуострове между Гольево и Захарково, в самом начале Новой Риги. Этот проект часто называют "Городом для миллионеров" – тут построят жилые дома высокого класса, спроектированные известнейшими мировыми архитектурными агентствами. <br>Начнется новая ветка от существующей станции метро «Шелепиха».*

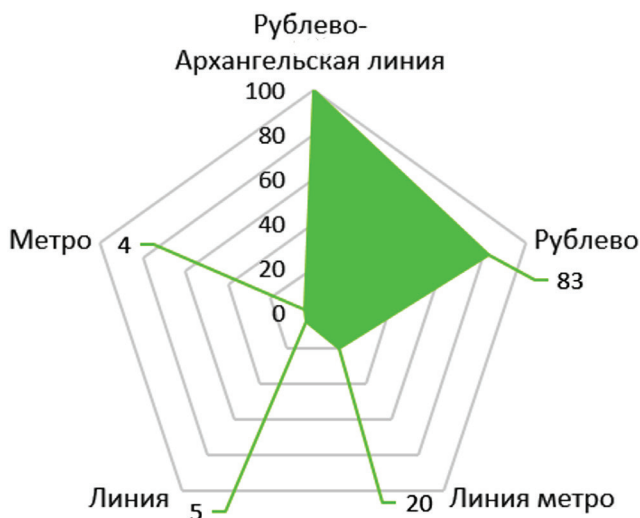


Рис. 28. Ядро ассоциативной сети стимула *Рублево-Архангельская линия метро* (10/1032) позитивного кластера.

Контексты с реакциями:

*Также будет построена Коммунарская линия от станции «Улица Новаторов» Большого кольца метро до Троицка. <br>Проектируются Рублево-Архангельская линия метро, а также Бирюлевская, которая придет в Новую Москву от станции ЗИЛ на МЦК. <br>До 2035 года в ТиНАО планируется построить 72 км линий метро, 33 станции и четыре электродепо. <br>Назад <br>К моменту присоединения новых округов к столице здесь было 598 км дорог, а сейчас длина дорожной сети достигла 790 км. <br>Всего к 2023 году в Новой Москве будет построено и реконструировано порядка 531,3 км автодорог, включая 1-й пусковой комплекс Центральной кольцевой автодороги (ЦКАД), проходящий по территории Новой Москвы.*

*«Это наш город»: строительство Рублево-Архангельской линии метро начнется до конца года <br>Новая ветка свяжет деловой центр Москва-Сити и «Рублево-Архангельское» 15.03 09:55*

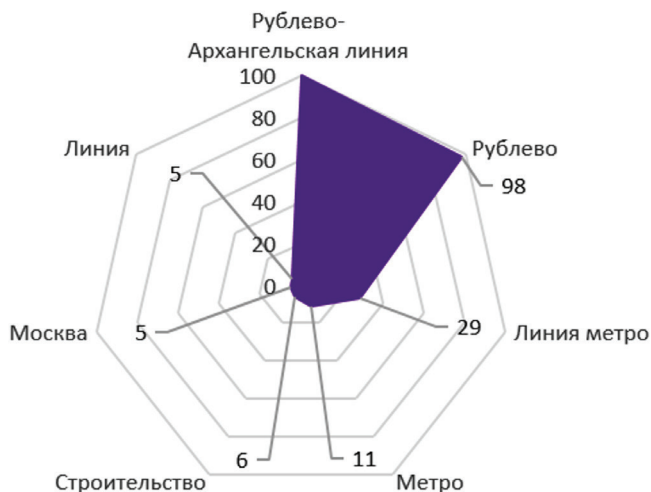


Рис. 29. Ядро ассоциативной сети стимула *Рублево-архангельская линия метро* (10/5732) нейтрального кластера.

Контексты с реакциями:

*Москва: Строительство станции метро «Карамышевская» Большой Кольцевой линии* Уже скоро откроют двери для пассажиров две новых станции Большой Кольцевой линии.

*Линия будет востребована жителями не только района Хорошёво-Мнёвники, но и территории Рублево-Архангельское в районе Кунцево на западе города*, — пояснил Бочкарев. [reality.interfax.ru](https://realty.interfax.ru) Строительство Рублево-Архангельской линии метро в Москве начнут до конца года.

#### **Выявление наличия/отсутствие социального напряжения в районах реализации проекта**

Можно констатировать значительный рост нейтральных сообщений (с 847 в 2020 году до 2 342 в 2021 году) и позитивных (с 211 в 2020 году до 361 в 2021 году). Рост негативных сообщений содержит крайне низкие показатели (с 4 в 2020 году до 6 в 2021) (см. Рис. 30).

Анализа цифровой агрессии, результат ассоциативного поиска и исследование ассоциативной сети показали нейтральное восприятие акторов, отсутствие социальной напряженности. Вычисления индексов социального стресса и благополучия подтвердили полученные результаты. Данные позволили сделать заключение об отсутствии социального стресса (индекс – 0,67) и выделить средний индекс социального благополучия – 10,65. Методика вычисления индексов социального стресса и социального благополучия описана в [Kharlamov, Pilgun 2020] (Рис. 31).

#### **Выводы**

Интеграция психолингвистических методик (в рамках Московской психолингвистической школы) и нейросетевых технологий с позиций теории речевой деятельности позволила построить когнитивную модель анализа социального

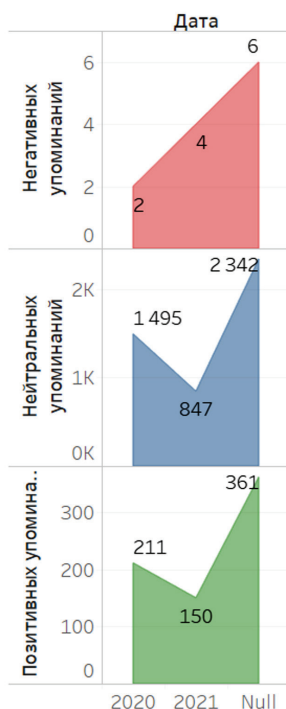


Рис. 30. Динамика сообщений с различной тональностью.

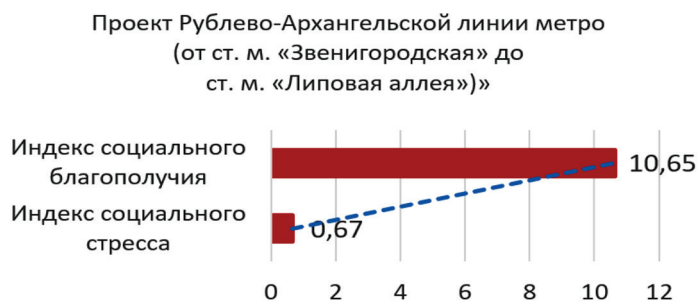


Рис. 31. Уровень социального стресса и благополучия.

напряжения в медиапространстве и провести исследование анализа мнений, восприятия речи акторов социальных медиа, которое показало отсутствие социальной напряженности вокруг строительства Рублево-Архангельской линии метро (от ст. м. «Звенигородская» до ст. м. «Липовая аллея»).

Контент-анализ, выделение тематической структуры и саммаризация позволили уточнить ключевые темы, определить рейтинг контента; сентимент-анализ, анализ семантической и ассоциативной сетей показали по большей части нейтральное восприятие акторов реализации проекта. Анализ лексических ассоциаций дал возможность выявить имплицитные оценки и мнения акторов. Анализ ассоциативных

сетей, динамики цифровой агрессии, вычисление индексов социального стресса и благополучия помогли определить отсутствие социального напряжения вокруг проекта и низкие риски его возникновения в будущем в ходе строительства. Развитие событий подтвердило корректность результатов исследования.

Наибольшее внимание пользователей вызывает информация о начале строительства, перспективах и сроках строительства. Потребности жителей в изменении транспортной ситуации в связи с реализацией проекта носят средние значения.

Перечень претензий жителей, связанный со строительством Рублево-Архангельской линии метро, состоит из ограниченного числа тем. Кроме того, следует отметить, что часть негативных сообщений носит традиционный характер и относится к стереотипам противников любого городского строительства:

- негативное восприятие жителей вызвало обсуждение выбора названия станции (Карамышевская vs Народное ополчение).
- Недовольство сроками общественного обсуждения названий станций.
- Строительство проекта объясняют корыстными мотивами, коррупцией строительной отрасли.
- Отрицательное отношение к выбору месторасположения новой линии метро.
- Уничтожение зеленой зоны Москвы при строительстве.

Проект Рублево-Архангельской линии метро (от ст. м. «Звенигородская» до ст. м. «Липовая аллея») на момент проведения исследования находился на начальной стадии реализации и в анализируемый период представлен только в определенных сегментах медиапространства. В качестве рекомендаций были сформулированы следующие положения: для повышения эффективности диалога с жителями, для предотвращения появления социального напряжения в будущем при расширении информационного сопровождения проекта, помимо традиционных описаний улучшения транспортной ситуации в районе реализации проекта, преимуществ, которые получают москвичи, обратить особое внимание на обсуждение названия станций, а также на обоснование выбора месторасположения новой линии метро и станций.

Кроме того, в информационном сопровождении проекта представляется целесообразным учитывать специфику цифровых платформ, на которых генерируется негативный контент. Нейтральные и позитивные официальные сообщения формировали контент в микроблогах, мессенджерах и социальных сетях, между тем пользовательский негативный контент был представлен только в социальных сетях. Рейтинг источников с различными типами тональности контента показал, что позитивный контент генерировался преимущественно на платформах ВКонтакте, telegram.ru, Телеграм; нейтральный: ВКонтакте, Телеграм, Твиттер<sup>4</sup>, а негативный – Фейсбук<sup>5</sup>, Zen.Yandex, ВКонтакте.

© Пильгун М.А., 2023

---

<sup>4</sup> Twitter заблокирован в России с 4 марта 2022 г. на основании требования Генпрокуратуры.

<sup>5</sup> Facebook заблокирован в России с 4 марта 2022 г. Роскомнадзор исключил Facebook (экстремистская организация, запрещенная на территории РФ) из реестра соцсетей.

### Литература

*Основы теории речевой деятельности* / Отв. ред. А.А. Леонтьев. М.: Наука, 1974. 368 с.

*Харламов А.А.* Ассоциативная память — среда для формирования пространства знаний. От биологии к приложениям. Дюссельдорф: Академическое издательство Palmarium, 2017.

*Demuth K.* Perception, production, and individual differences // *Applied Psycholinguistics*. 2018. 39 (4). P. 735–741.

*Diesner J., Carley K.M.* AutoMap1.2 – Extract, analyze, represent, and compare mental models from texts. URL: [http://www.casos.cs.cmu.edu/publications/papers/diesner\\_2004\\_automap12extract.pdf](http://www.casos.cs.cmu.edu/publications/papers/diesner_2004_automap12extract.pdf).

*Erkan G., Radev D.R.* LexPageRank: Prestige in Multi-Document Text Summarization. // *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. Barcelona, Spain. Association for Computational Linguistics, 2004. P. 365–371.

*Hickok G.* Functional Anatomy of Speech Perception and Speech Production: Psycholinguistic Implications // *Journal of psycholinguistic research*. 2001. №30. P. 225–235.

*Hickok G., Poeppel D.* Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*. 2000. №4. P. 131–138.

*Kreyßig N., Krautz A.E.* Lying and perception of lies by bilingual speakers // *Applied Psycholinguistics*. 2019. 40 (5). P. 1313–1329.

*Neuroinformatics and semantic representations. Theory and Applications* / Kharlamov A. A., Pilgun M. A. (eds). Newcastle upon Tyne: Cambridge, 2020.

*Phani K.P., Arulmozi S., Shiva R.M.* Sensory Perception in Blind Bilinguals and Monolinguals // *Journal of psycholinguistic research*. 2020. №49. P. 631–639.

*Pilgun M., Raskhodchikov A.N., Koreneva-Antonova O.* Effects of COVID-19 on Multilingual Communication // *Frontiers in Psychology*. 2022. №12. URL: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.792042/full>.

*Ransom T.G., Dale R., Kreuz R.J., Tollefsen D.* How Do Different Types of Alignment Affect Perceived Entity Status? // *Journal of psycholinguistic research*. 2019. 48. P. 961–985.

*Speech Perception* / L. Holt, J. Peelle, A.B. Coffin, A.N. Popper, R.R. Fay. Cham: Springer, 2021.

*Uhrig S.* Human Information Processing in Speech Quality Assessment. Cham: Springer, 2021.

#### Сведения об авторе:

**Пильгун Мария Александровна** – доктор филол. наук, профессор, Российский государственный социальный университет

#### Контактная информация:

129226, г. Москва, ул. Вильгельма Пика, д. 4, стр. 1

ORCID: 0000-0002-8948-7075

e-mail: pilgunm@yandex.ru

#### Для цитирования:

Пильгун М.А. Когнитивная модель анализа контента цифровых коммуникаций // *Вопросы психолингвистики* №2(56) 2023, С. 68–93. doi: 10.30982/2077-5911-2023-56-2-68-93



UDC 81'22

LBC 81

DOI 10.30982/2077-5911-2023-56-2-68-93

Research article

## COGNITIVE MODEL OF CONTENT ANALYSIS OF DIGITAL COMMUNICATIONS

**Maria A. Pilgun**

Russian State Social University,  
Moscow, Russia

### *Abstract*

The article is devoted to the development of a model for the integration of psycholinguistic techniques and neural network technologies for the analysis of speech data from the standpoint of the theory of speech activity. The results of the study of speech perception of social media actors for the analysis of social tension around the construction of the Rublevo-Arkhangelskaya metro line (from the Zvenigorodskaya metro station to the Lipovaya Alleya metro station) are presented. The material for the study was data from microblogs, blogs, instant messengers, social networks, video hosting, Internet media, thematic portals, map services and TV. Date of material collection: 01.07.2020 23:59:59 – 24.03.2021 23:59:59. The analysis of the database showed the absence of social tension around the project. For content analysis, a multimodal approach was used using neural network technologies, text analysis, semantic analysis, content analysis, sentiment analysis, analysis of word association. The results of the study showed that the integration of psycholinguistic methods and neural network technologies from the standpoint of the theory of speech activity made it possible to build a cognitive model for the analysis of social tension in the media space, to conduct a study of the analysis of opinions, speech perception of social media actors. Analysis of the data showed the absence of social tension around the construction of the Rublyovo-Arkhangelskaya metro line (from the Zvenigorodskaya metro station to the Linden Alley metro station). Content analysis, the selection of the thematic structure and summarization made it possible to clarify key topics, determine the content rating; sentiment analysis, analysis of semantic and associative networks confirmed for the most part a neutral perception of the actors of the project implementation. The analysis of lexical associations made it possible to identify and interpret the implicit assessments and opinions of actors that confirm the results of the study.

**Keywords:** theory of speech activity, speech perception, psycholinguistics, social media, neural network technologies

©Pilgun M.A., 2023

### **Bionotes:**

**Maria A. Pilgun** – Dr. Sci. in Philology, Professor, Russian State Social University

### **Contact information:**

4, Vil'gel'ma Pika St., Bd. 1. Moscow, Russian Federation, 129226

ORCID: 0000-0002-8948-7075

*e-mail:* pilgunm@yandex.ru

**For citation:**

Pilgun M.A. Cognitive model of content analysis of digital communications // Journal of Psycholinguistics. 2(56), 2023. P. 68–93. Available from: doi: 10.30982/2077-5911-2023-56-2-68-93 (In Russian)