

# Возможности семантического анализа ключевых биграмм для исследования дискурса соседского онлайн сообщества

А.Е. Ненько, Е.В. Недосека, А.А. Галактионова

**Аннотация** — В статье представлены возможности применения семантического анализа текста для исследования дискурса соседских онлайн сообществ. Дискурс, рождаемый онлайн соседскими сообществами в рамках общего публика, является смысловым полем, в котором отражаются нарративы о повседневной жизни, истории о среде проживания, формах взаимодействия между соседями, а также происходит концептуализация отличительных особенностей среды. С помощью автоматизированного семантического анализа, дополненного качественным анализом текста, были выявлены ключевые и интенсивно обсуждаемые концепты соседского публика центральной части Санкт-Петербурга “Пять Углов”, которые отражают особенности среды проживания. Данные текстового корпуса публика были собраны автоматизированным путем с помощью библиотек сетевого скраппинга Scrapy и Selenium в языке программирования Python. Исследование семантического поля ключевых концептов осуществлялось методом биграмм (устойчивых сочетаний) слов, находящихся в непосредственной связи друг с другом в тексте. Предлагаемый подход продемонстрирован на примере анализа биграмм ключевого концепта “улица”. По результатам анализа были выделены четыре измерения, которые связаны с ментальной репрезентацией улицы в представлении анализируемого соседского сообщества: а) улицы - топонимы, названия улиц, которые актуализированы в дискурсе данного сообщества и отражают действительно существующие на данной территории фрагменты городской среды; б) проблемные вопросы, связанные с организацией уличной жизни; в) ценностные характеристики улицы; г) ментальная репрезентация пространственной структуры улицы.

**Ключевые слова**— семантический анализ текста, метод биграмм, соседские онлайн сообщества, дискурс, концепт.

## ВВЕДЕНИЕ

Распространение новых информационных технологий в современном обществе приводит к

Статья получена 14 ноября 2021 года.

Работа выполнена при поддержке гранта РФФИ и АНО ЭИСИ 21-011-31874 опп "Исследование локальной низовой устойчивости соседских онлайн сообществ" 2021-2022 гг.

А.Е. Ненько, Институт дизайна и урбанистики Университета ИТМО, ЦИГЕ СПбГУ, Санкт-Петербург, Россия (al.nenko@itmo.ru)

Е.В. Недосека, Социологический институт РАН — филиал ФНИСЦ РАН, Санкт-Петербург, Россия (nedelena@socinst.ru)

А.А. Галактионова, Институт дизайна и урбанистики Университета ИТМО, Санкт-Петербург, Россия (aagalaktionova@itmo.ru)

формированию новых социокультурных и социополитических феноменов, таких как онлайн (цифровые) локальные сообщества, члены которых взаимодействуют и формируют социальную солидаризацию по поводу различных вопросов местного уровня, используя онлайн коммуникационные средства. Распространенность феномена цифровых городских сообществ, имеющих привязку к какой-либо территории, довольно велика: от любителей истории своего района до сообществ жителей конкретного дома, квартала или района.

Исследовательское внимание в данной статье обращено на соседские онлайн сообщества, как вида локальных городских сообществ, организованных по принципу территориальной привязки к определенной среде проживания. Процесс взаимодействия с конкретной средой проживания приводит к формированию у жителей ее образа, понимания уникальности места и его значения [1, 2].

В данном контексте, необходимо отметить, что городская среда не является объективной данностью, она формируется в процессе взаимодействия различных акторов и является социальным конструктом, формируемым в процессе субъективного восприятия, оценивания и социального действия. Субъективно воспринимаемое качество нельзя свести к набору объективных характеристик, например, имеющейся сервисной и социальной инфраструктуре, количеству зеленых насаждений или качеству транспортной инфраструктуры. Это те характеристики, которые воспринимаются как значимые самими обитателями данной среды и которые формируют резидентную удовлетворенность и привязанность к территории проживания [3]. В представлениях жителей об окружающей их городской среде проявляется их оценка уровня доступности, разнообразия и уникальности компонентов окружающего их контекста и именно данная оценка является для них решающей.

Для понимания качества городской среды важно анализировать субъективное восприятие среды самими жителями и выявлять те элементы и характеристики, которые определяются людьми как важные. Дискурс, формируемый жителями в результате обсуждения городской среды, является источником информации о ее субъективно значимых компонентах. Одним из контекстов, в котором формируется подобный дискурс, является онлайн коммуникация соседских сообществ, объединенных общей территорией проживания.

## СОСЕДСКИЕ СООБЩЕСТВА КАК СУБЪЕКТ ДИСКУРСА О ГОРОДСКОЙ СРЕДЕ

Выбор соседских сообществ, имеющих цифровую природу, создаваемых с помощью новых коммуникационных технологий, в качестве объекта исследования обусловлен несколькими причинами. По мнению ведущего теоретика цифровых сообществ Б. Веллмана, онлайн сообщества - характерный феномен современного мира, в котором львиная доля коммуникации перешла в онлайн режим с помощью ИКТ [4].

Социальные медиа, в том числе, коллективные обсуждения в facebook, уже более 10 лет анализируются как коммуникационные платформы для обсуждения городских проблем и участия в городском управлении. Онлайн платформы стали представлять новые возможности в коммуникации с точки зрения повышения интерактивности и сетевого эффекта взаимодействия [5]. Анализ социальных сетей рассматривается как платформа для общественного участия жителей города в управлении городской средой. Так, исследование Lappas et al. [6] показывают влияние контента коммуникации на facebook страницах муниципалитетов на социальное участие и вовлеченность горожан. Динамика социального участия в деятельности городского совета, отраженная в постах и комментариях в facebook, Twitter и Instagram, осмысливается как развитие практик "умного города", ориентированного на горожан [7]. Отдельные исследования анализируют семантику обсуждений в онлайн соседских сообществах. Например, Afzalan & Muller показывают, как общественное благо - размещение соседского парка - становится предметом дискуссии в соседской социальной сети [8]. На основании коммуникативной теории Ю. Хабермаса авторы оценивают аргументацию различных предложений по поводу нового места расположения парка.

В приведенных выше статьях фигурирует анализ содержания коммуникативного процесса в онлайн сообществах, в том числе в социальной сети facebook. Данный коммуникативный процесс рассматривается в основном как инструмент общественного участия в обсуждении управления городской средой. Новизна данной статьи заключается в анализе семантического поля соседского сообщества, которое формируется в таком социальном медиа, как facebook, методом полуавтоматизированного анализа биграмм с точки зрения отраженного в дискурсе воспринимаемого качества среды обитания.

Онлайн сообщества обладают качествами обычных сообществ, то есть, предлагают своим членам эмоциональную поддержку, материальную немонетарную и монетарную помощь, являются средой для формирования практик обмена, однако при этом обладают значительной гибкостью - возможностью постоянной коммуникации в режиме реального времени, способностью накопления ресурсов в виде информации, данных, продуктов деятельности в самом сообществе, возможностью формировать сопутствующие Интернет-инструменты для своих практик, например, форумы, фотоальбомы, страницы полезных ссылок, и даже

собственные приложения для участия в городском развитии.

Онлайн коммуникация позволяет соседям затрачивать меньше время на установление взаимодействия и начало коммуникации, находиться в фактически круглосуточном неограниченном потоке соседских новостей. Рубрикация тем обсуждения с помощью назначения отдельных веток позволяет удобно разделить для дискуссии отдельные вопросы соседской жизни. Возможности онлайн коммуникационных каналов транслировать и хранить информацию в различных модусах - тексты, фотографии - позволяет создавать архив жизнедеятельности сообщества, к которому соседи могут обращаться снова и снова. Возможность коммуникации отдельными сообщениями в виде публикаций на "стене" сообщества позволяет структурировать и глубоко прорабатывать ежедневную повестку: обмениваться новостями, формировать быстрые призывы к действию (просьбы о помощи, предупреждения об опасности), делиться эмоционально насыщенными переживаниями (фото дня с изображением района), вести долгосрочное обсуждение вопроса (регулярная рубрика). Возможность личной онлайн коммуникации между подписчиками одного соседского сообщества (сообщения в директ) позволяют перейти на индивидуальный уровень общения и планирования совместных действий. Российские ученые в своих работах доказывают "ресурсность" сетей онлайн сообществ для понимания структуры городского пространства [9] и в контексте решения городских конфликтов [10].

Каждое онлайн соседское сообщество так или иначе реагирует и взаимодействует с тем средовым контекстом, в котором оно сформировалось и действует. Дискурс, рождаемый онлайн соседским сообществом в рамках общего паблика или канала, является смысловым полем, в котором отражаются нарративы о повседневной жизни, истории места, формах взаимодействия между соседями, а также концептуализация отличительных особенностей среды, которые составляют чувство места. Дискурс онлайн сообщества, рождаемый им в процессе онлайн коммуникации, является вербализацией воспринимаемого окружения. Наиболее актуализированные средовые проблемы и ценности будут интенсивно обсуждаться и концептуализироваться.

В данной работе объектом исследования выступило соседское онлайн сообщество г. Санкт-Петербурга «Пять углов». Выбор сообщества был обоснован, во-первых, показателями активности группы (как администраторов, так и пользователей), во-вторых, значительным количеством участников; в-третьих, это группа наиболее укорененной части жителей города, транслирующей наибольшую идентификацию с местом проживания. Площадкой паблика выступает социальная сеть

Facebook (<https://www.facebook.com/groups/942323882576064>). Группа была основана 16 апреля 2017 года и в настоящий момент насчитывает 9,0 тыс. участников. Степень активности паблика в среднем 4 сообщения на стене в день. Контент создается и генерируется пользователями (user generated content), но регулируется и отслеживается администраторами посредством

установленных правил. За весь период существования группы генеральная совокупность всех сообщений составила 47307 (на момент 15 августа 2021 г.).

Соседское сообщество “Пять углов” представляет исторический район (Владимирский округ) города Санкт-Петербург. Это среда значимого общественного пространства, мест коллективной памяти, наполненных объектами исторического и культурного наследия, определяющих высокую материальную и символическую ценность для горожан. Это уникальная физическая среда, где происходит социализация жителей, именно она формирует сильную городскую идентичность по отношению к месту у коренных жителей и является способом познания города у приезжих и иммигрантов [11, 12]. Жители района «Пяти углов» живут в дореволюционных домах с уникальной архитектурой вблизи основных культурных аттракторов города. Спрос на территорию жителей «Пяти углов» высок со стороны иных социальных групп - туристов, представителей ритейла и индустрии гостеприимства, а также жителей других районов города. Анализ агрегированных анонимизированных данных о подписчиках сообщества показывает, что для его социального состава характерна большая доля творческих профессий, а также большее разнообразие возрастных групп. По словам главного администратора группы основным электоратом паблика является «питерская интеллигенция».

#### МЕТОДОЛОГИЯ СБОРА И АНАЛИЗА ДАННЫХ

Для формирования базы семантических данных собирался текстовый дискурс сообщества “Пять углов”. К текстовому корпусу были отнесены:

а) текстовые данные с публичной страницы “стены” сообщества, на которой от лица сообщества, администраторов и участников публикуются сообщения (“посты”);

б) специальные обсуждения внутри сообщества;

в) посты отдельных пользователей.

Данные текстового корпуса собраны автоматизированным путем с помощью библиотек сетевого скраппинга Scrapy и Selenium в языке программирования Python (Python html). Итоговый текстовый корпус составил 47307 текстовых комментариев, в составе которого было 820 582 слов. Корпус, прежде всего, был очищен от знаков препинания, эмодзи, а также латиницы. Собранный массив был подвергнут алгоритму лемматизации уникальных слов - приведению слов в нормализованный вид с помощью библиотеки Python `py morphology2`, которая также позволила определить часть речи, род, число и отношение слова к имени человека. Были исключены имена собственные пользователей, которые присутствовали в ответах пользователям в комментариях. Кроме того, был применен словарь стоп-слов для исключения “мусорных” слов (которые содержательно не несут смысла для анализа феноменов среды) – числительных, местоимений, междометий и наречий.

Дальнейший анализ выстраивался на основании изучения семантического поля ключевых концептов. Данный подход основывается на теории Л. Витгенштейна [13] об относительности смысла в языке

и тезисе о том, что смысл какого-либо понятия основывается на его интерпретации, которую можно проследить через сопутствующие слова, с которыми данное понятие встречается вместе в устной или письменной речи. Анализ семантического поля ключевых концептов включает поиск слов, которые колоцированы, или встречаются вместе, с концептом в заданной единице текста, фразе или письменном фрагменте [14-16]. Данный подход получил в последнее время распространение в лингвистике и социолингвистике [17-19]. Он представлен такими методами анализа, как анализ биграмм (устойчивых словосочетаний), анализ колоцированных концептов (концептов, которые встречаются вместе, однако в предложении могут находиться с определенным заданным шагом, например, в 2 или 3 слова), анализ семантических сетей (где узлами сети являются концепты, а связью между ними является частота их совместной упоминаемости в тексте без шага или с шагом в несколько слов). В данном исследовании был использован анализ биграмм для определения семантического поля концепта на основании слов, которые непосредственно встречаются в одном словосочетании с ним. Такая сила связи представляется наиболее высокой по сравнению с силой связи при сопоставлении слов с определенным шагом. При назначении шага в несколько слов необходимы серьезные основания для выбора так называемого “окна”, в которое попадает изучаемый концепт и другие содержательные слова, которые гипотетически с ним связаны; сделать подобный выбор обоснованно автоматизированным путем все еще представляет сложность для исследователей [20].

Для выделения ключевых концептов, семантическое поле которых будет изучено, был использован прием, который позволяет осуществить выбор на основании экспертного погружения в данные для понимания принципов существования соседского сообщества, и избежать ошибок выбора ключевых слов на основании частотного распределения слов, которое может иметь смещения в силу особенностей естественного языка. К одному из стандартных смещений данного рода относится высокая частотность концептов, которые, с одной стороны, имеют смысл с точки зрения анализа, но с другой стороны, часто употребляются для стандартного построения описательных фраз в естественном языке, например, концепт “человек” и его множественное число “люди” (Сравните суггестивность терминов в таких двух случаях: “Этот человек - хороший, потому что он соблюдает правила нашей улицы” - концепт воплощает характеристики нормативного поведения, разделяемого в сообществе; “Опять эти люди вышли толпой”, “Я не знаю, что сказали эти люди” - концепт имеет вспомогательную роль для построения фразы или отражает собирательный образ, который сложно проинтерпретировать). Экспертный подход к выделению ключевых концептов (*expert-driven sampling*) требует гораздо больших затрат на погружение в семантику исследуемого феномена, однако дает свои плоды с точки зрения выбора концептов, которые

являются максимально красноречивыми с точки зрения задач анализа.

В данном исследовании экспертный подход основывался на предварительном качественном анализе собранного текстового корпуса сообщества “Пять углов”, который был осуществлен на основании открытого и осевого кодирования данных [Ненько и др. в печати]. Качественный анализ позволил выделить топики, которые часто фигурируют в обсуждении соседей центрального района Санкт-Петербурга, в том числе, наиболее актуализированную проблематику - общественные споры по поводу наполнения и использования улицы Рубинштейна. Как было указано выше, улица Рубинштейна является аттрактором для услуг и заведений индустрии гостеприимства, гастрономии и туризма, а рестораторы, туристы и владельцы гостиниц и магазинов - значимыми агентами влияния. При этом улицей активно пользуются соседи Пяти Углов как повседневной пешеходной артерией, кроме того, на улицу выходят окна жилых домов, вследствие чего соседи находятся в конфликте с внешними “пришлыми” стейкхолдерами улицы. Качественный анализ подсказывает ключевой концепт, который был рассмотрен нами путем автоматизированного семантического анализа, а именно “улица”. Данный концепт также отражает отношение соседей к окружающей городской среде и является базовым социопрограммной единицей в ментальном образе городской среды, который формируется соседским сообществом в обсуждениях.

При выделении биграмм были использованы автоматизированные процедуры процессинга очищенного текстового корпуса с помощью библиотеки языка программирования Python nltk, которая предназначена для обработки текстовых данных. Отобранные слова нормальной формы были проиндексированы в соответствии с отзывами пользователей. С помощью функции bigrams, которая принимает на вход список слов, были получены пары из слов, стоящих рядом в отзыве. При формировании биграмм исследователи могут назначать те имена речи, которые будут фигурировать в итоговой выборке. При анализе полных биграмм остаются глаголы, содержательные наречия, имена существительные и прилагательные, однако распространен и метод анализа биграмм, включающих только имена существительные и прилагательные. В данном случае был использован второй путь. Данный выбор обоснован фокусом исследования: рассмотрением содержательных характеристик ключевых концептов, иными словами, основных черт изучаемого феномена, которые в естественном языке задаются через прилагательные. Имена существительные оставлены в биграмах, исходя из результатов качественного эмпирического исследования, которые говорят о том, ключевые концепты “улица” и “дом” могут быть связаны с действующими силами - социальными агентами, а также местами, которые в естественном языке задаются существительными.

В анализируемом текстовом корпусе общее количество биграмм с концептом “улица” составляет 1681 единиц, а количество биграмм с именами

существительными и прилагательными 1597 единиц. Так как методически было принято решение сконцентрироваться на описании базовых характеристик средовых концептов, в рамках дальнейшего анализа рассматривали только биграмы второго вида. После получения биграмм с концептами были произведены дополнительные процедуры полуавтоматизированной чистки: были определены часто встречаемые, но бессодержательные слова, которые были отфильтрованы. Итоговый корпус содержит 933 биграмы с концептом “улица”.

Далее список биграмм был подвергнут качественному анализу путем открытого кодирования тех характеристик улицы, которые отражаются в биграмах. Были выделены следующие четыре измерения, которые связаны с ментальной репрезентацией улицы в представлении соседского сообщества:

- улицы - топонимы, названия улиц, которые актуализированы в дискурсе данного сообщества и отражают действительно существующие на данной территории фрагменты городской среды;
- проблемные вопросы, связанные с организацией уличной жизни;
- ценностные характеристики улицы;
- пространственная структура улицы, иными словами, ментальное конструирование различных структурных частей “улицы”.

Кодировка биграмм была осуществлена в программе MS Excel. Дальнейшие количественные расчеты и визуализация диаграмм осуществлялась в этом же программном обеспечении. Визуализация облаков слов осуществлялась в программе Tableau.

#### МЕНТАЛЬНАЯ РЕПРЕЗЕНТАЦИЯ КОНЦЕПТА “УЛИЦА” В ПРЕДСТАВЛЕНИИ СОСЕДСКОГО СООБЩЕСТВА

Выделение измерений концепта “улица”, представленных выше, демонстрирует полисемантическую природу понятия и его взаимосвязь с различными когнитивными процессами, происходящими в “коллективном интеллекте” соседского сообщества.

Во-первых, в дискурсе сообщества активно проявляется топонимическое знание и географическое ориентирование в среде обитания. Биграмы, которые отражают улицы как топонимы, составляют примерно  $\frac{1}{5}$  всей выборки. На диаграмме представлены основные топонимы по упоминаемости названий улиц в биграмах (рис. 1).

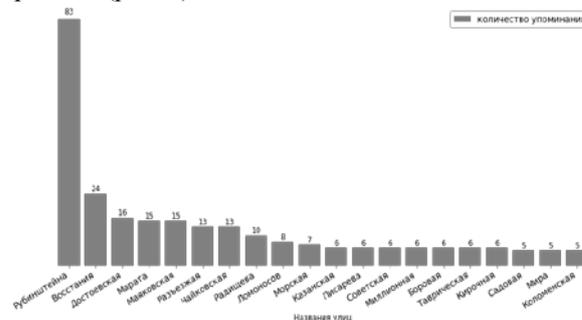


Рис. 1. Рейтинг 15 наиболее упоминаемых биграмм улиц-топонимов, по частоте упоминаний

Вне конкуренции с точки зрения упоминаний - улица Рубинштейна, ее упоминаемость почти в четыре раза выше, чем улицы, которая следует в рейтинге вслед за ней - улицы Восстания. Это объясняется актуализированностью данной улицы в коллективном сознании сообщества Пяти Углов как наиболее конфликтогенной ценности городской среды

Во-вторых, в семантическом поле концепта “улица” выражены проявлены проблемы использования улиц на территории проживания соседского сообщества. На рисунке 2 в виде облака слов представлены наиболее характерные с точки зрения упоминаемости проблемные характеристики улиц. Среди биграмм выявляются словосочетания по типу прилагательное-существительное, которые характеризуют уникальные проблемные черты центральной улицы, например, “ресторанная улица”. Соседи обсуждают присутствие ресторанов и иных аттракторов на улице Рубинштейна как источник беспорядка, шума и других проблем. Также проявлены биграммы, состоящие из двух существительных, например, “улица машина”, “улица фура”. Данные биграммы отражают остроту проблемы стихийной парковки на улицах Пяти Углов.

С точки зрения проблематики, в вызовах, с которыми сталкиваются жители улиц Пяти Углов, можно выделить как стандартные проблемы жилищно-коммунального хозяйствования, такие как “уборка”, “мусор”, так и проблемы использования улицы, в том числе, автомобилистами “парковка”, “машина”, “фура”. Уникальной проблематикой, характерной для улиц данного соседства, являются проблемы с апроприацией улицы “пришлыми” агентами, такими как бары, магазины, рестораны.

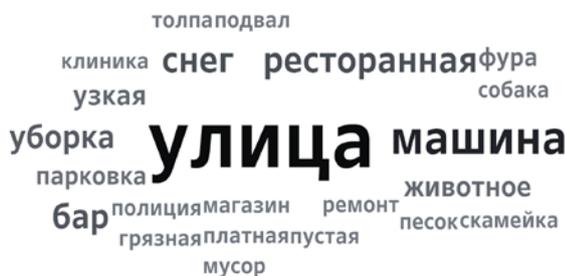


Рис. 2. Проблемные вопросы, связанные с концептом улицы, по частоте упоминания

В-третьих, в концептуализации “улицы” соседским сообществом выделяются ценностные характеристики этого компонента городской среды (рис. 3). Данные характеристики связаны, с одной стороны, с возможностью использовать улицу пешеходами, соседями, живущими здесь, а также комфортом улиц для детей и даже маленьких животных. Данные ценности можно связать с процессом “одомашнивания” улицы как элемента среды обитания ее жителями. С другой стороны, ценностное измерение улицы связано с ее характерным центральным местоположением в городе и с уникальным “парадным” характером, репрезентирующим Петербург в глазах соседей.

Можно говорить о том, что улицы Пяти Углов осмысляются соседями в семантических полюсах “дом - центр города”, и оба полюса непротиворечивы и важны для качества жизни и идентичности. Важно отметить, что данный аспект подтверждается и в самоназвании

паблика “Пять углов: соседский паблик жителей центра Петербурга”. Однако центральность улиц является одновременно и причиной проблем соседей, связанных с оспариванием права на данное пространство внешними агентами влияния.



Рис. 3. Ценностные характеристики улицы, по частоте упоминаемости прилагательных

В-четвертых, семантическое поле концепта “улица” демонстрирует процесс конструирования пространственной структуры данного элемента среды (рис. 4). Улица воспринимается и как публичное, и как приватное пространство. На улице расположены двор и дом жителя, на улицу выходят окна и балкон квартиры, вследствие чего к уличному пространству предъявляются требования личного комфорта и безопасности.



Рис. 4. Пространственная структура образа “улицы”

С другой стороны, улица - это часть центра города, часть района, пространственный объект, который

смыкается с другими местами, территориями и пространствами, выходит к проспектам и площадям, расположенным поблизости. С этой точки зрения, это публичное пространство, которое не оторвано от жизни города, и которое соединяет жителя с обширной городской средой. Как убедительно показывает анализ, концепт “улица” сочетает в себе противоречивые части, живущие по разным законам и требованиям - частную зону человеческой жизни и общественную городскую ткань. Следовательно, именно улицам как соединительным нитям между частным и общим пространствами необходимо уделять особое внимание при планировании городской среды.

#### ЗАКЛЮЧЕНИЕ

В рамках статьи было продемонстрировано, каким образом дискурс соседского сообщества, генерируемый в рамках онлайн обсуждений в социальных сетях, может быть использован для анализа качества городской среды, в которой проживают соседи.

В силу ограничения объема статьи авторы не стали рассматривать другие концепты, которые также могут составлять интерес с точки зрения анализа образа городской среды и коллективной идентичности соседского сообщества. В будущих исследованиях предлагается рассмотрение концептов, которые отражают процессы самоидентификации сообщества, такие как “соседи”, “житель”, “сообщество”. Кроме того, для рассмотрения интересны и другие пространственные концепты, в интерпретации которых может отражаться специфика окружающей среды, такие как “двор”, “район”, “город”. Как мы показали выше, пространственные концепты могут иметь несколько семантических измерений, например, частное и публичное. Также в рамках будущих исследований представляет интерес проверить вложенность семантических структур данных концептов, иными словами, наличия факта интерпретации концепта “улица” через концепт “двор”, концепта “город” через концепт “улица”.

С точки зрения специфики соседского сообщества, взятого в рассмотрение, которое находится в городской среде, конструируемой различными агентами влияния, представляет интерес более подробное рассмотрение семантических полей различных агентов через анализ персонифицированных концептов “туристы”, “приезжие”, “рестораторы”, “владельцы”, а также концептов, отражающих места и функции, которые вызывают дискуссию у соседей - “рестораны”, “гостиницы”, “магазины”, “экскурсии”.

#### БЛАГОДАРНОСТИ

Работа выполнена при поддержке гранта РФФИ и АНО ЭИСИ 21-011-31874 опн “Исследование локальной низовой устойчивости соседских онлайн сообществ” 2021-2022 гг.

#### БИБЛИОГРАФИЯ

[1] Jones C., Svejnova S. The architecture of city identities: A multimodal study of Barcelona and Boston // *Multimodality, meaning, and institutions (Research in the Sociology of Organizations, Vol. 54B)*. – Emerald Publishing Limited, 2017. – P.

- 203-234. DOI: <https://doi.org/10.1108/S0733-558X2017000054B007>.
- [2] De Biase C. The city as a place of ethnic and cultural fusion-how the growth of foreign population influenced the transformations of the territory. – *Gangemi Editore International*, 2019. – P. 203-216.
- [3] Bonaiuto M. et al. Multidimensional perception of residential environment quality and neighbourhood attachment in the urban environment // *Journal of environmental psychology*. – 1999. – Vol. 19 (4). – P. 331-352.
- [4] Wellman B., Wortley S. Different strokes from different folks: Community ties and social support // *American journal of Sociology*. – 1990. – Vol. 96 (3). – P. 558-588.
- [5] Evans-Cowley J., Hollander J. The New Generation of Public Participation: Internet-based Participation Tools // *Planning Practice & Research*. – 2010. – Vol. 25 (3). – P. 397-408.
- [6] Lappas G., Triantafyllidou A., Kani A. Harnessing the power of dialogue: examining the impact of facebook content on citizens’ engagement // *Local Government Studies*. – 2021. – DOI: 10.1080/03003930.2020.1870958.
- [7] Steinmetz C., Rahmat H., Marshall N., Bishop K., Thompson S., Park M., Corkery L., Tietz C. Liking, Tweeting and Posting: An Analysis of Community Engagement through Social Media Platforms // *Urban Policy and Research*. – 2021. – Vol. 39 (1). – P. 85-105.
- [8] Afzalan N., Muller B. The Role of Social Media in Green Infrastructure Planning: A Case Study of Neighborhood Participation in Park Siting // *Journal of Urban Technology*. – 2014. – Vol. 21 (3). – P. 67-83.
- [9] Nenko A., Koniukhov A., Petrova M. Areas of habitation in the city: improving urban management based on check-in data and mental mapping // *International Conference on Electronic Governance and Open Society: Challenges in Eurasia*. – Springer, Cham, 2018. – P. 235-248.
- [10] Чернышева Л. А. Онлайн и офлайн-конфликты вокруг городской совместности: забота о городском пространстве на территории большого жилого комплекса // *Журнал социологии и социальной антропологии*. – 2020. – Т. 23 (2). – С. 36-66.
- [11] Proshansky H. The city and self-identity // *Environment and behavior*. – 1978. – Т. 10 (2). – P. 147-169.
- [12] Морозова К. Как объединить соседей, чтобы сделать классное районное медиа? Инструкция // Собака. – 2020. – 10 февр. URL: <https://www.sobaka.ru/city/city/103159>
- [13] Wittgenstein L. *Philosophical Investigations*. Vol. 50. New York: Prentice Hall, 1953. – 321 p.
- [14] Basov N., Lee J. S., Antoniuk A. Social networks and construction of culture: A socio-semantic analysis of art groups // *Studies in Computational Intelligence*. – 2017. – Vol.693 – P. 785-796.
- [15] Carley K. Extracting culture through textual analysis // *Poetics*. – 1994. – Vol. 22 (4). – P. 291-312.
- [16] Cluley R. Art words and art worlds: The methodological importance of language use in Howard S. Becker’s sociology of art and cultural production // *Cultural sociology*. – 2012. – Vol. 6 (2). – P. 201-216.
- [17] Diesner J. From texts to networks: Detecting and managing the impact of methodological choices for extracting network data from text data // *KI-Künstliche Intelligenz*. – 2013. – Vol. 27 (1). – P. 75-78.
- [18] Sinclair J., Sinclair L. *Corpus, concordance, collocation*. – Oxford University Press, USA, 1991. – 170 pp.
- [19] Roth C., Cointet J. P. Social and semantic coevolution in knowledge networks // *Social Networks*. – 2010. – Vol. 32 (1). – P. 16-29.
- [20] Bekkerman R., Allan J. Using bigrams in text categorization. – Technical Report IR-408, Center of Intelligent Information Retrieval, UMass Amherst, 2004. – P. 1-10.

**Ненько Александра Евгеньевна**, кандидат социологических наук, доцент Института дизайна и урбанистики Университета ИТМО, координатор проектов ЦИГЕ СПбГУ, 199034, Санкт-Петербург, Биржевая линия, 14, лит.А, e-mail: [al.nenko@itmo.ru](mailto:al.nenko@itmo.ru), Центр изучения Германии и Европы СПбГУ, [elibrary.ru](http://elibrary.ru): authored=966941, ORCID: [orcidID=0000-0003-3436-1069](https://orcid.org/0000-0003-3436-1069).

**Недесека Елена Владимировна**, кандидат социологических наук, ст. научн. сотрудник Социологического института РАН — филиала ФНИСЦ РАН, 190005, Санкт-Петербург, ул. 7-я Красноармейская, д. 25/14, e-mail: [nedelena@socinst.ru](mailto:nedelena@socinst.ru), [elibrary.ru](http://elibrary.ru): authorID=64872, ORCID Id: [0000-0003-1944-0367](https://orcid.org/0000-0003-1944-0367).

**Галактионова Анастасия Алексеевна**, аспирант Института дизайна и урбанистики Университета ИТМО, 199034, Санкт-Петербург,

Биржевая линия, 14, лит.А, e-mail: aagalaktionova@itmo.ru, ORCID:  
orcidID=0000-0003-3767-7237.

# Possibilities of the Key Bigrams Semantic Analysis for Studying the Discourse of an Online Neighbor Community

Alexandra E. Nenko, Elena V. Nedoseka, Anastasiia A. Galaktionova

**Abstract** — The article presents the possibilities of semantic text analysis in studying the discourse of neighboring online communities. The discourse generated online by such a community in a social media group is a semantic field that contains narratives about everyday life, habitat, interaction between neighbors, as well as conceptualizes the distinctive features of the urban environment. With the help of an automated semantic analysis supplemented by a qualitative analysis of the text, the key intensively discussed concepts of the «Pyat' Uglov (Five Corners)» neighboring community located in the central part of St. Petersburg were identified. The text corpus of the community online discussions was collected with Scrapy and Selenium libraries of Python programming language. The semantic fields of the key concepts were studied with the method of bigrams (stable combinations) of words that are in direct connection with each other in the text. The proposed approach is demonstrated with analysis of the key concept “street” and its bigrams. Based on the results of the analysis, four dimensions, associated with the community’ mental representation of the street, were identified: a) streets - toponyms, or street names pronounced in the discourse of the community, which reflect the actually existing fragments of the urban environment; b) problematic issues related to the organization of street life; c) value characteristics of the street; d) the mental representation of the spatial structure of the street.

**Key words** - semantic text analysis, bigram method, neighboring online communities, discourse, concept.

## REFERENCES

- [1] Jones C., Svejenova S. The architecture of city identities: A multimodal study of Barcelona and Boston // *Multimodality, meaning, and institutions (Research in the Sociology of Organizations, Vol. 54B)*. – Emerald Publishing Limited, 2017. – P. 203-234. DOI: <https://doi.org/10.1108/S0733-558X2017000054B007>.
- [2] De Biase C. The city as a place of ethnic and cultural fusion-how the growth of foreign population influenced the transformations of the territory. – *Gangemi Editore International*, 2019. – P. 203-216.
- [3] Bonaiuto M. et al. Multidimensional perception of residential environment quality and neighbourhood attachment in the urban environment // *Journal of environmental psychology*. – 1999. – Vol. 19 (4). – P. 331-352.
- [4] Wellman B., Wortley S. Different strokes from different folks: Community ties and social support // *American journal of Sociology*. – 1990. – Vol. 96 (3). – P. 558-588.
- [5] Evans-Cowley J., Hollander J. The New Generation of Public Participation: Internet-based Participation Tools // *Planning Practice & Research*. – 2010. – Vol. 25 (3). – P. 397-408.
- [6] Lappas G., Triantafyllidou A., Kani A. Harnessing the power of dialogue: examining the impact of facebook content on citizens’ engagement // *Local Government Studies*. – 2021. – DOI: 10.1080/03003930.2020.1870958.
- [7] Steinmetz C., Rahmat H., Marshall N., Bishop K., Thompson S., Park M., Corkery L., Tietz C. Liking, Tweeting and Posting: An Analysis of Community Engagement through Social Media Platforms // *Urban Policy and Research*. – 2021. – Vol. 39 (1). – P. 85-105.
- [8] Afzalan N., Muller B. The Role of Social Media in Green Infrastructure Planning: A Case Study of Neighborhood Participation in Park Siting // *Journal of Urban Technology*. – 2014. – Vol. 21 (3). – P. 67-83.
- [9] Nenko A., Koniukhov A., Petrova M. Areas of habitation in the city: improving urban management based on check-in data and mental mapping // *International Conference on Electronic Governance and Open Society: Challenges in Eurasia*. – Springer, Cham, 2018. – P. 235-248.
- [10] Chernysheva L. Online and offline conflicts around urban commons: caring for urban space in the territory of a large housing estate // *The Journal of Sociology and Social Anthropology*. – 2020. – Vol. 23 (2). – P. 36-66.
- [11] Proshansky H. The city and self-identity // *Environment and behavior*. – 1978. – T. 10 (2). – P. 147-169.
- [12] Morozova K. How to unite neighbors to create a cool district media? Instruction // *Sobaka*. – 2020. – 10 February. URL: <https://www.sobaka.ru/city/city/103159>
- [13] Wittgenstein L. *Philosophical Investigations*. Vol. 50. New York: Prentice Hall, 1953. – 321 p.
- [14] Basov N., Lee J. S., Antoniuk A. Social networks and construction of culture: A socio-semantic analysis of art groups // *Studies in Computational Intelligence*. – 2017. – Vol.693 – P. 785-796.
- [15] Carley K. Extracting culture through textual analysis // *Poetics*. – 1994. – Vol. 22 (4). – P. 291-312.
- [16] Cluley R. Art words and art works: The methodological importance of language use in Howard S. Becker’s sociology of art and cultural production // *Cultural sociology*. – 2012. – Vol. 6 (2). – P. 201-216.
- [17] Diesner J. From texts to networks: Detecting and managing the impact of methodological choices for extracting network data from text data // *KI-Künstliche Intelligenz*. – 2013. – Vol. 27 (1). – P. 75-78.
- [18] Sinclair J., Sinclair L. *Corpus, concordance, collocation*. – Oxford University Press, USA, 1991. – 170 pp.
- [19] Roth C., Cointet J. P. Social and semantic coevolution in knowledge networks // *Social Networks*. – 2010. – Vol. 32 (1). – P. 16-29.
- [20] Bekkerman R., Allan J. Using bigrams in text categorization. – Technical Report IR-408, Center of Intelligent Information Retrieval, UMass Amherst, 2004. – P. 1-10.

**Nenko Alexandra Evgenievna**, PhD, Associate Professor of Institute of Design and Urban Studies, ITMO University; project coordinator of CGES SPbSU, 199034, St. Petersburg, Birzhevaya Line, 14, letter A, e-mail: [al.nenko@itmo.ru](mailto:al.nenko@itmo.ru), Centre for German and European Studies, St. Petersburg State University, [elibrary.ru: author=966941](mailto:author=966941), ORCID: [orcidID = 0000-0003-3436-1069](https://orcid.org/0000-0003-3436-1069).

**Nedoseka Elena Vladimirovna**, senior researcher, Sociological Institute of FCTAS RAS, 190005, Saint-Petersburg, ul. 7-ya Krasnoarmeyskaya, d. 25/14, e-mail: [nedelena@socinst.ru](mailto:nedelena@socinst.ru), [elibrary.ru: authorID=:664872](mailto:elibrary.ru:authorID=:664872), ORCID [ID: 0000-0003-1944-0367](https://orcid.org/0000-0003-1944-0367).

**Galaktionova Anastasiia Alekseevna**, PhD student of Institute of Design and Urban Studies, ITMO University, 199034, St. Petersburg, Birzhevaya line, 14, lit. A, e-mail: [aagalaktionova@itmo.ru](mailto:aagalaktionova@itmo.ru), ORCID: [orcidID = 0000-0003-3767-7237](https://orcid.org/0000-0003-3767-7237).