

© 2023 г.

А.В. СМИРНОВ

## РОССИЙСКАЯ СОЦИОЛОГИЯ В УСЛОВИЯХ ЦИФРОВИЗАЦИИ ОБЩЕСТВА: РЕЗУЛЬТАТЫ АНАЛИЗА КОРПУСА НАУЧНЫХ ТЕКСТОВ

---

СМИРНОВ Андрей Владимирович – кандидат экономических наук, старший научный сотрудник Института социально-экономических и энергетических проблем Севера ФИЦ Коми НЦ УрО РАН, Сыктывкар, Россия (av.smirnov.ru@gmail.com).

---

**Аннотация.** На массиве публикаций из восьми ведущих социологических журналов за период с 2000 по 2021 г. методами анализа текстов изучаются особенности влияния цифровизации общества на социологию. Частотный анализ 13,8 тыс. научных текстов позволил отследить введение в научный оборот концептов, связанных с цифровизацией. Выявлялись различия между журналами. Изучался опыт использования цифровых социальных платформ в качестве источника данных и объекта осмысления для социологии. Исследование показало, что социологи к ним обращаются все чаще: если с 2010 по 2012 г. они упоминались только в 2,7% публикаций, то с 2019 по 2021 г. – почти в каждой пятой. Исследование подтвердило, что благодаря внедрению новых цифровых технологий возрастает частота использования и сложность применяемого социологами программного инструментария. Было установлено, что с 2015 г. растет число исследований, авторы которых разрабатывают собственный программный инструментарий сбора и анализа данных. Обозначены перспективы исследования закономерностей развития социологической науки методами корпусной лингвистики.

**Ключевые слова:** цифровизация • инструментарий • цифровые платформы • источники данных • российская социология • корпусная лингвистика

DOI: 10.31857/S013216250022128-3

**Введение.** Цифровые технологии влияют на социологию в разных аспектах: возникают новые объекты изучения и концепты, появляются цифровые источники данных, эволюционирует исследовательский инструментарий, родилось понятие «цифровая социология» [Lupton, 2014], с 2018 г. выходит посвященный ей специализированный журнал<sup>1</sup>. Д. Булье выделяет в развитии социологии три этапа. На первом источником данных были статистические методы и переписи, на втором – опросы общественного мнения, а на третьем такими источниками становятся цифровые следы – отпечатки активности человека в цифровом пространстве [Boullier, 2017; Дудина, 2021: 4–5]. В статье предпринята попытка изучения изменений в российской социологии с использованием методов анализа корпуса научных текстов – публикаций в ведущих социологических журналах России

---

Исследование выполнено за счет гранта РФФИ, проект № 21-78-00081.

<sup>1</sup> Журнал *Цифровая социология* издается Государственным университетом управления (Москва) в печатном и электронном форматах. Его периодичность четыре номера в год. URL: <https://digitalsociology.giu.ru/jour> (дата обращения: 04.03.2023).

за 2000–2021 гг. Методы корпусной лингвистики [Ignatow, 2015; Бызов, 2019; Zinn, 2020] применимы не только для изучения социальных явлений, но и закономерностей развития науки, поскольку частотный и контент-анализ обширного корпуса текстов позволяет выявить изменения в тематическом репертуаре статей, источниках данных и инструментари исследований.

**Социология в условиях цифровизации.** Рассмотрим наиболее существенные аспекты влияния цифровизации на социологию.

**1. Цифровая среда как объект осмысления социологии.** Ранние социологические исследования цифровых технологий концентрировались в основном на осмыслении цифрового неравенства в доступе к технологиям, на навыках и мотивации людей [Dimaggio, Hargittai, 2001]. Сегодня предметом осмысления стал широкий набор тематик, среди которых влияние интернета на социальную жизнь, соотношение «реального» и «виртуального», действия людей в цифровой среде и многие другие [Колозариди, Макушева, 2018]. Разрабатываются теоретические концепты и модели цифрового общества [Katzenbach, Bächle, 2019], а количество публикаций в социальных науках, посвященных «цифровой» тематике, за последние годы увеличилось многократно [Смирнов, 2021: 132].

Анализ сети цитирований ведущих англоязычных социологических журналов показал, что наиболее влиятельными в контексте цитирований в последние годы были статьи, связанные с цифровыми технологиями и большими данными [Булычев, Мальцева: 129]. Это обзор Д. Лэйзера и Дж. Рэдфорда «Data ex Machina», посвященный применению больших данных в изучении социальной жизни [Lazer, Radford, 2017], статья С. Брейна о внедрении больших данных в полицейском управлении Лос-Анджелеса [Brayne, 2017] и статья А. Кристин о количественной оценке веб-журналистики в США и Франции [Christin, 2018]. В российской социологии тоже следует ожидать освещения цифровой тематики. Поэтому *первая гипотеза* исследования состоит в том, что концепты, связанные с цифровизацией общества, все чаще используются в российских социологических исследованиях.

На содержание журналов помимо самой логики развития науки оказывает влияние редакционная политика. Научные издания могут придерживаться меритократической линии, то есть принимать в печать публикации независимо от тематики, статуса и места работы авторов, либо ставить своей целью репрезентировать всю дисциплину, публикуя материалы по определенному набору тематик и обеспечивая представительство авторов из разных регионов [Губа, 2019: 20]. Поэтому анализ необходимо проводить в разрезе редакционной стратегии научных журналов, выявляя различия между ними.

**2. Цифровая среда как источник данных.** Ключевой характеристикой современной науки является нацеленность на анализ огромных объемов данных [Kitchin, 2014]. Применительно к социальным наукам их важнейшим источником является интернет-среда. Сегодня практически любой аспект повседневной жизни находит цифровое отражение [Дудина, 2016: 21]. Данные, получаемые в цифровой среде, являются одновременно огромными и микроскопическими – «огромными в том смысле, что число изучаемых людей может исчисляться миллионами, а данные измеряться терабайтами, и микроскопическими в том смысле, что регистрируются отдельные микровзаимодействия» [Golder, Macy, 2014: 131].

Источники цифровых данных обладают большим разнообразием: социальные сети, поисковые системы, логи звонков, показания датчиков GPS, оцифрованные тексты, масштабные административные данные [Богданов, Смирнов, 2021: 306]. Наибольший интерес для социологов представляют социальные интернет-платформы, где содержание генерируется самими пользователями. Они дают практически безграничный простор для сбора и анализа данных. «Развитый капитализм двадцать первого столетия постепенно выстроился вокруг задачи извлечения и использования особого типа сырья – данных... Подобно нефти, данные есть сырье, которое извлекают, очищают и используют самым различным образом. Чем больше у кого-то данных, тем больше различных возможностей их использования» [Срничек, 2020: 37].

Цифровые данные обладают как преимуществами, так и недостатками. «Одним из главных ограничений классических массовых опросов является их реактивность – респонденты и испытуемые практически всегда знают, что участвуют в исследовании, а это, в свою очередь, может оказать эффект на результаты исследования» [Богданов, Смирнов, 2021: 307]. Цифровые источники данных обычно лишены этого недостатка, в связи с чем получил распространение термин «незапрошенное общественное мнение», отделяющий данные, которые стихийно формируются и «извлекаются» из сети как цифровой след от общественного мнения, конструируемого в результате массового опроса [Дудина, Юдина, 2017: 65]. Разумеется, данные цифровых платформ имеют и свои недостатки. Масштаб и кажущаяся полнота данных часто скрывают важные проблемы, связанные с репрезентативностью и обобщаемостью. Интернет-сообщества отличаются по демографической структуре от сообществ реального мира. Более того, в интернете не все записи созданы людьми. Они могут создаваться чат-ботами, которые используются организациями для рассылок, приема платежей и т.д., а также злоумышленниками в мошеннических целях [Golder, Macy, 2014; Lazer, Radford, 2017]. Поэтому цифровые данные по возможности необходимо дополнять традиционными, а также сопоставлять и верифицировать полученные по ним результаты.

И все же исследователи отмечают, что «возрастающий объем цифровых данных и изменение структуры их использования начинают играть более значимую роль в современном научном познании» [Журавлева, 2012: 121]. Чтобы проверить, насколько это справедливо для российской социологии, сформулируем вторую гипотезу нашего исследования. Она состоит в том, что социальные интернет-платформы становятся важным объектом изучения и источником данных российской социологической науки. Для ее подтверждения будет проведен контент-анализ публикаций, в которых упоминаются названия наиболее популярных платформ, что позволит определить, какие интернет-платформы используются социологами чаще других и в каких исследовательских целях.

**3. Изменение инструментария социологов под влиянием цифровизации.** Задачи извлечения из цифровых источников и обработки больших объемов данных требуют от ученых овладения такими компетенциями, как программирование и анализ данных [Смирнов, 2015]. Навыки написания алгоритмов для выгрузки данных (парсинг) и взаимодействия с программными интерфейсами интернет-платформ (API) позволяют исследователям получать сгенерированные в цифровой среде данные. Новые инструменты могут применяться социологами как самостоятельно, так и в рамках междисциплинарного взаимодействия с представителями точных наук. Основные технологии, полезные социологу, развиваются в настоящее время в рамках четырех подходов: интеллектуальный анализ данных (data mining), большие данные (big data), цифровые гуманитарные науки (digital humanity) и наука о данных (data science) [Толстова, 2015: 4]. Эти подходы не имеют четких границ и взаимно пересекаются. Новые технологии позволяют получать нетривиальные выводы и генерировать знания из массивов данных несоциологического характера.

Логично выдвинуть *третью гипотезу* исследования, согласно которой со временем должна возрасть как частота применения новых инструментов, так и их сложность. Если раньше социологи применяли преимущественно офисное программное обеспечение (ПО) общего назначения и статистические пакеты типа SPSS (Statistical Package for the Social Science), то со временем должно применяться все более специализированное ПО, в том числе собственной разработки. Очень широк перечень конкретных методов, технологий, языков программирования и программных средств, которые могут использоваться в социологических исследованиях. Правила оформления статей в ведущих журналах требуют от авторов подробного изложения методики исследований. Поэтому в научных текстах могут фигурировать названия программных продуктов и технологий, которыми пользовались социологи.

Опубликован ряд работ, в которых на основе анализа сотен журнальных текстов, авторы делали выводы о развитии методов социологических исследований. А.В. Кинчарова и М.М. Соколов проанализировали 322 статьи в журнале «Социологические

исследования» за три года – 1981, 1996 и 2011. Они пришли к выводу, что за 30 лет «предпочитаемые методы сбора данных практически не изменились (исключение – распространение качественных методов), а использование статистических методов осталось на низком уровне». Причину этого они видят в невозможности долгосрочного планирования в кризисной ситуации, в которой оказалась российская наука. Затраты на изучение сложных методов и инструментов ученым казались не рационально высокими [Кинчарова, Соколов, 2015]. Г.Г. Татарова и А.В. Кученкова рассмотрели 178 публикаций в журнале «Социологические исследования» по методологическим направлениям за 2000–2018 гг., а также 362 статьи в журнале «Социология: 4М» за 1991–2018 гг. Они показали, что «методная» проблематика, в особенности вопросы математического моделирования, «полевых» проблем социолога и некоторых других направлений представлены в ведущих российских журналах довольно скупо [Татарова, Кученкова, 2020].

В приведенных работах анализируются только отдельные элементы публикаций: тематическая направленность, описание выборки, методологические особенности. Другой подход с анализом текстов целиком методами корпусной лингвистики представлен Ф. Биллари и Э. Загени, которые для выделения самых востребованных инструментов анализа демографических проблем в разные периоды времени использовали сервис Google Books Ngram Viewer. С его помощью были построены графики частотности таких терминов, как «life table» (таблица смертности) и «proportional hazards» (пропорциональные риски), на основе всех англоязычных текстов в Google Books с 1800 года [Billari, Zagheni, 2017]. Похожий анализ мы применим к российским социологическим текстам.

**Методы и данные.** Для анализа эволюции отечественных социологических публикаций в течение XXI в. требовалось отобрать ведущие научные журналы, соответствующие двум критериям. Во-первых, их тексты за последние 20 лет должны быть размещены в открытом доступе. Во-вторых, они должны быть востребованы научным сообществом. Поскольку разные наукометрические показатели дают совершенно разные перечни журналов<sup>2</sup>, было решено использовать показатель, полученный экспертным путем – результат общественной экспертизы РИНЦ, в которой принимали участие высокоцитируемые ученые. Всего восемь социологических журналов выходят на протяжении как минимум двух десятилетий и получили в ходе общественной экспертизы оценку более 3 баллов по 5-балльной шкале (табл. 1). Все они индексируются в международных научных базах данных. Представлены журналы, выпускаемые как академическими институтами, так и ведущими вузами. Один из журналов издается государственной организацией ВЦИОМ.

Все научные статьи, обзоры, рецензии и другие материалы, опубликованные в журналах за 2000–2021 гг., загружались с официальных сайтов журналов и преобразовывались в текстовый формат. Общее количество текстов корпуса составило 13 772 (60,1 млн слов). Из них 46% приходится на «Социологические исследования», который оценивается многими российскими социологами как наиболее читаемый в РФ [Губа, 2019: 21] и как лучшее социологическое издание [Соколов, 2021: 53]. Высокий удельный вес этого журнала (33% от общего числа слов в корпусе из 8 ведущих отечественных социологических журналов) вызван его ежемесячной периодичностью. Во времени тексты распределены довольно равномерно: от 536 до 697 текстов в год. Средний объем статьи вплоть до 2013 г. менялся незначительно и составлял около 4000 слов. С 2014 г. он начал расти, достигнув 5814 слов к 2020 г. Это вызвано главным образом переориентацией журналов на международные требования – добавились список литературы и аннотация на английском языке.

Все тексты прошли предварительную обработку, в ходе которой из них были исключены знаки препинания, цифры, заглавные буквы и другие элементы, затрудняющие анализ (токенизация). Для удобства сравнения все слова на русском языке были приведены

---

<sup>2</sup> Рассматривались такие показатели, как Science Index, двух- и пятилетние импакт-факторы журналов по РИНЦ и ядру РИНЦ с учетом и без учета самоцитирования. Во всех случаях перечни журналов различались, а у ряда журналов значения показателей отсутствовали.

Таблица 1

## Состав корпуса текстов ведущих российских журналов по социологии

| Название журнала   | Учредитель     | Входит в базы данных | Показатели за 2020 г. |       |       | Номеров в год | Число текстов *** | Слов, млн *** |
|--|----------------|----------------------|-----------------------|-------|-------|---------------|-------------------|---------------|
|  |                |                      | Science Index         | ОЭ *  | ИФ ** |               |                   |               |
| Социологические исследования   | РАН, ФНИСЦ РАН | WoS, Scopus, RSCI    | 3,531                 | 3,690 | 0,526 | 12            | 6 297             | 19,4          |
| Мониторинг общественного мнения: экономические и социальные перемены | ВЦИОМ          | Scopus, RSCI         | 2,688                 | 3,079 | 0,461 | 4–6           | 1 794             | 8,7           |
| Журнал социологии и социальной антропологии                          | СПбГУ          | RSCI                 | 0,924                 | 3,195 | 0,274 | 4–6           | 1 466             | 7,6           |
| Экономическая социология   | НИУ ВШЭ        | WoS, Scopus, RSCI    | 1,544                 | 3,052 | 0,561 | 5             | 1 266             | 6,6           |
| Мир России. Социология. Этнология                                    | НИУ ВШЭ        | WoS, Scopus, RSCI    | 3,470                 | 3,050 | 0,805 | 4             | 677               | 5,8           |
| Социологическое обозрение  | НИУ ВШЭ        | WoS, Scopus, RSCI    | 0,804                 | 3,073 | 0,370 | 1–4           | 811               | 5,4           |
| Социологический журнал   | ФНИСЦ РАН      | Scopus, RSCI         | 1,157                 | 3,357 | 0,421 | 2–4           | 1 138             | 5,0           |
| Социология: методология, методы, математическое моделирование (4М)   | ФНИСЦ РАН      | RSCI                 | 0,107                 | 3,200 | 0,235 | 1–2           | 323               | 1,6           |

*Примечания.* \*Результаты общественной экспертизы РИНЦ высокоцитируемыми учеными по 5-балльной шкале; \*\*импакт-фактор по ЯДРУ РИНЦ за 5 лет без самоцитирования; \*\*\* всего за 2000–2021 гг.

в нормальную/начальную форму (лемматизация) с помощью морфологического анализатора `rumorphy2` на языке Python [Korobov, 2015]. Анализ текстов производился методами корпусной лингвистики. Была составлена матрица терм-документ, позволяющая получить список текстов, в которых встречается любое слово (словосочетание). Использовалась модель, определяющая частоту встречаемости слов в корпусе текстов в разрезе временных периодов и журналов. Алгоритмы расчетов реализованы автором с использованием языка программирования Julia и пакетов обработки текстовых и табличных данных: `StringAnalysis.jl`, `Languages.jl`, `PDFIO.jl` и `DataFrames.jl`.

**Результаты и их обсуждение.** Частота встречаемости основных концептов цифрового общества многократно увеличилась (рис. 1). Если слово «интернет» в 2021 г. стало встречаться всего втрое чаще, чем в 2000 г., то «онлайн» – в 22 раза, «платформа» – в 24 раза, «цифровой» и «цифровизация» – в 80 раз. Применительно к последним особо быстрый прирост фиксируется после 2017 г. – в 4,9 раза всего за 4 года. Эти значения свидетельствуют о том, что социология обратила пристальное внимание на проблематику, связанную с процессами цифровизации общественной жизни.

При рассмотрении частоты появления слов «цифровой» и «цифровизация» в текстах журналов (табл. 2) выявлена следующая закономерность: частота их употребления многократно возросла во всех журналах. Раньше остальных цифровая проблематика

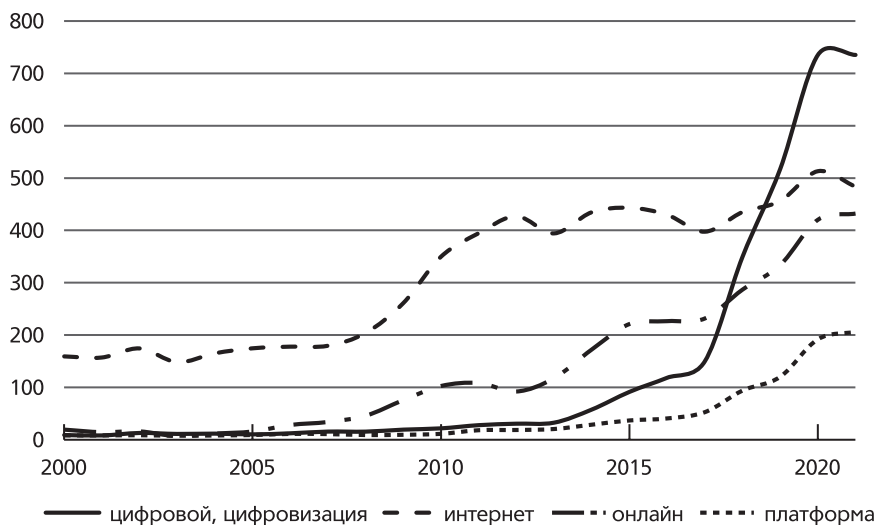


Рис. 1. Частота встречаемости терминов, связанных с цифровизацией, скользящее среднее за 3 года, 2000–2021 гг., ед. на 1 млн слов

стала широко освещаться в журнале «Социология: 4М» (2013–2015 гг.), что было связано с его методологической направленностью. Однако в последние годы самую высокую частоту упоминаемости вышеназванных терминов демонстрирует журнал «Мониторинг общественного мнения: экономические и социальные перемены». Это вызвано редакционной политикой данного издания. Цифровым исследованиям посвящены многие его специальные тематические выпуски: «Интернет как проблемное поле социальных наук» (№ 1, 2018), «Цифровизация экономики, политики, гражданского общества» (№ 5, 2019), «Искусственный интеллект и искусственная социальность» (№ 1, 2021), «Трансформация гражданского активизма под влиянием новых ИКТ» (№ 6, 2021). При их подготовке

Таблица 2

Частота встречаемости слов «цифровой» и «цифровизация» в корпусе текстов по периодам, 2001–2021 гг., ед. на 1 млн слов

| Журнал   | 2001–2003 | 2004–2006 | 2007–2009 | 2010–2012 | 2013–2015 | 2016–2018 | 2019–2021 |
|--|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Мониторинг общественного мнения: экономические и социальные перемены | 2         | 6         | 19        | 48        | 47        | 289       | 1268      |
| Журнал социологии и социальной антропологии                          | 21        | 17        | 7         | 42        | 55        | 78        | 748       |
| Социологические исследования   | 18        | 11        | 17        | 32        | 89        | 188       | 712       |
| Социологический журнал   | 12        | 9         | 14        | 34        | 18        | 68        | 464       |
| Экономическая социология   | 5         | 10        | 7         | 18        | 64        | 108       | 462       |
| Социологическое обозрение  | 3         | 34        | 23        | 6         | 15        | 100       | 405       |
| Социология: 4М   | 11        | 9         | 77        | 17        | 299       | 227       | 341       |
| Мир России   | 14        | 3         | 6         | 3         | 2         | 14        | 161       |

редакцией целенаправленно отбирались материалы соответствующей проблематики с помощью объявлений на сайте и других сетевых механизмов. Частотный анализ текстов подтверждает первую гипотезу исследования. Однако необходимо учитывать, что скорость распространения в научных публикациях тематики, связанной с цифровизацией общества, неодинакова по журналам и зависит от их редакционной политики.

Названия цифровых социальных платформ упоминаются в 783 публикациях (5,7% от общего числа текстов в корпусе), из которых 387 приходятся на 2019–2021 гг. (19,5% от общего числа текстов за эти годы). То есть в последние годы названия платформ упоминаются в каждой пятой статье, что подтверждает вторую гипотезу исследования. Для поиска использовались названия пятнадцати наиболее популярных в России на 2020 г. цифровых социальных платформ<sup>3</sup>, а также различные варианты их транслитераций: YouTube, ВКонтакте, WhatsApp, Instagram, Одноклассники, Viber, Facebook<sup>4</sup>, TikTok, Telegram, Skype, Twitter, Pinterest, Snapchat, Twitch, Likee (табл. 3). Чаще других в статьях упоминаются Facebook (386 текстов), ВКонтакте (321), YouTube (233) и Twitter (203). Все тексты, в которых упоминаются цифровые платформы, были классифицированы по месту и характеру упоминания (табл. 3).

В 33 публикациях названия платформ вынесены в названия. Чаще всего это социальная сеть ВКонтакте (19 публикаций) и видеохостинг YouTube (5). Из 33 статей 25 были опубликованы в 2018–2021 гг. В основном тексте публикаций названия социальных платформ встречаются в 665 публикациях, а в списке литературы, гиперссылках и сносках на источники – в 328. В 101 публикации название цифровой платформы упоминается как средство для поиска и рекрутинга респондентов (в основном используются социальные сети ВКонтакте, Одноклассники и Facebook) или проведения интервью (чаще всего Skype).

Таблица 3

**Число статей, в которых упоминаются наиболее популярные в России социальные цифровые платформы по периодам и категориям, 2000–2021 гг., ед.**

| Категория текстов  | 2000–2009 | 2010–2012 | 2013–2015 | 2016–2018 | 2019–2021 | Всего |
|--|-----------|-----------|-----------|-----------|-----------|-------|
| Всего текстов в корпусе  | 5835      | 1924      | 1966      | 2062      | 1985      | 13772 |
| Текстов с упоминанием цифровых платформ  | 4         | 53        | 126       | 213       | 387       | 783   |
| В них:   |           |           |           |           |           |       |
| 1. Упоминания в названии публикации  | 0         | 3         | 3         | 12        | 15        | 33    |
| 2. Упоминания в основном тексте  | 4         | 49        | 104       | 181       | 327       | 665   |
| 3. Упоминания в списке литературы или гиперссылках                                       | 1         | 17        | 47        | 87        | 176       | 328   |
| 4. Цифровая платформа использовалась для рекрутинга респондентов или проведения интервью | 0         | 4         | 17        | 26        | 54        | 101   |
| 5. Цифровая платформа является источником эмпирических данных                            | 1         | 9         | 18        | 37        | 82        | 147   |
| 6. Цифровая платформа является объектом исследования                                     | 2         | 17        | 17        | 33        | 74        | 143   |

<sup>3</sup> Leading social media platforms in Russia as of 3rd quarter of 2020, by penetration rate. URL: <https://www.statista.com/statistics/867549/top-active-social-media-platforms-in-russia/>

<sup>4</sup> Социальные сети Facebook и Instagram принадлежат компании Meta, деятельность которой в 2022 г. признана в России экстремистской и запрещена.

В 147 научных статьях цифровые платформы используются как источники эмпирических данных для исследования. Это может быть анализ сообщений, учетных записей и сообществ в социальных медиа, изучение сетей взаимодействия людей, контент-анализ или анализ статистики популярности материалов. В 143 публикациях сам объект изучения является цифровой платформой или тесно связанным с ней явлением: сетевые коммуникации и дискурсы, стратегии использования соцсетей и интернет-практики, онлайн-цензура, цифровой активизм, распространение мнений и фейков, мемы, цифровые следы и извлечение веб-данных, социальные боты и искусственные профили, виртуальная самопрезентация, интернет-знаменитости. Эта группа публикаций сильно пересекается с предыдущей.

Для проверки третьей гипотезы разделим социологический программный инструментарий на три группы и выберем в каждой несколько популярных представителей. В первую группу входят статистические пакеты общего назначения (SPSS, Stata, Statistica), которые могут применяться не только в социологии, но и в других науках, включая точные и естественные. Во вторую группу входят инструменты анализа качественных данных (ATLAS.ti, NVivo, MAXQDA), которые решают более узкий класс задач, связанных с разметкой и изучением текстов, изображений, аудио- и видеозаписей. В третью группу (Python, API, GitHub) вошли технологии, упоминание которых может свидетельствовать о том, что авторы самостоятельно разрабатывают или адаптируют программное обеспечение под свою исследовательскую задачу. Python – один из наиболее популярных языков программирования, API (Application Programming Interface) – инструменты для выгрузки данных приложений или цифровых платформ, GitHub – самый популярный сервис для размещения IT-проектов в интернете.

Число текстов, в которых упоминаются статистические пакеты, было высоким на протяжении всего периода и выросло в 2,5 раза (рис. 2). Упоминания пакетов для анализа качественных данных начали увеличиваться в 2013 г. и за весь изучаемый период выросли в 13,3 раза. Это может быть связано как с увеличением популярности качественных методов, так и с повышением доступности специализированного программного обеспечения для научных организаций и ученых. Третья группа слов почти не упоминалась вплоть до

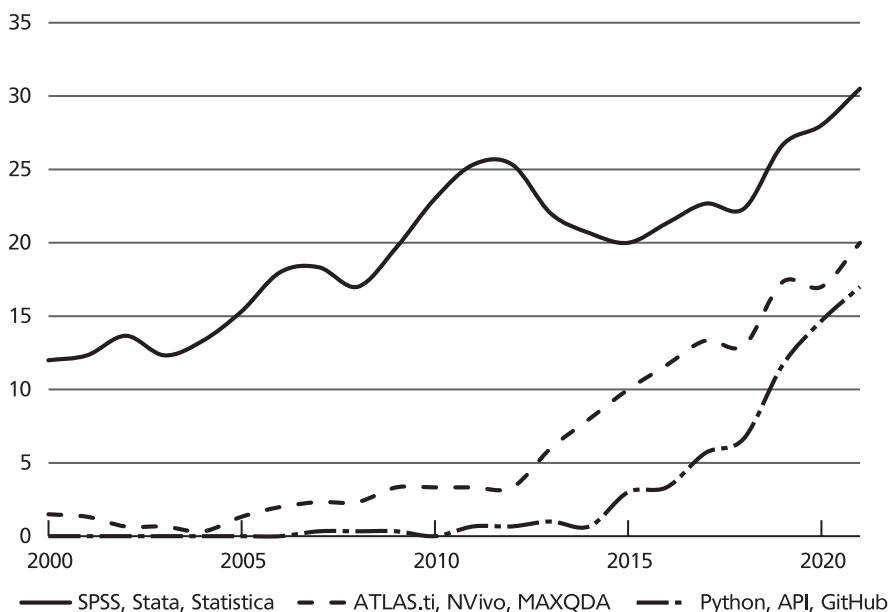


Рис. 2. Число публикаций с упоминанием разных видов программного обеспечения и технологий, скользящее среднее за 3 года, 2000–2021 гг., ед.



2015 г., но затем их употребление стало резко увеличиваться. К 2021 г. количество статей, в которых они были упомянуты, почти сравнялось со второй группой. Рассмотрим эти статьи и особенности применения авторами каждой технологии в отдельности.

Из 29 статей, в которых упоминается язык программирования Python, в 20 случаях их авторы реализовывали на нем алгоритмы для своих исследований. Хотя первая такая статья появилась в 2012 г., 13 из 20 статей приходится на 2020–2021 гг. Слово «GitHub» впервые встречается в корпусе в 2016 г. Из 21 статьи, в которых упоминается GitHub, в 8 даются ссылки на используемые алгоритмы и методы чужой разработки, а еще в 8 – на собственные данные, методы и результаты. Чаще всего авторы размещают код на языках Python и R. Аббревиатура «API» встречается в 29 публикациях, наиболее ранняя из которых датирована 2014 г. Однако широкое распространение API-социология в России получила в 2018 г., когда статьи стали появляться регулярно. Из 20 статей, в которых описан опыт применения авторами API, в 15 случаях использовался API социальной сети ВКонтакте. Похожие закономерности демонстрируют и другие маркеры. Так, слово «парсер» начинает использоваться в статьях из ведущих социологических журналов в 2015 г. Словосочетание «большие данные» не встречалось в корпусе текстов ранее 2013 г., а относительно высокой частоты употребления оно достигло в 2015 г. Таким образом, в последние годы наблюдается активизация разработки собственных алгоритмов, но она еще не носит массовый характер, число статей исчисляется десятками.

Хотя выявленные закономерности и могут свидетельствовать о справедливости третьей гипотезы исследования, необходимо учитывать, что далеко не во всех статьях указывается используемый программный инструментарий. Вероятно также, что авторы более склонны явно указывать на использование экзотического программного обеспечения, а не простых табличных редакторов и статистических пакетов. Поэтому для проверки этой гипотезы требуется проведение опроса ученых-социологов об используемых ими средствах и инструментах обработки данных.

**Заключение.** Контент-анализ публикаций подтвердил, что цифровые социальные платформы становятся важным объектом изучения и источником данных российской социологии. За пределами нашего анализа остались многие другие цифровые сервисы, используемые учеными, такие как анализаторы поисковых запросов, геоинформационные системы, тематические сайты.

Анализ корпуса научных текстов продемонстрировал многочисленные изменения в публикациях российских социологов, ставшие результатом цифровизации и появления новых информационных технологий. Есть основания полагать, что выявленные тенденции сохранятся в будущем, так как все рассмотренные показатели демонстрируют устойчивый рост во времени. Генеральная совокупность была сознательно ограничена текстами из восьми ведущих журналов и двумя первыми десятилетиями XXI в., чтобы их удалось охватить целиком. Однако в дальнейших исследованиях корпус текстов может быть расширен как во времени, так и по охвату изданий, включать журналы по смежным наукам. С помощью анализа текстов могут изучаться и другие вопросы развития социологической мысли, совершенствования методов социологических исследований. Помимо выявления частотных характеристик методы анализа текстов могут использоваться для их классификации, анализа стилистических особенностей, создания тематических подборок литературы.

## СПИСОК ЛИТЕРАТУРЫ

- Богданов М.Б., Смирнов И.Б. Возможности и ограничения цифровых следов и методов машинного обучения в социологии // Мониторинг общественного мнения: экономические и социальные перемены. 2021. № 1. С. 304–328. DOI: 10.14515/monitoring.2021.1.1760.
- Булычева Е.Е., Мальцева Д.В. Выделение актуальных тематик в социологии: взгляд сквозь призму анализа сети цитирований // Мониторинг общественного мнения: экономические и социальные перемены. 2020. № 6. С. 113–140. DOI: 10.14515/monitoring.2020.6.971.

- Бывов А.А. Интеллектуальный анализ текстов в социальных науках // Социология: методология, методы, математическое моделирование. 2019. № 49. С. 131–160.
- Губа К.С. Быть главным журналом в российской социологии: когда миссия имеет значение // Экономическая социология. 2019. Т. 20. № 4. С. 14–38. DOI: 10.17323/1726-3247-2019-4-14-38.
- Дудина В.И. «Пересборка социологии»: цифровой поворот и поиски новой теоретической оптики // Социологические исследования. 2021. № 11. С. 3–11. DOI: 10.31857/S013216250016829-4.
- Дудина В.И. Цифровые данные – потенциал развития социологического знания // Социологические исследования. 2016. № 9. С. 21–30.
- Дудина В.И., Юдина Д.И. Извлекая мнения из сети Интернет: могут ли методы анализа текстов заменить опросы общественного мнения? // Мониторинг общественного мнения: Экономические и социальные перемены. 2017. № 5. С. 63–78. DOI: 10.14515/monitoring.2017.5.05.
- Журавлёва Е.Ю. Эпистемический статус цифровых данных в современных научных исследованиях // Вопросы философии. 2012. № 2. С. 113–123.
- Кинчарова А.В., Соколов М.М. Исследовательские практики российских социологов // Социологические исследования. 2015. № 6. С. 58–68.
- Колозариди П.В., Макушева М.О. Интернет как проблемное поле социальных наук // Мониторинг общественного мнения: Экономические и социальные перемены. 2018. № 1. С. 1–11. DOI: 10.14515/monitoring.2018.1.01.
- Смирнов А.В. Цифровое общество: теоретическая модель и российская действительность // Мониторинг общественного мнения: экономические и социальные перемены. 2021. № 1. С. 129–153. DOI: 10.14515/monitoring.2021.1.1790.
- Смирнов В.А. Новые компетенции социолога в эпоху больших данных // Мониторинг общественного мнения: экономические и социальные перемены. 2015. № 2. С. 44–54. DOI: 10.14515/monitoring.2015.2.04.
- Соколов М.М. Академические репутации в российской социологии: опыт измерения // Социологические исследования. 2021. № 3. С. 44–56. DOI: 10.31857/S013216250013728-3.
- Срничек Н. Капитализм платформ. 2-е изд. М.: ВШЭ, 2020.
- Татарова Г.Г., Кученкова А.В. «Методная» проблематика на страницах журнала «Социологические исследования» (2000–2018) // Социологические исследования. 2020. № 7. С. 47–56. DOI: 10.31857/S013216250009486-7.
- Толстова Ю.Н. Социология и компьютерные технологии // Социологические исследования. 2015. № 8. С. 3–13.
- Billari F., Zagheni E. Big Data and population processes: A revolution? // SIS 2017. Statistics and data science: new challenges, new generations. Florence: Firenze University Press, 2017. P. 167–178. DOI: 10.36253/978-88-6453-521-0.
- Boullier D. Big data challenges for the social sciences: from society and opinion to replications // ISA e-Symposium for Sociology. 2017. Vol. 7 (2). URL: <https://www.boullier.bzh/wp-content/uploads/EBul-Boullier-Jul2017.pdf> (дата обращения: 01.06.2022).
- Brayne S. Big data surveillance: The case of policing // American Sociological Review. 2017. Vol. 82 (5). P. 977–1008. DOI: 10.1177/0003122417725865.
- Christin A. Counting Clicks: Quantification and Variation in Web Journalism in the United States and France // American Journal of Sociology. 2018. Vol. 123 (5). P. 1382–1415. DOI: 10.1086/696137.
- Dimaggio P., Hargittai E. From the 'digital divide' to 'digital inequality': Studying Internet use as penetration increases. Working Paper #15. Princeton University: Center for Arts and Cultural Policy Studies, 2001.
- Golder S.A., Macy M.W. Digital footprints: opportunities and challenges for online social research // Annual Review of Sociology. 2014. Vol. 40 (1). P. 129–152. DOI: 10.1146/annurev-soc-071913-043145.
- Ignatow G. Theoretical foundations for digital text analysis // Journal for the Theory of Social Behaviour. 2015. Vol. 46 (1). P. 104–120. DOI: 10.1111/jtsb.12086.
- Katzenbach C., Bächle T.C. Defining concepts of the digital society // Internet Policy Review. 2019. Vol. 8 (4). DOI: 10.14763/2019.4.1430.
- Kitchin R. Big data, new epistemologies and paradigm shifts // Big Data & Society. 2014. Vol. 1 (1). P. 1–12. DOI: 10.1177/2053951714528481.
- Korobov M. Morphological analyzer and generator for Russian and Ukrainian languages // International Conference on Analysis of Images, Social Networks and Texts. Cham: Springer, 2015. P. 320–332.
- Lazer D., Radford J. Data ex machina: Introduction to big data // Annual Review of Sociology. 2017. Vol. 43 (1). P. 19–39. DOI: 10.1146/annurev-soc-060116-053457.
- Lupton D. Digital sociology. London, New York: Routledge, 2015.
- Zinn J.O. The UK at risk. A corpus approach to social change 1785–2009. Cham: Palgrave Macmillan, 2020. DOI: 10.1007/978-3-030-20238-5.

## RUSSIAN SOCIOLOGY IN THE CONTEXT OF SOCIETY DIGITALIZATION: RESULTS OF A CORPUS ANALYSIS OF SCIENTIFIC TEXTS

SMIRNOV A.V.

*Institute for Socio-Economic and Energy Problems of the North, Komi Science Centre of the UB RAS, Russia*

*Andrey V. SMIRNOV, Cand. Sci. (Econ.), Senior Research Fellow, Institute for Socio-Economic and Energy Problems of the North, Komi Science Centre of the UB RAS, Syktyvkar, Russia (av.smirnov.ru@gmail.com).*

**Acknowledgements.** The paper has been prepared with the support of the RSF, the project No. 21-78-00081.

**Abstract.** Using the analysis of a corpus of texts from eight leading Russian sociological journals, the article examines the impact of the digitalization of society on sociology in 2000–2021. Frequency analysis of 13.8 thousand scientific texts tracked the introduction of concepts related to digitalization into academic circulation. The article reveals the differences between the journals, due to their editorial policy in the selection of articles. Through content analysis of texts, the author studied the experience of using digital social platforms as a source of data and an object of reflection. The text shows that they are being used more and more over time, and since 2019, one in five publications mentions digital platforms. The study confirmed that due to the introduction of new digital technologies, the frequency and complexity of the software tools used by sociologists is increasing. Since 2015, there has been an increase in the number of studies whose authors develop their own algorithms for data collection and analysis. The identified trends are likely to continue in the future, as all the indicators considered show a steady increase over time. In the conclusions of the article, the author formulates the prospects for studying the patterns of development of sociological science using the methods of corpus linguistics.

**Keywords:** digitalization, tools, digital platforms, data sources, Russian sociology, corpus linguistics.

## REFERENCES

- Billari F., Zagheni E. (2017) Big Data and population processes: A revolution? *SIS 2017. Statistics and data science: new challenges, new generations*. Florence: Firenze University Press: 167–178. DOI: 10.36253/978-88-6453-521-0.
- Bogdanov M.B., Smirnov I.B. (2021) Opportunities and limitations of digital footprints and machine learning methods in sociology. *Monitoring obshchestvennogo mneniya: ekonomicheskiye i sotsial'nyye peremeny* [Monitoring of Public Opinion: Economic and Social Changes]. No. 1: 304–328. DOI: 10.14515/monitoring.2021.1.1760. (In Russ.)
- Boullier D. (2017) Big data challenges for the social sciences: from society and opinion to replications. *ISA eSymposium for Sociology*. Vol. 7 (2). URL: <https://www.boullier.bzh/wp-content/uploads/EBul-Boullier-Jul2017.pdf> (accessed 01.06.2022).
- Brayne S. (2017) Big data surveillance: The case of policing. *American Sociological Review*. Vol. 82 (5): 977–1008. DOI: 10.1177/0003122417725865.
- Bulycheva E.E., Maltseva D.V. (2020) Highlighting key topics in sociology: A glance through the prism of citation network analysis. *Monitoring obshchestvennogo mneniya: ekonomicheskiye i sotsial'nyye peremeny* [Monitoring of Public Opinion: Economic and Social Changes]. No. 6: 113–140. DOI: 10.14515/monitoring.2020.6.971. (In Russ.)
- Byzov A. (2019) Text mining in social sciences. *Sotsiologiya: metodologiya, metody, matematicheskoye modelirovaniye* [Sociology: methodology, methods, mathematical modeling]. No. 49: 131–160. (In Russ.)
- Christin A. (2018) Counting Clicks: Quantification and Variation in Web Journalism in the United States and France. *American Journal of Sociology*. Vol. 123 (5): 1382–1415. DOI: 10.1086/696137.
- Dimaggio P., Hargittai E. (2021) *From the 'digital divide' to 'digital inequality': Studying Internet use as penetration increases*. Working Paper #15. Princeton University: Center for Arts and Cultural Policy Studies, 2001.
- Dudina V.I. (2016) Digital data potentialities for development of sociological knowledge. *Sotsiologicheskie issledovaniya* [Sociological Studies]. No. 9: 21–30. (In Russ.)
- Dudina V.I. (2021) Reassembling sociology: Digital turn and searching for new theoretical optics. *Sotsiologicheskie issledovaniya* [Sociological Studies]. No. 11: 3–11. DOI: 10.31857/S013216250016829-4. (In Russ.)

- Dudina V.I., Iudina D.I. (2017) Mining opinions on the Internet: can the text analysis methods replace public opinion polls? *Monitoring obshchestvennogo mneniya: ekonomicheskiye i sotsial'nyye peremeny* [Monitoring of Public Opinion: Economic and Social Changes]. No. 5: 63–78. DOI: 10.14515/monitoring.2017.5.05.
- Golder S.A., Macy M.W. (2014) Digital footprints: opportunities and challenges for online social research. *Annual Review of Sociology*. Vol. 40 (1): 129–152. DOI: 10.1146/annurev-soc-071913-043145.
- Guba K.S. (2019) To be the flagship journal of Russian sociology: When the mission matters. *Ekonomicheskaya sotsiologiya* [Journal of Economic Sociology]. Vol. 20 (4): 14–38. DOI: 10.17323/1726-3247-2019-4-14-38. (In Russ.)
- Ignatow G. (2015) Theoretical foundations for digital text analysis. *Journal for the Theory of Social Behaviour*. Vol. 46 (1): 104–120. DOI: 10.1111/jtsb.12086.
- Katzenbach C., Bächle T.C. (2019) Defining concepts of the digital society. *Internet Policy Review*. Vol. 8 (4). DOI: 10.14763/2019.4.1430.
- Kintcharova A.V., Sokolov M.M. (2015) Research practices of Russian sociologists. *Sotsiologicheskie issledovaniya* [Sociological Studies]. No. 6: 58–68. (In Russ.)
- Kitchin R. (2014) Big data, new epistemologies and paradigm shifts. *Big Data & Society*. Vol. 1 (1): 1–12. DOI: 10.1177/2053951714528481.
- Kolozaridi P.V., Makusheva M.O. (2018) The Internet as a problematic field of study in social sciences. *Monitoring obshchestvennogo mneniya: ekonomicheskiye i sotsial'nyye peremeny* [Monitoring of Public Opinion: Economic and Social Changes]. No. 1: 1–11. DOI: 10.14515/monitoring.2018.1.01. (In Russ.)
- Korobov M. (2015) Morphological analyzer and generator for Russian and Ukrainian languages. *International Conference on Analysis of Images, Social Networks and Texts*. Cham: Springer: 320–332.
- Lazer D., Radford J. (2017) Data ex machina: Introduction to big data. *Annual Review of Sociology*. Vol. 43 (1): 19–39. DOI: 10.1146/annurev-soc-060116-053457.
- Lupton D. (2015) *Digital sociology*. London, New York: Routledge.
- Smirnov A.V. (2021) Digital society: Theoretical model and Russian reality. *Monitoring obshchestvennogo mneniya: ekonomicheskiye i sotsial'nyye peremeny* [Monitoring of Public Opinion: Economic and Social Changes]. No. 1: 129–153. DOI: 10.14515/monitoring.2021.1.1790. (In Russ.)
- Smirnov V.A. (2015) Sociologist's new competences in the times of "big data". *Monitoring obshchestvennogo mneniya: ekonomicheskiye i sotsial'nyye peremeny* [Monitoring of Public Opinion: Economic and Social Changes]. No. 2: 44–54. DOI: 10.14515/monitoring.2015.2.04. (In Russ.)
- Sokolov M.M. (2021) Academic recognition in Russian sociology: a study using reputation surveys. *Sotsiologicheskie issledovaniya* [Sociological Studies]. No. 3: 44–56. (In Russ.)
- Srnicek N. (2020) *Platform capitalism*. Moscow: VSHE. (In Russ.)
- Tatarova G.G., Kuchenkova A.V. Methodological issues on the pages of the journal "Sociological studies" (2000–2018). *Sotsiologicheskie issledovaniya* [Sociological Studies]. No. 7: 47–56. (In Russ.)
- Tolstova Yu.N. (2015) Sociology and computer technologies. *Sotsiologicheskie issledovaniya* [Sociological Studies]. No. 8: 3–13. (In Russ.)
- Zhuravleva E.Yu. (2012) Epistemic status of digital data in modern scientific research. *Voprosy Filosofii* [Questions of Philosophy]. No. 2: 113–123. (In Russ.)
- Zinn J.O. (2020) *The UK at risk. A corpus approach to social change 1785–2009*. Cham: Palgrave Macmillan. DOI: 10.1007/978-3-030-20238-5.

Received: 25.07.22. Accepted: 14.03.23.