

Социология ценностей: опыт построения таксономии путем использования технологии анализа естественного языка

УДК 303.1 DOI 10.26425/2658-347X-2023-6-1-48-58

Получено 07.02.2023

Доработано после рецензирования 09.03.2023

Принято 22.03.2023

Кашина Марина Александровна

Д-р полит. наук, проф. каф. социальных технологий

ORCID: 0000-0003-4509-8657

E-mail: kashina-ma@ranepa.ru

Северо-Западный институт управления – филиал Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации, г. Санкт-Петербург, Россия

Ткач Сергей

Программист

ORCID: 0000-0001-5016-9187

E-mail: sergey@ngostellit.ru

Научный парк Санкт-Петербургского государственного университета, г. Санкт-Петербург, Россия

АННОТАЦИЯ

Современные исследования в области социологии науки становятся все более сложными из-за постоянно растущей публикационной активности авторов. Для отслеживания трендов в отраслевых социологиях ученые обращаются к методам наукометрии, однако их оказывается недостаточно. Предметом исследования выступила тенденция развития социологии ценностей как отрасли социологии. Цель статьи – оценка возможностей использования методов анализа естественного языка NLP/NLA (англ. natural language processing – обработка текста на естественном языке, natural language assessment – оценка естественного языка) для тематической и теоретической кластеризации исследований в социологии ценностей. Дизайн исследования количественно-качественный, работа проводилась в два этапа. На первом этапе с помощью интеллектуального анализа текста проанализирована 121 аннотация научной статьи, после чего их общий массив был разбит на кластеры. На втором этапе результаты машинной кластеризации

проверялись методом качественного анализа текста, на основе чего выделялись ограничения и возможности метода NLP/NLA для решения задачи кластеризации научных текстов. В ходе анализа обнаружено, что статьи с более консервативным ядром теоретических категорий (гендерные исследования, миграционные исследования, теория глобализма) лучше поддаются кластеризации, а теории со слабо структурированным теоретическим ядром (теории с использованием экологической терминологии, теории неравенства) поддаются кластеризации значительно хуже. Полученные результаты позволяют сформировать новое направление работы с большими массивами научных текстов, связанное с их кластеризацией с помощью NLP/NLA. Построение кластеров дает возможность исследователям изучать все тексты данной тематической области, а не только самые цитируемые. Это, в свою очередь, обеспечивает видимость всех научных идей, в том числе не получивших популярности/известности.

Ключевые слова

Смешанные методы, научный текст, кластеризация, социогуманитарное знание, обработка естественного языка, анализ естественного языка, NLP/NLA, метод Варда, искусственный интеллект, структурализм, метафора, акторная теория

Для цитирования

Кашина М.А., Ткач С. Социология ценностей: опыт построения таксономии путем использования технологии анализа естественного языка // Цифровая социология. 2023. Т. 6, № 1. С. 48–58.

© Кашина М.А., Ткач С., 2023.

Статья доступна по лицензии Creative Commons «Attribution» («Атрибуция») 4.0. всемирная (<http://creativecommons.org/licenses/by/4.0/>).



Sociology of values: experience of building a taxonomy by using natural language analysis technology

Received 07.02.2023

Revised 09.03.2023

Accepted 22.03.2023

Marina A. Kashina

Dr. Sci. (Polit.), Prof. at the Social Technologies Department

ORCID: 0000-0003-4509-8657

E-mail: kashina-ma@ranepa.ru

North-West Institute of Management – Branch of the Russian
Presidential Academy of National Economy and Public
Administration, St. Petersburg, Russia

Sergey Tkach

Programmer

ORCID: 0000-0001-5016-9187

E-mail: sergey@ngostellit.ru

St. Petersburg University Research Park, St. Petersburg, Russia

ABSTRACT

Modern research in the field of sociology of science is becoming more complicated due to the constantly growing publication activity of authors. To track trends in sectoral sociology, scientists turn to scientometric methods, but they are not enough. Trends in the development of the sociology of values as a branch of sociology are the subject of the study. The purpose of the work is an assessment of the possibilities of using natural language analysis methods (NLP/NLA) for thematic and theoretical clustering of research in the sociology of values. The design of the study was quantitative and qualitative, it was carried out in two stages. At the first stage, 121 abstracts of a scientific articles were analyzed using text mining, after which their total array was divided into clusters. At the second stage, the results of machine clustering were examined by the method of qualitative text analysis, on the basis of which the limitations

and capabilities of the NLP/NLA method were identified for solving the problem of clustering scientific texts. It was found that articles with a more conservative core of theoretical categories (gender studies, migration studies, the theory of globalism) are more amenable to clustering, while theories with a loosely structured and fluid theoretical core (theories using environmental terminology, theories of inequality) are much less amenable to explicit clustering. The results obtained allow us to form a new direction of work with large arrays of scientific texts, associated with their clustering using NLP/NLA. Building clusters enables researchers to work with all texts in a given subject area, and not just with the most cited ones. This, in turn, provides the visibility of all scientific ideas, including those that have not gained popularity/notability.

Keywords

Mixed methods, scientific text, clustering, socio-humanitarian knowledge, natural language processing, natural language assessment, Ward method, artificial intelligence, structuralism, metaphor, actor theory

For citation

Kashina M.A., Tkach S. (2023) Sociology of values: experience of building a taxonomy by using natural language analysis technology, *Digital Sociology*, vol. 6, no. 1, pp. 48–58. DOI: 10.26425/2658-347X-2023-6-1-48-58



ВВЕДЕНИЕ / INTRODUCTION

В условиях лавинообразного роста научной информации в цифровизирующемся обществе с неизбежностью возникает вопрос об эффективных способах работы с ней. Ответ на этот вопрос можно найти, обратившись к новым для социологии методам, а именно, методам интеллектуального анализа текстов. Традиционная методология предполагает качественный анализ научных текстов. Цифровая методология – это количественный анализ с использованием методов машинной обработки информации. Такое решение лежит в основе четвертой технологической революции [Юдина, 2017]. В настоящее время разработано немало технологий анализа естественного языка [Young et al., 2018]. Однако задача автоматизации процессов научного анализа информации все еще не решена.

Традиционным подходом к анализу результатов научной деятельности является анализ содержания опубликованных трудов. Однако сегодня даже в сравнительно узкой тематической области их может оказаться очень много. В связи с этим исследователи стараются ограничиться анализом только ключевых работ, так или иначе конструируя понятие ключевой работы: либо через выделение некоего теоретического фронта науки, либо через обращение к наукометрическим методам. Каждый из подходов имеет свои ограничения: фронт науки предполагает глубокое понимание внутренней кухни исследуемой дисциплины, а наукометрические методы, как отмечают J.W. Schneider [2013] и Е.И. Кононенко [2021], нередко становятся предметом для спекуляций.

Среди первых попыток решить эту проблему можно назвать структуралистскую интервенцию в социологию науки, предпринятую во второй половине XX в. [Collyer, 2011]. Это решение заключалось в отказе от анализа собственно содержания в пользу анализа отношений авторов статьи и коллег, устройству университетов, освещению статьи в прессе и т.д. Среди значимых представителей этого направления можно назвать П. Бурдьё и его агентную теорию поля науки. В его оптике наука – это в первую очередь борьба за политическое преимущество и ресурсы [Bourdieu, 2020].

Попытку проанализировать науку, не обращаясь к ее содержанию, можно также найти в работах, посвященных концепции экспертности, например, в работах Н. Коллинса [Collins, Evans, 2007]. Он отмечает, что наука есть все, что признается наукой экспертным сообществом. Такая теоретическая уловка позволяет избежать непосредственного соприкосновения с корпусами

текстов, оставив для анализа только выводы ученых-экспертов о них. Но этот подход не позволяет оценить само содержание и качество экспертных выводов. Ж.-Ф. Лиотар в своей критике концепции экспертности говорит о том, что выводы ученых-экспертов могут строиться не только на самом содержании материала, но и через привлечение всего доступного им арсенала языковых игр: обмана, сокрытия правды, софистических приемов, шантажа и другого [Lyotard, 1997]. Это же прослеживается и, например, в работах С. Жижека о критике экспертности в современную эпоху [Žižek, Damià, 2020]. Проявлением этих тенденций является появление псевдонауки, фейковых исследований, различного рода «научных сенсаций».

Другой версией структурализма применительно к анализу науки выступает акторная теория Б. Латура. Автор использует идею М. Вебера о методе как разграничительном рифе науки и не-науки: а именно, «наука разрабатывает методы мышления, рабочие инструменты и вырабатывает навыки обращения с ними» [Вебер, 1990, с. 729]. Исходя из этого, исследование метода (буквально совокупности действий ученого) должно составить предмет социологии науки. Б. Латур идет в этом отношении дальше М. Вебера и предлагает понимать лабораторию науки как совокупность практик, осуществляемых людьми, приборами, роботами, компьютерными программами и прочим [Latour, 1987]. Решение проблемы приближения к содержательному анализу возрастающего массива научного знания может находиться в поиске тематических кластеров, которые можно рассматривать как научные синтагмы.

Современные методы обработки/анализа естественного языка (англ. natural language analysis, далее – NLA) позволяют эффективно справляться с задачами кластеризации текстов обыденной речи, что демонстрируется в их использовании для решения практических задач. Как отмечают М.А. Саламбас и соавторы, такие методы показывают свою эффективность в анализе текстов судебных решений [Calambas et al., 2015]. В работе I.Ö. Arnarsson и коллег показано, как NLA оказывается эффективным при анализе текстовой инженерной документации [Arnarsson et al., 2021]. Способность обрабатывать большие массивы научного текста продемонстрирована, например, в работе В. Зиннатуллина и С. Коледина, которые использовали NLA и кластеризацию для анализа научной карьеры четырех профессоров: трех профессоров кафедры математики и одного – кафедры истории [Zinnatullin, Koledin, 2020].

Однако при обсуждении возможностей применять методы NLA к анализу научных текстов социо-гуманитарного знания следует упомянуть про трудности, которые могут возникнуть. В первую очередь, речь идет о понимании абстрактных метафор. Как отмечают G. Lakoff, и M. Johnson [Lakoff, Johnson, 1980] и M. Black [Black, 1955] научный язык практически всегда базируется на приеме метафоры. Явления обыденности осмысляются или понимаются через метафору теоретической модели во время концептуализации. Концептуализация же является неотъемлемой частью научного исследования.

Другой трудностью может стать сама структура научного языка. Научный язык являет собой разновидностью дискурса. Как отмечает М. Фуко, текст никогда не заканчивается физическими границами книги или статьи, продолжаясь за ее пределами через ссылки, цитаты и т.д. [Foucault, 2002]. Научный текст в этом смысле особенно далеко заходит в своей связанности с внешними источниками, как отмечает S.P. McDonald, G.J. Kelly [2012].

Задача оценки использования NLA в социологии науки очень масштабна, наше исследование является скорее пилотным, позволяя при этом высветить важные аспекты рассматриваемой темы.

Целью статьи является описание ограничений и преимуществ метода кластеризации технологии анализа естественного языка в построении таксономии исследовательского поля на примере поля социологии ценностей. Цель реализована посредством решения следующих задач: описание метода и процедуры его применения в рамках исследования, описание результатов исследования (обнаруженных кластеров), их содержательный анализ с позиции выявления общих ограничений метода и его преимуществ в плоскости содержательного наполнения кластеров.

МЕТОДОЛОГИЯ И ДИЗАЙН ИССЛЕДОВАНИЯ / RESEARCH METHOD AND DESIGN

Поскольку анализ всего социологического дискурса невозможно выполнить в рамках одного эмпирического исследования, необходимо было выбрать какую-то одну отрасль социологии, при этом она должна была входить в мейнстрим социологического знания и быть тематически разнообразной. Этим критериям отвечала социология ценностей. Под ценностью в социологии понимается «значимое для индивида, либо группы явление духовной культуры или предмет материальной культуры, воспринимаемые

субъектом оценки в качестве нормы, идеала, блага, пользы и выгоды» [Максименко, 2011]. Соответственно, социология ценностей – область социологической науки, предметом исследования в которой выступают ценности и их роль в социальной жизни. И сами ценности, и их роль в жизни общества трактуются весьма по-разному, исходя из теоретической парадигмы исследования, но в то же время все без исключения социологические теории включают в себя аксиологическую проблематику, что превращает социологию ценностей в мейнстрим социологии.

Исследование выполнено с использованием смешанных методов (англ. mixed methods) и состояло из двух этапов. Эмпирическим материалом стали аннотации научных статей как наиболее типичная (устоявшаяся) форма научного языка. На первом этапе с помощью количественного анализа решалась задача кластеризации естественного языка. Для этих целей использовались Python-библиотеки NLTK, SKLearn и SciPy. Среди апробированных методов иерархической кластеризации библиотеки SciPy наиболее релевантным выступил метод Ward. Одиннадцать из тридцати трех выделенных им кластеров мы включили в дальнейший анализ.

Далее перечислим критерии отбора кластеров. С одной стороны, задача описания исследовательской повестки требует соблюдения критерия полноты: необходимо охватить повестку целиком. С другой стороны, полнота получаемой информации может быть достигнута и с использованием методов выборочной репрезентации, к которым можно отнести метод основного массива – 50 % и более от всех работ. Мы выбрали размерность кластеров, которая позволила закрыть «основной массив». Это оказались кластеры, включавшие в себя от 4 до 7 работ. Второй этап исследования был качественным: тексты аннотаций прочитывались и выделялись их общие тематически и семантически черты.

Были поставлены следующие задачи. Представить с помощью методов NLA множество аннотаций статей по ценностям в виде набора четко очерченных кластеров. Проанализировать наиболее объемные кластеры с точки зрения их тематического наполнения и оценить уровень однородности тематики.

На первом этапе из общего массива высокорейтинговых научных журналов по социологии были отобраны журналы, в которых тематически в период с 2016 г. по 2020 г. могли бы быть размещены статьи по исследованию ценностей. Уровень журнала – 1-й квартиль Scopus. Таких журналов оказалось 25.

На втором этапе проводился отбор статей из этих журналов. Такой отбор с содержательной стороны являлся достаточно нетривиальной задачей, поскольку англоязычное слово *value* многозначно, и редко используется в значении «ценность» применительно к теме или теории исследования. То же справедливо и для выражения *value(s) of* и *value(s) in*. Поэтому необходимо было ввести дополнительные реперные точки, позволяющие идентифицировать статью как исследование ценностей. В качестве таких точек были использованы фамилии социологов, которые считаются классиками в исследовании ценностей и/или предложившими собственные подходы (теории) ценностей. Проведенный обзор литературы позволил определить авторов, чьи работы составляют современные методологические и литературные основания анализа ценностей. Ими оказались следующие ученые: F. Adler, G. Allport, H. Becker, R. Benedict, J. Blake, C. du Bois, W. Catton, R. Cochrane, L. Coleman, K. Davis, D. Fave, J. Gillin, J. Kahl, C. Kluckhohn, V. Lindsey, A. Marsh, K. Marx, C. Morris, G. Myrdal, T. Parsons, D. Riesman, A. Smith, H. Spencer, W. Thomas, R. Williams. Логично предположить, что авторы современных статей по ценностям так или иначе должны были упоминать эти фамилии в аннотациях своих работ. Результатом такого отбора стал список, включающий 121 статью (подробнее: [Кашина и др., 2021]), их аннотации и были подвергнуты машинному анализу.

АНАЛИЗ РЕЗУЛЬТАТОВ / ANALYSIS OF RESULTS

На рисунке 1 представлено разбиение всего множества анализируемых аннотаций на кластеры. Кластеров с объемом более четырех статей оказалось 11:

- 1) технологии и неравенство;
- 2) методология/капитализм;
- 3) метафора экологии;
- 4) ценности глобализма;
- 5) гендерные проблемы;
- 6) занятость и рынок труда;
- 7) культура и культурная политика;
- 8) семья и родительство;
- 9) молодежь и образование;
- 10) благополучие мигрантов;
- 11) здоровье мигрантов.

В кластер № 1 вошли четыре статьи. В кластеры №3, 5, 8, 9, 11 – по 5 статей. В № 6 – шесть статей. В кластеры № 2, 4, 7, 10 – по семь статей. Всего анализировались аннотации 63 научных статей из общего массива в 121 статью.

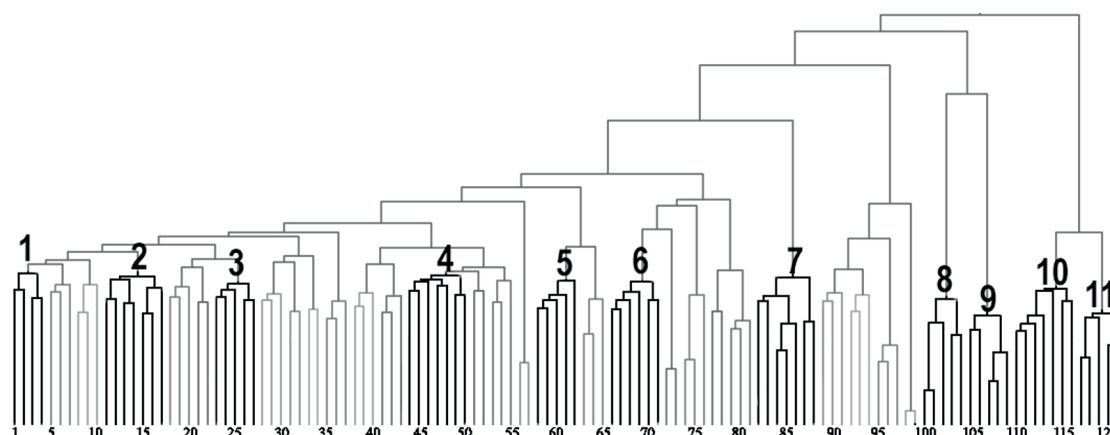
Наиболее однородным в тематическом отношении оказался кластер 8, посвященный проблемам

семьи, воспитания детей и родительства. Все статьи в нем так или иначе затрагивают эту тему, а именно, связь миграции и воспитания детей, ненадлежащее исполнение родительских обязанностей, гендерные особенности бремени домашних обязанностей в семьях с большим количеством недееспособных, проблема отцовства, в частности использование отпуска по уходу за детьми. В этом кластере представлены статьи из 4-х журналов, анализируется опыт Норвегии, Англии, Эквадора.

Тематически однороден и кластер 9, посвященный проблемам молодежи, а именно получению молодыми людьми образования и выбору ими жизненных траекторий, влиянию на это их представлений о своей будущей взрослой жизни, семейной принадлежности и т.д. Несколько особняком стоит статья об экономической отдаче от межрегиональной миграции молодежи в Китае, но решение переехать в другой регион – это тоже решение о своей жизненной пути и карьере. Три статьи этого кластера вышли в *Journal of Youth Studies*, а две – в *Sociological Research Online*. Здесь только одна статья имеет отсылку к страновой специфике.

Очень однороден кластер статей по гендерным проблемам (№ 5). Его особенность заключается в том, что все 5 статей являются национальными. Представлены Камбоджа, Италия, Великобритания и Танзания. Различия в странах в связи с гендерно-ролевыми установками и центральной ролью брака анализируются на опыте 24-х стран. Причем в статье про Великобританию рассматривается опыт китайских и корейских матерей, то есть представлено кросс-культурное исследование. Гендерные проблемы анализируются во всех сферах жизни общества: STEM-образование (англ. *science, technology, engineering and mathematics* – наука, технологии, инженерия и математика), гендерные квоты в бизнесе, семейные отношения, материнство и занятость, гендерные аспекты здоровья, а именно лечение кровоизлияния в связи с абортом и без него. Это единственный кластер, где политика журнала не оказывает влияния, потому что все пять статей опубликованы в пяти разных журналах.

В анализируемом множестве аннотаций выделяются еще два очень сходных кластера по проблемам мигрантов 10-й и 11-й. Но в отличие от гендера здесь тематическое единство определяется политикой журнала: все статьи блока 11 опубликованы в одном журнале – *International Journal of Migration, Health and Social Care*. Их тематика связана со здоровьем мигрантов, причем в 3-х



Составлено авторами по материалам исследования / Compiled by the authors on the materials of the study

Рис. 1. Результат машинной кластеризации множества из 121 аннотации

Fig. The result of machine clustering of a set of 121 annotations

из них представлен опыт Канады. 10-й кластер более разнообразен. В нем есть статья, опубликованная в другом журнале (*Sociological Perspectives*) Сама тематика шире. Помимо барьера в доступе к услугам здравоохранения появляется тема причин смертности, отсутствия продовольственной безопасности, расовой дискриминации, поддержки контактов с друзьями и семьей в странах исхода, питания иммигрантов в домах престарелых, особенности отцовства в семьях мигрантов. Обобщая, эту тематику можно назвать «Благополучие мигрантов в принимающих странах». Речь идет о Бельгии, Австралии и Канаде. Отметим важный момент. Несмотря на то, что в обоих кластерах речь идет об исследованиях мигрантов, программа отделяет тему благополучия от темы здоровья, при этом в 11-м кластере ярче выражен гендерный аспект, которого в 10-м практически нет.

Достаточно однородным получился кластер 7, посвященный проблемам культуры. Как и в случае с исследованиями мигрантов здесь также присутствует заметное влияние политики журнала. Из семи статей пять опубликованы в *International Journal of Cultural Policy* и две – в *International Journal of Sociology and Social Policy*. В двух статьях акцентируется национальная культурная повестка – городские легенды в Мексике и культурная политика в Квебеке, есть статья по географическому региону – развитие третьего сектора в Европе.

6-й кластер посвящен проблемам занятости, в том числе мотивации к работе. В целом он однороден: четыре из шести статей опубликованы в профильном журнале *Journal of Leadership and Organizational Studies*, одна – в *Work, Employment and*

Society, и одна в *Psychology, Society and Education*. Рассматриваются роли soft-skills (с англ. – «мягкие навыки»), проблемы лояльности работников в транснациональных компаниях, мотивации через обогащение труда, молодых «синих воротничков» и другое.

Кластер № 3 можно назвать «ценности глобализации». Интересно отметить, что, в отличие от кластера по экологии, он, как и кластер по гендерным проблемам, абсолютно неоднороден по составу журналов. Все четыре вошедшие в него статьи опубликованы в разных журналах. Стараясь реконструировать логику выделения этого кластера, к которой обращалась модель, мы можем отметить общие термины/категории, которые используют авторы работ. Первая из них (автор F. Haskaj) рассматривает в рамках парадигмы некроэкономики современную глобальную экономическую политику. Важно отметить, что в таком исследовательском вопросе есть две точки опоры: собственно, некроэкономика и взгляд на проблему в глобальном контексте. Некроэкономика – это обоснование экономической выгоды смерти, в том числе геноцида. Вторая работа (автор S. Perkiss) посвящена приложению теории текущего мира З. Баумана к глобальной картине мира. Здесь можно выделить такие же две опоры для исследовательского вопроса: концепция текущего мира (современности) и концепция глобализма. Третья работа (N.X.M. Tadiar) интереснее. Автор обращается к проблеме мегагородов, разрастающихся мегаполисов. Однако ключевым критерием для включения становится снова контекст для рассмотрения городов. Автор использует термин

«Global South» (Глобальный Юг). С одной стороны, прочтение может быть тривиальным и буквальным: Южное полушарие. Однако, как видно далее, речь идет о секторе в глобальном мире. Об этом свидетельствует, например, более позднее обращение к «движениям глобальной политической городской жизни» (англ. *movements of globopolitical urban life*). Перед нами опять концептуализация проблемы с позиции ее глобального контекста. Четвертая и пятая работы более прямо соотносятся с ценностями глобализма: в первой (B. Gorman) речь про политический язык демократии в глобальной политике, вторая (D. Vandegrift) посвящена взрослению молодежи в Российской Федерации и роли тренда глобализации в этом процессе.

Нам осталось рассмотреть кластеры № 1, 2, 4. В отличие от предыдущих кластеров определить их общее тематическое поле заметно сложнее. Кластер 1 вообще выглядят как хаотичный набор тем. В него вошли статьи о филиппинской идентичности на Гавайях, о влиянии инноваций и лидерства на эффективность технологической диверсификации, о проблемах вспомогательных репродуктивных технологий и о роли меритократических ценностей в социальной мобильности в Чили. В некотором приближении можно обозначить две темы этого кластера: технологии и неравенство. Однако эти темы не будут объединяющими для всех статей: технологии для первой (Y.-C. Huang и др.) и второй статьи (E. Blyth и др.) кластера, а неравенство – для третьей (D.B. Eisen) и четвертой (M. Fercovic). Понятно, что все эти статьи, имея столь разную тематику, вышли в разных журналах. Тем не менее программа нашла близость используемого языка в их аннотациях достаточной, чтобы объединить их в один кластер. Существует ли такая смысловая связь на самом деле, или это проблемы метафоричности научного языка, о которых мы писали в литературном обзоре, требует отдельного исследования.

Более очевидные взаимосвязи прослеживаются в 4-м кластере, статьи в котором так или иначе посвящены экологической проблематике. Хотя здесь тоже есть ряд статей, которые, судя по названию, смотрятся в разделе по классической экологии достаточно неожиданно. Например, «Социальный анализ элитной группировки: пример Формулы-1» (англ. *Social Analysis of an Elite Constellation: The Case of Formula 1*). Сама статья посвящена формированию новых видов элит, но использование авторами экологических метафор привело к тому, что она попала в кластер по экологии. При этом 5 статей

из 7 опубликованы в одном журнале – *Theory, Culture and Society*, в нем же опубликована и статья про гонщиков Формулы-1 как представителей новой элиты. А это явно не журнал по экологическим проблемам, когда рассматривается связь человека и природы. Обобщая, можно сказать, что в одном кластере оказались работы по воздействию человека на природу, политике и религии.

Считать ли это ошибкой программы или новым видением экологии в социологии – вопрос, требующий размышлений. В любом случае можно утверждать, что вряд ли бы эти статьи при традиционном анализе были объединены в одну группу, исходя просто из их названий. Можно попробовать воспроизвести логику, в соответствии с которой они были выделены в единый кластер. Все статьи обращаются к допущению о существовании властных отношений и иерархии, которая на их основе выстраивается. Это и теория элит в Формуле-1, и политические элиты Калифорнии и их власть над подземными водами, и городская элита, обращающаяся к практикам *amoral familism* (с англ. – аморальное семейничество), и *fashion-элита* в Кувейте, а также три статьи, посвященных выстраиванию иерархических макротеорий. Кроме того, авторы активно обращаются к естественно-научным метафорам в текстах: уже упоминавшаяся выше экология, Антропоцен, теории габитуса. Экологическая терминология оказывается омонимичной для разных теорий и подходов. Например, биогеоценоз используется и бурдианской традиции и у Б. Латура.

В заключение обсудим самый сложный кластер 2. В таблице 1 представлены названия статей, которые в него вошли, и журналов, где они были опубликованы.

Таблица. Перечень статей, вошедших в кластер 2

Table. List of articles included in cluster 2

Название статьи (перевод на русский язык)	Название журнала (перевод на русский язык)
The case of Black Millennials (Кейс телешоу Black Millennials)	Sociological Perspectives (Социологические перспективы)
Making things possible (Делая вещи возможными)	Sociological Methods and Research (Социологические методы и исследования)
On guilt and the depoliticization of downsizing practices (О вине и деполитизации практики сокращения штатов)	International Journal of Sociology and Social Policy (Международный журнал социологии и социальной политики)

Окончание табл.

Название статьи (перевод на русский язык)	Название журнала (перевод на русский язык)
Abandon canon in American sociology (Отказ от канона в американской социологии)	International Journal of Sociology and Social Policy (Международный журнал социологии и социальной политики)
Religion and social class: Theory and method after Bourdieu (Религия и социальный класс: теория и метод по Бурдьё)	Sociological Research Online (Социологические исследования онлайн)
Sense of belonging as an indicator of social capital (Чувство принадлежности как показатель социального капитала)	International Journal of Sociology and Social Policy (Международный журнал социологии и социальной политики)
Pharmaceutical capitalism and its logistics: access to Hepatitis C treatment (Фармацевтический капитализм и его логистика: доступ к лечению гепатита С)	Theory, Culture and Society (Теория, культура и общество)

Составлено авторами по материалам исследования /
Compiled by the authors on the materials of the study

Пять статей этого кластера так или иначе можно отнести к рубрике методологии для исследования ценностей, однако в нее никак не вписываются статьи про вину при сокращении персонала и о фармацевтическом капитализме, которые можно было бы отнести к капиталистическим ценностям. Другими словами, здесь, как и в случае с кластером 1, мы можем зафиксировать неудачу в выделении кластера с помощью машинный технологий обработки/анализа естественного языка.

ДИСКУССИЯ / DISCUSSION

Полученные в ходе исследования результаты, в целом, согласуются с более ранними работами, посвященными NLA. В частности, в работе R. Tao и соавторов [Tao et al., 2021] или в более ранней работе E. Shutova и коллег [Shutova et al., 2016] показано, что проблема понимания метафор в естественном языке является актуальной проблемой интерпретации данных. При решении задачи кластеризации речь идет о совмещении метафорического и буквального употребления терминов. Ярким примером является «экология» как состояние окружающей среды и как метафора для описания социальной реальности.

Отрицать значимость и актуальность проблематики повестки ценностей экологии сложно – это популярная, объемная тема. Одновременно с этим S. Liu и M. Emirbayer [2016] отмечают, что в социальной теории последних лет наблюдается экологическая экспансия: общества начинают пониматься в категориях экосистем.

В социологии науки есть множество исследований, описывающих инструменты регулирования работы ученых со стороны сообщества, к которому они принадлежат. Можно отметить работы Д.В. Александрова [1998] о влиянии научных школ на работу ученых; Р. Мертон о власти авторитета научного руководителя в коллективе или власти учебного заведения [Merton, 1979]. Однако в таких объяснениях объем анализируемой институции очень ограничен: у научной школы речь идет о двух-трех поколениях научных коллективов, у института – о нескольких десятках работников администрации и десятке-другом коллег-ученых. Остальные группы же еще меньше. Методы NLA, как показывает настоящая статья, позволяют расширить границы анализа для целой тематической области, которой занимаются ученые не только разных локальных коллективов, но и разных университетов и стран, поскольку здесь уже нет ограничений, связанных с объемом анализируемого материала.

ЗАКЛЮЧЕНИЕ / CONCLUSION

Проведенный выше анализ результатов исследования позволяет сделать ряд выводов.

1. Наблюдается заметное пересечение кластеров, выделяемых традиционными методами, и кластеров, выделенных машинной программой. В частности, в сравнении с выделяемыми тематическими группами в программной статье J.L. Spates об истории развития социологии ценностей в XIX-XX вв. (труд, образование, политика, гражданственность и прочее) [Spates, 1983] появляются и новые темы, например, здоровье мигрантов, устойчивое развитие, прямое и метафорическое употребление термина «экология».

2. NLA позволяет привязывать, объединять в кластеры статьи, которые из-за метафор, использованных в названиях, могут выпасть из поля зрения исследователей. Возможно, их можно рассматривать в качестве точек роста или предвестников появления новых исследовательских полей.

3. Косвенной верификацией эффективности машинной программы выступает то, что в большинстве случаев выделенные кластеры совпали с профильными журналами. Яркий пример – тематика мигрантов и культурная политика.

Следует принять во внимание следующие ограничения исследования.

1. Характер эмпирического материала. Кластеризации подвергались только тексты аннотаций, результаты, полученные авторами статей, никак не анализировались.

2. Объектом анализа выступает только социологическое знание. Насколько применима предлагаемая технология к анализу естественно-научного знания – вопрос, требующий специального изучения.

3. Ограниченность анализируемой тематики. В анализ включены лишь тексты, посвященные исследованию ценностей. Будут ли получены аналогичные результаты на иных по тематике текстах, требует дополнительной проверки.

4. Методика выборки. Данная методика является авторской, она разработана специально для анализа крайне неоднородного и обширного проблемного поля исследования ценностей. В частности, были выбраны работы, в аннотациях которых есть отсылки к авторам, которые писали ключевые работы в рамках четырех констелляций научных школ: социальный статус, идеалы, экономическая теория, антропологическая

теория (англ. social status, ideal views, economic theory, antropological theory). Критерии выборки могли быть и другими, например, страновыми или языковыми.

5. Способность к интерпретации текстов аннотаций опирается на собственные социологические знания и компетенции исследователей. Поэтому нельзя исключить ситуации, когда нами были неверно поняты мысли, которые авторы статей хотели донести до своих читателей. Тем более что аннотации, к сожалению, не всегда аутентично передают содержание самих статей.

Проведенное пилотное исследование позволяет обозначить контуры нового направления работы с большими массивами социологических научных текстов, связанное с их кластеризацией с помощью NLA. Построение кластеров дает возможность исследователям изучать все тексты данной тематической области, а не только самые цитируемые. Это, в свою очередь, обеспечивает видимость всех научных идей, в том числе не получивших популярности/известности. Тем самым можно выделить возможные точки роста и новые направления исследований.

СПИСОК ЛИТЕРАТУРЫ

- Александров Д.А. (1998). Научные школы как социальные сети // Академические научные школы Санкт-Петербурга. К 275-летию Академии наук. СПб.: СПбНЦ РАН. С. 11–18.
- Вебер М. (1990). Избранные произведения. Пер.с нем., общ. ред. Давыдов Ю.Н. М.: Прогресс. 808 с.
- Кашина М.А., Ляшко С.В., Ткач С. (2021). Актуальная повестка исследования ценностей в зарубежной социологии: тематика, теории, методологии // Управленческое консультирование. №10. С. 62–82. <https://doi.org/10.22394/1726-1139-2021-10-62-82>
- Конonenko Е.И. (2021). Еще раз о симуляции научной деятельности в гуманитарных дисциплинах // Вестник Санкт-Петербургского университета. Искусствоведение. Т. 11, № 2. С. 327–349. <https://doi.org/10.21638/spbu15.2021.209>
- Максименко А.А. (2011). Социологическая интерпретация понятия «ценность» // Вестник Костромского государственного университета. Т. 17, № 2. С. 284–291.
- Юдина М.А. (2017). Индустрия 4.0: перспективы и вызовы для общества // Государственное управление. Электронный вестник. № 60. С. 197–215.
- Arnarsson I.Ö., Frost O., Gustavsson E., Jirstrand M., Malmqvist J. (2021). Natural language processing methods for knowledge management – Applying document clustering for fast search and grouping of engineering documents // Concurrent Engineering Research and Applications. V. 29, no. 2. Pp. 142–152. <https://doi.org/10.1177/1063293X20982973>
- Black M. (1955). Metaphor. Proceedings of the Aristotelian Society // New Series. V. 55. Pp. 273–29.
- Bourdieu P. (2020). Habitus and Field: General Sociology. V. 2 (1982–1983). NJ: Wiley. 446 p.
- Calambas M.A., Ordonez A., Chacon A., Ordonez H. (2015). Judicial precedents search supported by natural language processing and clustering // 10th Colombian Computing Conference, 10CCC (2015). Art. no. 7333448. Pp. 372–377. <https://doi.org/10.1109/ColumbianCC.2015.7333448>
- Collins H., Evans R. (2007). Rethinking expertise. University Chicago: Chicago Press. 159 p.
- Collyer F. (2011). Reflexivity and the sociology of science and technology: The invention of “Eryc” the antibiotic // Qualitative Report. V. 16, no. 2. Pp. 316–340.
- Foucault M. (2002). Archaeology of knowledge. London New York: Routledge. 256 p.

- Lakoff G., Johnson M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press. 242 p.
- Latour B. (1987). *Science in action: how to follow scientists and engineers through society*. Cambridge: Harvard University Press. Print.
- Liu S., Emirbayer M. (2016). Field and ecology // *Sociological Theory*. V. 34, no. 1. Pp. 62–79.
- Lyotard J.F. (1997). *The postmodern condition*. Minnesota: University of Minnesota. 144 p.
- McDonald S.P., Kelly G.J. (2012). “Beyond argumentation: Sense-making discourse in the science classroom” // *Perspectives on Scientific Argumentation: Theory, Practice and Research*. V. 9789400724709. Pp. 265–281. https://doi.org/10.1007/978-94-007-2470-9_13
- Merton R.K. (1979). *The Normative Structure of Science // The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press. 636 p.
- Schneider J.W. (2013). Caveats for using statistical significance tests in research assessments // *Journal of Informetrics*. V. 7, no. 1. Pp. 50–62.
- Shutova E., Kiela D., Maillard J. (2016). “Black holes and white rabbits: Metaphor identification with visual features” // 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 – Proceedings of the Conference. Pp. 160–170. <http://dx.doi.org/10.18653/v1/N16-1020>
- Spates James L. (1983). The sociology of values // *Annual Review of Sociology*. V. 9. Pp. 27–49.
- Tao R., Wei Y., Yang T. (2021). Metaphor analysis method based on latent semantic analysis // *Journal of Donghua University (English Edition)*. V. 38, no. 1. Pp. 83–90. [http://dx.doi.org/10.1016/S1532-0464\(02\)00004-7](http://dx.doi.org/10.1016/S1532-0464(02)00004-7)
- Young T., Hazarika D., Poria S., Cambria E. (2018). Recent trends in deep learning based natural language processing [Review Article] // *IEEE Computational Intelligence Magazine*. V. 13, no. 3. Art. no. 8416973. Pp. 55–75. <https://doi.org/10.48550/arXiv.1708.02709>
- Zinnatullin V., Koledin S. (2020). Analysis of scientists work directions based on natural language processing and clustering // *CEUR Workshop Proceedings*. No. 2667. Pp. 57–61.
- Žižek S., Damià A. (2020). Пандемия: Covid-19 сотрясает мир [Pandemia: la covid-19 estremece al mundo]. Barcelona: Editorial Anagrama. 120 p. (Исп. яз.).

REFERENCES

- Alexandrov D.A. (1998), “Scientific schools as social networks” [“Nauchnye shkoly kak sotsial’nye seti”], *Academic scientific schools of St. Petersburg. To the 275th anniversary of the Academy of Sciences*, St. Petersburg Scientific Center of the Russian Academy of Sciences, St. Petersburg, Russia, pp. 11–18 (in Russian).
- Arnarsson I.Ö., Frost O., Gustavsson E., Jirstrand M., Malmqvist J. (2021), “Natural language processing methods for knowledge management – Applying document clustering for fast search and grouping of engineering documents”, *Concurrent Engineering Research and Applications*, vol. 29, no. 2, pp. 142–152, <https://doi.org/10.1177/1063293X20982973>
- Black M. (1955), “Metaphor. Proceedings of the Aristotelian Society”, *New Series*, vol. 55, pp. 273–29.
- Bourdieu P. (2020), *Habitus and Field: General Sociology, volume 2 (1982–1983)*, Wiley, NJ, USA.
- Calambas M.A., Ordonez A., Chacon A., Ordonez H. (2015), “Judicial precedents search supported by natural language processing and clustering”, *10th Colombian Computing Conference, 10CCC (2015)*, art. no. 7333448, pp. 372–377, <https://doi.org/10.1109/ColumbianCC.2015.7333448>
- Collins H., Evans R. (2007), *Rethinking expertise*, University of Chicago Press, Chicago, US.
- Collyer F. (2011), “Reflexivity and the sociology of science and technology: The invention of ‘Eryc’ the antibiotic”, *Qualitative Report*, vol. 16, no. 2, pp. 316–340.
- Foucault M. (2002), *Archaeology of knowledge*, Routledge, London, New York.
- Kashina M.A., Lyashko S.V., Tkach S. (2021), “Topical agenda for the study of values in foreign sociology: topics, theories, methodologies”, *Administrative Consulting*, no. 10, pp. 62–82, <https://doi.org/10.22394/1726-1139-2021-10-62-82> (in Russian).
- Kononenko, E.I. (2021), “Once again about simulation of scientific activity in the humanities”, *Vestnik of Saint Petersburg University. Arts*, vol. 11, no. 2, pp. 327–349, <https://doi.org/10.21638/spbu15.2021.209> (in Russian).
- Lakoff G., Johnson M. (1980), *Metaphors we live by*, University of Chicago Press, Chicago, US.
- Latour B. (1987), *Science in action: how to follow scientists and engineers through society*, Harvard University Press, Cambridge, Mass., US.
- Liu S., Emirbayer M. (2016), “Field and ecology”, *Sociological Theory*, vol. 34, no. 1, pp. 62–79.
- Lyotard J. F. (1997), *The postmodern condition*, University of Minnesota, Minnesota, US.
- Maksimenko A.A. (2011), “Sociological interpretation of ‘value’ concept” [“Sotsiologicheskaya interpretatsiya ponyatiya ‘Tsennost’”], *Vestnik of Kostroma State University*, vol. 17, no. 2, pp. 284–291 (in Russian).

- McDonald S.P., Kelly G.J. (2012), “Beyond argumentation: Sense-making discourse in the science classroom”, *Perspectives on Scientific Argumentation: Theory, Practice and Research*, vol. 9789400724709, pp. 265–281, https://doi.org/10.1007/978-94-007-2470-9_13
- Merton R.K. (1979), “The Normative Structure of Science”, In: *The Sociology of Science: Theoretical and Empirical Investigations*, University of Chicago Press, Chicago, US.
- Schneider J.W. (2013), “Caveats for using statistical significance tests in research assessments”, *Journal of Informetrics*, vol. 7, no. 1, pp. 50–62.
- Shutova E., Kiela D., Maillard J. (2016), “Black holes and white rabbits: Metaphor identification with visual features”, *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 – Proceedings of the Conference*, pp. 160–170, <http://dx.doi.org/10.18653/v1/N16-1020>
- Spates James L. (1983), “The sociology of values”, *Annual Review of Sociology*, vol. 9, pp. 27–49.
- Tao R., Wei Y., Yang T. (2021), “Metaphor analysis method based on latent semantic analysis”, *Journal of Donghua University (English Edition)*, vol. 38, no. 1, pp. 83–90, [http://dx.doi.org/10.1016/S1532-0464\(02\)00004-7](http://dx.doi.org/10.1016/S1532-0464(02)00004-7)
- Weber M. (1990), *Selected works [Izbrannye proizvedeniya]*, Trans. from Germ. and Ed. Davydov Yu.N. Progress, Moscow, Russia (in Russian).
- Young T., Hazarika D., Poria S., Cambria E. (2018). “Recent trends in deep learning based natural language processing” [Review Article], *IEEE Computational Intelligence Magazine*, vol. 13, no. 3, art. no. 8416973, pp. 55–75, <https://doi.org/10.48550/arXiv.1708.02709>
- Yudina M.A. (2017), “Industry 4.0: prospects and challenges for Society” [“Industriya 4.0: perspektivy i vyzovy dlya obshchestva”], *Public Administration. E-journal (Russia)*, no. 60, pp. 197–215 (in Russian).
- Zinnatullin V., Koledin S. (2020), “Analysis of scientists work directions based on natural language processing and clustering”, *CEUR Workshop Proceedings*, no. 2667, pp. 57–61.
- Žižek S., Damià A. (2020), *Pandemic: covid-19 shakes the world [Pandemia: la covid-19 estremece al mundo]*, Editorial Anagrama, Barcelona (in Spanish).