

АВТОМАТИЗАЦИЯ ДИСКУРС-АНАЛИЗА: МЕТОДОЛОГИЯ И ПРАКТИКА

Турусина Т.В., студентка группы СГНЗ-21М

*Московский государственный технический университет им. Н.Э. Баумана
ttv19s056@student.bmstu.ru*

*Научный руководитель: Галаганова С.Г., кандидат философских наук,
доцент кафедры «Информационная аналитика и политические технологии»
galaganovasg@bmstu.ru*

Аннотация: На основе сравнительного анализа основных методологических подходов к дискурс-анализу автор выявляет и оценивает потенциальные возможности автоматизации данного метода информационной аналитики в рамках каждой методологической парадигмы. В статье исследуется сущность дискурс-анализа, особенности современного цифрового дискурса, влияние интенсификации процесса обмена информацией на информационные контексты.

Ключевые слова: дискурс, дискурс-анализ, цифровой дискурс, дискурсивная практика, социолингвистика, контекст, коммуникация.

Понятие дискурс-анализа объединяет несколько различных подходов к качественному анализу как устных, так и письменных форм языковых практик. Дискурс-анализ используется в прикладной лингвистике, социальной психологии, политическом анализе, социологии.

Данный метод может использоваться для анализа общественного мнения, получения обратной связи при принятии управленческих решений. В этих случаях необходимо сфокусироваться на анализе социолингвистических особенностей дискурсивных практик, которые содержат разделяемые неким коллективом убеждения, порождают либо усиливают их и предполагают ту или иную позицию в дискурсном поле. Такой подход предполагает рассмотрение участников коммуникации как представителей единой социальной группы и анализ контекста, то есть обстоятельств общения в широком социокультурном контексте.

Что касается трактовок роли языковых средств в дискурсе, то они существенно разнятся: одни исследователи рассматривают язык в качестве действенного инструмента формирования социальных практик, другие подчёркивают наличие возможностей для манипуляций языковыми практиками с целью формирования идеологизированного дискурса.

С появлением Интернета и социальных сетей дискурс претерпел значительные изменения. Интенсивный информационно-коммуникационный обмен стал распространяться на все сферы жизни, смешивая не только разнообразные дискурсы, но и расширяя контексты обсуждения. Например, портал по обмену рецептами, который изначально задумы-

вался как средство коммуникации на бытовые темы, может стать площадкой для обсуждения социальных вопросов, таких как размеры пенсий, безработица и другие. При этом контекст обсуждения будет дополняться экономическими и политическими концепциями.

Понятие контекста становится центральным в эпоху цифровизации дискурса. Контекст – не только фон или обстоятельства коммуникации, но и комплекс факторов, включающий общие знания, опыт, ценности, убеждения и социальные нормы, которые разделяют участники коммуникационного процесса. Для эффективного взаимопонимания необходимо, чтобы субъекты коммуникации обладали схожим контекстом, позволяющим им интерпретировать сообщения друг друга с учётом всех нюансов и подтекстов. Так, в повседневной жизни мы склонны говорить с приятелями неполными предложениями, делать отсылки на предыдущие разговоры, использовать сленг. Однако при этом общение остаётся действенным, так как знакомые люди обладают общим коммуникативным контекстом. В противном случае, когда контексты различаются, коммуникация сводится к пониманию лишь поверхностного, имплицитного содержания.

В цифровом мире обмен информацией настолько интенсифицируется, что приводит к следующим, важным для данной работы, последствиям:

1. Быстрое устаревание информационных контекстов: инфоповоды и отношение общественности к ним могут меняться за считанные часы, что затрудняет использование автоматизированного анализа дискурса. Это требует постоянного обновления корпуса текстов, на которых проводится анализ. В случае необходимости анализа реакций целесообразно выявление тональности сообщений.

2. Конфликт контекстов: из-за большого количества разнородной информации об одном и том же объекте в Интернете формируются группы людей с противоположными убеждениями. Данную проблему только подкрепляет феномен «пузыря фильтров»: ситуации, когда человек оказывается в замкнутом информационном пространстве, где алгоритмы и фильтры предлагают контент, соответствующий его предпочтениям, создавая иллюзию объективности и ограничивая доступ к альтернативным точкам зрения. Таким образом, при проведении автоматизированного дискурс-анализа в социальных сетях возможно формирование когорт респондентов в зависимости от источников получаемой ими информации.

3. Контексты как набор знаков: в условиях информационной перегрузки люди начинают воспринимать контекст не как сложный комплекс факторов, а как набор знаков, декодирование которых помогает быстро определить суть сообщения, его направленность и ожидаемую

реакцию аудитории. Таким образом, для каждого пользователя формируется список слов-триггеров: ярлыков, специфической лексики, стилистического оформления. Следовательно, при анализе дискурса необходимо рассматривать не только дискурсивный акт, но и его взаимодействие с потенциальной аудиторией [10].

4. Контексты как правила: негласные нормы Интернет-коммуникации, правила сообществ и платформ, так называемая «культура отмены» ограничивают разнообразие дискурсивных практик, зато несут полезную для исследователя информацию о нормативных и идеологических контекстах сообщества. С учётом данного контекста становится возможным отследить по изменениям дискурса сообществ их социальную динамику – радикализацию, демократизацию и т. д.

Рассмотрим различные подходы к дискурс-анализу в работах основоположников данного метода.

1. М. Фуко описывал структуру дискурса как совокупность его объектов, модальности, концептов и тематического единства. В рамках концепции Фуко дискурс-анализ происходит следующим образом: выделяются субъекты коммуникации, определяются концепты, используемые ими; затем отбираются высказывания особой модальности, после чего определяется частота нахождения объекта в массиве высказываний [14; 15].

2. Французский философ Р. Барт рассматривал дискурс в качестве своего рода синонима социального мифа. Что касается дискурс-анализа как семиологической процедуры, то он, по мнению учёного, направлен на трактовку и декодирование мифов. Это позволяет определить функциональные зависимости означающего и означаемого, которые гарантируют социальному знаку постоянство формы в разных вариантах реализации. Барт обосновывает необходимость семиологического подхода к дискурс-анализу тем, что в силу установок говорящего любое его высказывание формируется под влиянием социального мифа, а это искажает содержание, приводя коммуникативный акт к определённом социокультурному сценарию [1].

3. В области теории дискурс-анализа «классическими» считаются работы Тёна ван Дейка. Лингвист утверждал, что на структуру текста и высказывания влияет не сама социальная ситуация, а, скорее, определение участниками дискурса соответствующих свойств коммуникативной ситуации. По этой причине ван Дейк предлагал использовать «контекстуальные модели» [3]. Динамические качественные модели описывают социальную ситуацию в текущий момент, определяя социальные отношения между автором и публикой, политический и исторический контекст, а также отношения между различными дискурсами. При анализе текста высказывания в рамках концепции ван Дейка исследуется выбор лексических средств, структура текста (взаимное расположение тезиса и аргумен-

тов), умышленно опущенные факты, описательные высказывания об объектах и субъектах дискурса, использование поляризаций и обобщений.

Что касается автоматизации дискурс-анализа, то сложность решения этой задачи обусловлена тем, что большинство методов дискурс-анализа являются «качественными», с трудом поддающимися алгоритмизации, и поэтому требуют привлечения экспертов. Использование некоторых подходов невозможно алгоритмизировать в принципе – например, семиотического подхода Р. Барта. Также требуется учитывать описанные выше специфические особенности современного цифрового дискурса и цифрового контекста.

Исследование существующих методов дискурс-анализа позволяет выделить и описать как общие требования к процессу его поэтапной автоматизации, так и особенности, обусловленные спецификой концепций различных исследователей. Общее касается первых двух этапов:

1. По мнению всех исследователей, необходимо определить принцип отбора документов дискурса для исследования. Определяется список авторов, тематик и других критериев отбора. В случае, если исследователю необходимо рассмотреть все дискурсы, связанные с определённой тематикой, он может привлекать данные, практически не задумываясь об их репрезентативности (социальной типичности или симптоматичности), поскольку в данном случае портрет автора реконструируется по результатам исследования. Автоматизация в данном случае позволит собрать большую выборку при помощи поиска в Интернете по ключевым словам.

Если же исследователь хочет реконструировать устойчивую дискурсивную практику (например, при исследовании тематик юмористического дискурса), большую роль начинают играть социально-демографические характеристики и идеологические ориентации автора. В данном случае имеет смысл отталкиваться от источника информации, на который ссылается автор, ведь даже у похожих по множеству параметров пользователей могут быть разные информационные пространства. С точки зрения автоматизации дискурс-анализа социальных сетей, это будет означать требование устанавливать подписки на сообщества у того или иного актора.

2. Собранную информацию нужно подвергать первичной обработке и регулярно обновлять. Автоматизация поможет очистить данные от шумов, например, использование процедур лемматизации и стемминга позволит привести слова к их базовой форме, что упрощает анализ и поиск по корпусу текстов.

Этим общие требования исчерпываются – далее начинаются различия, обусловленные особенностями различных методологических подходов.

Так, М. Фуко предлагает:

3. Выделить концепты, что согласно его теории, позволит описать смыслы, составляющие область дискурса [16]. Для этого может быть использовано тематическое моделирование – технология статистического анализа текстов для автоматического выявления тематик и ключевых слов.

4. Анализ модальности сводится к выявлению семантического ядра текста и предикатов – установок субъекта коммуникации [13]. Существует несколько инструментов анализа естественного языка, которые могут быть использованы для анализа предикатов. Нейронные модели семантической разметки ролей (SRL) могут применяться для определения семантических ролей участников действия, выраженного глаголом (агент, пациент, инструмент и т.д.). Следует, однако, иметь в виду, что данная процедура может затруднена из-за отсутствия наборов русскоязычных текстов для обучения нейронной модели.

5. Выявить пересечения между объектами анализа и корпусом высказываний. Это задача поиска по массивам, которая легко может быть автоматизирована.

В рамках концепции Тёна ван Дейка:

6. В случае необходимости создания контекстной модели основной задачей исследователя становится описание имплицитных смыслов, закладываемых в языковые практики [19].

Несмотря на очевидную сложность задачи автоматизации контекстного анализа, существует способ отслеживания контекстной близости выражений, то есть определения языкового окружения, в котором употребляется та или иная единица языка в тексте. Другими словами, существует возможность определения подтекстов, закладываемых автором посредством слов-ярлыков, метафор, образных выражений. Для этого необходимо сделать срез контекста – получить максимальное количество текстов, циркулирующих в социальной группе, члены которой разделяют искомый контекст. Эти данные будут использованы для обучения нейронной модели Word2Vec, которая на больших массивах данных позволяет «уловить» скрытые связи между словами [18]. В результате при анализе других текстов Word2Vec будет сопоставлять их с уже описанным контекстом. Алгоритмически модель преобразует слова в числовые векторы, отражающие семантические связи: слова с близким значением будут иметь близкие векторы в векторном пространстве.

Таким образом, большинство методов, использующихся в рамках подходов к дискурс-анализу Мишеля Фуко и Тёна ван Дейка, может быть автоматизировано. Однако необходимо предварительно оценить точность результатов работы алгоритма, сравнивая их с результатами экспертного анализа.

Литература и источники

1. Барт Р. Мифологии / пер. с фр. М.: Академический проект, 2000. 351 с.
2. Водак Р. Критическая лингвистика и критический анализ дискурса // Политическая лингвистика. 2011. № 4.
3. Дейк Т. Дискурс и власть. Репрезентация доминирования в языке и коммуникации / пер. с англ. М.: URSS, 2014. 352 с.
4. Галаганова С.Г., Турусина Т.В. Технологии анализа социальных сетей с целью выявления социальных трендов // Человеческий капитал, 2023, № 1 (169). С. 121–136.
5. Информационная аналитика и информационно-аналитические технологии в контексте социального управления. МГТУ имени Н.Э. Баумана. Москва, 2023.
6. Йоргенсен М.В., Филлипс Л. Дискурс-анализ. Теория и метод / пер. с англ. М.: Генезис, 2008. 352 с.
7. Квадратура смысла. Французская школа анализа дискурса / пер. с фр. португ. / сост. и общ. ред. П. Серио. – М.: Прогресс, 1999. 413 с.
8. Макаров М.Л. Основы теории дискурса. – М.: Гнозис, 2003. 276 с.
9. Маклашова Е.Г. Дискурсный анализ в социальных науках: обзор теорий // Северо-восточный гуманитарный вестник. 2013. № 2 (7).
10. Молодыхенко Е.Н. Уровни контекста: как анализ текста становится анализом дискурса? (На материале новых лайфстайл-медиа в сети Интернет) // Слово.ру: Балтийский акцент. 2023. № 1.
11. Ремарчук В.Н. Информационная аналитика: теория, методология, технологии : учебник для вузов. СПб.: Лань, 2022. 224 с.
12. Современные теории дискурса: мультидисциплинарный анализ. – Екатеринбург: Дискурс-Пи, 2006. 177 с.
13. Тичер С., Мейер М., Водак Р., Веттер Е. Методы анализа текста и дискурса / пер. с англ. – М.: Гуманитарный центр, 2017. 356 с.
14. Фуко М. Порядок дискурса / пер. с фр. – М.: 1996.
15. Фуко М. Археология знания / пер. с фр. – СПб.: Гуманитарная академия, 2012. 416 с.
16. Шаляпина С.К. Построение категориальной модели в контент-анализе : дисс ... канд.социол.наук : 22.00.01. – Москва, 1998.
17. Blei D., Nag A., Jordan M.I. Latent Dirichlet Allocation // Journal of Machine Learning Research, 2003.
18. Content of the process of formation of students' speech abilities at the university / Ovsyannikova O.A., Mishcherina M.A., Bocharnikov I.V. В сборнике: E3S Web of Conferences. 8. Сер. "Innovative Technologies in Science and Education, ITSE 2020" 2020. С. 18106.
19. Mikolov T., Chen K., Corrado G., Dean J. Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781, 2013
20. Wodak R., Meyer M. Methods of Critical Discourse Analysis // Sage Publications, 2015.