Юлия Владимировна Балакина

Национальный исследовательский университет Высшая школа экономики (Нижний Новгород)

[julianaumova@gmail.com](mailto:julianaumova@gmail.com)

**Автоматизированный анализ текстовых данных в изучении цифровых СМИ: простые инструменты для сложных теорий (на примере изучения массмедийного дискурса пандемии)**

Исследование выполнено за счет гранта Российского научного фонда № 22-28-00015.

В тезисах представлен алгоритм изучения биополитики государства как политического феномена при помощи базовых инструментов автоматизированной обработки текстов. Материалом для исследования послужили тексты российских и немецких медиа, публикующих материалы о пандемии в течение четырех «волн» (январь 2020 – март 2022 гг.) общим объемом 35,07 млн слов.

Ключевые слова: пандемия, медиа, идеологема, компьютерная лингвистика

Необходимость изучения современного медиадискурса подталкивает исследователей к непрерывному поиску новых методов исследования. Особенно востребованными являются инструменты компьютерной лингвистики для анализа больших массивов данных. Сложные инструменты автоматизированной обработки требуют наличия компетенций в таких сферах, как программирование, статистический анализ и текстовая аналитика, которыми большинство исследователей гуманитарных направлений (политологов, социологов, психологов, филологов и др.) зачастую не обладают.

Предлагаемый алгоритм анализа является примером использования простых инструментов компьютерной лингвистики для изучения комплексных социальных/политических феноменов и требует минимальных пререквизитов.

Алгоритм включает в себя следующие шаги [2]:

1) описание социального/политического феномена;

2) выбор социальной /политической теории для научной артикуляции феномена;

3) характеристика вербального контекста феномена в теории;

4) лингвистический анализ и определение лингвистических маркеров для данного феномена;

5) выбор инструментов компьютерной лингвистики (частотные n-граммы, коллокаты, ключевые слова);

6) поиск и отбор языковых единиц на больших массивах текстов;

7) характеристика феномена на большом и реальном материале.

Следуя перечисленным выше этапам, представляемое исследование изучает биополитику государства как политический феномен с опорой на теорию биополитики в понимании Мишеля Фуко. С точки зрения теории медиа, данное исследование базируется на модели пропаганды Хермана и Хомского [3], а на инструментальном уровне ‒ на алгоритме идентификации и анализа контекстуальных идеологем. Теоретическое поле биополитики в данном исследовании используется как основа для понимания действий правительств и сообществ, представленных государственными и негосударственными медиа, во время пандемии Covid-19.

Материалами исследования выступили три российских издания и четыре немецких СМИ, публикующие материалы о пандемии в течение четырех «волн» общим объемом 35,07 млн слов.

Тексты о пандемии были идентифицированы на основе ключевых слов («пандемия», «covid», «коронавирус»), извлечены из архивов и обработаны. Отправной точкой анализа послужила идеологема «вакцинация», которая реконструирует биополитическое измерение и содержит отсылку к «общественному телу» и защите общества от пандемии. Далее, с использованием меры относительной частотности (ipm), был составлен перечень частотных лексем. Согласно исследованиям в области культуромики (новая область текстологии), частотность лексем (включая и контекстуальные идеологемы) является эффективным современным способом изучения общества, так как указывает на культурные явления, отраженные в текстах массмедиа и не только [1]. В итоге анализ частотных лексем позволил ответить на вопрос, как идеологема «вакцинация» интегрируется в изучаемый медийный дискурс.

Результаты исследования показали, что государственные и негосударственные медиа в совокупности представляют неоднородную площадку репрезентации биополитических идей, однако российские СМИ – в критике или поддержке – более ориентированы на позицию правительства в создании биополитической повестки дня. Согласно проведенному исследованию, медиа (не будучи правительственными изданиями) могут активно участвовать в продвижении государственных биополитических проектов, поскольку, как и государство, заинтересованы в «сохранении населения».

Представляемое исследование продемонстрировало возможность, используя один из базовых инструментов автоматизированной обработки текстов, получить знание не только о том, как феномен проявляется в тексте, но и о социальных/политических процессах в стране.

Предлагаемый алгоритм может быть успешно использован для решения задач в области гуманитарного знания, однако для больших исследовательских проектов видится необходимым привлекать специалистов разного профиля, так как общее видение проблемы, усиленное компетенциями в технических вопросах, представляется наиболее эффективным.

Литература

1. Бонч-Осмоловская А. А. Культуромика в национальном корпусе русского языка, к постановке задачи: три века русских дорог // Труды института русского языка им. В. В. Виноградова. 2015. Т. 4. № 6. С. 605‒641.

2. Радина Н. К. Методика идентификации контекстуальных идеологем в цифровом медиадискурсе (на примере медиадискурса о пандемии Covid-19) // Вестник Московского университета. Серия 10: Журналистика. 2021. № 5. С. 116–136.

3. Pedro-Carañana J., Broudy D., Klaehn J. (eds.) (2018). The Propaganda Model Today: Filtering Perception and Awareness. London: University of Westminster Press. DOI: https://doi.org/10.16997/book27.