

Научная статья
УДК 80*81
doi: 10.17223/19986645/89/4

О прошлом, но в разное время: компьютерный анализ текстов учебников по истории СССР / России для шести поколений студентов

Анастасия Владимировна Колмогорова¹,
Полина Алексеевна Колмогорова², Елизавета Романовна Куликова³

^{1, 2, 3} *Национальный исследовательский университет «Высшая школа экономики»-Санкт-Петербург, Санкт-Петербург, Россия*

¹ *akolmogorova@hse.ru*

² *pakolmogorova@edu.hse.ru*

³ *ekulikova@edu.hse.ru*

Аннотация. Анализируются тексты о семи периодах российской истории в шести вузовских учебниках: 1946, 1983, 1997, 2001, 2006, 2010 гг. С помощью методов компьютерного анализа текстовых данных выявляется специфика поколенческого нарратива об истории страны в указанные периоды. Результаты демонстрируют, что тексты разных лет имеют различный эмоциональный «размах», разные тематические доминанты и фокусные тональности для разных исторических сюжетов. Данные отличия связаны с «духом» соответствующего периода – послевоенное время, период «развитого социализма», постперестроечное время реформ, становления новой российской государственности – и формируют специфический способ рассказывать истории «об истории», который можно назвать поколенческим историческим нарративом.

Ключевые слова: тексты учебников, идеологический дискурс, сентимент-анализ, тематическое моделирование, российская история

Благодарности: в данной научной работе использованы результаты проекта «Текст как Big Data: методы и модели работы с большими текстовыми данными», выполненного в рамках Программы фундаментальных исследований НИУ ВШЭ в 2024 году.

Для цитирования: Колмогорова А.В., Колмогорова П.А., Куликова Е.Р. О прошлом, но в разное время: компьютерный анализ текстов учебников по истории СССР / России для шести поколений студентов // Вестник Томского государственного университета. Филология. 2024. № 89. С. 73–103. doi: 10.17223/19986645/89/4

Original article

doi: 10.17223/19986645/89/4

About the past, but at different times: Computer analysis of texts in textbooks on the history of the USSR/Russia for six generations of students

**Anastasia V. Kolmogorova¹, Polina A. Kolmogorova²,
Elizaveta R. Kulikova³**

*^{1, 2, 3} National Research University Higher School of Economics,
St. Petersburg, Russian Federation*

¹ akolmogorova@hse.ru

² pakolmogorova@edu.hse.ru

³ ekulikova@edu.hse.ru

Abstract. In this article, we focus on the analysis of the texts of three history textbooks for university students published at different times: in 1946, 1983, 2001, 2006 and 2010. As a material, we use texts in each of the textbooks describing seven historical topics since the beginnings of the Principality of Kiev till the Reforms of Peter I. In our research, we tried to move away from the tradition banalized in discursive research to analyze history textbooks as a kind of ideologically labeled discourse. Instead, we consider the analyzed texts as a form of manifestation of a certain generational narrative. The authors of the textbooks, being not only institutional narrators, but also representatives of their generation, color the historiographical canvas, which remains, in principle, unchanged, with a certain emotional tone, and, when telling about the same events, shift the focus of thematic attention based on the spirit of their time. To solve this problem, we use computational linguistics methods: sentiment analysis, clusterization and topic modeling. Their use in combination with interpretive analysis allowed us to draw a number of conclusions: (1) there are historical subjects that are told, in general, within the same dominant tonality, while for others there is an ambivalence of evaluation; (2) the ranges even within the same tonality may vary greatly from textbook to textbook; for example, this is characteristic of texts about Ivan the Terrible; (3) each textbook is characterized by its own “tonal range”: 1946 is the most restrained, while 1983, 2001 and 2006 are the most altitudinal; (4) even in the textbooks of the same author team, published 9 years apart, the tonality of the texts of the same sections is not identical: from 1997 to 2006 it becomes, on the whole, noticeably more positive; (5) within the generational narrative, historical stories are revealed through the prism of a certain dominant idea – it is different for each time: for the post-war narrative it is the idea of protecting the state and paying attention to its geopolitical neighbours; for the post-perestroika period of the formation of the “young” Russian democracy (2001) – the idea of paysan community and veche as the forms of original democracy “of the people”; for the time of the formation of the modern Russian vertical of power (2010) it is the idea of centralization of power and its stability.

Keywords: textbook texts, ideological discourse, sentiment analysis, topic modeling, Russian history

Acknowledgements: This article uses the results of the Text as Big Data: Methods and Models for Working with Large Text Data project carried out within the framework of the Fundamental Research Program of the National Research University Higher School of Economics in 2024.

For citation: Kolmogorova, A.V., Kolmogorova, P.A. & Kulikova, E.R. (2024) About the past, but at different times: Computer analysis of texts in textbooks on the history of the USSR/Russia for six generations of students. *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology*. 89. pp. 73–103. (In Russian). doi: 10.17223/19986645/89/4

Введение

Данная публикация посвящена анализу учебных текстов по истории, осуществляемому с помощью инструментов автоматической обработки текстовых данных.

Материалом для анализа стали тексты шести учебников истории СССР/России для высших учебных заведений, изданных в 1946, 1983, 1997, 2001, 2006 и 2010 гг. государственными издательствами.

Мы не случайно взяли данные временные срезы. Они, в общем, последовательно отражают смену картин мира советского / российского общества: послевоенное время, так называемую эпоху советского застоя, эпоху рыночных реформ в послеперестроечное время, формирование «нового российского капитализма» в нулевые и дальнейшее политическое становление российского общества в следующее десятилетие.

При анализе данного материала мы фокусируемся на специфике не столько содержания, сколько способа рассказывания исторического нарратива [1]. Причем этот нарратив касается периодов истории страны, очень отдаленных по времени от институциональных рассказчиков, но построен так, что позволяет последним связывать описываемое время со «своим» временем посредством суждений, интерпретативный базис которых находится в настоящем рассказчиков [2]. В частности, нам интересно, какая эмоциональная тональность окрашивает рассказ о тех или иных периодах истории страны в учебных текстах, изданных в разные эпохи и предназначенных молодым гражданам, родившимся в разное время; как тематические цепочки, через которые конструируется общая «главная тема» рассказа о разных периодах истории, смещают тематический фокус в нем.

Тексты по истории в целом представляют довольно уязвимую дискурсивную субстанцию: в них легко увидеть идеологическую предвзятость, элементы пропаганды, рудименты политических установок разного времени. Однако в данном исследовании мы хотели бы уйти от традиции дискурсивного анализа непосредственного текста и предложить иной ракурс – дискурсивную интерпретацию данных, полученных на основе принципов «дальнего чтения» [3] и методов компьютерной лингвистики, для описания латентных, неявных, но от этого не менее значимых характеристик текстов учебников как «историй», которые рассказываются институциональными нарраторами.

Таким образом, целью данной публикации стало выявление общего и отличного в поколенческих нарративах институциональных авторов учебников по истории, принадлежащих шести разным временным срезам и расска-

зывающих студентам своего времени о событиях и деятелях, далеко отстоящих во времени от актуального рассказчика и слушателям момента. Подобный анализ осуществляется через призму компьютерного анализа характеристик текстов, в центре которых семь тем, две из которых касаются исторических личностей, а пять – наиболее удаленных во времени исторических периодов становления Российского государства, играющих, тем не менее важнейшую роль в национальной самоидентификации русских.

Наша гипотеза состоит в том, что при общей идентичности фактологической канвы в изложении исторических тем и периодов в анализируемых учебниках эмоциональная тональность текстов (сентимент), внутренние тематические цепочки (топики и термы) в них будут в той или иной мере отличаться, «выдавая» не столько идеологические «соскальзывания» авторов, сколько специфику поколенческого нарративного стиля институциональных авторов.

1. История: два значения слова, два подхода к анализу текстов учебников истории

В культовом для русской идентичности романе М. Булгакова «Мастер и Маргарита» есть игра слов, построенная на многозначности слова *история*. На вопрос Берлиоза «Так вы историк?» Воланд отвечает: «Я – историк... Сегодня вечером на Патриарших прудах будет интересная история!».

Если для Берлиоза *историк* – от слова *история* в значении 1 ‘научная дисциплина, занимающаяся изучением прошлого’, то для Воланда *историк* – от *история* в значении 2 ‘рассказ о событии или само событие’.

В политической лингвистике и дискурс-анализе существует богатая традиция изучения текстов учебников, в частности учебников по истории. Но эта традиция, если следовать сформулированной нами дихотомии *история 1* и *история 2*, соотносится с первой ее частью – в фокусе внимания исследователей тексты по истории как дисциплине, как совокупности транслируемых в текстовой форме знаний.

Данный подход получил развитие, прежде всего, в западноевропейском анализе дискурса, для которого тексты учебников истории интересны тем, что они, как и любые другие образцы педагогического дискурса, призваны легитимировать ценности, институционализированные в данном обществе на текущий момент его развития [4]. Так, например, в своей основополагающей работе С. Ситрон, анализируя французские школьные учебники по истории, издававшиеся начиная с XIX в., констатирует, что такие институциональные тексты – идеальная почва для возвращения «национального мифа» [5].

Несколько десятилетий спустя эта тема нашла продолжение в нашумевшем сборнике Жереми Дюбуа и Патрисии Легрис «Школьные предметы и политические культуры. Изменения в национальных моделях после 1945 года» [6]. Авторы постулируют, что национализация знания, произошедшая благодаря всеобщему государственному образованию, привела к

тому, что программы учебных дисциплин, учебники для их преподавания стали прямым инструментом формирования национальной идентичности, оказавшейся по этой причине чрезвычайно чувствительной к политической конъюнктуре. Эволюция идеологических установок в учебниках по истории показана авторами коллективного труда на примере языковых средств, используемых в текстах для развенчания культов личностей Гарибальди в Италии [7] и Сталина в России [8]. Автор последнего раздела, О. Конкка, в другом своем труде [9] показывает, как на лексическом уровне в постсоветских учебниках по истории сохраняются рудиментарные клише советских учебников: *триумф советской науки и техники, милитаристские круги, сельские труженики*. Данный факт интерпретируется исследователем как свидетельство глубинной дискурсивной преемственности между постсоветскими учебниками и их предшественниками: «За повторением «советизмов», устойчивых формулировок, многократно использованных в эпоху СССР, скрывается желание воспроизводить устойчивый *позитивный* образ отечественного прошлого» (курсив наш. – Авт.) [9. С. 63]. В этой связи возразим, что идеологически маркированные формулы содержат некоторую оценочную ловушку – их позитивность или негативность напрямую зависят от текущей дискурсивной формации и дискурсивной позиции автора [10]. Так, сочетание *сельские труженики* будет считываться как имеющее положительную тональность в дискурсах учебников 1940–1960-х гг., но уже в 1980-е и 1990-е гг., став признаком советского канцелярита, данную тональность потеряет, а в некоторых более поздних контекстах приобретет негативный оттенок. Подобные идеологические «реперные точки» для выстраивания актуальной моменту оценочной канвы довольно легко считываются методами классического экспертного дискурс-анализа.

Однако мы предлагаем оттолкнуться от второго значения лексемы *история* – историческое повествование, даже если речь идет о педагогическом тексте институционального автора, является одним из вариантов «рассказывания историй о событиях». В этом смысле нам близко определение А. Стибби: «Истории – это когнитивные структуры в мышлении индивидов. Эти истории определяют то, как мы мыслим, говорим и действуем. Истории, которыми мы живем, – это истории в сознании всех тех, кто принадлежит какой-либо культуре. Истории – это ментальные модели, упрощенные схемы, которыми мыслят люди» [11. С. 6]. Истории об «истории», которые рассказывают институциональные авторы, несмотря на общую фактологическую идентичность, укорененную в историографии, и обязательную четкую идеологическую рамку, тем не менее отражают укорененные в подсознании нарраторов как представителей своего поколения, в том числе поколения историков, ментальные модели прошлого, бытующие в лингвокультурном сообществе и уже – в рамках профессиональной среды, в актуальный нарраторам период. Обсуждая данный феномен, можно провести параллель с понятием социального настроения в социологии – «фундаментальная психологическая связь между предшествующим опытом социальной или этнической группы и ее способностью чувствовать, воспринимать

и оценивать» [12. С. 112]. Такая психологическая связь, как мы полагаем, должна обнаруживать себя в тексте за счет общей эмоциональной тональности, не связанной напрямую с оценочными идеологизированными клише, а также при помощи тематических фокусов и сдвигов, не проявляющих себя открыто в структуре тем и подтем, эксплицитно артикулируемой авторами учебников. Как ни странно, хорошей иллюстрацией того, что мы называем тематическим фокусом эпохи здесь может быть комментарий Н.И. Бухарина относительно одного из учебников истории для средней школы, представленных на конкурс, объявленный ЦК ВКП(б) в марте 1934 г.: «Народы (народы СССР. – *Авт.*) эти тракуются почти исключительно как объекты захватов; между тем нужно, группируя материал вокруг формирования и эволюции России, как государства, всё же давать материал... так, чтоб была и диалектика развития: самостоятельные народы – превращаются в посидельцев «тюрьмы народов» – превращаются затем вновь в самостоятельные народы, но уже на общей братской основе социализма» [13].

Отчасти в силу тех же «поколенческих» мотивов, о которых мы писали выше, исследователями, работающими в рамках самой исторической науки, предлагается рассматривать учебники истории не как авторские тексты, а как особый тип массовой литературы определенного времени: «...обстоятельства их происхождения связаны прежде всего с государственным заказом, содержательные линии определяются на основании конкретных документов (к примеру, стандартов, учебных программ и т.п.), требования к ним формулируются в соответствии с решениями конкретных органов (Министерство просвещения / образования, ЦК ВКП(б) / КПСС, Федеральный экспертный совет и т.д.)» [14. С. 266–267].

Несомненно, у предлагаемого нами конструкта «поколенческих историй об истории» есть ограничения: 1) тексты учебников истории отражали и внутридисциплинарные отличия исторических школ, научные концепции авторских коллективов и их руководителей (ср., например, «формационную» и «цивилизационную» парадигмы в практике написания учебников истории, упоминаемые в [15]); 2) отличие между «поколенческим умонастроением» и «идеологической рамкой» не во всех случаях легко выявить. Тем не менее у конструкта есть и свои существенные черты, позиционирующие его как феномен мезоуровня, находящийся глубже, чем идиостилевые черты или идеологические клише, но залегающий не так глубоко, как, скажем, глубинные нарративные падежи А. Греймаса [16]: 1) он не является непосредственным продуктом идеологии, хотя не исключает ее влияния; 2) он невольно отвечает социальному настроению общества «текущего момента», хотя и пропущенному через призму теоретических концепций историографии; 3) он не является индивидуальным стилем речи, письма автора текста, поскольку первоначальный текст проходит через серию социальных фильтров, становясь в итоге феноменом массовой литературы.

Для выявления специфики таких «историй об истории» мы решили использовать ряд методов компьютерной лингвистики, получивших широкое применение на разнообразных массивах текстовых данных.

Материал и методы

Материал исследования составил корпус из 42 текстов на русском языке общим объемом 575 000 токенов. Данная коллекция состоит из текстов разделов шести учебников по истории, выпущенных для студентов высших учебных заведений. Из каждого учебника были взяты тексты по следующим семи темам: Докиевская Русь, Киевская Русь, Феодальная раздробленность, Татаро-монгольское иго, Объединение русских земель Иваном Калитой, Правление Ивана Грозного, Реформы Петра I.

Список учебников, ставших источником текстового материала:

1. Базилевич К.В., Новицкий Г.А. История СССР. Ч. 1. От древнейших времён до конца XVIII века. М. : Издательство Высшей партийной школы, 1946. 748 с.

2. Белявский М.Т., Рыбаков Б.А., Леонтьев А.К., Новицкий Г.А., Сахаров А.М. История СССР с древнейших времён до конца XVIII в. / под ред. Б.А. Рыбакова. М. : Высшая школа, 1983. 415 с.

3. Орлов А.С., Георгиев В.А., Георгиева Н.Г., Сивохина Т.А. История России с древнейших времен до наших дней : учеб. М. : Проспект, 1997. 544 с.

4. Павленко Н.И., Андреев И.Л., Кобрин В.Б., Федоров В.А. История России с древнейших времен до 1861 года / под ред. Н.И. Павленко: учеб. 2-е изд., перераб. М. : Высшая школа, 2001. 536 с.

5. Орлов А.С., Георгиев В.А., Георгиева Н.Г., Сивохина Т.А. История России : учеб. 3-е изд., перераб. и доп. М. : ТК Велби, 2006. 568 с.

6. Дворниченко А.Ю. Российская история с древнейших времен до падения самодержавия: учеб. М. : Весь мир, 2010. 949 с.

Подчеркнем, что относительно небольшой объем анализируемого материала связан с пилотным характером исследования – мы проводим своего рода «разведывательный поиск» для того, чтобы на следующем этапе, проанализировав большие данные, получить в той или иной степени прогнозируемые результаты.

Тексты были преобразованы в машиночитаемый формат и преобразованы: удалены стоп-слова, тексты токенизированы при помощи пакета *quanteda* в *r* (*quanteda tokens*), приведены к строчному написанию, очищены от знаков препинания, цифр, небуквенных символов и лишних пробелов, наконец, для ряда задач использовалась лемматизация при помощи морфологического анализатора для русского языка *MyStem*.

Основными методами анализа выступили сентимент-анализ, тематическое моделирование и кластеризация текстов на основе их косинусного расстояния.

Необходимо отметить, что в современной лингвистике наблюдается всплеск внимания к категории эмотивности как неотъемлемой части текстов разного жанра – по словам В.И. Шаховского, «весь язык эмотивен... не существует нейтральной лексики» [17. С. 25]. Сентимент-анализ, или анализ

тональности, – получивший широкое распространение компьютерный метод для выявления выраженных лексически эмоциональной оценки и субъективного отношения автора текста к некоторому объекту [18]. Вместе с расширением числа подходов, используемых для подобного анализа (словарный, основанный на правилах, нейросетевой), расширяется и спектр типов текстов, подвергающихся сентимент-анализу. Если традиционно метод использовался применительно к постам в социальных сетях [19, 20], отзывам на товары и услуги [21], кинокритикам [22], то сегодня уже появляются работы по сентименту в художественных [23] и академических текстах. Так, в [24] представлены результаты анализа тональности словарным методом 66 учебников по психологии, выпущенных за последние одиннадцать десятилетий, – эмоциональная окраска текстов с каждым годом приобретает все более негативный оттенок. Аналогичное исследование с той же методологией было проведено на материале современных отечественных учебников по истории и обществознанию для школьников 5–11-х классов в [25]. Авторы приходят к выводам, что в учебниках по истории превалирует негативный лексикон и что доля оценочной лексики в учебниках для старшей школы оказывается ниже, чем в учебниках для средних классов.

Опираясь на методику работы, представленную в [24, 25], мы также использовали словарный подход к работе с тональностью в том числе и потому, что имеющиеся дообученные модели [26] не позволяют анализировать большие по объему тексты (до 300 слов) и, будучи обучены на достаточно разнородных датасетах из социальных сетей, нечувствительны к «сдержанной» тональности не-сетевых текстов. Словарь же тональности уже содержит в себе ограниченный список слов некоторого языка, как правило сформированный авторами словаря самостоятельно и размеченный по шкале «позитивность – нейтральность – негативность» информантами, ангажированными разработчиками словаря [27]. Для русского языка на сегодняшний момент существуют три оригинальных, т.е. непереводаемых, достаточно широко используемых словаря: «Круг слов» [28], PolSentiLex [29] и RuSentiLex [30, 31]. У словарей много общего, например использование краудсорсинга для разметки, но есть и отличия: PolSentiLex – доменно-специфичный словарь, созданный для анализа межэтнической напряженности в социальных сетях, поэтому для нас он был нерелевантен; «Круг слов», кроме среднего значения сентимента по данным разметки, дает еще и для каждого такого значения коэффициент согласия аннотаторов – исключив из списка негативных и позитивных слов те из них, которые имеют низкие коэффициенты согласия, мы получили слишком ограниченный список слов; RuSentiLex имеет характер «общего» словаря, где размечены не только униграммы, но и n-граммы, поэтому мы воспользовались именно им.

Чтобы учесть при анализе n-граммы (они даны в лемматизированной форме), мы разбивали тексты для анализа не только на униграммы, но и n-граммы длиной от 2 до 6 слов, используя при этом инструмент `quanteda.tokens_ngrams` в пакете `quanteda` в г. N-граммы извлекались последова-

тельно: например, предложение «основной театр военный действие переноситься на юг» представлялось в виде униграмм (*основной, театр, военный, действие, переноситься, на, юг*), биграмм (*основной_театр, театр_военный, военный_действие...*), триграмм (*основной'_театр_военный, театр_военный_действие, ...*) и т.д. После этого производилось сопоставление текстов с полным вариантом словаря RuSentiLex. С помощью метода *quanteda::dfm_lookup* было подсчитано количество положительных и отрицательных слов из словаря, которые встретились в наших текстах. Список этих слов был проанализирован: для многих слов в RuSentiLex предложено несколько вариантов аннотации (в зависимости от контекста). Для подобных слов было уточнено контекстуальное значение, которое они имеют в текстах из нашей выборки. Например, из трех возможных вариантов значений для слова *великий* (1) *великий, Adj, великий, negative, opinion, "ВЕЛИК ПО РАЗМЕРУ"*; (2) *великий, Adj, великий, neutral, opinion, "БОЛЬШОЙ ПО РАЗМЕРУ"*; (3) *великий, Adj, великий, positive, opinion, "ВЫДАЮЩИЙСЯ, ВЕЛИКИЙ"* для текстов учебников истории релевантным оказалось только третье – с положительной коннотацией. Слово же *первобытный* в исторических текстах употребляется только в коллокации *первобытно-общинный строй*, поэтому маркировалось как нейтральное и было исключено из списка слов с позитивной тональностью (такая тональность валидна для значения 'девственный, нетронутый'). Таким образом, итоговый список позитивных слов включил 873 (–89 ед. по отношению к оригинальному списку из словаря) единицы, а список негативных – 1502 (–82 ед.).

Для вычисления итогового значения сентимента для текста рассчитывалась доля положительных или отрицательных слов по отношению к общему числу эмоциональных слов в этом тексте. Для аналитики применялся ряд частных методов, каждый из которых уточняется непосредственно в соответствующих пассажах текста статьи.

Кластеризация текстов производилась на основе вычисления косинусного расстояния: каждая биграмма текста преобразовывается в числовой вектор при помощи предобученной модели *fastText*, вычисляется среднее косинусное расстояние для всех биграмм текста – сходство между текстами оценивается через косинус между их усредненными векторами по формуле $1 - \|u\|_2 \cdot \|v\|_2 u \cdot v$. Для иерархической кластеризации использовался метод дальнего соседа (*complete linkage*), при котором в ходе объединения кластеров в один большой кластер попадают пары, демонстрирующие наименьшее расстояние между их наиболее далекими членами.

Облака слов строились на основе предобработанных указанными выше методами текстов, к которым применялся метод *Bag of words*, а его результаты репрезентировались в виде «облака слов»: чем крупнее шрифт, тем частотнее слово.

Для тематического моделирования применена модель *Structural Topic Modeling (STM)*, поскольку на материале художественных текстов (анализируемые нами академические тексты все-таки ближе к ним, чем к текстам социальных сетей), как показал сравнительный анализ моделей Латентного

размещения Дирехле (LDA), неотрицательной матричной факторизации (NMF) и STM, представленный в [31], именно данная модель позволяет построить дистинктивный словарь для каждой темы.

Последовательность методов выстроена следующим образом: сначала мы используем кластеризацию, чтобы провести самую первую проверку гипотезы и ответить на вопрос, какие факторы могут влиять на «похожесть» текстов: время написания, авторский коллектив, тема; получив предварительно некоторые результаты, мы проверяем их, используя sentiment-анализ текстов; выявив отличия в тональности текстов, посвященных одной и той же исторической теме, но в учебниках разных лет издания, осуществляем тематическое моделирование, чтобы получить ответ на вопрос, может ли изменение сентимента быть связано с изменением тематического фокуса «рассказывания истории об историях» или возможны случаи, когда тематический фокус константен, но меняется эмоционально-оценочное отношение к нему.

Результаты и обсуждение

На рис. 1, 2 представлены одни и те же результаты кластеризации: на рис. 1 семь кластеров визуализированы при помощи метода мультимодального шкалирования (Multidimensional Scaling), на рис. 2 – аггломеративной кластеризации, при которой построение кластеров идет от мелких к крупным и делается по принципу «схожести» объектов.

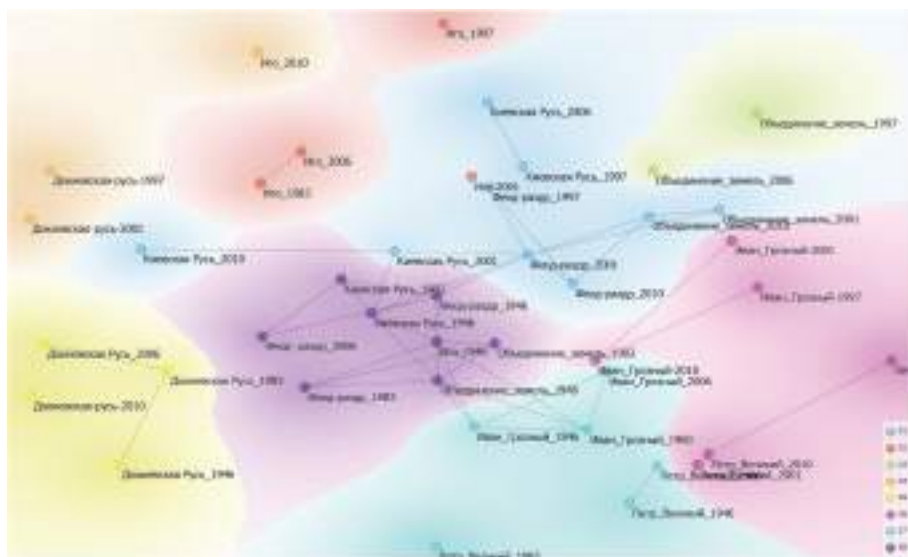


Рис. 1. Кластеризация текстов на основе их косинусного расстояния, визуализированная при помощи мультимодального шкалирования

одинаково – Петр Великий и Иван Грозный. Проверять эти гипотезы мы продолжили при помощи сентимент-анализа.

После проведения сентимент-анализа текстов методом, описанным выше, чтобы сравнить тексты нашего корпуса между собой, мы перевели полученные значения доли позитивных и негативных слов в шкалу от -1 до $+1$ с помощью функции `rescale` в пакете `Scales` для R. Результаты визуализированы, см. приложение.

Такой метод перевода значений не говорит о том, что текст с тональностью $+1$ исключительно позитивный – он просто самый позитивный из наших текстов, поскольку там самая большая доля положительных слов.

Рассмотрим некоторые общие тенденции.

В табл. 1 приведено количество учебников в выборке, в которых та или иная тема была представлена с преобладанием позитивных или негативных слов.

Таблица 1

Количество вхождений тем в категории «с преимущественно позитивной тональностью» / «с преимущественно негативной тональностью» по выборке из шести учебников

Количество учебников, в которых превалирует тональность	Докиевская Русь	Киевская Русь	Феодальная раздробленность	Иго	Объединение земель	Иван Грозный	Петр Великий
Позитивная	4	2	4	1	3	0	0
Негативная	2	4	2	5	3	6	6

Заметно, что только две темы из семи в большинстве учебников представлены нарративами с преобладанием (в разной степени) позитивной тональности – это темы «Докиевская Русь» и «Феодальная раздробленность». Если посмотреть на специфику распределения сентимента для первой темы по учебникам (см. приложение), то с наибольшей позитивностью (0,8) она представлена в 2001 г., 0,33 – в 2006, 0,24 – в 1997, 0,09 – в 1983; с незначительной долей преобладания негативного вокабуляра ($-0,008$ и $-0,1$) в 1946 и 2010 гг. соответственно. Принимая во внимание тот факт, что учебники 1997 и 2006 гг. написаны одним и тем же авторским коллективом, представляется интересным, что за 9 лет, разделяющие эти тексты, Докиевская Русь и проблема происхождения славян стали видеться авторам в более положительном ключе. Кроме того, в рамках данной темы несколько изменились акценты в описании этнического «котла», в котором возникла русская национальная идентичность. На рис. 3 представлены в сопоставлении реберные графы относительных частот (`ipm`) всех встретившихся в данных

разделах этнонимов + имя собственное *Византия*, полученные в корпусном менеджере Voyant. Заметно, что в тексте 2006 г. по сравнению с 1997 г. сократилась частотность лексемы *славяне*, сократилось также лексическое «присутствие» Византии, совсем исчезли греки и гунны, но появились скифы.

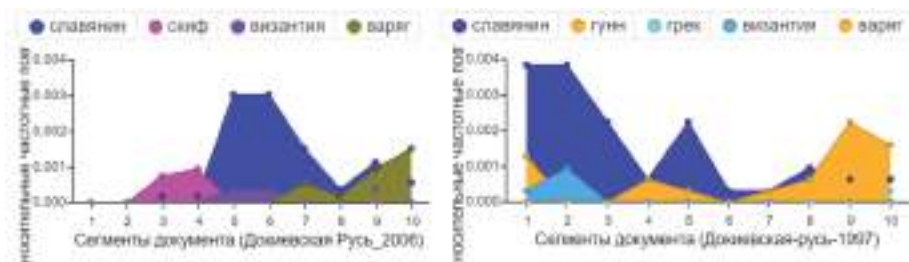


Рис. 3. Реберные графы относительных частот лексем-этнонимов и имени собственного *Византия* в текстах учебников 1997 и 2006 гг. (авторы: А.С. Орлов, В.А. Георгиев, Н.Г. Георгиева, Т.А. Сивохина), посвященных теме «Докиевская Русь»

Тенденцию к повышению сентимента с годами наблюдаем в обсуждаемых двух учебниках и в теме «Феодальная раздробленность»: в 2006 г. – это самый позитивный текст в выборке (1,0), хотя в 1997 г. у тех же авторов позитивность была минимальной (0,26).

Возвращаясь к общей характеристике динамики сентимента в выборке, отметим, что абсолютное преобладание негативного вокабуляра над позитивным характерно для тем «Иван Грозный» и «Петр Великий». Несомненно, в целом тексты об Иване Грозном содержат больше негативных слов (униграмм и биграмм), чем тексты о Петре Великом. Тексты об Иване Грозном располагаются в интервале от $-0,47$ в учебнике 1946 г. до $-0,98$ и $-1,0$ в учебниках 1983 и 2001 гг. соответственно. При этом текст 2001 г. – самый негативный текст в выборке.

Что касается Петра Великого, то здесь значения преобладающего сентимента варьируют не так заметно и не в столь широком диапазоне: от $-0,03$ в учебнике 2001 г. до $-0,35$ в учебнике 2006 г.

Другая тема с преимущественно негативным вокабуляром – «Иго». Единственный учебник, рассматривающий эту тему с позитивным вокабуляром, минимально преобладающим над негативным, – это учебник 1946 г., но, как мы покажем далее, такая тенденция связана, скорее, с общей эмоциональной сдержанностью текста всего учебника.

Примечательно, что две наиболее противоречивые по сентименту темы – это «Киевская Русь» и «Объединение земель». Для темы «Киевская Русь» минимальное преобладание позитивного вокабуляра зафиксировано в учебниках 1983 и 2010 гг., где становление государственности подается через идеи власти, порядка, а частотными словами становятся *государство*, *культура*, *Русь*, *церковь* (см. облака слов в приложении) в таких, например, контекстах:

1) Гарнизоны новых пограничных крепостей были набраны из дружинников далеких северных земель (кривичей, вятичей и словен), чтобы привлечь к обороне государства все силы новой державы. Опираясь на эти линии, Владимир оградил Русь от новых набегов (учебник 1983 г., с. 65);

2) Годы его правления были плодотворны для развития русской культуры: в Киеве возводится целый ряд монументальных храмов (учебник 2010 г., с. 127).

В других учебниках преобладает негативный вокабуляр, что становится особенно заметно в 1997 и 1946 гг. – красной нитью проходят образы кровопролитной борьбы молодого государства за самостоятельность, что проявляется и в выдвигании на первый план имена князей, таких слов, как дружина, дань (см. приложение):

3) Наиболее мощным было восстание в Киеве в 1068 г. Оно вспыхнуло в результате поражения, которое потерпели сыновья Ярослава (Ярославичи) – Изяслав (ум. 1078 г.), Святослав (ум. 1076 г.) и Всеволод (ум. 1093 г.) – от половцев (учебник 1997 г., с. 29);

4) С дружиной князь ходил в походы, подавлял восстания отложившихся племен, захватывал новые земли, делил военную добычу и собирал дань (учебник 1946 г., с. 84).

Объединение же земель имеет достаточно значительное доминирование позитивного вокабуляра (см. приложение) в 1946 г., а далее в порядке убывания его значимости – в 2001 и 2010 гг. и переходит, в порядке возрастания, в доминирование негативного вокабуляра в 1997, 2006 и 1983 гг..

Для того чтобы посмотреть «размах» эмоциональной тональности по каждому из учебников в сравнении, мы использовали метод подсчета Compound score:

$$\frac{\text{Sum(pos)} - \text{Sum(neg)}}{\text{TextLength}} \times 100,$$

где Sum(pos) – количество положительных слов; Sum(neg) – количество негативных слов; TextLength – общее количество слов в тексте. Данный метод используется для подсчета значений сентимента и их сравнения в таких известных инструментах, как, например, Orange [32].

Визуализация (рис. 4) демонстрирует, что тональный «размах» учебников неодинаков: самый сдержанный диапазон сентимента характерен для учебника 1946 г. – от 0, до –1,84; самые выраженные «качели» из позитива в негатив демонстрирует учебник 2006 г. (1,24...–2,06), самый большой диапазон, но со скосом в негативные значения – учебники 2001 г. (0,74...–3,43) и 1983 г. (0,17...–3,41). Учебник 1997 г., по-видимому, самый негативный, для него по методу Compound score нет значений метрики в положительную сторону ни по одной теме. Подтверждается и отмеченная нами ранее тенденция в изменении сентимента в написанных одним и тем же авторским коллективом учебниках 1997 и 2006 гг.

За исключением Ивана Грозного, вокабуляр для всех тем становится более позитивным.

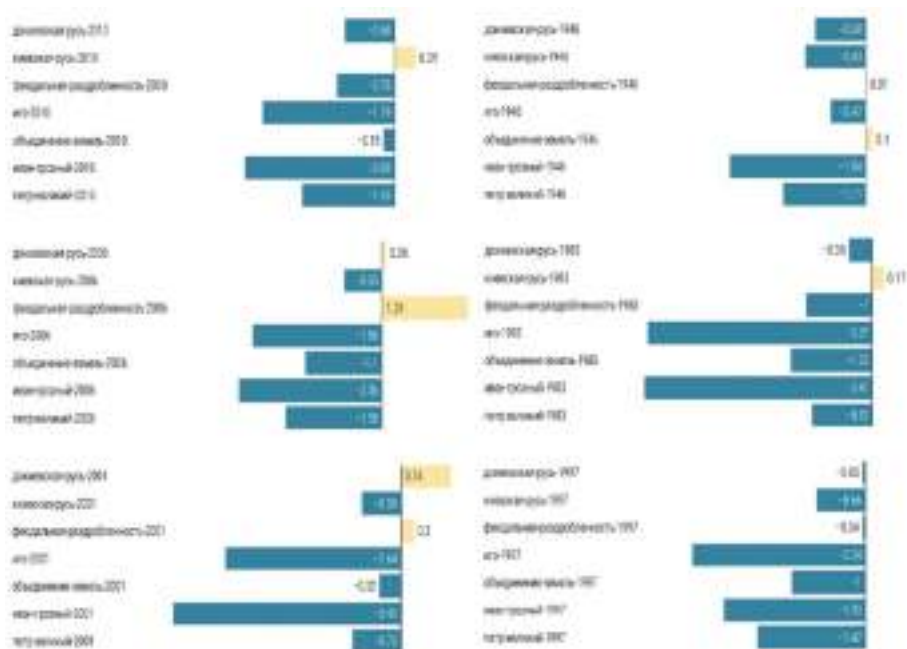


Рис. 4. Колебания значения сентимента в каждом из шести учебников, рассчитанные методом Compound Score

Рассмотрим показательный фрагмент текста о Петре Великом, начинающий подпараграф «Социальные противоречия в первой четверти XVIII в.» в двух учебниках в режиме «было – стало»:

5) **Социальные противоречия в первой четверти XVIII в.** Как заметил А.С. Пушкин, Петр I показал все черты «нетерпеливого, самовластного помещика», многие указы которого «писаны кнутом». В 1705–1706 гг. восстали стрельцы, работные и посадские люди, беглые крестьяне в Астрахани, которые более семи месяцев удерживали город. На подавление восстания Петр направил своего лучшего полководца фельдмаршала Б.П. Шереметева. Почти семь лет продолжались волнения в Башкирии (1705–1711) (учебник 1997 г., с. 154).

6) **Социальные противоречия в первой четверти XVIII в.** Вся тяжесть петровских преобразований легла на плечи трудового населения. Это вызвало ряд мощных народных волнений. В 1705–1706 гг. восстали стрельцы, работные и посадские люди, беглые крестьяне в Астрахани, которые более семи месяцев удерживали город. На подавление восстания Петр направил своего лучшего полководца фельдмаршала Б.П. Шереметева (учебник 2006 г., с. 142).

В первом фрагменте (пр. 5) негативная оценочность выражена более резко за счет эпитетов (*нетерпеливого, самовластного помещика*), образности, создающей экспрессивность (*писаны кнутом*) в первом предложении.

Во втором фрагменте та же идея передается уже более сдержанно (*тяжесть преобразований, мощных волнений*) и при помощи не одного, а двух предложений.

Таким образом, проведенный анализ сентимента позволяет сделать несколько предварительных выводов:

1) есть исторические сюжеты, которые рассказываются в пределах одной и той же доминирующей тональности, а для других – налицо амбивалентность оценки: Докиевская Русь и Феодальная раздробленность в большинстве учебников тяготеют к позитивному вокабуляру; Иго, Иван Грозный и Петр Великий – к негативному; доминирование того или иного тонального вокабуляра в темах Объединение земель и Киевская Русь в большей степени зависит от года издания учебника;

2) однако диапазоны даже в рамках одной тональности могут сильно различаться от учебника к учебнику, например, это характерно для текстов об Иване Грозном;

3) для каждого учебника характерен свой «тональный размах»: 1946 г. – самый сдержанный, а 1983, 2001 и 2006 гг. – наиболее амплитудные; 1997 г. – самый негативный; 2010 приближается по сбалансированности тональности к 1946 г.;

4) даже в учебниках одного и того же авторского коллектива, изданных с интервалом в 9 лет, тональность текстов одних и тех же разделов не идентична: от 1997 к 2006 г. она становится в целом заметно позитивнее.

Следующим этапом стало проведение процедуры **тематического моделирования** на основе алгоритмов машинного обучения. Как уже отмечалось выше, использовалась модель Structural Topic Modeling. Определяя количество тем для моделирования, мы остановились на 14. Хотя максимальные значения метрики Held-Out Likelihood (вероятность хорошо интерпретируемых тем) и минимальные значения метрики Residuals (вероятность неинтерпретируемых тем) достигаются на 10 темах (рис. 5), мы решили использовать 14 тем, поскольку значения метрик в данном случае ухудшаются незначительно, а объем данных для интерпретации расширяется существенно. Кроме того, установлено [33], что в подавляющем большинстве существующих сегодня моделей для тематического моделирования оптимальным является количество метрик выше 10, но меньше 15.

Для тематического моделирования мы использовали не лемматизированные варианты текстов, разбитые на токены и приведенные к нижнему регистру; для получения более интерпретируемых термов в корпус одиночных токенов дополнительно добавлены частотные коллокации.

При формулировании названий тем мы опирались на метрику FREX (Frequency+Exclusivity – сочетание значимости и исключительности данного термина для всей темы): для каждой темы термы ранжируются в порядке убывания значений метрики – чем ближе слово к началу ряда, тем выше значения метрики (табл. 2).

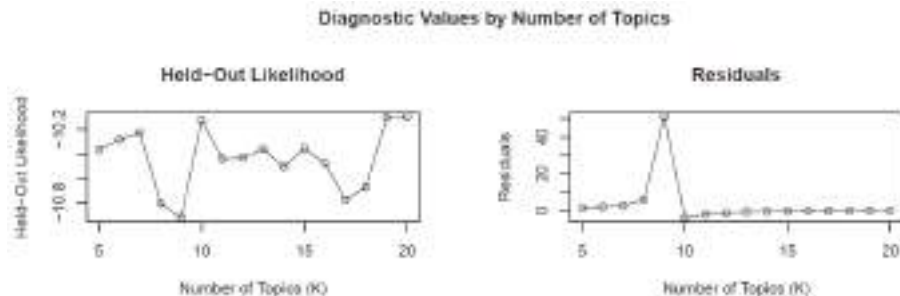


Рис. 5. Динамика значений метрик Held-Out Likelihood и Residuals в зависимости от количества выделяемых моделью тем (для предложенной выборки текстов)

Таблица 2

Темы и их термы, полученные при помощи модели структурного тематического моделирования на корпусе всех проанализированных текстов из шести учебников

№	Название темы	Термы темы в порядке убывания значений FREX ¹
1	Государственное устройство Киевской Руси	руссы, киевского_государства, по-видимому, ленин, верви, киевское_государство, руссов, владимир_мономах, киевского_князя, лев_диакон, дунае, славянские, воины, славянского_населения, киевском_государстве, киевской, восточно-славянских_племен, святослава, киевских_князей, век
2	Принятие христианства на Руси	погосты, гривен, апостол_андрей, первая_половина, принятие_христианства, ольги, игорь, ярославе_мудром, аскольд, правда_ярославичей, варяжской, глава_государство_русь, окаянного, игоря, пространной, древлянам, правда_ярослава, дружинники, христианство, принятия_христианства
3	Политика консолидации Ивана III	ивана_третьего, калиты, иван_данилович, шемяка, иван_третий, московских_князей, дмитрия, дмитрий, мамай, великое_княжение, ахмат, московской_земли, палеолог, мамая, василий, куликовская_битва, василию, единого_государства, феодалная_война
4	Реформы Петра	посошков, сената, балтике, регулярной_армии, казенных, новшества, прокоповича, концу_века, губернии, промышленности, мануфактур, учреждений, преобразований, петровского_времени, дворянства, булавина, просвещения
5	Защита русских земель от захватчиков	Захватчики, монголо-татар, русских_полков, монголами, внук_чингисхана, монголы, александр, золотая_орда, бату, монголов, монгольских, монголам, крестоносцев, завоевания_земель, нойонов, монгольские, ледовое_побоище, захватчиков
6	Происхождение славян и их этнические соседи	сардури, рабовладельцев, формации, гунны, антами, всей_вероятности, племени, учебника, ваварами, славянской, анты, племен, культура, римской_империи, наука, эпос, урарту, славянские_племена

¹ Значения округлены до 4-го знака после точки.

7	Политическая раздробленность	xii_xiii вв, владими́ро-суздальское княжество, галицко-во-лынско́й, андре́я, новгородско́й земли, респу́блики, полити́ческая раздробленность, новгородско́я земля, усоби́цы, киевско́й держа́вы, ростово-суздальско́я земля, большо́е гнездо, своего́ княжества, княжеств, княжества, роман_мстисла́вич, соборы, андре́я_боголю́бского, польскими
8	Внешняя и внутренняя политика Ивана Грозного. Опричнина	иван_iv, пересветов, избранная_рада, ивана_пересветова, ченслер, годы_боярского_правления, избранной_рады, курбский, царь_иван, ливонско́й войны, опри́чнину, опри́чина, опри́чины, курбско́го, опри́чнине, казани, опри́чного, сиби́рское ханство, курбско́му
9	Территория нашей страны в древности	восточно-славянского, лет_назад, объединений, артании, карпатских, освоили, пахотная, рюген, союзами, греческие, территории_нашей_страны, свидетельства, варяг, объединялись, племенные_княжения, орудий_труда булгария, зерно, восточно-славянских, археологическим_данным
10	Культура геополитических соседей Руси	руставели, низами, геродота, альберт, урарту, александр_ярославич, аракса, ванского_озера, тамары, экономическую, греческих_колоний, область, малой, поэме, азербайджан, греции, греческой, грузии, народного_творчества
11	Российские самодержцы	пётр, самодержавного, пётр_i, грозного, московского_государства, грозный, астрахань, причём, детей_боярских, фёдора, иван_грозный, всё-таки, российских, российским, ивана_грозного, священно́й, подписа́н, российско́й
12	Борьба Московского княжества с внешними врагами	тверскому_князю, юрий_данилович, чингисхан, александр_михайлович, вольных_слуг, ягайло, витовта, иван_калита, московского_князя, тимур, татары, витовт, литовского_княжества, состояли, некоторых_случаях, великокняжескую_власть, феода́льного хозяйства, ольгерд, войсках
13	Славянские племена и установление государственности на Руси	вождества, волости, городов-государств, язычества, столе, города-государства, волость, мономаха, долгое_время, пле́мён, вождество, галичей, кривичей, киевлян, изяслава, мономах, языческих, ла́дье, народно́й культуры, сажень
14	Армия и военные походы при Петре	второй_половине_xvii, никон, заводы, петровских_реформ, русского_флота, сухопутная_армия, регулярные, крымские_походы, австрией, подразделялись, предприняла, столп, петровского, потрясения, петр_i, карелии, стрельцов, флот, азов

К сожалению, пакет для STM не позволяет визуализировать матрицу значений метрики для каждого термина, поэтому мы опираемся на доказательства эффективности использования данной метрики, представленные в [34]. Над формулировкой самих названий тем на основе отобранных моделью термов работала группа экспертов-лингвистов, после обсуждения и гармонизации мнений выносилось окончательное решение о формулировке темы. Например, для темы 4 экспертной группой были предложены варианты названия: Армия при Петре Великом, Промышленность и экономика при Петре Великом, Реформы Петра. После обсуждения был выбран последний вариант,

поскольку среди термов есть токены, которые отсылают как к политическим, так и к экономическим реформам Петра, а также к военным преобразованиям, проведенным им. Кроме того, наибольшие значения по метрике FREX получила фамилия Посошков (И.Ф. Посошков, современник Петра I и ярый сторонник его реформ), униграмма *сената*, которая связана с фактом учреждения Петром Сената как высшего органа государственной власти, подчиняющегося императору; биграмма *регулярной_армии*, отсылающая к военным реформам императора, и т.д.

На рис. 6 представлена диаграмма распределения тем по учебникам. Каждый цвет маркирует определенную тему согласно легенде, где номера тем соответствуют номерам тем, представленных развернуто (сформулировано название и приводятся термы) в табл. 2. Например, топик 2 (рис. 6) – это тема 2 «Принятие христианства на Руси» (табл. 2).

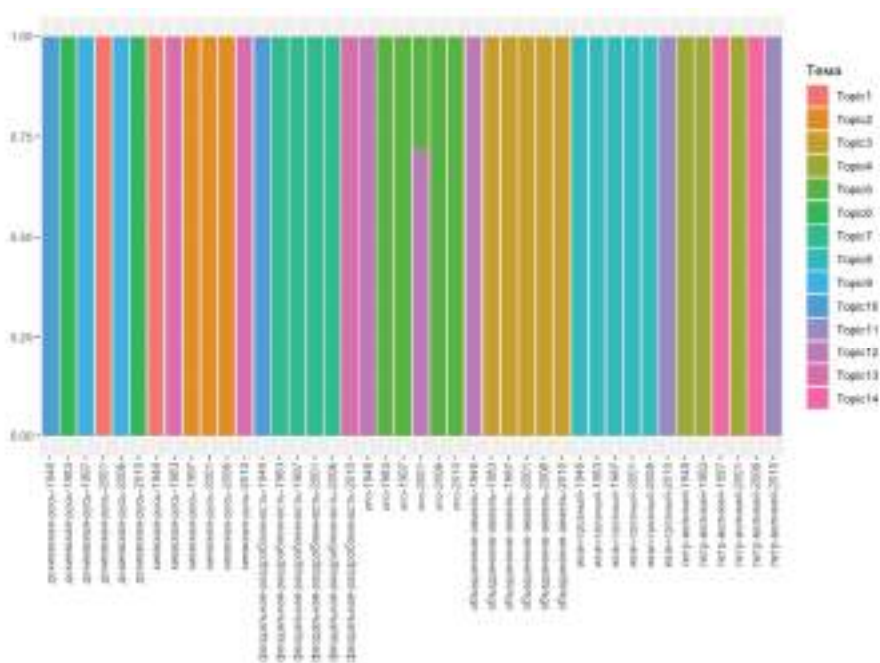


Рис. 6. Распределение тем по коллекции текстов учебников

Визуализация хорошо демонстрирует (см. рис. 6), что все исторические темы можно разделить на две большие группы: 1) стремящиеся к тематической гомогенности по годам (Феодальная раздробленность, Объединение земель, Иван Грозный) и 2) характеризующиеся тематической гетерогенностью, когда доминирующие темы обнаруживают значительную зависимость от года издания учебника (Докиевская Русь, Киевская Русь, Иго, Петр Великий).

Рассмотрим первую группу. В группе текстов о феодальной раздробленности доминирует тема 7 «Политическая раздробленность», в которой в качестве термов присутствуют названия княжеств, существовавших на территории Руси в XII–XIII вв. Исключение составляют учебники 1946 и 2010 гг. В первом доминирует тема «Культура геополитических соседей Руси», в которой важное место занимают имена великих литераторов (*Низами, Руставели*), покровителей искусств (*Тамара*), являющихся преимущественно представителями культуры народов Закавказья (к таковым в учебнике относятся не только Грузия, но и Азербайджан), а также существительные, относящиеся к сфере искусства: *поэма, народное творчество*. Такой акцент далеко не случаен: в тексте учебника раздел «Феодальная раздробленность (XII–XV вв.)» содержит подразделы, в которых история России рассматривается с высоты масштабного геополитического «полета» – в тесной связи с процессами, происходившими в Средней Азии и в особенности – в Закавказье, например, это хорошо заметно по названиям подразделов учебника «Феодальная раздробленность в Восточной Европе и феодальные объединения в Средней Азии и Закавказье в XI–XIII вв.» и «Закавказье в XI–XIII вв.». Идея параллельности исторических процессов, происходивших в России и в Грузии, проходит красной нитью через данный раздел. Ср., например:

7) *В.И. Ленин первый дал научное определение русскому феодализму, исходя из способа производства и системы общественных отношений. Феодальное натуральное хозяйство неизбежно приводит к политической обособленности. И.В. Сталин на примере феодальной Грузии показал, что экономическая раздробленность препятствует политическому объединению страны* (учебник 1994 г., с. 114).

Интересно, что в данном учебнике при рассмотрении периода феодальной раздробленности фокус смещается с усобиц на развитие культуры в перспективе «Русь – Закавказье». С этим, очевидно, связан и практически позитивный сентимент (см. рис. 4) данной темы в учебнике.

В учебнике же 2010 г. доминирует тема 13 «Славянские племена и установление государственности на Руси», которая была бы более уместна для предыдущего раздела о Киевской Руси. Но экспертный анализ текста показывает, что при изложении данной темы авторы фокусируются на специфике перехода «от вождеств к городам-государствам», подробно описывая сложившиеся в XI–XII вв. княжества, структуру органов власти в этих городах-государствах, но не останавливаются подробно на усобицах.

Сюжет же объединения земель во всех учебниках за исключением 1946 г. раскрывается через тему 3 «Политика консолидации Ивана III», главными терминами которой становятся имена собственные самого Ивана III, его жены Софьи Палеолог, их сына Василия, заговорщика Дмитрия Шемяки, ордынских ханов, а также биграмм, имеющие собирательную семантику: *московские князья, московские земли, единое государство*. В учебнике же 1946 г. в обсуждаемом разделе доминирует тема 12 «Борьба Московского княжества с внешними врагами», которая реализуется такими терминами, как

имена польских, литовских, ордынских князей, а также именем самого Ивана Калиты.

Рассказывая об Иване Грозном, авторы всех учебников, кроме авторов учебника 2010 г., акцентируют внимание на теме «Внешняя и внутренняя политика Ивана Грозного. Опричнина» (топик 8 на рис. 6), где основными терминами становятся разные формы лексемы *опричнина*, а также имена собственные Ивана Курбского; автора двух челобитных, в которых была предвосхищена политическая программа Ивана Грозного, Ивана Пересветова, а также *Казань*, взятие которой стало важной вехой деятельности Ивана IV. В стоящем особняком учебнике 2010 г. ведущей становится тема «Русские самодержцы», в которой фигуры Петра I и Ивана IV сближаются. Основывается такое сближение на их активной и жесткой внешней политике, стремлении расширить территорию страны, «прорубить окно в Европу», а также укрепить собственную власть внутри страны:

8) *В конечном итоге главная направленность опричнины заключалась в том, чтобы неизмеримо усилить власть и оторвать, отделить ее от массы населения. Это, безусловно, удалось. При этом сформировалась народная доктрина «доброто», истинного царя, произошла сакрализация царской власти (учебник 2010 г., с. 350).*

Рассмотрим вторую группу исторических сюжетов – тех, где наблюдаем больший разброс топиков от учебнику к учебнику.

Так, сюжет о Докиевской Руси в изложении институциональных авторов 1946, 1997 и 2006 гг. видится через призму темы 10 «Культура геополитических соседей Руси»; в учебниках 1983 и 2010 гг. – через призму темы 6 «Происхождение славян и их этнические соседи», а в учебнике 2001 г. ведущее место занимает тема 1 «Государственное устройство Киевской Руси». Причем экспертный текстовый анализ демонстрирует, что авторы смещают акцент на общинный характер этого устройства, на роль вервей, в которых общинная самоорганизация достигла высокого уровня и предопределила законы народной демократии, которые затем проявились, например, в Новгородской республике:

9) *Анализ социально-политических структур позволяет говорить о трех центрах притяжения, влиявших на общественное развитие: это прежде всего княжеская власть, набиравшая силу дружина (боярство), народное вече. В дальнейшем именно соотношение этих властных элементов станет определять тот или иной тип государственности, который возобладает на территориях, некогда входивших в состав державы Рюриковичей.*

Тема 2 «Принятие христианства на Руси» становится определяющей для рассказывания «истории» о Киевской Руси в трех постсоветских учебниках 1997, 2001 и 2006 гг., тема 13 «Славянские племена и установление государственности на Руси» – в учебниках 1983 и 2010 гг., тема 1 «Государственное устройство Киевской Руси» – в учебнике 1946 г.

Разделы о татаро-монгольском иге в анализируемых учебниках обнаруживают, в общем, либо тему 5 «Защита русских земель от захватчиков»

(1983, 1997, 2001 (частично), 2006, 2010), либо 12 – «Борьба Московского княжества с внешними врагами» (1946, 2001 (частично)).

Аналогичным образом, через достаточно близкие темы раскрывается исторический раздел о Петре Великом: тема «Армия и военные походы при Петре» (1997, 2006) и тема «Реформы Петра» (1946, 1983, 2001). Специфическая для учебника 2010 г. тема 11 «Российские самодержцы», проявившаяся в разделе об Иване Грозном, структурирует и рассказ о Петре Великом.

Обсуждение

Проведенный анализ позволяет сконцентрироваться на отличиях в текстовой ткани исторических нарративов в учебниках истории, рассмотренных нами.

По-видимому, самыми константными объектами нарративизации в педагогическом дискурсе являются темы об исторических персонажах – Иване Грозном и Петре I: данные тексты в шести проанализированных учебниках имеют очень схожие векторы, идентичную для всех учебников тональность (Грозный – более негативную, Петр – менее) и аналогичную тематическую наполненность. Но даже в этом случае заметны отличия: например, и ту и другую фигуры учебник 2010 г. подает преимущественно через одну и ту же тему русского самодержавия, а другие учебники – через идею реформ и побед.

Учебник 2010 г. в целом отдает предпочтение темам (топикам), связанным с идеями власти и государства: Докиевская Русь – «Происхождение славян и их этнические соседи» (идея национальной идентичности); Киевская Русь и Феодальная раздробленность – «Славянские племена и установление государственности на Руси» (истоки государственности); Иго – «Защита русских земель от захватчиков», Объединение земель – «Политика консолидации Ивана III» (централизация государственной власти); Иван Грозный и Петр Великий – «Российские самодержцы» (централизация государственной власти). Вместе с учебником 1946 г. это самый эмоционально сдержанный учебник, показывающий один из двух наименьших «размахов» сентимента.

Учебник 1946 г., кроме того, что он самый эмоционально сдержанный, еще и тяготеет к геополитическому взгляду на все исторические сюжеты, чем отличается от всех других учебников: тема «Культура геополитических соседей Руси» доминирует при рассказывании сюжетов и о Докиевской Руси, и о феодальной раздробленности, при этом акцентируется идея о связи Руси и государств Закавказья; «Борьба Московского княжества с внешними врагами» проявляется и в текстах об иго, и об объединении земель; даже раздел о Киевской Руси с доминирующей темой «Государственное устройство Киевской Руси» сфокусирован на теме защиты Отечества и молодой русской государственности от внешних врагов, что предопределяет одно из самых негативных значений сентимента для данного текста среди всех текстов по этой исторической тематике в выборке.

При кластеризации тексты учебника 1946 г. объединяются преимущественно вместе с текстами учебника 1983 г., что указывает, вероятно, на некоторую стилистическую преемственность, но по доминирующим топикам они мало пересекаются, к тому же учебник 1983 г. демонстрирует большую долю негативного вокабуляра в темах об Иване Грозном и иге, чем учебник 1946 г.

Интересным является, например, тот факт, что учебники, написанные одним и тем же авторским коллективом с разницей в 9 лет (1997 и 2006 гг.), совпадая полностью для каждого из семи исторических сюжетов по выделенным при помощи тематического моделирования топикам, значительно отличаются по значениям сентимента: учебник 1997 г. «рассказывает» все темы, кроме темы про Ивана Грозного, с большим преобладанием негативного вокабуляра, чем учебник 2006 г. Последний демонстрирует самый большой «размах» эмоциональной тональности между полюсами «позитив» и «негатив». Сравнение частотности этнонимов в разделе «Докиевская Русь» в данных двух учебниках показывает, что их авторы спустя 9 лет смягчают выраженность идеи панславизма.

Для учебника 2001 г. характерен фокус на идее общинности как исконной народной демократии, являющейся первоосновой русской государственности, – этому посвящены тексты раздела Докиевская Русь, которые имеют самый выраженный позитивный вокабуляр для этой исторической темы по всей выборке, а вот отношение к фигуре «узурпатора» Ивана Грозного, наоборот, одно из самых негативных в выборке.

Таким образом, сентимент-анализ и тематическое моделирование вкупе с элементами экспертного анализа позволили нам получить, на наш взгляд, ряд доказательств в пользу гипотезы о существовании присущего институциональным авторам некоего поколенческого нарратива, манифестируемого в текстах учебников по истории и становящегося доступным для обнаружения при помощи компьютерного инструментария. Основные признаки такого нарратива – специфический эмоциональный взгляд на исторические сюжеты, рассказываемые историографически идентично, но через призму одной или нескольких доминирующих идей, которые передают неуловимую доминанту времени, не имеющую прямой связи с идеологией. Так, для послевоенного времени это идеи защиты государства и внимания к его геополитическим соседям; для постперестроечного периода становления «молодой» российской демократии (2001 г.) – идея общины и вече как исконной демократии «из народа»; для времени становления современной российской вертикали власти (2010 г.) – идеи централизации власти и устойчивости государственных структур.

Но при этом не стоит забывать и об изменчивости самой исторической науки и внутринаучных тенденциях, течениях и трендах, которые, очевидно, тоже могли внести свою лепту в выявленные текстовые различия.

Заключение

Подводя итоги, отметим, что, с одной стороны, проведенный анализ не претендует на всеохватность – использованы лишь шесть учебников и семь тем, а с другой – этот своеобразный разведочный результат важен для развертывания более масштабного исследования, поскольку показывает, на наш взгляд, эффективность применяемого инструментария и наличие интересных качественных отличий текстовой субстанции в разных поколенческих нарративах об истории. Эти отличия находятся между собственно языковыми формами и содержанием текста, являя собой некий *мезоуровень*, который не считывается в рамках классического лингвистического анализа и остается невидимым при дискурс-анализе. Однако инструментарий хорошо зарекомендовавших себя методов компьютерного анализа, в том числе методов на основе машинного обучения, позволяет этот уровень обнаружить и сделать объектом дискурсивной интерпретации.

Перспективу работы составляет не только увеличение частоты временных срезов и длины анализируемой «стрелы времени», но и числа исследовательских гипотез, а именно: действительно ли исторические персонажи обладают большей стабильностью как объект поколенческого нарратива, нежели переходные эпохи в истории государства? отличаются ли по параметрам сентимента и распределения тематических фокусов тексты учебников истории для студентов и школьников? изменятся ли результаты, если применить для анализа другие инструменты или если оценивать тональность текстов в ходе эксперимента будут представители разных поколений (информанты-студенты 1980-х, 1990-х гг. и, скажем, сегодняшние студенты).

Заключая, отметим, что академические тексты в целом имеют не меньшее, а может быть, и большее влияние на мотивационный уровень личности, ее картину мира, чем социальные сети, например. Тексты учебников уже в силу контекста знакомства с ними (учебное учреждение, детство, молодость, авторитет родителей и учителей, психологическая и гносеологическая незрелость учеников) являются тем заделом, который формирует мировоззрение поколения. Тональность, тематические «рефрены», распределение ключевых слов – важные инструменты этого процесса, маркеры которого, в силу их идеологической немаркированности, легко пропускает «исследовательское сито» дискурс-анализа.



Облако слов для текста по теме «Киевская Русь» в учебнике 1946 г.



Облако слов для текста по теме «Киевская Русь» в учебнике 1997 г.

Список источников

1. *Ginburg C.* Clues, Myths and the Historical Method. Baltimore, 1989. 231 p.
2. *Megill A.* Historical Knowledge, Historical Error. Chicago, 2007. 288 p.
3. *Moretti F.* Distant Reading. London : Verso, 2013. 254 p.

4. *Козачина А.В.* Реализация мифопоэтической стратегии легитимации институционализированных ценностей в японском педагогическом дискурсе // Вестник Томского государственного педагогического университета. 2020. Вып. 6 (212). С. 20–26.

5. *Citron S.* Le mythe national : L'histoire de France en question. Paris : Ed. ouvrières: Etudes et doc. intern., 1987.

6. *Dubois J., Legris P.* (dir.). Disciplines scolaires et cultures politiques. Des modèles nationaux en mutation depuis 1945. Rennes : Presses universitaires de Rennes, 2018.

7. *Dubois J.* Garibaldi dans les manuels scolaires d'histoire en Italie : les usages pédagogiques et politiques. In : Jérémie Dubois et Patricia Legris (dir.), Disciplines scolaires et cultures politiques. Des modèles nationaux en mutation depuis 1945. Rennes : Presses universitaires de Rennes, 2018. P. 81–91.

8. *Konkka O.* Le dictateur ou le chef de la nation victorieuse ? L'évolution de la présentation de Joseph Staline dans les manuels scolaires d'histoire de la Russie postsoviétique In : Jérémie Dubois et Patricia Legris (dir.), Disciplines scolaires et cultures politiques. Des modèles nationaux en mutation depuis 1945. Rennes : Presses universitaires de Rennes, 2018. P. 105–120.

9. *Konkka O.* Язык постсоветских школьных учебников истории: два уровня легитимации политического в дискурсе о прошлом // Дискурс легитимации: язык и политика в эпоху глобальных вызовов / под общ. ред. А.В. Колмогоровой. Красноярск : Сиб. федер. ун-т, 2019. С. 50–69.

10. *Wodak R., Meyer M.* Methods of Critical Discourse Analysis. London : SAGE Publications Ltd, 2001. 200 p.

11. *Stibbe A.* Ecolinguistics: Language, ecology, and the stories we live by. New York : Routledge, 2015.

12. *Ерунов Б.Л.* Мнение и умонастроение в историческом аспекте // История и психология / под ред. Б.Ф. Поршнева, Л.И. Анциферовой. М., 1971. С. 110–126.

13. *Бухарин Н.И.* Рецензия на учебник, подготовленный Московским государственным педагогическим институтом им. А.С. Бубнова, 23 октября 1936 г. // РГАСПИ. Ф. 17. Оп. 120. Д. 361. Л. 9–13.

14. *Огановская И.С.* Школьный учебник отечественной истории: Учебные издания как исторический источник // Документ. Архив. История. Современность. 2011. № 12. С. 264–282.

15. *Повалява Н.Е.* История Отечества до начала XX в. в современных учебниках : автореф. ... канд. ист. наук. М., 2004. 24 с.

16. *Greimas A.J., Courtés J.* Adjuvant // Sémiotique. Dictionnaire raisonné de la théorie du langage. Paris : Hachette, 1979. P. 10–11.

17. *Шаховский В.И.* Обоснование лингвистической теории эмоций // Вопросы психолингвистики. 2019. № 1 (39). С. 22–37.

18. *Beigi Gh., Hu X., Maciejewski R., Liu H.* An overview of sentiment analysis in social media and its applications in disaster relief. In: Pedrycz W, Chen SM, editors. Sentiment analysis and ontology engineering. Berlin : Springer Cham, 2016. P. 313–340. doi: 10.1007/978-3-319-30319-2_13

19. *Negi S., Buitelaar P.* Suggestion Mining From Opinionated Text // Sentiment Analysis in Social Networks / eds by F.A. Pozzi, E. Fersini, E. Messina, B. Liu, M. Kaufmann. Elsevier, 2017. P. 129–139. doi: <https://doi.org/10.1016/B978-0-12-804412-4.00008-5>

20. *Koltsova O., Alexeeva S., Pashakhin S., Koltsov S.* PolSentiLex: Sentiment Detection in Sociopolitical Discussions on Russian Social Media, in: Artificial Intelligence and Natural Language. AINL 2020 Communications in Computer and Information Science Book 1292: Communications in Computer and Information Science. Cham : Springer, 2020. P. 1–16.

21. *Tiwari P., Yadav P., Agnihotri S., Mishra B., Nhu N., Gochhayat S., Singh J., Prasad M.* Sentiment Analysis for Airlines Services Based on Twitter Dataset // Social Network Analytics.

Computational Research Methods and Techniques. 2019. P. 149–162. doi: 10.1016/B978-0-12-815458-8.00008-6

22. *Yasen M., Tedmori S.* Movies Reviews Sentiment Analysis and Classification // Proceedings of the IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT). Amman, Jordan, 2019. P. 860–865. doi: 10.1109/JEEIT.2019.8717422

23. *Sherstinova T., Moskvina A., Kirina M., Karysheva A., Kolpashchikova E., Maksimenko P., Seinova A., Rodionov R.* Sentiment Analysis of Literary Texts vs. Reader's Emotional Responses // Proceedings of the 33rd Conference of Open Innovations Association (FRUCT) / IEEE. 2023. P. 243–249.

24. *Sell J., Farreras I.* LIWC-ing at a Century of Introductory College Textbooks: Have the Sentiments Changed? // Procedia Computer Science. 2017. № 118. P. 108–112.

25. *Solovyev V.D., Solnyshkina M.I., Gafiyatova E.V., McNamara D.S., Ivanov V.* Sentiment in Academic Texts // Proceedings of the 24th Conference of Open Innovations Association (FRUCT). P. 408–414.

26. *Smetanin S., Komarov M.* Deep transfer learning baselines for sentiment analysis in Russian // Information Processing & Management. 2021. Vol. 58, Is. 3. P. 102484. doi: 10.1016/j.ipm.2020.102484

27. *Николаев И.С., Митренина О.В., Ландо Т.М.* Прикладная и компьютерная лингвистика. М. : ЛЕНАНД, 2016. 320 с.

28. *Кулагин Д.И.* Открытый тональный словарь русского языка КартаСловСент // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог». Вып. 20. М., 2021. С. 1106–1119.

29. *Koltsova O.Yu., Alexeeva S.V., Kolcov S.N.* An Opinion Word Lexicon and a Training Dataset for Russian Sentiment Analysis of Social Media // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference «Dialogue-2016». 2016. Vol. 15 (22). P. 277–287.

30. *Loukachevitch N., Levchik A.* Creating a General Russian Sentiment Lexicon // Proceedings of Language Resources and Evaluation Conference LREC-2016. 2016.

31. *Кирина М.А.* Сравнение тематических моделей на основе LDA, STM и NMF для качественного анализа русской художественной прозы малой формы // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2022. Т. 20, № 2. С. 93–109. doi: 10.25205/1818-7935-2022-202-93-109

32. *Demsar J., Curk T., Erjavec A., Gorup C., Hocevar T., Milutinovic M., Mozina M., Polajnar M., Toplak M., Staric A., Stajdohar M., Umek L., Zagar L., Zbontar J., Zitnik M., Zupan B.* Orange: Data Mining Toolbox in Python // Journal of Machine Learning Research. № 14 (Aug). P. 2349–2353.

33. *Sciandra A., Trevisani M., Tuzzi A.* Diagnostics for topic modelling. The dubious joys of making quantitative decisions in a qualitative environment // Proceedings of the Statistics and Data Science Conference. Pavia University Press, 2023. P. 61–66.

34. *Bischof J.M., Airolidi E.M.* Summarizing topical content with word frequency and exclusivity // International Conference on Machine Learning. 2012. № 29. P. 201–208.

References

1. Ginzburg C. (1989) *Clues, Myths and the Historical Method*. Baltimore: Johns Hopkins University Press.
2. McGill, A. (2007) *Historical Knowledge, Historical Error*. Chicago: University of Chicago Press.
3. Moretti F. (2013) *Distant Reading*. London: Verso.
4. Kozatchina, A.V. (2020) The implementation of the mythopoetic strategy of legitimation in Japanese pedagogical discourse. *Vestnik Tomskogo gosudarstvennogo pedagogicheskogo universiteta*. 6 (212). pp. 20–26. (In Russian).

5. Citron, S. (1987) *Le mythe national: L'histoire de France en question*. Paris: Ed. ouvrières: Etudes et doc. intern.

6. Dubois, J. & Legris, P. (dir.) (2018) *Disciplines scolaires et cultures politiques. Des modèles nationaux en mutation depuis 1945*. Rennes: Presses universitaires de Rennes.

7. Dubois, J. (2018) Garibaldi dans les manuels scolaires d'histoire en Italie: les usages pédagogiques et politiques. In: Dubois, J. & Legris, P. (dir.), *Disciplines scolaires et cultures politiques. Des modèles nationaux en mutation depuis 1945*. Rennes: Presses universitaires de Rennes. pp. 81–91.

8. Konkka O. (2018) Le dictateur ou le chef de la nation victorieuse ? L'évolution de la présentation de Joseph Staline dans les manuels scolaires d'histoire de la Russie postsoviétique. In: Dubois, J. & Legris, P. (dir.), *Disciplines scolaires et cultures politiques. Des modèles nationaux en mutation depuis 1945*. Rennes: Presses universitaires de Rennes. pp. 105–120.

9. Konkka O. (2019) Yazyk postsovetskikh shkol'nykh uchebnikov istorii: dva urovnya legitimatsii politicheskogo v diskurse o proshlom [The language of post-Soviet school textbooks on history: two levels of political legitimation in discourse about the past]. In: Kolmogorova, A.V. (ed.) *Diskurs legitimatsii: yazyk i politika v epokhu global'nykh vyzovov* [Discourse of legitimation: language and politics in the era of global challenges]. Krasnoyarsk: SFU. pp. 50–69.

10. Wodak, R. & Meyer, M. (2001) *Methods of Critical Discourse Analysis*. London: SAGE Publications Ltd.

11. Stibbe, A. (2015) *Ecolinguistics: Language, ecology, and the stories we live by*. New York: Routledge.

12. Erunov, B.L. (1971) Mnenie i umonastroenie v istoricheskom aspekte [Opinion and sentiment in historical perspective]. In: Porshnev, B.F. & Antsiferova, L.I. (eds) *Istoriya i psikhologiya* [History and psychology]. Moscow: "Nauka". pp. 110–126.

13. Russian State Archive of Socio-Political History (RGASPI). Fund 17. List 120. File 361. Pages 9–13. Bukharin, N.I. (1936) *Retsenziya na uchebnik, podgotovlennyy Moskovskim Gosudarstvennym Pedagogicheskim Institutom im. A.S.Bubnova* [Review of the textbook prepared by Moscow State Pedagogical Institute named after A.S. Bubnov]. Moscow State Pedagogical Institute. 23 October 1936.

14. Oganovskaya, I.S. (2011) Shkol'nyy uchebnik otechestvennoy istorii. Uchebnye izdaniya kak istoricheskiy istochnik [School textbook of national history. Textbooks as a historical source]. *Dokument. Arkhiv. Istoriya. Sovremennost'*. 12. pp. 264–282.

15. Povalyava, N.E. (2004) *Istoriya Otechestva do nachala XX v. v sovremennykh uchebnikakh* [History of the Fatherland up to the beginning of the twentieth century in modern textbooks]. Abstract of History Cand. Diss. Moscow.

16. Greimas, A.J. & Courtés, J. (1979) *Sémiotique. Dictionnaire raisonné de la théorie du langage*. Paris: Hachette. pp. 10–11.

17. Shakhovskiy, V.I. (2019) Obosnovanie lingvisticheskoi teorii emotsii [Foundations of linguistic theory of emotions]. *Voprosy psikholingvistiki*. 1. pp. 22–37.

18. Beigi, Gh., Hu, X., Maciejewski, R. & Liu, H. (2016) An overview of sentiment analysis in social media and its applications in disaster relief. In: Pedrycz, W. & Chen, S.M. (eds) *Sentiment analysis and ontology engineering*. Berlin: Springer Cham. pp. 313–340. doi: 10.1007/978-3-319-30319-2_13

19. Negi, S. & Buitelaar, P. (2017) Suggestion Mining From Opinionated Text. In: Pozzi, F.A. et al. (eds) *Sentiment Analysis in Social Networks*. Elsevier. pp. 129–139. doi: 10.1016/B978-0-12-804412-4.00008-5

20. Koltsova, O., Alexeeva, S., Pashakhin, S. & Koltsov, S. (2020) PolSentiLex: Sentiment Detection in Sociopolitical Discussions on Russian Social Media. In: *Artificial Intelligence and Natural Language. AINL 2020 Communications in Computer and Information Science Book 1292: Communications in Computer and Information Science*. Cham: Springer. pp. 1–16.

21. Tiwari, P. et al. (2019) Sentiment Analysis for Airlines Services Based on Twitter Dataset. *Social Network Analytics. Computational Research Methods and Techniques*. pp. 149–162. doi: 10.1016/B978-0-12-815458-8.00008-6
22. Yassen, M. & Tedmori, S. (2019) Movies Reviews Sentiment Analysis and Classification. *Proceedings of the IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, Amman, Jordan. pp. 860–865. doi: 10.1109/JEEIT.2019.8717422
23. Sherstinova T., Moskvina A., Kirina M., Karysheva A., Kolpashchikova E., Maksimenko P., Seinova A., Rodionov R. (2023) Sentiment Analysis of Literary Texts vs. Reader's Emotional Responses. *Proceedings of the 33rd Conference of Open Innovations Association (FRUCT) / IEEE*. pp. 243–249.
24. Sell, J. & Farreras, I. (2017) LIWC-ing at a Century of Introductory College Textbooks: Have the Sentiments Changed? *Procedia Computer Science*. 118. pp. 108–112.
25. Solovyev, V.D. et al. (2019) Sentiment in Academic Texts. *Proceedings of the 24th Conference of Open Innovations Association (FRUCT)*, pp. 408–414.
26. Smetanin, S. & Komarov, M. (2021) Deep transfer learning baselines for sentiment analysis in Russian. *Information Processing & Management*. 58 (3). Art. 102484. doi: 10.1016/j.ipm.2020.102484
27. Nikolaev, I.S., Mitrenina, O.V. & Lando, T.M. (2016) *Prikladnaya i komp'yuternaya lingvistika* [Applied and Computational Linguistics]. Moscow: LENAND.
28. Kulagin, D.I. (2021) [Publicly available sentiment dictionary for the Russian language KartaSlovSent]. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii* [Computational Linguistics and Intellectual Technologies]. Dialogue International Conference Proceedings. Vol. 20. Moscow: RSUH. pp. 1106–1119. (In Russian).
29. Koltsova, O.Yu., Alekseeva, S.V. & Kol'tsov, S.N. (2016) An Opinion Word Lexicon and a Training Dataset for Russian Sentiment Analysis of Social Media. *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue-2016"*. 15 (22). pp. 277–287.
30. Loukachevitch, N. & Levchik, A. (2016) Creating a General Russian Sentiment Lexicon. *Proceedings of Language Resources and Evaluation Conference LREC-2016*.
31. Kirina, M.A. (2022) A Comparison of Topic Models Based on LDA, STM and NMF for Qualitative Studies of Russian Short Prose. *Vestnik NSU. Series: Linguistics and Intercultural Communication*. 20 (2). pp. 93–109. (In Russian). doi: 10.25205/1818-7935-2022-20-2-93-109
32. Demsar, J. et al. (2013) Orange: Data Mining Toolbox in Python (<http://jmlr.org/papers/volume14/demsar13a/demsar13a.pdf>), *Journal of Machine Learning Research*. 14 (Aug). pp. 2349–2353.
33. Sciandra, A., Trevisani, M. & Tuzzi, A. (2023) Diagnostics for topic modelling. The dubious joys of making quantitative decisions in a qualitative environment. *Proceedings of the Statistics and Data Science Conference*. Pavia University Press. pp. 61–66.
34. Bischof, J.M. & Airoldi, E.M. (2012) Summarizing topical content with word frequency and exclusivity. *International Conference on Machine Learning*. 29. pp. 201–208.

Информация об авторах:

Колмогорова А.В. – д-р филол. наук, профессор департамента филологии Санкт-Петербургской Школы гуманитарных наук и искусств Национального исследовательского университета “Высшая школа экономики”, академический руководитель магистерской программы “Языковые технологии в бизнесе и образовании”, заместитель руководителя научно-исследовательской лаборатории языковой конвергенции (Санкт-Петербург, Россия). E-mail: akolmogorova@hse.ru

Колмогорова П.А. – стажер-исследователь научно-исследовательской лаборатории языковой конвергенции департамента филологии Санкт-Петербургской Школы гуманитарных наук и искусств Национального исследовательского университета “Высшая школа экономики” (Санкт-Петербург, Россия). E-mail: pakolmogorova@edu.hse.ru

Куликова Е.Р. – стажер-исследователь научно-исследовательской лаборатории языковой конвергенции департамента филологии Санкт-Петербургской Школы гуманитарных наук и искусств Национального исследовательского университета “Высшая школа экономики” (Санкт-Петербург, Россия). E-mail: ekulikova@edu.hse.ru

Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

A.V. Kolmogorova, Dr. Sci. (Philology), professor; academic director of the Language Technologies in Business and Education master’s program; deputy head of the Research Laboratory of Language Convergence, National Research University Higher School of Economics (HSE University) (St. Petersburg, Russian Federation). E-mail: akolmogorova@hse.ru

P.A. Kolmogorova, intern researcher, Research Laboratory of Language Convergence, National Research University Higher School of Economics (HSE University) (St. Petersburg, Russian Federation). E-mail: pakolmogorova@edu.hse.ru

E.R. Kulikova, intern researcher, Research Laboratory of Language Convergence, National Research University Higher School of Economics (HSE University) (St. Petersburg, Russian Federation). E-mail: ekulikova@edu.hse.ru

The authors declare no conflicts of interests.

*Статья поступила в редакцию 25.07.2023;
одобрена после рецензирования 21.03.2024; принята к публикации 27.05.2024.*

*The article was submitted 25.07.2023;
approved after reviewing 21.03.2024; accepted for publication 27.05.2024.*