

ЦИФРОВЫЕ ГУМАНИТАРНЫЕ
ИССЛЕДОВАНИЯ

2024 № 1 (001)

ИНСТИТУТ РУССКОЙ ЛИТЕРАТУРЫ (ПУШКИНСКИЙ ДОМ) РАН
ЦИФРОВЫЕ ГУМАНИТАРНЫЕ ИССЛЕДОВАНИЯ. 2024 № 1.

ISSN (Online) XXXX–XXXX

Главный редактор
Орехов Б. В.
(Москва–Санкт-Петербург)

РЕДАКЦИОННЫЙ СОВЕТ

Алиева О. В. (Москва)
Акимова М. В. (Москва)
Беляк Г. Н. (Санкт-Петербург)
Балакин А. Ю. (Санкт-Петербург)
Белоусова А. С. (Богота, Колумбия)
Бонч-Осмоловская А. А. (Москва)
Вдовин А. В. (Москва)
Володин А. Ю. (Москва–Красноярск)
Гагарина Д. А. (Бишкек, Кыргызстан; Эрланген, Германия)
Кижнер И. А. (Хайфа, Израиль)
Колозариди П. В. (Санкт-Петербург)
Ляшевская О. Н. (Москва)
Маслинский К. А. (Санкт-Петербург)
Павлова Л. В. (Смоленск)
Полилова В. С. (Москва)
Пучковская А. А. (Лондон, Великобритания)
Романова И. В. (Смоленск)
Северина Е. М. (Ростов-на-Дону)
Сенаторова Е. Е. (Нью-Йорк, США)
Скоринкин Д. А. (Потсдам, Германия)
Шерстинова Т. Ю. (Санкт-Петербург)

© Авторы статей, 2024;

.....
Санкт-Петербург

ОГЛАВЛЕНИЕ

От редакции	4
-----------------------	---

ИССЛЕДОВАНИЯ

Ольга Алиева

Меры расстояния для определения авторства древнегреческих текстов	8
---	---

Борис Орехов

Нестиховедческий ритм в романе Н. Г. Чернышевского «Что делать?»	34
--	----

Инна Кижнер

Коллекции как данные: готовы ли мы к научным исследованиям с помощью цифровых коллекций культурно-значимых данных?	45
--	----

ДИСКУССИЯ

Борис Орехов, Андрей Володин

Digital Humanities в России и конец истории	63
---	----

ХРОНИКА

Динара Гагарина

Digital Humanities в Центральной Азии: заметки на полях круглого стола	86
--	----

РЕЦЕНЗИЯ

Дарья Артемьева

Джулия Томпсон Кляйн «Междисциплинарные цифровые гуманитарные науки: работа с границами в развивающейся сфере»	101
--	-----

ОТ РЕДАКЦИИ

Использование цифровых методов в гуманитарных науках имеет свою историю, восходящую к усилиям отдельных энтузиастов рубежа XIX–XX веков и первой трети XX века. Но именно в последние десятилетия оно получило широкое распространение благодаря доступности компьютеров, аккумулярованию значительных массивов данных и развитию технологий обработки этих данных.

В западной традиции применение цифровых методов к изучению объектов, традиционно являвшихся предметом интереса гуманитарных дисциплин (литературы, истории, художественной культуры, социологии творчества), объединилось под общим названием *Digital Humanities* (О вариантах перевода термина на русский см.: [Цифровые 2023, 6]), которое имеет несколько конкурирующих переводов на русский язык: *цифровая гуманитаристика*, *цифровые гуманитарные науки*; для исторических дисциплин сформировались обособленные наименования: историческая информатика, цифровая история.

Чувствительная новизна тематики *Digital Humanities* для гуманитарного поля приводит к тому, что исследователи испытывают известные трудности с публикацией работ, ориентированных на цифровые подходы. С особенной силой эта проблема ощущается для русскоязычной научной периодики. Традиционные журналы работают с подобными статьями неохотно, часто не способны найти рецензентов, готовых профессионально оценить уровень предлагаемого текста. Кроме того, специфика цифрового поля подразумевает необходимость публикации текстов, отличных по своим жанровым признакам от привычных для филологических, культурологических или исторических журналов: описание исследовательских ресурсов (корпусов, баз данных, электронных архивов, веб-сервисов, облегчающих работу ученого), наборов данных. Такие тексты воспринимаются редакциями и рецензентами с недоумением.

В западной науке, кроме сборников трудов конференций, существует считанное число мест, в которых применяющий цифровые методы исследователь-гуманитарий способен опубликовать свой текст, получив при этом качественную обратную связь. Речь идет о *Digital scholarship in the humanities*, *Digital humanities quarterly*, отчасти *Cultural analytics*. Научное сообщество с очевидностью

ощущает недостаточность этого публикационного ресурса, поэтому на наших глазах появляются новые издания: *International Journal of Digital Humanities* (с 2019), *Journal of Digital History* (с 2021), *Digital Orientalia* (с 2021), *Journal of Computational Literary Studies* (с 2023).

Возможности публикации работ в рамках digital humanities на русском языке еще более ограничены. Исторические науки с 1990 года получили такую площадку в виде «Информационного бюллетеня Ассоциации история и компьютер» (1–2 номера в год). С 2021 года в Смоленском университете выходит журнал «Квантитативная филология» (2 номера в год). Очевидно, что эти издания не могут обеспечить потребности растущего сообщества русскоязычных ученых, заинтересованных в применении цифровых методов в гуманитарных науках.

В последние годы профильные образовательные программы были созданы в Москве (НИУ ВШЭ, Финансовый университет), Санкт-Петербурге (ИТМО), Перми (НИУ ВШЭ), Екатеринбурге (УрФУ), Красноярске (СФУ), Томске (ТГУ), Калининграде (БФУ). Это означает, что число специалистов в области цифровых гуманитарных наук обязательно будет расти. По очевидным причинам следует предоставить ученым возможность развивать эту отрасль, одновременно развивая и соответствующий сегмент русского научного языка, традиция которого восходит к Б. Томашевскому, Б. Ярхо и др.

Для этого необходим русскоязычный журнал, в процесс работы над которым были бы вовлечены специалисты, хорошо ориентирующиеся в современных тенденциях в этой области, то есть способные на качественную экспертизу редакционного портфеля. Именно они составят ядро редколлегии. Поддержку и развитие цифровых гуманитарных наук на русском языке можно назвать главной задачей такого издания.

ИРЛИ (Пушкинский Дом) РАН представляется лучшей базовой институцией для такого издания, поскольку здесь успешно действует Лаборатория цифровых исследований литературы и фольклора.

Журнал называется «Цифровые гуманитарные исследования», выдерживая строгий академический стиль изданий ИРЛИ РАН и не ограничиваясь при этом дисциплинарной принадлежностью потенциальных авторов внутри содружества гуманитарных наук.

Периодичность выхода — 2 раза в год.

Круг авторов будет формироваться из сотрудников отечественных научных и учебных подразделений, вовлеченных в цифровые

гуманитарные исследования, а также представителей более традиционных гуманитарных дисциплин, готовых к междисциплинарным экспериментам.

Журнал должен стать площадкой для публикации главным образом русскоязычных исследований, но не обязательно посвященных данным, ассоциированным с русской культурой. Ключевым признаком подходящих для журнала статей будет использование цифровых исследовательских методов.

Основным разделом журнала станет рубрика «Исследования», призванная отразить круг актуальных цифровых подходов к гуманитарному материалу. Помимо нашедших применение в науке с пришествием компьютерной эры методов, журнал готов предоставить площадку и исследованиям, осуществленным в рамках традиционного для отечественной науки количественного стиховедения.

Не регулярной, но важной для методологических дискуссий должна стать рубрика «Теория», в которую будут помещаться материалы, отражающие неизбежные для молодой науки тексты авторефлексии, уточнения методологической сетки координат и дисциплинарных границ.

По мере наполнения редакционного портфеля возможно также и введение рубрики «Архив», в которой предполагается публикация архивных материалов, проливающих свет на историю применения количественных методов в отечественной гуманитарной науке.

Важной с точки зрения инфраструктуры отечественных Digital Humanities рубрикой должен стать раздел «Инструменты и наборы данных», в который предполагается определить материалы, в меньшей степени ориентированные на представление исследовательских результатов, а в большей степени — на обеспечение знакомства сообщества с новыми цифровыми ресурсами. В традиционной научной печати опубликовать такие материалы бывает особенно сложно, хотя польза от таких публикаций представляется даже большей, чем от конкретных исследований.

Рубрика «Рецензии и обзоры» может включать обзорные статьи об отечественных и зарубежных публикациях на близкие цифровым гуманитарным наукам темы, а рубрика «Дискуссия» предоставит площадку для полемики по общим и частным вопросам научного поля.

Литература

Цифровые 2023 — Цифровые гуманитарные исследования : монография / А. Б. Антопольский, А. А. Бонч-Осмоловская, Л. И. Бородкин [и др.]; рец. С. А. Баканов; Сибирский федеральный университет. Электрон. текстовые дан. (27,3 Мб). Красноярск : СФУ, 2023.

Б. В. Орехов

ИССЛЕДОВАНИЯ

Ольга Алиева

МЕРЫ РАССТОЯНИЯ ДЛЯ ОПРЕДЕЛЕНИЯ АВТОРСТВА ДРЕВНЕГРЕЧЕСКИХ ТЕКСТОВ

Хотя классическая филология была одной из первых гуманитарных дисциплин, обратившейся к количественным исследованиям, стилометрия никогда не рассматривалась филологами как самодостаточный метод. Стилистические и языковые особенности — лишь один, не главный и не единственный, инструмент исследователя, а из этих особенностей лишь некоторые могут быть описаны количественно. Более того, количественные методы пригодны только для отрицательного решения вопроса о подлинности, в то время как несомненных оснований для утвердительного решения вообще не существует. В настоящем исследовании сравниваются методы, в англоязычной литературе известные как *distance-based approaches*, то есть подходы, основанные на векторизации текстов через частотности токенов и измерении расстояний между векторами. Оценивается эффективность подобных методов в зависимости от длины отрывка и количества предикторов. Для сравнения привлекается корпус древнегреческой прозы объемом 694 тыс. слов. Наиболее эффективные метрики затем привлекаются для классификации трех спорных текстов.

Ключевые слова: стилометрия, авторство, межтекстовые расстояния, метрики

1. Постановка вопроса¹

Хотя классическая филология была одной из первых гуманитарных дисциплин, обратившейся к количественным исследованиям, стилометрия никогда не рассматривалась филологами как самодостаточный метод. Суждение об авторстве текста, как писал

Ольга Валерьевна Алиева
НИУ «Высшая школа экономики»
oalieva@hse.ru

Фридрих Бласс, должно учитывать данные рукописной традиции, свидетельства современников, соответствие бытовых и исторических реалий времени жизни автора, а также соответствие идей, тем и жанров тому, что известно по подлинным сочинениям автора [Blass, 1892. S. 289–295 = Бласс, 2016. С. 186–194]. Стилистические и языковые особенности — лишь один, не главный и не единственный, инструмент исследователя, а из этих особенностей лишь некоторые могут быть описаны количественно. В этом смысле и современные методы «атрибуции авторства», опирающиеся на статистику и технологии, могут играть только вспомогательную роль и не представляют угрозы традиционным подходам [Holmes, 1994. P. 104]. Более того, эти методы (впрочем, как и остальные) «пригодны только для отрицательного решения вопроса о подлинности», а несомненных оснований для утвердительного решения «вообще не существует» [Blass, 1892. S. 292 = Бласс, 2016. С. 189]. Говорят, что Пауль Фридендер как-то возразил Виламовицу: «Невозможно доказать, что Платон написал “Пир”» [Chambers, 1996. P. 220]². По умолчанию традиция достоверна, а бремя доказательства лежит на том, кто ее отвергает: *quivis praesumitur genuinus liber, donec demonstretur contrarium* [Boeckh, 1886. S. 240]³.

Но даже для отвержения авторства стилометрия может быть несовершенным инструментом. Кеннет Довер заметил, что длина предложения в греческом тексте больше говорит о редакторе, чем об авторе [Dover, 1968. P. 108]; как издатель Аристофана, он прекрасно понимал, что наши источники — результат множества решений, причем не только авторских. Требование избавиться от всех посторонних вмешательств понятно [Rudman, 2005], но применительно к древним текстам едва ли выполнимо⁴. С другой стороны, сам автор способен менять свой стиль в зависимости от темы, жанра и драматических особенностей произведения, и подобные стилизации неизбежно ставят под вопрос осмысленность наших экспериментов. «Менексен» почти во всех количественных исследованиях — и новых, и старых — не похож на остальные диалоги Платона [Koentges, 2020], но это мало смущает тех, кто считает диалог подлинным (у Платона нет других сочинений в жанре надгробной речи). Стилистическая близость «Послезакония» к поздним диалогам Платона [Ledger, 1989. P. 150] никого не убеждает: автором этого небольшого сочинения по-прежнему признается Филипп Опунтский [Tarán, 1975].

Доверять или не доверять количественным данным — решает исследователь. Древние авторы не могут, как Джоан Роулинг,

подтвердить или опровергнуть результаты автоматической классификации текста [Juola, 2013]; а если бы и могли, нам, возможно, не стоило бы их слушать: представления об авторстве меняются со временем [Пешков, 2016]. Феномен «школьной аккумуляции» в античности указывает на то, что «автор» и «не автор» — это, возможно, не бинарная переменная, а лишь два значения непрерывной величины, между которыми есть еще «почти автор» и «не вполне автор» [Thesleff, 2009. P. 351; Thesleff, 2023]. Это ставит нас перед более общим вопросом о том, кто или что является носителем «авторской функции» [Фуко, 2008] и что обеспечивает тождественность этого носителя. Если ранние «Европейцы» Генри Джеймса и его же поздние «Послы» не опознаются как сочинения одного автора, считать ли это ошибкой классификации [Hoover, 2004. P. 457]? Как в известной загадке о корабле Тесея, авторский стиль, постепенно меняясь, может обновиться настолько, что перестает быть «тем же». Эта загадка не решается количественными способами.

Любой метод имеет ограниченное применение, но это не значит, что мы должны признать стилометрию видом «шаманизма» [Love, 2002. P. 159]. Впрочем, если бы 80% ударов в бубен достигали своей цели, у нас были бы основания присмотреться и к шаманским практикам — а некоторые способы автоматической классификации текстов позволяют добиться даже большей точности. Лучшему пониманию и возможностей, и ограничений количественных методов способствуют их испытания на известном материале, и именно такое испытание документируется в этой статье. Разумеется, нет недостатка в сравнениях на материале других языков, прежде всего современных европейских [Stanikūnas et al., 2017; Eder, 2011; Jannidis, Lauer, 2014; Rybicki, Eder, 2011], восточных [Ahmed, 2019], а также латинском [Kestemont et al., 2016; Eder, 2015]. Но универсальных подходов нет, и то, что показывает хорошие результаты на одном корпусе, не обязательно работает на другом. Эксперименты с древнегреческими текстами до сих пор проводились лишь в довольно ограниченных масштабах [Фоминых, 2017; Алиева, 2022], и это, как представляется, оправдывает наше усилие.

2. Меры расстояния

Для сравнения мы выбрали группу методов, в англоязычной литературе известных как *distance-based approaches*, то есть «подходы, основанные на расстоянии» [Savoy, 2020. P. 33]. В их основе лежит идея о том, что текст или группа текстов могут быть пред-

ставлены в виде вектора — упорядоченного множества значений, которые называются координатами или компонентами вектора. Для каждой пары векторов может быть вычислено расстояние или сходство между ними; минимальное расстояние или максимальное сходство будут указывать на возможного автора. Обратим внимание, что функция считается метрикой расстояния, только если она удовлетворяет критериям неотрицательности, идентичности и симметричности, а также отвечает дополнительному условию — неравенству треугольника [Хачумов, 2012]. В прочих случаях используется понятие «расхождение» [Cha, 2007. P. 300].

Для сравнения мы отобрали несколько функций, уже применявшихся в стилометрических исследованиях и показавших хорошие результаты. Манхэттенское расстояние, оно же расстояние городских кварталов, лежит в основе метода Берроуза [Burgows, 2002]. В качестве альтернативы предлагалось использовать евклидово расстояние [Argamon, 2008], а также косинусное сходство⁵ [Smith, Aldridge, 2011. P. 79–80], в том числе с предварительной стандартизацией [Evert, Proisl et al., 2017]. Стандартизация признаков по z-оценке показывает, на сколько стандартных отклонений значения признака больше или меньше среднего арифметического. Высокая точность классификации достигались с использованием сходства Ружечки, оно же *minmax* [Koppel, Winter, 2014; Kestemont et al., 2016]. 1 – *minmax* эквивалентно расстоянию Танимото, и наоборот [Cha, 2007. P. 302]. Канберрское расстояние рекомендовалось для использования на арабском корпусе [Ahmed, 2019], а расстояние Кларка, среди прочих, — на английском и французском [Kocher, Savoy, 2019]. Из семейства энтропийных расстояний мы возьмем расхождение Джеффриса [Деза, Деза, 2008. С. 221], которое представляет собой симметричную версию расхождения Кульбака-Лейблера; последнее называют также относительной энтропией [Savoy, 2020. P. 39–42]. Поскольку перекрёстная энтропия $H(P, Q)$ для распределений P и Q определяется как сумма энтропии $H(P)$ и относительной энтропии $D_{KL}(P, Q)$, отдельно перекрёстную энтропию мы в этой работе не рассматриваем⁶. Кроме того, протестировано расстояние Лаббе [Labbé, Labbé, 2006; Labbé, 2007; Cortelazzo et al., 2013]. Формулы приведены в Табл. 1.

3. Задачи эксперимента и программные средства

В ходе эксперимента мы ставили перед собой следующие задачи:

1	DManhattan	$\sum_{i=1}^n P_i Q_i$
2	DEuclidean	$\sum_{i=1}^n P_i^2 + \sum_{i=1}^n Q_i^2$
3	SCosine	$\frac{\sum_{i=1}^n P_i Q_i}{\sqrt{\sum_{i=1}^n P_i^2} \sqrt{\sum_{i=1}^n Q_i^2}}$
4	DTanimoto	$\frac{\sum_{i=1}^n P_i Q_i}{\sum_{i=1}^n \max(P_i, Q_i)}$
5	DCanberra	$\frac{\sum_{i=1}^n P_i - Q_i }{\sum_{i=1}^n P_i + \sum_{i=1}^n Q_i}$
6	DClark	$\frac{\sum_{i=1}^n P_i - Q_i }{\sum_{i=1}^n P_i + \sum_{i=1}^n Q_i}$
7	DJeffreys	$\frac{\sum_{i=1}^n (P_i + Q_i) \ln P_i}{\sum_{i=1}^n (P_i + Q_i)}$
8	DLabbe	$\frac{\sum_{i=1}^n P_i - Q_i }{2NP}$

Таблица 1. Меры расстояния и сходства

выяснить, какие меры расстояния дают наибольшую точность на отрывках разной длины с использованием разного числа переменных;

установить, обнаруживаются ли различия при использовании стандартизированных и нестандартизированных значений частотности (для тех методов, которые это допускают);

сравнить точность атрибуции при использовании словоформ или трехсимвольных энграм;

сделать выводы о том, на каких текстах классификатор чаще ошибается;

Все вычисления и визуализации выполнены в R, при этом токенизация (деление на слова или энграммы) выполнена с использованием библиотеки Stylo; для составления матриц расстояния или сходства привлекалась библиотека Philentropy, для других вычислений и визуализаций использовались Tidyverse и Tidymodels. Переход от квадратной матрицы расстояния или сходства к классификации осуществлялся при помощи «доморожденных» (в смысле *home-brew*) функций `get.pred.min` и `get.pred.max`, для каждого ряда матрицы извлекающих имя столбца с минимальным (для расстояний) или максимальным (для сходства) значением в корпусе. При этом, разумеется, не учитывался столбец с тем же самым произведением⁷.

4. Корпус

В корпус, подготовленный для этого эксперимента, вошли сочинения древнегреческих прозаиков (историков, врачей, философов и ораторов) из библиотеки Perseus⁸. Для извлечения текстов ис-

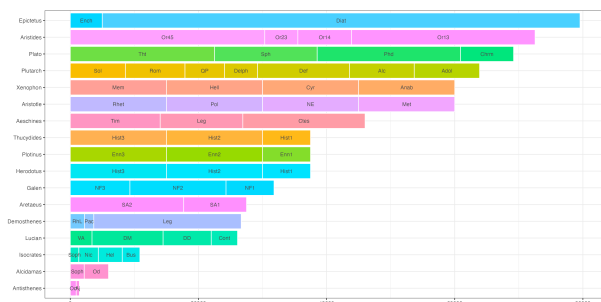


Рис. 1. Количество слов и текстов на автора

пользовались пакеты: XML (<https://cran.r-project.org/web/packages/XML/index.html>) и RPerseus (<https://github.com/ropensci/rperseus>). Объем корпуса — 694 тыс. слов. Корпус является несбалансированным в двух отношениях: разные авторы представлены разным количеством текстов, а сами эти тексты неравномерны по объему. Диспропорцию хорошо видно на Рис. 1. Две декламации Антисфена («Одиссей» и «Аякс») в сумме дают меньше полуторы тысячи слов; максимальное число слов у Эпиктета, но из них лишь около 5000 приходится на «Энхиридион». Аристид представлен четырьмя речами (13 «Панафинейская речь», 14 «Похвала Риму», 23 «Священные речи а'», 45 «К Платону о риторике»), а Платон — четырьмя диалогами («Хармид», «Федон», «Софист», «Теэтет»), как и Лукиан («Харон, или Наблюдатели», «Разговоры мертвых», «Разговоры богов», «Продажа жизней»). Среди сочинений Плутарха — жизнеописания («Ромул», «Солон», «Алкивиад»), «Платоновские вопросы», трактат «О том, как юноше слушать поэтические произведения», а также два диалога: «Об упадке оракулов» и «О “Е” в Дельфах».

Аристотель представлен выборками из «Метафизики», «Политики», «Риторики» и «Никомаховой этики»; Ксенофонт — выборками из «Анабасиса», «Меморабилей», «Греческой истории» и «Киропедии». Также среди историков Геродот (три выборки из «Истории») и Фукидид (три выборки из «Истории»); к философам мы добавили Плотина (три выборки из «Энеад»). Среди текстов Галена в репозитории библиотеки Perseus нам был доступен единственный трактат «О естественных способностях», из которого также взято три выборки; из сочинений врача II в. Аретея — пер-

вая и вторая книги «Этиологии и симптомов острых заболеваний». К этому мы добавили две небольшие речи Алкидаманта («Против софистов» и «Одиссей»), три речи Эсхина («Против Ктесифонта», «Против Тимарха», «О преступном посольстве»), четыре речи Исократ («Бусири», «Елена», «К Никоклу», «Против софистов») и три речи Демосфена («О преступном посольстве», «О мире», «О свободе родосцев»)⁹. Итого 57 текстов (включая выборки) и 17 авторов.

Тексты не подвергались лемматизации, стеммингу, частеречной или синтаксической разметке¹⁰. Для анализа они были только разделены на токены: словоформы и трехсимвольные энграммы (с сохранением диакритики).

5. Оценивание

Для сравнения каждый метод опробовался на отрывках 1000–7000 токенов (с шагом 500, всего 13) с разным количеством предикторов (mfw) — от 100 до 1000 (с шагом 100, всего 10). Для каждой длины отрывка и mfw проведено 10 итераций, при этом использовались повторные выборки (с замещением из-за наличия в корпусе очень коротких текстов). Таким образом, для оценки каждого метода выполнено $57101310 = 74100$ классификаций без стандартизации и столько же со стандартизацией (там, где это допускает метрика). Это было сделано сначала на словоформах, потом на энграммах. Исключение составляет метод Лаббе, не предполагающий отбора mfw, а задействующий все данные об абсолютной частотности слов в отрывке (поэтому здесь всего 7410 классификаций). Точность рассчитывалась как количество верных классификаций к общему числу выполненных классификаций.

6. Результаты классификации для всех методов

На всех отрывках и mfw наилучшие результаты показали расстояние Лаббе (на абсолютных значениях частотности), косинусное сходство на стандартизированных значениях, а также расстояние Танимото (относительная частотность без стандартизации). Чуть хуже показали себя Джеффрис и манхэттенское расстояние. Методы, давшие менее 0.6 процентов точности, исключены из дальнейшего рассмотрения.

method	scale	.estimate
cosine	TRUE	0.934
labbe	FALSE	0.919
tanimoto	FALSE	0.915
jeffreys	FALSE	0.899
manhattan	FALSE	0.893
clark	FALSE	0.847
manhattan	TRUE	0.84
cosine	FALSE	0.836
canberra	FALSE	0.835
euclidean	FALSE	0.826
euclidean	TRUE	0.814
clark	TRUE	0.228
tanimoto	TRUE	0.208
canberra	TRUE	0.045

Таблица 2. Средние показатели точности для всех методов. Слово-формы

method	scale	.estimate
labbe	FALSE	0.857
cosine	TRUE	0.818
tanimoto	FALSE	0.783
manhattan	FALSE	0.769
manhattan	TRUE	0.751
jeffreys	FALSE	0.749
euclidean	TRUE	0.739
cosine	FALSE	0.735
euclidean	FALSE	0.734
canberra	FALSE	0.694
clark	FALSE	0.658
clark	TRUE	0.152
tanimoto	TRUE	0.11
canberra	TRUE	0.042

Таблица 3. Средние показатели точности для всех методов. Энграммы

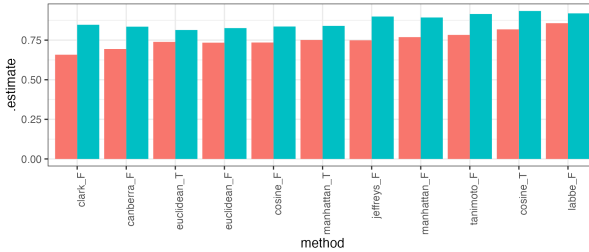


Рис. 2. Точность классификации в зависимости от использования словоформ (бирюзовый) или энграм (розовый); Т (True) указывает на стандартизацию данных

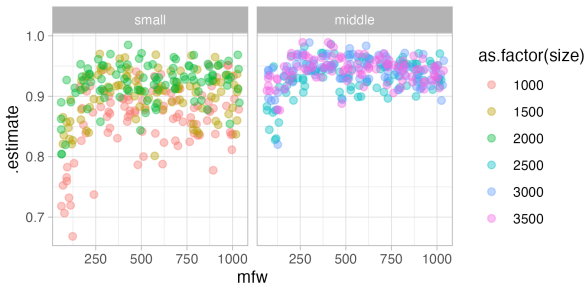


Рис. 3. Точность классификации для небольших и средних отрывков (Delta) в зависимости от числа mfw

Классификация с использованием трехсимвольных энграм стабильно дает худший результат (Рис. 2), что может быть связано с особенностями древнегреческой диакритики.

Как видно в Табл. 2, 3, только косинусное сходство улучшает показатели со стандартизацией; таким образом, знаменитая Delta Берроуза [Орехов, 2020; Skorinkin, Orekhov, 2023] проигрывает не только альтернативным метрикам, но и манхэттенскому расстоянию без стандартизации, то есть фактически более примитивной версии самой себя.

Однако на Рис. 3 можно заметить, что показатели точности снижаются главным образом в тех случаях, когда количество mfw приближается к размеру отрывка, для чего Delta изначально не предназначена. На отрывках длиной 1000–2000 слов (слева) нега-

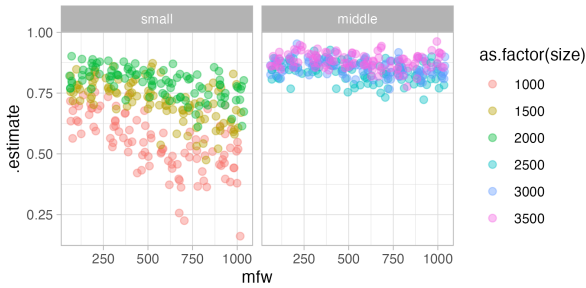


Рис. 4. Точность классификации для небольших и средних отрывков (COS_S) в зависимости от числа *mfw*

тивный эффект очевиден; для средних (справа) и тем более больших отрывков увеличение или уменьшение *mfw* уже не играет такой роли. Можно предположить, что в случае с манхэттенским расстоянием без стандартизации увеличение числа *mfw* на небольших отрывках, хотя и ухудшает результат, но не так сильно — сказывается избыточное влияние гиперчастотных слов (согласно закону Ципфа). Для сравнения: косинусное сходство с применением стандартизации (далее COS_S) на небольших отрывках гораздо более устойчиво к изменениям *mfw* (Рис. 4), что отражается и на общей эффективности метрики.

7. Результаты классификации в зависимости от числа *mfw*

Рис. 5¹¹ и 6 подтверждают, что небольшие *mfw* имеют преимущество лишь при классификации с использованием расстояния Манхэттена на стандартизированных значениях (что соответствует подходу Берроуза). Однако это не распространяется на классификации с использованием трехсимвольных энграм. Для большинства методов точность либо возрастает вместе с числом *mfw*, либо не очень сильно от него зависит.

Наиболее стабильный результат достигается на словоформах при *mfw* > 200 и применении COS_S, расстояния Танимото (далее TAN), а также расхождения Джеффриса (далее JEFF). Для этих методов уже на отрывках в 2000-2500 токенов точность классификации стабилизируется на высоком уровне, что подтверждает выводы Эдера [Eder, 2017]. В случае с расстоянием Лаббе (далее

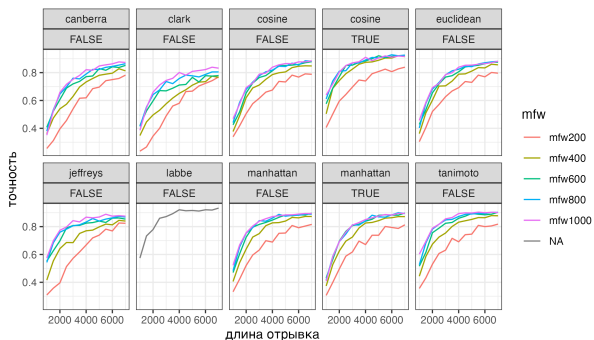


Рис. 5. Точность классификации в зависимости от числа mfw (словоформы)

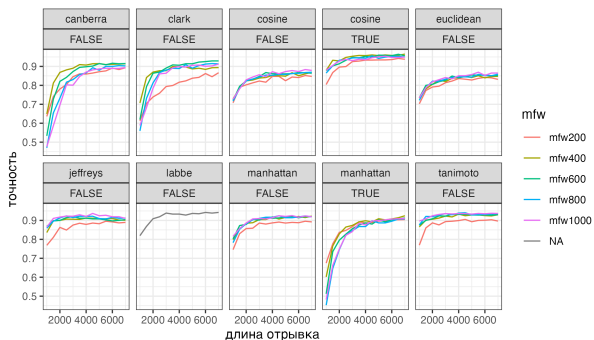


Рис. 6. Точность классификации в зависимости от числа mfw (трех-символьные энграммы)

LAB) максимальная точность достигается начиная примерно с 3000 токенов.

При сопоставимой средней точности TAN, LAB и COS_S (см. выше Табл. 2) наилучшие результаты достигаются с применением именно последнего метода. В топе 50 лучших результатов (начало списка в Табл. 4) — только COS_S, причем на средних и больших отрывках, от 2500 токенов и выше.

method	scale	mfw	size	mean
cosine	TRUE	500	6500	0.968
cosine	TRUE	300	7000	0.965
cosine	TRUE	400	7000	0.965
cosine	TRUE	500	7000	0.965
cosine	TRUE	600	5500	0.965
cosine	TRUE	300	4500	0.963
cosine	TRUE	400	6500	0.963
cosine	TRUE	300	5500	0.961
cosine	TRUE	300	6000	0.961
cosine	TRUE	400	3500	0.96
cosine	TRUE	400	4000	0.96
cosine	TRUE	400	5000	0.96
cosine	TRUE	400	5500	0.96
cosine	TRUE	500	5000	0.96
cosine	TRUE	600	5000	0.96
cosine	TRUE	600	6500	0.96
cosine	TRUE	200	6500	0.958
cosine	TRUE	400	4500	0.958
cosine	TRUE	500	6000	0.958

Таблица 4. Наилучшие классификации с использованием слово-форм (среднее значение для 10 итераций)

8. Методы на малых выборках

В исследовательской работе нередко приходится иметь дело с текстами меньшего размера, поэтому оптимальные параметры отобраны для отрывков в 1000, 1500 и 2000 слов. На отрывках 1000 точность нигде не превышает 0.9, причем TAN превосходит COS_S. На отрывках в 1500 и 2000 слов COS_S показывает весьма достойные результаты (Табл. 5).

Наименьший разброс значений при высокой точности достигается, вполне ожидаемо, на отрывках 2000 слов. Это хорошо видно на Рис. 7, где отчетливо выделяются две «косинусные» вершины.

9. Матрицы ошибок

Поскольку ни в одном случае классификация не достигает стопроцентной точности, интересно узнать, на каких текстах чаще всего ошибается классификатор. Мы визуализировали матрицы

method	scale	mfw	size	.estimate	.sd
tanimoto	FALSE	900	1000	0.9	0.027
tanimoto	FALSE	1000	1000	0.893	0.04
cosine	TRUE	300	1000	0.882	0.032
cosine	TRUE	400	1500	0.933	0.022
cosine	TRUE	300	1500	0.932	0.023
tanimoto	FALSE	700	1500	0.925	0.019
cosine	TRUE	700	2000	0.935	0.019
cosine	TRUE	400	2000	0.933	0.025
cosine	TRUE	900	2000	0.932	0.021

Таблица 5. Точность и стандартное отклонение точности на небольших отрывках (на основе 10 классификаций с повторными выборками)

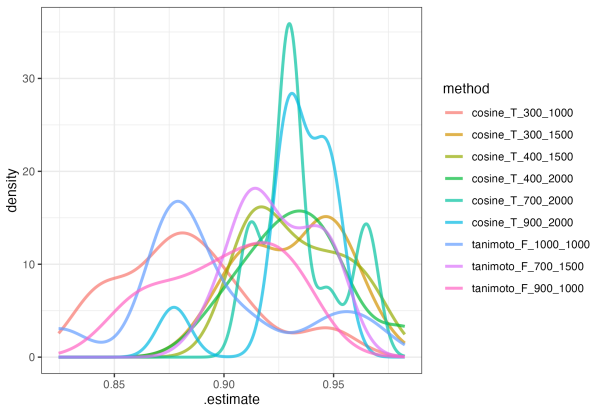


Рис. 7. Кривые плотности для лучших методов на малых выборках

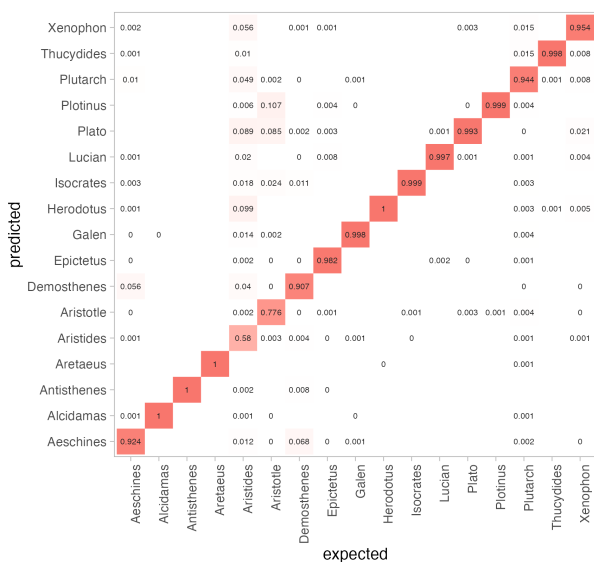


Рис. 8. Матрицы ошибок. COS_S

ошибок для четырех методов: COS_S, TAN, JEFF, а также для Delta (Рис. 8, 9, 10, 11).

Для представленных методов наиболее «непредсказуемые» авторы совпадают — это Аристотель, Демосфен и Аристид; Delta и Джеффрис, кроме того, плохо справились с небольшими текстами Антисфена, отдав их отчасти Алкидаманту¹². Delta также путает Исократ с Алкидамантом (что для обоих было бы очень обидно), Ксенофонта — с Фукидидом и Плутархом, а Эпикета — с Плутархом. Из сочинений Плутарха в нашем корпусе были как моральные, так и исторические сочинения, и можно предположить, что смешение происходит между текстами близкой жанровой природы, что в целом характерно для Delta [Алиева, 2022]. Эти ошибки сохраняются, даже если удалить выборки размером 1000 токенов (см. выше раздел 6).

Ошибки, связанные с атрибуцией текстов Аристида, известного мастера стилизаций, и с путаницей между Демосфеном и Эскином, более или менее понятны. Но судить о том, почему Аристотель оказывается похож на Плотина в каждом десятом случае (COS_S) или даже в каждом пятом (TAN), сложнее. С одной стороны, из-

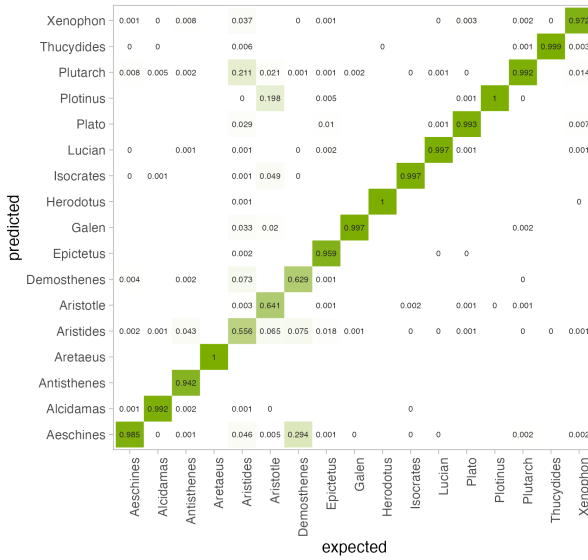


Рис. 9. Матрицы ошибок. TAN

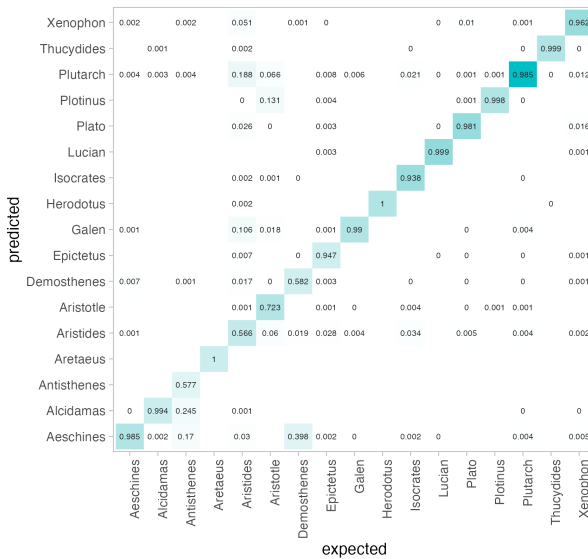


Рис. 10. Матрицы ошибок. JEFF

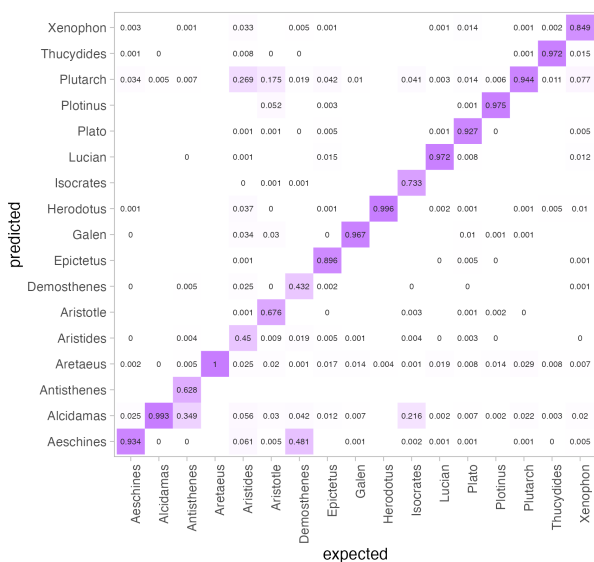


Рис. 11. Матрицы ошибок. Delta (MANH_S)

вестно, что сочинения Аристотеля и его комментатора Александра Афродисийского читались в кружке Плотина (об этом сообщает Порфирий в Vita Plotini 14.13), и это должно было сказаться на философской терминологии «Эннеад». С другой стороны, отдельных общих терминов недостаточно для того, чтобы повлиять на классификатор, который пользуется сотнями предикторов¹³. В целом, вопрос как о стиле самого Аристотеля, так и о его отражении в позднеантичной литературе требует дополнительного изучения.

10. Тестирование

Наш небольшой эксперимент показал, что при классификации древнегреческих текстов наилучшие результаты достигаются при использовании косинусного сходства на стандартизированных значениях частотности, расстояния Танимото на относительных значениях частотности и расстояния Лаббе на абсолютных значениях частотности. Количество признаков для первых двух методов должно быть больше двухсот. Использование трехсимвольных энграм вместо словоформ ухудшает результат. Лаббе показывает высокую

size	.estimate
1000	0.875
1500	0.911
2000	0.926
2500	0.938
3000	0.944
3500	0.946
4000	0.948
4500	0.953
5000	0.952
5500	0.955
6000	0.954
6500	0.956
7000	0.955

Таблица 6. Точность классификации (все mfw) для отрывков разной длины. COS_S

эффективность на отрывках от 3000 токенов, а на самых коротких отрывках лучше работает Танимото.

Можно ли говорить о статистически значимой разнице в точности предсказаний между этими тремя методами? Для сравнения трех векторов, хранящих данные о точности классификации на отрывках разной длины, мы использовали парный непараметрический тест Уилкоксона [Мастицкий, Шитиков, 2015. С. 157]. Для этого были удалены классификации с использованием mfw 100 и 200, показавшие наихудший результат, а результаты сохранены в виде трех векторов, хранящих сведения о средней точности классификации по итогам десяти итераций на всех mfw (Табл. 6).

Тесты показывают отсутствие статистически значимой разницы между LAB и TAN, но на уровне значимости 0.05 отвергнута гипотеза об отсутствии различия между COS_S и LAB, а также TAN и COS. Таким образом, косинусное сходство оказывается наиболее эффективным методом на данном корпусе.

11. Применение к новым данным

Используя все три метода, мы попробовали классифицировать послание «К Демонику». Еще в древности высказывались сомнения в авторстве Исократу; сегодня принято считать, что текст создан одним из учеников Исократу в ответ на «Протрептик» Аристоте-

size	.estimate
1000	0.819
1500	0.868
2000	0.909
2500	0.919
3000	0.939
3500	0.933
4000	0.933
4500	0.928
5000	0.937
5500	0.935
6000	0.942
6500	0.939
7000	0.942

Таблица 7. Точность классификации (все mfw) для отрывков разной длины. LAB

size	.estimate
1000	0.878
1500	0.908
2000	0.915
2500	0.924
3000	0.929
3500	0.929
4000	0.935
4500	0.933
5000	0.93
5500	0.931
6000	0.93
6500	0.931
7000	0.933

Таблица 8. Точность классификации (все mfw) для отрывков разной длины. TAN

ля¹⁴. Так, Фридрих Бласс полагал, что стиль «Демоника» недостойн (nicht würdig) Исократы [Blass, 1874. S. 257–258]. При этом Бласс учитывал то, что сам он называл *mikroskopische Eigenheiten*, такие как количество зияний или частотность предлога *syn*¹⁵, однако в случае с «Демоником» руководствовался, кажется, больше исследовательским чутьем, чем статистикой. Даг Хатчинсон, автор недавней реконструкции «Протрептика», настаивает, что убедительных доводов в пользу отвержения нет¹⁶. Используя три метода (COS_S size 6500, mfw 500; TAN size 6000 mfw 1000; LAB size 6000) и десять повторных выборок, мы сравнили это послание со всеми текстами в нашем корпусе. В результате все три метода отдали текст Исократу. Эти результаты следует интерпретировать осторожно: в нашем корпусе может не быть подлинного автора текста, и, если подражатель достаточно искусен, то атрибуция Исократу не удивительна.

Еще один пример — трактат «О воспитании детей» из корпуса Плутарха, который, согласно авторитетному суждению Виттенбаха, не может быть творением херонейского философа [Wytttenbach 1820. P. 1–30]. В этом случае наши метрики скорее подтверждают сомнения Виттенбаха: COS_S колеблется между Исократом и Плутархом (3/7); TAN — между Аристидом, Аристотелем и Плутархом (1/1/8); LAB — между Аристотелем и Аристидом (9/1). Это ничего не говорит нам о личности автора (подражателя скорее всего его нет в нашей коллекции), но в достаточной мере свидетельствует о том, что текст не очень похож на Плутарха.

И здесь уместно снова вернуться к Фридлиндеру и его словам о «Пире», процитированным в начале этой статьи. Семь речей «Пира» написаны каждая в своей в узнаваемой стилистике, что должно создавать некоторые трудности при автоматической классификации текста¹⁷. Речь Агафона — блестящая имитация Горгия [Dover, 1968. P. 90–91], речь Федра, большого поклонника Лисия, похожа на речь самого Лисия из «Федра» [Tarrant, 2016. P. 86], а голос Алквиада у Платона звучит во многом так же, как у Фукидида [Sansone, 2018]. Тем не менее, все три классификатора отдают «Пир» Платону: COS_S в 8 случаях из 10 (еще 2 — автору подложного «Феага», которого мы тоже добавили в подборку), LAB и TAN в 10 случаях из 10. Но справедливости ради надо сказать, что и «Феаг» оказывается ближе всего к Платону в целом или к «Пиру» в частности (лишь в двух случаях TAN отдает текст Ксенофону). С точки зрения статистики, это вполне удачная имитация, хотя opinio

communis утверждает, что текст Платону не принадлежит, даже если и написан еще при жизни Платона [Thesleff, 2009. P. 364].

Таким образом, оптимистичные показатели точности не должны нас вводить в заблуждение, причем это касается не только описанных в этой статье методов, но и других, вероятно более совершенных. Авторитет традиции действителен до тех пор, пока не опровергнут, в то время как подтвердить его статистически не представляется возможным.

Примечания

- 1 Я благодарю Бориса Орехова и Даниила Скоринкина за замечания к первым версиям этой статьи, позволившие существенно доработать как сам текст, так и код. Я также признательна Артему Юнусову, который обратил мое внимание на две работы Р. Ойкена о служебных словах у Аристотеля. — ОА.
- 2 О важных импликациях этого тезиса применительно к античной эпистографии см.: [Forcignandò, Martinelli Tempesta, 2023].
- 3 «Всякая книга признается подлинной, пока не доказано обратное» (лат.).
- 4 О проблеме «неустойчивости» античных и средневековых текстов и подходах к их реконструкции см.: [Шумилин, 2020. С. 33 с прим. 11 et *passim*].
- 5 Также называемое подобность Орчини, угловая подобность или нормированное скалярное произведение [Деза, Деза, 2008. С. 264].
- 6 Об энтропии в целом см.: [Shannon, 1948]; перекрестная энтропия для классификации текстов: [Juola 1997; Juola, 1998; Juola, Baayen, 2005] относительная энтропия для классификации текстов: [Zhao, Zobel, Vines, 2006] и [Zhao, Zobel, 2007].
- 7 Весь код доступен на GitHub: <https://github.com/locusclassicus/compareDist/tree/master>
- 8 Perseus, Canonical Greek Literature: <https://github.com/PerseusDL/canonical-greekLit>
- 9 В вопросе о подлинности речей Демосфена мы опирались на [Harris, 2013. P. 401–402], а также [Trevett, 2018]. Гиппократ и Лисия мы были вынуждены исключить из-за спорного статуса большинства сочинений корпуса. Так, [Dover, 1968. P. 193] признает безусловно подлинной лишь XII речь Лисия; критика у [Winter, 1973], но вопрос остается открытым.
- 10 О способах обработки текстов в целом см.: [Stamatatos et al., 2001; Stamatatos, 2009].
- 11 Для удобства на графике линия, обозначенная как *mfw* 200, объединяет испытания на 100 и 200 *mfw*, и т.д.
- 12 LAV справляется с Антисфеном, но тоже спотыкается на Аристотеле, Демосфене и Аристиде (точность примерно 0.7–0.75 для всех троих).

¹³ Интерпретировать этот результат тем сложнее, что стиль Аристотеля (в отличие от стиля Платона) изучен очень плохо: помимо двух работ Ойкена [Eucken, 1866; Eucken 1868] можно упомянуть лишь исследования Кенни [Kenney, 1978. P. 70–160], но они посвящены почти исключительно «Этикам». Кроме того, как следует из диссертации Ойкена [Eucken, 1866. P. 15, 31 *et passim*], рукописные и издательские разночтения нередко касаются именно тех служебных слов, которые особенно важны как показатели авторского стиля (что, впрочем, верно для большинства античных авторов).

¹⁴

См.: [Mathieu, Brémond, 1928. P. 110–119] и [Düring, 1961. P. 23, 226]. [Wendland, 1905. S. 81–101] был первым, кто отметил параллели с Аристотелем; он также считает «Демоника» творением подражателя.

¹⁵ [Blass, 1892. S. 295 = Бласс 2016. С. 193]: «Самыми убедительными являются мелкие признаки, те свойственные каждому писателю микроскопические особенности, которых подражатель или не мог заметить, или подражание которым представилось бы слишком трудным. Таковы особенности относительно допущения зияния и соблюдения ритма, употребления известных частиц, избегания других, и т.п.». О вкладе Фридриха Бласса в количественные исследования стиля Платона см.: [Brandwood, 1990. P. 9–10].

¹⁶ Письмо автору от 4.12.2017. Реконструкция доступна по ссылке: <http://www.protrepticus.info/>

¹⁷ На подобные трудности указывалось и в связи с речами «Федра»; см., напр.: [Robinson, 1992. P. 381]. Вообще способность автора сознательно контролировать свой «почерк» — одно из главных возражений противников платоновской стилеметрии в ее традиционном виде, то есть направленной на установление сравнительной хронологии диалогов; см., напр.: [Howland, 1991. P. 203] и [Waterfield, 1980. P. 275]. В то же время сознательная «смена регистра» все чаще становится предметом специальных количественных исследований; см., напр.: [Tarrant, Benitez, Roberts, 2011].

Литература

Исследования

Алиева 2022 — Алиева О.В. Delta Берроуза для древнегреческих авторов: опыт применения // Scholē. Философское антиковедение и классическая традиция. 2022. Т. 16. № 2. С. 693–705.

Деза 2008 — Деза Е., Деза М.М. Энциклопедический словарь расстояний. М.: Наука, 2008.

Мастицкий, Шитиков, 2015 — Мастицкий С.Э., Шитиков В.К. Статистический анализ и визуализация данных с помощью R. М.: ДМК, 2015.

- Орехов 2020* — Орехов Б.В. «Илиада» Е.И. Кострова и «Илиада» А.И. Любжина: стилеметрический аспект // *Аристей*. Т. 21. 2020. С. 282–296.
- Пешков 2016* — Пешков И.В. Зарождение категории авторства в золотом веке английской литературы: дисс. . . . докт. филол. наук: 10.01.08. М.: РГГУ, 2016.
- Фоминых 2017* — Фоминых С.В. Автоматический независимый от языка анализ авторства патристических текстов на основании статистики частот переходов // *Исторический журнал: научные исследования*. Т. 5. 2017. С. 70–79.
- Фуко 2008* — Фуко М. Что такое автор? // *Эстетика и теория искусства XX в.: Хрестоматия* / Н.А. Хренов, А.С. Мигунов (сост.). М.: Прогресс-традиция, 2008.
- Хачумов 2012* — Хачумов М.В. Расстояния, метрики и кластерный анализ // *Искусственный интеллект и принятие решений*. Т. 9. № 1. 2012. С. 81–89.
- Шумилин 2020* — Шумилин М.В. Ошибка против варианта: «новая филология», латинистика и «плохой язык» // *Vox medii aevi*. Т. 1–2. 2020. С. 28–75. <<http://voxmediaevi.com/volumens/2020-1-2/2020-1-2-shumilin/>>
- Ahmed 2019* — Ahmed H. Distance-Based Authorship Verification Across Modern Standard Arabic Genres // *Proceedings of the Third Workshop on Arabic Corpus Linguistics*. 2019. P. 89–96. <<https://aclanthology.org/W19-5611.pdf>>.
- Argamon 2008* — Argamon S. Interpreting Burrows' Delta: Geometric and Probabilistic Foundations // *Literary and Linguistic Computing*. Vol. 23. No. 2. 2008. P. 131–147.
- Blass 1874* — Blass F.W. *Attische Beredsamkeit*. 2. Abtheilung: Isokrates und Isaios. Leipzig: B.G. Teubner, 1874.
- Blass 2016* — Blass F.W. *Hermeneutik und Kritik* // *Handbuch der klassischen Altertums-Wissenschaft in systematischer Darstellung mit besonderer Rücksicht auf Geschichte und Methodik der einzelnen Disziplinen* / I. von Müller (hrsg.). Bd. 1: Einleitende und Hilfs-Disziplinen. 2. Aufl. München: C.H. Beck, 1892. Бласс Ф.В. Герменевтика и критика: искусство понимания произведений классической древности и их литературная оценка / Л.Ф. Воеводский (пер.). 2-е изд. М.: Ленанд, 2016.
- Boeckh 1886* — Boeckh A. *Encyklopädie und Methodologie der philologischen Wissenschaften* / Ernsts Bratuscheck (hrsg). 2. Aufl. Wiesbaden: Springer Fachmedien, 1886.
- Brandwood 1990* — Brandwood L. *The Chronology of Plato's Dialogues*. Cambridge: Cambridge University Press, 1990.
- Burrows 2002* — Burrows J.F. 'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship // *Literary and linguistic computing*. Vol. 17. No. 3. 2002. P. 267–287.

- Chambers 1996* — Chambers M. The Athenaeon Politeia after a Century // Transitions to Empire. Essays in Greco-Roman History, 360–140 B.C., in honor of E. Badian / R. Wallace and E. Harris (eds.). Norman and London: University of Oklahoma Press, 1996. P. 211–225.
- Cortelazzo et al., 2013* — Cortelazzo M.A., Nadalutti P., Tuzzi A. Improving Labbé's Intertextual Distance: Testing a Revised Version on a Large Corpus of Italian Literature // Journal of Quantitative Linguistics. Vol. 20. No. 2. 2013. P. 125–152.
- Dover 1968* — Dover K.J. Lysias and the Corpus Lysiacum. Berkeley and Los Angeles: University of California Press, 1968.
- Eder 2011* — Eder M. Style-Markers in Authorship Attribution A Cross-Language Study of the Authorial Fingerprint // 2011. No 6. P. 99–114.
- Eder 2015* — Eder M. Taking Stylometry to the Limits: Benchmark Study on 5,281 Texts from “Patrologia Latina” // Digital Humanities 2015: Book of Abstracts, University of Western Sydney. : Alliance of Digital Humanities Organizations (ADHO), 2015.
- Eder 2017* — Eder M. Short Samples in Authorship Attribution: A New Approach // Digital Humanities 2017. Montreal, 2017. <https://dh2017.adho.org/abstracts/341/341.pdf>
- Eucken 1866* — Eucken R. De Aristotelis dicendi ratione: Pars prima. Observations de particularum usu. Dissertatio inauguralis. Gottingae: Typis expressit officina Hoferiana, 1866.
- Eucken 1868* — Eucken R. Ueber den Sprachgebrauch des Aristoteles: Beobachtungen ueber die Praepositionen. Berlin: Weidmansche Buchhandlung, 1868.
- Evert et al. 2017* — Evert S., Proisl Th., Jannidis F., Reger I., Pielström S., Schöch Ch., Vitt Th. Understanding and explaining Delta measures for authorship attribution // Digital Scholarship in the Humanities. Vol. 32. Supp. 2. 2017. P. ii4-ii16.
- Forcignanò et al. 2023* — Forcignanò F., Martinelli Tempesta S. Comparing Corpora, Rethinking Authenticity: Why Are Platonic Letters “Platonic”? // The Making of the Platonic Corpus / Alieva O., Nails D., Tarrant H. (eds.). Paderborn: Brill, 2023. P. 203–221.
- Harris 2013* — Harris E.M. The Rule of Law in Democratic Athens. Oxford: Oxford University Press, 2013.
- Holmes 1994* — Holmes D.I. Authorship Attribution // Computers and the Humanities. Vol. 28. No. 2. 1994. P. 87–106.
- Hoover 2004* — Hoover D.L. Testing Burrows's Delta // Literary and Linguistic Computing. Vol. 19. No. 4. 2004. P. 453–475.

Howland 1991 — Howland J. Re-Reading Plato: The Problem of Platonic Chronology // *Phoenix*. Vol. 45. No. 3. 1991. P. 189–214.

Jannidis, Lauer 2014 — Jannidis F., Lauer G. Burrows's Delta and Its Use in German Literary History // *Distant Readings. Topologies of German Culture in the Long Nineteenth Century Studies in German Literature Linguistics and Culture* / M. Erlin, L. Tatlock (eds.). Rochester: Camden House, 2014. P. 29–54.

Juola 1998 — Juola P. Cross-Entropy and Linguistic Typology // *New Methods in Language Processing and Computational Natural Language Learning* / D.M.W. Powers (ed.). Sydney: ACL, 1998. P. 141–149. <<https://aclanthology.org/W98-1217.pdf>>

Juola 2013 — Juola P. How a Computer Program Helped Show J.K. Rowling write *A Cuckoo's Calling*: Author of the Harry Potter books Has a Distinct Linguistic Signature // *Scientific American* 20.08.2013 <<https://www.scientificamerican.com/article/how-a-computer-program-helped-show-jk-rowling-write-a-cuckoos-calling/>>

Juola 1997 — Juola P. What Can We Do with Small Corpora? Document Categorization via Cross-entropy // *Proceedings of an Interdisciplinary Workshop on Similarity and Categorization*. Edinburgh: University of Edinburgh, 1997.

Juola, Baayen 2005 — Juola P., Baayen H. A Controlled-corpus Experiment in Authorship Identification by Cross-Entropy // *Literary and Linguistic Computing*. Vol. 20 Suppl. 2005. P. 59–67.

Kenny 1978 — Kenny A. *The Aristotelian Ethics: A Study of the Relationship between the Eudemian and Nicomachean Ethics of Aristotle*. Oxford: Clarendon Press: 1978.

Kenny 1982 — Kenny A. *The Computation of Style: An Introduction to Statistics for Students of Literature and Humanities*. Oxford, Pergamon Press.

Kestemont et al. 2016 — Kestemont M., Stover J., Koppel M., Karsdorp F., Daelemans W. Authenticating the writings of Julius Caesar // *Expert Systems with Applications*. Vol. 63. 2016. P. 86–96.

Kocher, Savoy 2019 — Kocher M., Savoy J. Evaluation of Text Representation Schemes and Distance Measures for Authorship Linking // *Digital Scholarship in the Humanities*. Vol. 34. No. 1. 2019. P. 189–207.

Koentges 2020 — Koentges Th. The Un-Platonic Menexenus: A Stylometric Analysis with More Data // *Greek, Roman, and Byzantine Studies*. Vol. 60. 2020. P. 211–241.

Koppel, Winter 2014 — Koppel M., Winter Y. Determining If Two Documents are Written by the Same Author // *Journal of the Association for Information Science and Technology*. Vol. 65. No. 1. 2014. P. 178–187.

- Labbé 2007* — Labbé D. Experiments on Authorship Attribution by Intertextual Distance in English // *Journal of Quantitative Linguistics*. Vol. 14. No. 1. 2007. P. 33–80.
- Labbé 2006* — Labbé C., Labbé D. A Tool for Literary Studies: Intertextual Distance and Tree Classification // *Literary and Linguistic Computing*. Vol. 21. No. 3. 2006. P. 311–326.
- Ledger 1989* — Ledger G. *Recounting Plato: A Computer Analysis of Plato's Style*. Oxford: Oxford University Press, 1989.
- Love 2002* — Love H. *Attributing Authorship: An Introduction*. Cambridge: Cambridge University Press, 2002.
- Mathieu, Brémond 1928* — Mathieu G., Brémond E. eds. 1928. *Isocrate: Discours*. T. 1. Paris: Les belles lettres.
- Robinson 1992* — Robinson T.M. Plato and the Computer // *Ancient Philosophy*. Vol. 12. 1992. P. 375–382.
- Rudman 2005* — Rudman J. Unediting, De-Editing, and Editing in Nontraditional Authorship Attribution Studies: With an Emphasis on the Canon of Daniel Defoe // *The Papers of the Bibliographical Society of America*. Vol. 99. No. 1. 2005. P. 5–36.
- Rybicki, Eder 2011* — Rybicki J., Eder M. Deeper Delta across Genres and Languages: Do We Really Need the Most Frequent Words? // *Literary and Linguistic Computing*. Vol. 26. No 3. 2011. P. 315–321.
- Sansone 2018* — Sansone D. Stylistic Characterization in Plato: Nicias, Alcibiades, and Laches // *Greek, Roman, and Byzantine Studies*. Vol. 58. 2018. P. 156–176.
- Savoy 2020* — Savoy J. *Machine Learning Methods for Stylometry: Authorship Attribution and Author Profiling*. Cham: Springer, 2020.
- Shannon 1948* — Shannon C.E. A Mathematical Theory of Communication // *The Bell System Technical Journal*. Vol. 27. No. 3. 1948. P. 379–423.
- Skorinkin, Orekhov 2023* — Skorinkin D., Orekhov B. Hacking stylometry with multiple voices: Imaginary writers can override authorial signal in Delta // *Digital Scholarship in the Humanities*. 2023 <<https://doi.org/10.1093/llc/fqad012>>
- Smith, Aldridge 2011* — Smith P.W.H., Aldridge W. Improving Authorship Attribution: Optimizing Burrows' Delta Method // *Journal of Quantitative Linguistics*. Vol. 18. No. 1. 2011. P. 63–88.
- Stamatatos 2009* — Stamatatos E. A Survey of Modern Authorship Attribution Methods // *Journal of the American Society for Information Science and Technology*. Vol. 60. No. 3. 2009. P. 538–556.

- Stamatatos et al. 2000* — Stamatatos E., Kokkinakis G., Fakotakis N. Automatic Text Categorization in Terms of Genre and Author // Computational linguistics. Vol. 26. No. 4. 2000. P. 471–495.
- Stanikūnas et al 2017* — Stanikūnas D., Mandravickaitė J., Krilavičius T. Comparison of distance and similarity measures for stylometric analysis of Lithuanian texts // ICYRIME 2017: Proceedings of the Symposium for Young Researchers in Informatics, Mathematics and Engineering. Aachen: CEUR-WS, 2017. P. 1–7 <<https://ceur-ws.org/Vol-1852/p01.pdf>>
- Tarán 1975* — Tarán L. *Academica: Plato, Philip of Opus, and the Pseudo-Platonic Epinomis*. Philadelphia: American Philosophical Society, 1975.
- Tarrant 2016* — Tarrant H. Stylistic Difference in the Speeches of the Symposium // Plato in Symposium: Selected Papers from the Tenth Symposium Platonicum / M. Tulli and M. Erler (eds.). Sankt Augustin: Academia Verlag, 2016. P. 84–90.
- 2016 et al. 2011* — Tarrant H., Benitez E.E., Roberts T. The Mythical Voice in the Timaeus-Critias: Stylometric Indicators // Ancient Philosophy. Vol. 31. 2011. P. 95–120.
- Thesleff 2009* — Thesleff H. *Platonic Patterns: A Collection of Studies* by Holger Thesleff. Las Vegas, Zurich, Athens: Parmenides Publishing, 2009.
- Thesleff 2023* — Thesleff H. Afterthoughts on “School Accumulation” in Plato’s Academy // The Making of the Platonic Corpus / Alieva O., Nails D., Tarrant H. (eds.). Paderborn: Brill, 2023. P. 1–14.
- Trevett 2018* — Trevett J. 2018. Authenticity, Composition, Publication // The Oxford Handbook of Demosthenes / G. Martin (ed.). Oxford. P. 419–430.
- Waterfield 1980* — Waterfield R. A. H. The Place of the Philebus in Plato’s Dialogues // Phronesis. Vol. 25, No. 3. 1980. P. 270–305.
- Wendland 1905* — Wendland P. *Anaximenes von Lampsakos: Studien zur ältesten Geschichte der Rhetorik* // Festschrift für die XLVIII. Versammlung deutscher Philologen und Schulmänner in Hamburg, von Paul Wendland. Berlin, 1905.
- Winter 1973* — Winter T.N. On the Corpus of Lysias // The Classical Journal. Vol. 69. No. 1. 1973. P. 34–40.
- Wytttenbach 1820* — Wytttenbach D. *Animadversiones in Plutarchi Opera moralia*. Lipsiae, 1820.
- Zhao, Zobel 2007* — Zhao Y., Zobel J. Entropy-Based Authorship Search in Large Document Collections // Lecture Notes in Computer Science. Vol. 4425. 2007. P. 381–392.
- Zhao, Zobel 2006* — Zhao Y., Zobel J., Vines Ph. Using Relative Entropy for Authorship Attribution // Lecture Notes in Computer Science. Vol. 4182. 2006. P. 92–105.

Борис Орехов

НЕСТИХОВЕДЧЕСКИЙ РИТМ В РОМАНЕ Н. Г. ЧЕРНЫШЕВСКОГО «ЧТО ДЕЛАТЬ?»

Несмотря на неоднозначную репутацию, самый знаменитый роман Чернышевского нуждается в исследовании поэтики. В статье предлагается анализ самого заметного при чтении плана – ритмического. Он настойчиво задается автором при помощи разнообразных языковых средств, и именно цифровой подход позволяет оценить масштаб применения этого приема.

Ключевые слова: Чернышевский, ритм, sequential pattern mining, паттерны

Роман Н. Г. Чернышевского «Что делать?» уже давно получил среди исследователей литературы неоднозначную репутацию.

С одной стороны, в советское время этот текст рассматривался как важная веха в общественно-политической борьбе, итогом которой стало установление социалистической власти. Таким образом, роман получал преимущественный статус принадлежности к литературе первого ряда, которой можно и нужно посвящать исследовательские тексты. Правда, такое положение дел не стимулировало исследование формы романа, поскольку советская идеология поощряла именно содержательную, «идейную» критику, сочетавшуюся с контекстно-биографическими работами. Это хорошо иллюстрирует библиографический указатель, составленный О. В. Ломовой [Ломова 2015].

С другой стороны, во многом насаждаемый характер канонизации «Что делать?» вызывал ответную негативную реакцию среди исследователей художественной формы, например, среди представителей московско-тартуской структурно-семиотической школы.

Соответствующую иллюстрацию приводит в своих воспоминаниях А. К. Жолковский: «Характерной была реакция на доклад Б.М. Гаспарова и И.А. Паперно (перед самой их эмиграцией) о революционно-демократическом романе Чернышевского „Что делать?“, со школьных лет опостылевшем каждому советскому человеку. Авторы, привлеченные замечательной главой о Чернышевском в „Даре“ В. Набокова, решились подойти к этому роману не как к идейному, а как к чисто литературному явлению и открыли в нем интереснейшую систему приемов сюжетной иронии — главным образом, восходящих к Стерну.

После доклада наступила минутная тишина, а потом чей-то неуверенный голос: „Так вы хотите сказать, что “Что делать?” — хорошая книга?“ — и почти извиняющийся ответ Б.М. Гаспарова: „Пожалуй, да“. В научном кружке, где в рассуждениях принципиально не допускалось никакой оценочности, это звучало почти трогательно»¹.

В своем весьма эмоциональном эссе, посвященном Чернышевскому А. К. Жолковский добавляет: «Однако на сегодняшний день самое интересное в ЧД — это его бесспорная и сокрушительная прагматическая эффективность»; «Роман по праву знаменит своей „плохописью“. Он столь откровенно и безобразно антихудожествен (да и сам его повествователь непрерывно настаивает на ненужности хорошего стиля), что царские цензоры сознательно пропустили его в печать, рассчитывая повредить таким образом репутации и политической линии автора» [Жолковский 1994].

Аналогичная судьба постигла и других важных для советского проекта авторов, например, о Некрасове современный исследователь пишет: «Поэтика и проблематика некрасовской лирики остается настолько плохо описанной, что говорить о ней приходится практически заново» [Зубков 2022].

Идеологическая нагруженность текста порождала негативную реакцию со стороны эстетствующих читателей. Одной из риторических задач романа «Дар» Набокова состояла в уничтожении художественной репутации Чернышевского. В свете темы, которую мы рассматриваем далее, особенно любопытен стиховедческий угол зрения на эту проблему [Орлицкий 2015]. Иллюстративная стратегия автора также состояла в самоумалении: «У меня нет ни тени художественного таланта. Я даже и языком-то владею плохо. Но это все-таки ничего: читай, добрейшая публика! прочтешь не без пользы. Истина — хорошая вещь: она вознаграждает недостатки писателя, который служит ей. Поэтому я скажу тебе: если б я не

предупредил тебя, тебе, пожалуй, показалось бы, что повесть написана художественно, что у автора много поэтического таланта. Но я предупредил тебя, что таланта у меня нет, — ты и будешь знать теперь, что все достоинства повести даны ей только ее истинностью»². По всей видимости, Чернышевский рассматривал такие высказывания как способ обойти цензурные ограничения (и этот расчет оказался успешным).

При этом беспристрастный взгляд, которого до сих пор не могла себе позволить русскоязычная наука о литературе, действительно открывает ряд небанальных особенностей организации текста, в частности, связанных с ритмизацией прозы. Маскирующие слова Чернышевского не стоит воспринимать всерьез: автор предисловия к современной публикации диссертации Чернышевского отмечает, что писатель «оказывается не так уж семиотически наивен и невнимателен к языку как таковому» [Арсеньев 2022, 34]. Внимание к форме обнаруживается и в авторских метакомментариях: «Я употребил обыкновенную хитрость романистов: начал повесть эффектными сценами, вырванными из середины или конца ее, прикрыл их туманом».

Феномен ритмизованной прозы обычно связывается с опытами Андрея Белого, прежде всего, романом «Петербург». Однако ритм как эстетическая категория так или иначе присутствует в любом художественном произведении, независимо от вида искусства, стилистики и жанра. Если рассматривать ритм как регулярное чередование однопорядковых элементов, то такого рода элементы будут встроены и в музыкальную композицию, и в архитектурную проектировку, и, разумеется, в произведение словесности.

Вот характерный пример узуса: «Присутствие ритма <...> допустимо и там, где вообще нет ни колонн, ни антаблемента. Его могут “отбить” и простые окна и двери, и зубцы крепостных стен, и скульптурные украшения, и даже сами объемы, формирующие композицию здания» [Кавтарадзе 2016, 25].

В живописи ритм проявляет себя в соотношении цвета и элементов композиции, в скульптуре — пластическими формами. Очевиден и визуализированный ритм орнамента, оформленный упорядоченным чередованием его элементов. Таким образом, ритм присутствует даже в «пространственных» (по Лессингу) искусствах, но более очевиден во «временных».

Отечественная школа литературоведения традиционно сильна стиховедческими исследованиями. Стиховедение в применении к русскому классическому стиху рассматривает текст как упорядо-

ченное чередование ударных и безударных слогов. В смысле применения статистики к такого рода исследованиям Чернышевский выступает для стиховедов как прямой предшественник: «Сама по себе мысль Чернышевского — возвести законы чередования метрических ударений к естественной языковой акцентовке и для этого запастись статистическими данными из анализа немерной речи — была и смелой, и в конечном счете плодотворной. Подлинное осуществление она нашла только в стихологической литературе нашего времени, в работах Б. В. Томашевского и Г. Шенгели. Новейшие стиховеды — через 60 с лишним лет после попытки Чернышевского, естественно, могли обставить свою работу всеми необходимыми научными гарантиями» [Гиппиус 1966, 284].

Такая оптика позволяет в рамках стиховедческой методологии рассматривать прозаический текст как содержащий фрагменты, ритмизованные по правилам классического русского стиха, называемые случайными метрами [Орехов 2022].

Однако не следует забывать, что ритмизация не ограничивается стихоподобной силлаботоникой. Если ритм — это регулярное чередование однопорядковых элементов, то эти элементы можно найти на любом уровне организации текста. Об этом ясно пишет М. М. Гиршман, указывающий, что ритм может быть обнаружен как в чередовании фразовых компонентов, целых фраз, абзацев, так и в повторах тем, мотивов, ситуаций, в закономерностях сюжетного движения, в соотношении различных композиционно-речевых единиц, в развертывании всей системы образов и каждого из них [Гиршман 1982].

Действительно, даже при беглом чтении романа хорошо заметна используемая Чернышевским ритмизация, задаваемая через повторение и варьирование однотипных конструкций:

«А у сестрицы жених-то богатый! А маменька говорит: жених-то глупый!» «А уж маменька как за женихом-то ухаживает!»

Я думала об этом и решилась. Я тогда не останусь здесь. Я могу быть актрисой.

Как я хорошо придумала! Как я вам благодарна! Как я буду счастлива!

Скажи: я против жены был. Скажи: нам в угоду сделал, потому что видел, не было вашего желания. Скажи: моя жена была одна виновата, я вашу волю исполнял. Скажи: я сам их и свел.

Как долго ты любил меня, и молчал! Как ты благороден! Как он благороден, Саша!

Ты узнаешь ли мой голос? Ты узнаешь ли лицо мое? Ты видела ли лицо мое?

Задача настоящей работы в том, чтобы измерить степень подобной ритмизации текста и сопоставить эти измерения с современным «Что делать?» (1863) текстом.

Еще раз подчеркнем, что речь идет не о том предмете, с которым обычно работают стиховеды, мы совсем не касаемся вопроса регулярности чередования ударных и безударных слогов и случайно формируемых отрезков текста, которые можно интерпретировать как силлаботонику («случайные метры в прозе»). Мы анализируем ритмичность, формируемую другими способами, прежде всего, на уровне лексических единиц и конструкций.

Текст романа состоит из приблизительно 153 тыс. слов, это крупная романная форма, хотя «Война и мир» достигает приблизительно 587 тыс. слов. С помощью встроенного в пакет `natasha` для Python сплиттера мы разбили текст романа на предложения, их оказалось около 9200. Хороший сопоставительный (и сопоставимый) материал предоставляет роман «Обломов» (1859): 156 тыс. слов и 12 тыс. предложений.

Для начала пойдем по простому пути и посчитаем количество случаев, когда соседние предложения начинаются с одного и того же слова таким образом, чтобы таких предложений было не менее трех. Именно такие случаи приведены в примерах выше.

Вопреки ожиданиям, число такого рода эффектов в романе невелико, всего 41, при этом не все они претендуют на экспликацию ритма. Так, нижеследующий фрагмент, формально подпадающий под наши критерии, демонстрирует единоначалие на довольно длинном текстовом отрезке, к тому же разорванным абзацем:

«Я передаю слова Дмитрия Сергеича с буквальной точностью, как уже сказал.

Я человек совершенно чужой вам: но корреспонденция, в которую я вступаю с вами, исполняя желание погибшего Дмитрия Сергеича, имеет такой интимный характер, что, вероятно, интересно будет вам узнать, кто этот чуждый вам корреспондент, совершенно посвященный во внутреннюю жизнь погибшего Дмитрия Сергеича. Я отставной медицинский студент — больше ничего не умею сказать вам о себе».

Более того, описанным формальным критериям соответствует сопоставимое число случаев в «Обломове»: 44.

— Что? Что? — вдруг с изумлением спросил Илья Ильич, приподнимаясь с кресел. — Что ты сказал?

Может быть, когда дитя еще едва выговаривало слова, а может быть, еще вовсе не выговаривало, даже не ходило, а только смотрело на все тем пристальным немым детским взглядом, который взрослые называют тупым, оно уж видело и угадывало значение и связь явлений окружающей его сферы, да только не признавалось в этом ни себе, ни другим.

Может быть, Илюша уж давно замечает и понимает, что говорят и делают при нем: как батюшка его, в плисовых панталонах, в коричневой суконной ваточной куртке, день-деньской только и знает, что ходит из угла в угол, заложив руки назад, нюхает табак и сморкается, а матушка переходит от кофе к чаю, от чая к обеду; что родитель и не вздумает никогда поверить, сколько копен скошено или сжато, и взыскать за упущение, а подай-ка ему не скоро носовой платок, он накричит о беспорядках и поставит вверх дном весь дом.

Может быть, детский ум его давно решил, что так, а не иначе следует жить, как живут около него взрослые.

Поищем ритмизацию внутри предложений. Варьируемые элементы у Чернышевского (по крайней мере, в современных изданиях) скорее всего будут разделены знаками препинания.

Значимой представляется такая графическая граница синтагмы, как точка с запятой. В «Что делать?» 43 предложения, в которых части, отделяемые точкой с запятой, начинаются с одного и того же слова, и такие предложения в самом деле несут на себе отчетливые признаки ритмизации.

У одних от скрытности; у других от застенчивости; у третьих от расположения хандрить, задумываться; у четвертых от недостатка симпатии к людям.

Когда она становилась его женою, она становилась его подданною; она должна была трепетать его; он запирал ее; он переставал любить ее.

На этот раз сопоставление с «Обломовым» показывает значимую разницу: подобных предложений в романе Гончарова обнаруживается всего 19:

События его жизни уменьшились до микроскопических размеров, но и с теми событиями не справится он; он не переходит от одного к другому, а перебрасывается ими, как с волны на волну; он не в силах одному противопоставить упругость воли или увлечься разумом вслед за другим.

Опять у него мурашки поползли по сердцу; опять что-то лишнее оказалось там; опять ее ласковый и любопытный взгляд стал жечь его.

Последний пример также отличается сознательной ритмизацией, но очевидно, что такого рода примеров у Гончарова значительно меньше.

Поиск нужных нам структур и подструктур обеспечивается специальной областью знания, которая называется *Sequential pattern mining* [Han J. et al 2001].

Используя концепции из этой области знания и специальное программное обеспечение³, попробуем отыскать повторяющиеся паттерны (именно они и создают ритм в тексте романа). Возьмем в качестве элемента последовательности частеречный тег (например, *NOUN*), но так как частей речи слишком мало и они слишком часто повторяются, ограничим вариации с ними первыми буквами соответствующих слов. Таким образом, предложение «Верочка усиливалась победить в себе отвращение, но не могла» предстанет в виде последовательности:

'вер-PROPN', 'уси-VERB', 'поб-VERB', 'в-ADP', 'себ-PRON',
'отв-NOUN', 'но-CCONJ', 'не-PART', 'мог-VERB'

Чтобы не искать самые частотные паттерны такого рода в романе вообще, а находить их в ограниченном регионе текста, где они ощущаются читателем как повторение и фактор ритмизации, настроим поиск таким образом, чтобы находить паттерны только в пределах 5 предложений. Кроме того, ограничим наш поиск паттернами, состоящими как минимум из трех элементов и имеющих в пределах заявленной области поиска частотность 3. Такие ограничения позволят избавиться от изрядного количества мусора.

Методика поиска позволила выявить очевидно ритмизованные фрагменты в количестве 168 единиц. Так, последовательность

'на-ADP', 'мос-NOUN', 'на-ADP', 'мос-NOUN'

соответствует фрагменту:

Этот удовлетворительный для всех результат особенно прочен был именно потому, что восторжествовали консерваторы: в самом деле, если бы только пошалил выстрелом **на мосту**, то ведь, в сущности, было бы еще сомнительно, дурак ли, или только озорник. Но застрелился **на мосту**, — кто же стреляется **на мосту**? как же это **на мосту**? зачем **на мосту**? глупо **на мосту**! — и потому, несомненно, дурак.



Рис. 1. Распределение ритмизованных фрагментов в тексте романа

Такие фрагменты распределены по тексту романа следующим образом (Рис. 1).

Распределение более-менее постоянное, за исключением отрезка текста от 2000-го до 3000-го предложения. Эта часть приходится на венчание и устройство совместной жизни Веры Павловны с Лопуховым.

Симптоматично, что ритмизованных фрагментов, найденных по той же методике, в «Обломове» обнаруживается на порядок меньше (при большем объеме романа): 64.

К тому же сами фрагменты не читаются как ритмизованные, хотя формально подпадают под паттерн:

Впрочем, **он не** был педант в этом случае и **не** стал бы настаивать на своем; **он** только **не** умел бы начертать в своем уме другой дороги сыну.

Таким образом, нам удалось обнаружить методику поиска последовательностей, которая дает количественно различающиеся результаты для «Что делать?», где ритмизация является частью авторской манеры, и для «Обломова», где она в достаточной степени случайна.

Как мы видим, роман Чернышевского оказывается сложно организованной прозой, нетривиально использующей ритмизацию, и в этом противоречит авторской репрезентации и поверхностным оценкам, сделанным в пылу литературной полемики.

Параметр нестиховедческой ритмизации прозы кажется нам перспективным, так что стоит иметь его в виду при проектировании большой интегральной формально-количественной истории русской литературы, а сама практика поиска паттернов последовательностей должна быть расширена на частные морфологические и просодические характеристики, а также морфемы.

Примечания

¹ Цит. по: [Гаспаров 2006]

² Цитаты даются по изданию: [Чернышевский 2023].

³ PrefixSpan-ru. URL: <https://github.com/chuanconggaio/PrefixSpan-ru>

Литература

Источники

Чернышевский 2023 — Чернышевский Н. Г. Что делать? М.: Азбука, 2023.

Исследования

Арсеньев 2022 — Арсеньев П. Беззащитная диссертация // Чернышевский Н. Г. Эстетические отношения искусства к действительности. М.: V-A-C Press, 2022. С. 5–45

Гаспаров 2006 — Гаспаров М.Л. Семинар А.К. Жолковского - Е.М. Мелетинского: Из истории филологии в Москве 1970-1980-х гг. // Новое лит. обозрение. М., 2006. № 77. С. 113-125.

Гиппиус 1966 — Гиппиус В. В. Чернышевский-стиховед // Гиппиус В. В. От Пушкина до Блока. М., Л.: Наука, 1966. С. 276–294.

Гиришман 1982 — Гиришман М. М. Ритм художественной прозы. М.: Сов. писатель, 1982.

Жолковский 1994 — Жолковский А. К. О пользе вкуса // Золотой век. 1994. № 5. С. 17–22.

Зубков 2022 — Зубков К. Воздвиг ты памятник: Парадоксы авторства и субъектная структура стихотворения Н. А. Некрасова «Баюшки-баю» // Прекраснейшей: сборник памяти Елены Душечкиной. СПб. : Нестор-История, 2022. С. 139-145.

Кавтарадзе 2016 — Кавтарадзе С. Анатомия архитектуры. Семь книг о логике, форме и смысле. М.: Изд. дом Высшей школы экономики, 2016.

Ломова 2015 — Ломова О. В. Роман Н. Г. Чернышевского «Что делать?»: библиографический указатель // Н. Г. Чернышевский. Статьи, исследования и материалы: Сборник научных трудов. Саратов: изд-во «Техно-Декор», 2015. С. 156–195.

Орехов 2022 — Орехов Б. В. Случайные метры в русской прозе XIX века // Вещество поэзии: К 70-летию Юрия Борисовича Орлицкого: Сборник научных статей. – М.: РГГУ, 2022. – С. 24–30.

Орлицкий 2015 — Орлицкий Ю. Эстетическая функция метра в полемике Набокова с Чернышевским // Набоковский сборник 2/2015. СПб.: Музей Набокова, 2015. С. 89–100.

Han J. et al 2001 — . Han, J., Pei, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., & Hsu, M. Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth // Proceedings of the 17th international conference on data engineering. Piscataway, NJ, USA : IEEE, 2001. P. 215–224.

Инна Кижнер

КОЛЛЕКЦИИ КАК ДАННЫЕ: ГОТОВЫ ЛИ МЫ К НАУЧНЫМ ИССЛЕДОВАНИЯМ С ПОМОЩЬЮ ЦИФРОВЫХ КОЛЛЕКЦИЙ КУЛЬТУРНО-ЗНАЧИМЫХ ДАННЫХ?

Задача настоящей работы — показать ограничения, которые существуют на пути к использованию цифровых коллекций культурных учреждений как источника данных. Цифровые коллекции предстают инструментом познания. Они удивительно редко используются как данные. У этого обстоятельства есть ряд обсуждаемых в статье причин. Кроме того, существенно, что объекты в цифровых коллекциях распределены неравномерно и подобраны субъективно. Все эти особенности влияют на результаты научного анализа и их необходимо учитывать в ходе цифрового исследования.

Ключевые слова: цифровые коллекции, данные, цифровое неравенство, музейные коллекции

Введение

Еще недавно цифровые коллекции библиотек, архивов и музеев воспринимались только как инструмент поиска, ускоренного доступа к первичным источникам и чтения с экрана. Недавний перелом в отношении к цифровым коллекциям ведет к переосмыслению роли собраний цифровых копий артефактов (текстов, изображений и аудиофайлов). Представление о цифровых коллекциях как об источнике данных — «коллекции как данные» — делает собрания библиотек музеев и архивов одним из немногих мест, где большие датасеты могут использоваться как эвристический инструмент. При

Инна Александровна Кижнер
Университет Хайфы
inna.kizhner@gmail.com

этом возникает возможность пользоваться коллекциями в рамках разных подходов, используя разные инструменты. Использование коллекций как данных в разных проектах и разными научными коллективами полезно не только для верификации предшествующих результатов, распространения полученного знания и планирования дальнейших исследований, но и для расширения границ применимости инструментов и интерпретации результатов [Borgman 2012, Sköld et al. 2023]. Это выгодно отличает коллекции библиотек, архивов и музеев от закрытых личных или проектных коллекций с ограниченным использованием. Можно ли приблизиться к объективному знанию, используя цифровые коллекции музеев, архивов и библиотек как данные? Какие препятствия существуют на этом пути? Должны ли ограничения, присущие цифровым коллекциям, быть эксплицированы и объяснены исследователю?

Возможность неоднократно обращаться к данным существует и для открытых репозиторий в академических учреждениях, таких как университеты и научно-исследовательские институты. Репозитории научных организаций или проектные репозитории выполняют функцию агрегирования и курирования данных в отдельной научной области, такой как археология или литературоведение. Однако при анализе разных способов классификации, кластеризации и представления исторических объектов и документов имеет смысл рассмотреть библиотеки, музеи и архивы. Такие организации, предназначены для долгосрочного хранения и обработки информации. Они получают постоянное финансирование, а некоторые из них существуют многие десятки и сотни лет. Там работают специально обученные люди, квалификация которых позволяет объединить разные данные в коллекции и разработать способы анализа. Несомненно личные, проектные и институциональные цифровые коллекции обладают дополнительными преимуществами навыков обработки данных в узкой области или высокой квалификации инициатора и организатора репозитория, но они подвержены разным рискам, связанным с прекращением финансирования или сменой интересов организатора.

Задача настоящей работы — показать ограничения, которые существуют на пути к использованию цифровых коллекций культурных учреждений как источника данных. Интересно, что анализ препятствий показывает возможности цифровых коллекций как инструмента познания. При этом анализ данных, полученных при изучении самого предмета, осложняется тем контекстом, который в цифровой коллекции определяет место предмета и документа.

Имеют значение не только обстоятельства поступления предмета в коллекцию, но и обстоятельства, при которых коллекции были сформированы, а предметы и документы классифицированы и каталогизированы. Принципы создания физических и цифровых коллекций определяются политическими, экономическими и социальными причинами, которые влияют на академические традиции и научные интересы коллекционеров.

Цифровые коллекции культурно-значимых данных вошли в нашу жизнь вместе с цифровыми инструментами анализа. Преимущества анализа больших датасетов, такие как возможность обобщений и воспроизводимость исследований, поддерживаются значительно большим количеством примеров чем анализ, основанный на небольшом количестве примеров в случае медленного чтения. Количественный и статистический анализ, который приводит к кластеризации данных, выигрывает от возможности создать стратифицированную выборку. Инструменты познания, ранее привязанные к воображению и риторическим приемам, при наличии больших датасетов и цифровых коллекций становятся инструментами, которые приводят к созданию универсальных теорий, - разумеется, в рамках своих возможностей [Piper 2020].

Возможность воспроизводимых исследований с использованием коллекций как данных часто связывают с возможностью публиковать данные с лицензиями, которые разрешают неограниченное использование (см., например, [Wallace and Euler 2020, Vezina et al. 2022, Candela et al. 2023]). Другие необходимые условия могут включать наличие интерфейса прикладного программирования (API), который обеспечивает возможность переноса большого количества данных на машины научного коллектива и дополнительные возможности анализа. Еще одно важное условие - предоставление ссылок (возможности цитирования и повторного обращения к наборам данных), таких например как DOI [Маслинский 2022, Candela 2023]. Последнее обстоятельство обеспечивает валидацию результатов исследований, поскольку при смене способов публикации или при переходе на другую платформу, датасет все равно будет обнаружен. При этом учреждение, которое публикует данные несет ответственность за предоставление доступа и возможность обнаружения датасета.

Действительно, некоторые из этих условий соблюдаются ограниченным количеством библиотек, архивов и музеев. Некоторые из этих институций обладают значительным влиянием и обширными коллекциями. Для других учреждений важно использовать

открытые лицензии для небольшой части коллекции. В июле 2023 года список библиотек, музеев, архивов и агрегаторов культурно-значимых данных, которые используют открытые лицензии включал около 1650 учреждений из 64 стран мира [Wallace and McCarthy 2023]. Географическое распределение стран отражает (разумеется, с некоторыми исключениями) распределение стран с высокими и средними доходами на душу населения. Ряд музеев и библиотек действительно использует интерфейс прикладного программирования (API) для предоставления доступа к своим данным¹. Подобные датасеты, в частности, используются для обучения алгоритмов компьютерного зрения для классификации и генерации произведений искусства (см., например, [Conde and Kurgutlu 2021, Castellano and Vessio 2021, Rei et al. 2023]). Эти источники данных включают изображения в разных форматах и жанрах и сопровождаются текстовыми метаданными, которые характеризуют период, формат и культуру, в рамках которой создано изображение. Это важно, потому что дает возможность создания обучающей выборки для последующей работы алгоритмов и решения задач классификации и генерации изображений.

Однако несмотря на соблюдение ряда условий некоторыми библиотеками, музеями и архивами, удивляет редкое использование цифровых коллекций культурных учреждений как источника данных. Так например, из тринадцати статей журнала “Journal of Computational Literary Studies” 2022 год только три статьи ссылаются на цифровые коллекции научных коллективов, институций и фондов (Project Gutenberg², Perseus Digital Library³, European Literary Text Collection⁴). Ни одна из статей не ссылается на цифровые коллекции библиотек и архивов. В пятидесяти статьях сборника расширенных тезисов конференции Digital Humanities 2022⁵ только шесть статей упоминают такие коллекции в качестве источника данных. В чем же причины предпочтения, которое ученые оказывают личным и проектным коллекциям. Почему обширные цифровые коллекции, опубликованные с открытыми лицензиями, а иногда и с ключом API, не используются исследователями?

Почему коллекции не используются как данные? Причины создания авторских или проектных корпусов

Среди препятствий, который осложняют использование цифровых коллекций как источника данных, можно выделить следующие обстоятельства:

- Непоследовательность в индексации данных даже тогда, когда в культурном учреждении есть внутренний стандарт метаданных;
- Отсутствие данных, опубликованных в форматах, которые позволяют проводить машинный анализ;
- Неоднозначность и нечеткость культурно-значимых данных, уникальные описания объектов, которые включают точку зрения, характерную для исследователя, который изучал этот объект в прошлом;
- Пропуски в значениях метаданных, для значительной части объектов некоторые поля пусты;
- Вопрос исследования требует добавочного поля или нескольких полей;
- Коллекция музея, архива или библиотеки недостаточно репрезентативна и сбалансирована;
- Коллекция музея, архива или библиотеки не включает художественные произведения того канона, который интересует исследователя.

Отсутствие стандартизованных и нормализованных данных

Отсутствие единого стандарта данных на международном, национальном и даже институциональном уровне приводит к тому, что сложно объединить данные, которые относятся к одному типу, периоду или географической локации. Это обстоятельство является важным препятствием на пути к количественной обработке и обобщениям. В первую очередь неравномерное распределение результатов анализа данных в коллекции может быть вызвано пропусками в данных, которые могут достигать больших значений, например, в поле метаданных «Место издания книги» в цифровой коллекции Библиотеки Эдинбургского университета стоят пропуска в 70% случаев [Navens et al. 2022]. Помимо пропусков в полях метаданных большую проблему составляет разнообразие типов, жанров и форматов гуманитарных данных. Описание и документация такого разнообразия проводятся по-разному в зависимости от той академической традиции, к которой принадлежат сотрудники

библиотек, музеев и архивов. Даже в одном и том же учреждении формат записи даты, географической локации или решения принятые относительно жанра, техники или темы работы могут (и должны) отличаться для значений полей метаданных, а иногда и для полей метаданных в системе управления коллекциями, если система позволяет добавлять новые поля. Совместимость полей и/или значений полей метаданных, а также технологические решения, обеспечивающие совместимость данных цифровых коллекций, обсуждаются в информационных исследованиях, начиная с последних десятилетий 20 века (см., например, [Bearman 1995, Besser 2002, Doerr 2014, Tolfo et al. 2021]). Отдельной задачей совместимости данных является нормализация значений метаданных. Эта задача, в частности, основана на совместимости методов записи дат, географических локаций и персоналий. Например, в археологических коллекциях способы обозначения дат, и тем более записи текстовых выражений, обозначающих приблизительность датировки, значительно варьируются в зависимости от принятого метода или академической традиции [Binding and Tudhope 2023]. Несмотря на многочисленные попытки ввести словари, онтологии и объединенные списки для унифицированной записи культурно-значимых данных, такие как словари Гетти [Harping, 2013] или словари, принятые внутри одного учреждения⁶, а также попытки перевода словарей Гетти на другие языки (см., например, [da Silva 2022]), проблема нормализации данных все еще остается нерешенной даже в академических культурах с широкими традициями обработки гуманитарных данных. Она остается особенно важной в контексте анализа культурно-значимых данных из коллекций Музейного фонда РФ, опубликованных в связи с созданием Государственного каталога Музейного фонда, который на момент написания статьи включает около 40 миллионов учетных записей [Глазунов, Орехов 2020, Костенко, Козлова 2021].

Одна из причин отсутствия стандарта метаданных или неохотного следования стандарту в случае, когда справочники и словари доступны, — неоднозначность и нечеткость культурно-значимых данных. Разнообразие точек зрения на объект или явление, разница в академических традициях и идеологической позиции приводит к тому, что данные цифровых коллекций представляют собой не только разные слои подходов к коллекциям при переходе от физической коллекции к цифровой [Мак 2014], но и разные варианты описаний объектов. При этом варианты описаний могут быть созданы в разные эпохи, разными людьми, при разных социальных обстоя-

тельствах [Parry 2007, Cameron 2010, Gnoli 2011, Zhitomirsky-Geffet 2019]. С одной стороны эффект “пограничного объекта” [Star and Griesemer 1989] осложняет создание инфраструктуры, в которой коллекции могут рассматриваться как данные и/или эвристический инструмент. С другой стороны нечеткость и неопределенность места в классификационной схеме присущи и естественно-научным объектам и тоже определены контекстом, социальными обстоятельствами и академическими традициями. Эти проблемы, вероятно, могут быть решены с помощью развернутых онтологий и новых способов классификации [Bowker 2000]. Использование Wikidata как ресурса, который не связан с языком описания и использует URI для идентификации нужного значения поля метаданных, сталкивается с рядом трудностей, которые связаны с точностью и качеством данных в Wikidata [Zhao 2023]. Часто обсуждаемая неравномерная представленность данных в Wikidata тоже связана с социальным контекстом и условиями создания датасета [Fischer et al. 2023, Zhitomirsky-Geffet and Minster 2023]

Форматы данных

Следующая возможная причина предпочтения, которое исследователи отдают личным или проектным коллекциям, — отсутствие нужных форматов данных на сайтах библиотек, музеев и архивов. Действительно, по разным причинам, в том числе из-за ограничений, связанных с авторским правом, а иногда от того, что тексты в рамках известных канонов лучше известны ученым и лучше документированы, исследователи предпочитают работать с текстами, созданными до второй четверти двадцатого века, то есть с теми текстами, для которых не существует копий, которые легко могут быть обработаны машинным способом. Исключение составляют англоязычные тексты, которые хранятся в собрании Проекта Гутенберг (Project Gutenberg) и представлены, в том числе, в формате Plain Text и тексты, доступ к которым предоставляют некоторые другие проекты и институции. Одним из способов решения этой проблемы является подготовка цифровых академических изданий исследуемых текстов в соответствии со стандартом Text Encoding Initiative⁷ с последующей конвертацией в нужные форматы данных, но такое решение является трудозатратным и дорогостоящим. Еще один способ подготовки текста в нужном формате — работа с одной из платформ, таких как Transcribus⁸, которые дают возможность автоматического распознавания, транскрибирования и

конвертации текста в нужный формат, как для рукописных, так и для печатных текстов [Nockels et al. 2022]. С другой стороны ряд крупных библиотек и музеев с активно работающими исследовательскими лабораториями публикуют некоторую часть данных в форматах пригодных для машинной обработки [Candela 2023]. Однако даже и в этом случае публикуются скорее датасеты, в которых представлены метаданные цифровых коллекций. Публикация текстов в форматах, предназначенных для машинной обработки, является скорее исключением, чем правилом.

Неравномерная представленность данных разного типа в цифровой коллекции

Как известно, любой корпус не свободен от скоса и нерепрезентативного представления данных [Chomsky 1957, Clear 1992, Raineri and Debras 2020]. Во многих случаях это может привести к неправильному представлению о распределении в популяции. Это явление много раз обсуждалось при анализе естественно-научных коллекций данных [Daru et al. 2028, Meineke and Daru 2021, Daru and Rodriguez 2023] и агрегаторов, публикующих коллекции культурно-значимых и естественно-научных данных [Kizhner et al. 2021, Raja et al. 2022].

Неравномерное распределение культурно-значимых данных является в результате анализа результатов скошенной выборки, которая возникает при неравномерной представленности культур разного типа и разных периодов в датасетах или коллекциях. Такую неравномерность несложно оценить, сравнив метаданные для разных географических локаций и для разных периодов в отдельной цифровой коллекции или агрегаторе. Неравномерность представления данных и скос в сторону западных культур и столичных регионов объясняется разными причинами. Чаще всего коллекции культурно-значимых данных являются зеркалом, которое отражает разные периоды в создании инфраструктур, производящих знание (“infrastructures of knowledge-making” [Mak 2014]). Онтологические пропуски возникают при отсутствии знания об объекте (пропуски или нечеткие формулировки в значениях полей метаданных), в то время как эпистемические скосы – результат заполнения полей или переноса текстов из физических каталогов. Предпочтения, которые отдаются формулировкам и номенклатурным терминам в этом случае вызваны политическими и социальными обстоятельствами и академическими традициями. Терминологические предпочтения

или выбор точки зрения при индексации в большой степени определяются социальным влиянием [Bar-Ilan et al. 2010]. Это приводит к тому, что скос в представлении данных появляется в результате субъективных представлений куратора, хранителя или создателя коллекций, вызванных социальными конвенциями эпохи [Ortolja-Baird and Nyhan 2022, Kizhner et al. 2022]. Цифровые коллекции и агрегаторы, которые не сопровождаются указанием на обстоятельства создания данных, вызывают неожиданные эффекты, которые не предполагались создателями коллекций. Такие эффекты возникают в ситуации, когда объект определенного типа оказывается в неожиданных контекстах и приобретает дополнительные значения из-за связи с другими объектами. В случае, когда это делается с помощью алгоритмов доминирующие связи создают основное значение, которое может быть несвойственно этому объекту в других контекстах. Так например, поиск по запросу “Казахстан” в агрегаторе Google Arts and Culture весной 2023 года приводил к получению фотографий, на которых была представлена деятельность космодрома “Байконур” в Казахстане (6 004 из 7,899 изображений или 75% датасета).

Таким образом, социальные предпосылки создания знания во многом определяют ограничения цифровых коллекций как источников данных. В этом контексте очень важно понимать, “кто финансирует проекты оцифровки, какие критерии отбора применяются, как документируются предметы и тексты, какие классификационные схемы используются” [Zaagsma 2022]. Мы не можем обеспечить равномерное распределение данных в цифровой коллекции, но мы можем отнестись к цифровым коллекциям как к первичным источникам [Liu, 2017]. Эпистемическая и научная ценность коллекций как данных — следствие последовательной, долгой и высококвалифицированной работы, основанной на анализе предшествующих практик [Leonelli 2016, Wyatt 2022], в том числе текстологическом и библиографическом анализе, исследованиях в области истории книги, сравнительных медийных исследованиях и критическом анализе цифровой инфраструктуры [Vode 2020b]. Таким образом, возникает необходимость явным образом представить обстоятельства, при которых формировались физические и цифровые коллекции и объяснить “характер и происхождение онтологических пропусков и эпистемических скосов в данных” [Vode 2020], *перевод мой*). Первым шагом на этом пути может быть явное сообщение о контексте, истории и ограничениях создания цифровой коллекции, которое сопровождает публикацию цифровой

коллекции [Houswedell et al. 2020, Vode 2020a, Zaagsma 2022]. Более подробным образом обстоятельства создания и документирования коллекции могут быть зафиксированы в параданых. Процесс создания параданых — важного типа информации об условиях и контексте создания данных — влечет за собой проблемы, похожие на те, с которыми сталкиваются при разработке метаданных. Разнообразие практик и типов данных в гуманитарных исследованиях приводит к сложностям разработки стандартов и принятия решений относительно того, в какой степени и с какой точностью должны быть зафиксированы такие данные. Как и многие другие данные о данных, параданые никогда не будут зафиксированы полностью [Huvila 2022].

Скосы в цифровых коллекциях и источники неравномерного распределения данных

Основные скосы в цифровых коллекциях связаны с преобладанием культур стран с высоким и средним уровнем дохода, то есть географическим смещением в результатах анализа. Следующий скос связан с преобладанием данных, созданных в девятнадцатом и двадцатом веке, то есть со смещением во времени (см., например, [Zhitomirsky-Geffet and Minster 2023]). Скос, связанный с предпочтениями коллекционеров, руководителей экспедиций, кураторов и хранителей является следующим заметным фактором, влияющим на результаты анализа. Это приводит к тому, что предметы “неслучайным образом распределены в пространстве и времени, представляя смещенные (biased) наборы данных истинного распределения” [Meineke and Daru 2021], перевод мой).

Можно выделить несколько уровней создания источников неравномерности представления данных:

- Физические коллекции (политические, социальные, экономические ограничения, предпочтения коллекционеров и руководителей экспедиций)
- Метаданные физических коллекций (степень изученности коллекций, степень неопределенности при интерпретации музейных предметов)
- Данные о выставках, опубликованные в физических каталогах (политические ограничения, повестка научных исследований, предпочтения куратора)

- Текстовые описания в физических каталогах (повестка научных исследований, академические традиции)
- Оцифрованные предметы (политические, экономические и технические ограничения)
- Опубликованные цифровые коллекции, метаданные цифровых коллекций (субъективность куратора, экономические ограничения, социальные ограничения, такие как наличие квалифицированных сотрудников)

При этом организация данных в датасеты, вследствие процедур, созданных алгоритмическим образом на основе общности некоторых полей метаданных или значений полей метаданных имеет большое значение, учитывая все большее количество датасетов, созданных с участием алгоритмов. Алгоритмы представления данных разных типов и выравнивания исходного распределения могут играть важную роль в искажении истинного распределения, создавая следующий уровень смещения. Например, данные из локаций с малой представленностью оказываются в результатах поиска в Google Arts and Culture и создают впечатление разнообразия культур в датасете, несмотря на малое количество объектов из этих локаций.

Способы оценки неравномерного распределения

Оценить неравномерное распределение связанное с разными форматами представления данных, вариантами записи значений метаданных и разной степенью глубины и точности отображения для каждого типа данных является сложной задачей. Такие исследования только ищут подходы к решению проблемы нормализации метаданных и обнаружения соответствий между данными разного типа. Решение таких задач осложняется тем, что классификационные схемы принятые при поиске соответствия между данными тоже не предлагают единственное возможное решение. Даже в рамках узкой исследовательской области может существовать несколько возможных классификационных схем (например, при датировке керамики в археологических исследованиях). В ситуации, когда надо принять решение об обнаружении соответствия между разными типами объектов из разных культур, эта задача становится еще сложнее.

Особенно трудно оценить уровень искажения в распределении объектов цифровой коллекции по сравнению с исходным уровнем

распределения в популяции артефактов. Поскольку истинный уровень распределения неизвестен, остается полагаться на косвенные способы определения истинного уровня, подобные тем, которые используются, например, в археологии (см., например, [Drennan et al. 2015, Palmisano et al., 2017, Palmisano et al. 2021]). Для обеспечения валидности результатов могут использовать сочетание нескольких косвенных способов определения истинного уровня [Palmisano et al. 2017]. Сложности и проблемы использования косвенных методов могут быть связаны с нечетким/переменчивым соотношением между косвенным показателем и показателями исходного распределения, например, в наблюдениях за динамикой популяции [Bevan and Crema 2021]. Сами данные, собранные для анализа косвенных показателей могут демонстрировать неравномерное распределение и, таким образом опять приведут к скосу в распределении результатов.

Еще один вариант определить истинное распределение для идентификации скоса в цифровой коллекции — сравнить данные цифровой коллекции с агрегированным мнением экспертов для определения распределения, например, по периодам или для пространственного распределения [Daru and Rodriguez 2023].

При отсутствии данных об исходном распределении можно определить скос, используя данные о распределении в физической коллекции объектов [Kizhner et al. 2019, Kizhner et al. 2021]. Такой подход покажет распределение, связанное с производством артефактов, которые были включены в коллекции физических музеев, архивов и библиотек по разным причинам. Эти институции могли, например, выполнять роль центров, которые фиксировали факт существования небольшого поселения или документировали те виды деятельности, которые в нем происходили [Kelly, 2006]. С другой стороны этот подход не покажет распределение, основанное на восприятии канонических произведений, которое можно оценить, например, используя сравнение распределения в цифровой коллекции с распределением в списках канонических произведений художественных текстов, визуальных произведений и музыкальных произведений. Нечеткость и субъективность подобных списков делают задачу поиска распределения и скоса для такого типа распределения сложной проблемой.

Заключение

В настоящий момент нам доступны датасеты, в которых опубликованы самые частые типы объектов с самыми подробными атрибутами из самых заметных коллекций. Применение алгоритмов, выделяющих доминирующие типы данных, которые показывают самые частые объекты для пространственного распределения, распределения во времени и других типов распределения (например, по жанрам), приведет нас в мир унифицированных культур, где культурное разнообразие сведено к тем типам культур, которые рекомендованы алгоритмами.

Таким образом, можно утверждать, что объекты в цифровых коллекциях распределены неравномерно и подобраны субъективно. Это связано с отбором предметов в физические коллекции, часто под влиянием предпочтений сотрудников или под влиянием обстоятельств, а также субъективным отбором при оцифровке и публикации. Неравномерное распределение предметов связано с национальными идеологиями и повестками, традициями учреждений, принципами комплектования коллекций в конкретном учреждении, академическими традициями и разнообразием точек зрения. Разные уровни или слои неравномерности (неравномерность при оцифровке, неравномерность при заполнении полей метаданных, неравномерность при выборе точек зрения при индексации объектов) дополнительно осложняют анализ. Онтологические пропуски редко объясняются явным образом, и мы не знаем с какими пропусками нам придется иметь дело, когда собираем данные для исследования. Будущие исследования, по-видимому, будут развиваться в сторону определения популяции объектов (baseline), оценки неравномерности (bias), определения онтологических и эпистемических пропусков в данных, выяснения политических, социальных и экономических причин пропусков, определения стадий и этапов создания неравномерности, предложения решений для смягчения неравномерности (mitigating bias).

Адаптируя вопросы, предложенные относительно естественнонаучных цифровых коллекций [Meineke and Daru 2021], можно предложить следующие направления для дальнейших исследований:

- Как объекты с разнообразной семантикой и многочисленными смыслами могут быть преобразованы в ‘открытые наборы данных’?

- В какой мере объекты и метаданные (учетные записи) отражают реальное знание истории культуры и изменений в культуре во времени и пространстве?
- Действительно ли политические, экономические и социальные обстоятельства вводят неравномерность (bias), которая свойственна конкретным коллекциям, регионам, периодам времени и типам объектов?
- Если на формирование коллекций влияют политические обстоятельства и финансирование, то представляют ли коллекции те каноны, которые приняты политиками и финансирующими организациями? Если да, то как слои канонов представлены в наборах оцифрованных культурно-значимых данных?
- На какие периоды времени, регионы и типы объектов должны быть направлены усилия по формированию коллекций и оцифровке коллекций, чтобы смягчить неравномерность?

Даже если использование цифровых коллекций как источников данных не приведет к возможности делать широкие обобщения в ближайшее время, дальнейшие улучшения, приведут к необходимости интерпретировать результаты анализа с учетом обстоятельств создания коллекций. Возможно, в этом случае мы сможем получать более надежные результаты, а не только воспроизводимые результаты.

Примечания

¹ См., например, <http://museum-api.pbworks.com/w/page/21933420/Museum%C2%A0APIs>

² <https://www.gutenberg.org/>

³ <http://www.perseus.tufts.edu/hopper/>

⁴ <https://www.distant-reading.net/eltec/>

⁵ <https://dh2022.adho.org/>

⁶ см., например, <https://data.rijksmuseum.nl/controlled-vocabularies/download/>

⁷ <https://teibyexample.org/exist/>

⁸ <https://readcoop.eu/transkribus/>

Литература

Исследования

Глазунов, Орехов 2020 — Глазунов, Е. В., Орехов Б. В. (2020). Унификация данных Музейного Госкаталога РФ. Сибирский антропологический журнал, 4 (3), 154-168.

Костенко & Козлова 2021 — Костенко, В. В., & Козлова, А. С. (2021). Госкаталог музейного фонда России: первый подход к прикладному анализу данных. Скиф. Вопросы студенческой науки, (9 (61)), 34-38.

Маслинский 2022 — Маслинский К. “О культуре работы с данными в ДН, или роль Репозитория открытых данных”, Семинар DHRI, DHRI, Сибирский федеральный университет, Красноярск, 16.02.2022. <https://www.youtube.com/watch?v=18BUQBh2P5E>

Bar-Ilan et al. 2010 — Bar-Ilan, J., Zhitomirsky-Geffet, M., Miller, Y., & Shoham, S. (2010). The effects of background information and social interaction on image tagging. *Journal of the American Society for Information Science and Technology*, 61 (5), 940-951.

Besser 2002 — Besser, Howard. "Moving from isolated digital collections to interoperable digital libraries.-First Monday 7.6 (2002).

Bevan & Crema 2021 — Bevan, A., & Crema, E. R. (2021). Modifiable reporting unit problems and time series of long-term human activity. *Philosophical Transactions of the Royal Society B*, 376 (1816), 20190726.

Binding et al. 2023 — Binding, Ceri, and Douglas Tudhope. "Automatic Normalization of Temporal Expressions.-Journal of Computer Applications in Archaeology 6.1 (2023).

Bode 2020a — Bode, K. (2020a). Why you can't model away bias, *Modern Language Quarterly*, 81: 1.

Bode 2020b — Bode, K. (2020b). The Archive, in *The Cambridge Companion to Literature in the Digital Age* (Ed. Adam Hammond). Cambridge: Cambridge University Press.

Borgman 2012 — Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059–1078.

Castellano et al. 2021 — Castellano, Giovanna, and Gennaro Vessio. "Deep learning approaches to pattern extraction and recognition in paintings and drawings: An overview.-Neural Computing and Applications 33.19 (2021): 12263-12282.

Candela 2023 — Candela, Gustavo. "Towards a semantic approach in GLAM Labs: The case of the Data Foundry at the National Library of Scotland.-Journal of Information Science (2023).

- Candela et al. 2023* — Candela, Gustavo, et al. "A Checklist to Publish Collections as Data in GLAM Institutions.-arXiv preprint arXiv:2304.02603 (2023).
- Chomsky 1957* — Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Clear 1992* — Clear, J. (1992). Corpus sampling. In Leitner, G. (ed.) *New Directions in English Language Corpora*. Berlin: Mouton-de-Gruyter, pp. 21–31.
- Conde et al. 2021* — Conde, Marcos V., and Kerem Turgutlu. "CLIP-Art: Contrastive pre-training for fine-grained art classification." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.
- da Silva 2022* — da Silva, Camila. "The Ongoing Translation of the Getty Art & Architecture Thesaurus® into Portuguese: An Art Information Access and Retrieval Tool for Cultural Institutions in Portuguese-Speaking Countries.-*Getty Research Journal* 16.1 (2022): 209-225.
- Daru et al. 2018* — Daru, B. H., Park, D. S., Primack, R. B., Willis, C. G., Barrington, D. S., Whitfield, T. J., ... & Davis, C. C. (2018). Widespread sampling biases in herbaria revealed from large-scale digitization. *New Phytologist*, 217 (2), 939-955.
- Daru & Rodriguez 2023* — Daru, B. H., & Rodriguez, J. (2023). Mass production of unvouchered records fails to represent global biodiversity patterns. *Nature Ecology & Evolution*, 1-16.
- Doerr et al. 2014* — Doerr, Martin, et al. "Realizing lessons of the last 20 years: A manifesto for data provisioning & aggregation services for the digital humanities (a position paper).-*D-lib magazine* 20.7/8 (2014).
- Drennan et al. 2015* — Drennan, Robert D., Berrey, C. Adam, and Christian E. Peterson. *Regional settlement demography in archaeology*. ISD LLC, 2015.
- Fischer et al. 2023* — Fischer, F., Blakesley, J., Wojcik, P., & Jäschke, R. (2023). Preface: World Literature in an Expanding Digital Space. *Journal of Cultural Analytics*, 8 (2).
- Gnoli 2011* — Gnoli, Claudio. "Animals belonging to the emperor: enabling viewpoint warrant in classification.-*Subject access: Preparing for the future* 42 (2011): 91.
- Harpring 2013* — Harpring, Patricia. *Introduction to controlled vocabularies: terminology for art, architecture, and other cultural works*. Getty Publications, 2013.
- Hauswedell et al.* — Hauswedell, T., Nyhan J., Beals M., Terras, M., Bell E. (2020). Of global reach yet of situated contexts: an examination of the implicit and explicit selection criteria that shape digital archives of historical newspapers. *Archival Science*, 20: 139–65.

- Havens 2022* — Havens, Lucy, et al. "The Library Catalogue as Dataset: Exploring Data Science Approaches to Analyse Collections at Scale." (2022).
- Huvila 2022* — Huvila, Isto. "Improving the usefulness of research data with better paradata.-Open Information Science 6.1 (2022): 28-48.
- Kelly 2006* — Kelly, L. (2006). Measuring the impact of museums on their communities: The role of the 21st century museum. *Intercom*, 2 (4).
- Kizhner et al. 2021* — Kizhner, I., Terras, M., Rumyantsev, M., Khokhlova, V., Demeshkova, E., Rudov, I., & Afanasieva, J. (2021). Digital cultural colonialism: measuring bias in aggregated digitized content held in Google Arts and Culture. *Digital Scholarship in the Humanities*, 36(3), 607-640.
- Kizhner et al. 2022* — Kizhner, Inna, et al. "The Culture of the Very Rich and Very Poor: Do Digital Museum Collections Tell us Anything about Jewish Culture.-Studies in Digital History and Hermeneutics 43 (2022).
- Liu 2017* — Liu, A. *Toward Critical Infrastructure Studies: Digital Humanities, New Media Studies, and the Culture of Infrastructure* / A. Liu. — University of Connecticut, 23 February, 2017
- Mak 2014* — Mak, B. (2014). Archaeology of a digitization. *Journal of the Association for Information Science and Technology*, 65(8): 1515–26.
- Meineke & Daru 2021* — Meineke, E. K., & Daru, B. H. (2021). Bias assessments to expand research harnessing biological collections. *Trends in Ecology & Evolution*, 36 (12), 1071-1082.
- Nockels et al. 2022* — Nockels, J., Gooding, P., Ames, S., Terras, M. (2022). Understanding the application of handwritten text recognition technology in heritage contexts: a systematic review of Transkribus in published research. *Archival Science*, 22 (3), 367-392.
- Ortolja-Baird & Julianne 2022* — Ortolja-Baird, Alexandra, and Julianne Nyhan. "Encoding the haunting of an object catalogue: on the potential of digital technologies to perpetuate or subvert the silence and bias of the early-modern archive.-Digital Scholarship in the Humanities 37.3 (2022): 844-867.
- Palmisano et al. 2017* — Palmisano, A., Bevan, A., & Shennan, S. (2017). Comparing archaeological proxies for long-term population patterns: An example from central Italy. *Journal of Archaeological Science*, 87, 59-72.
- Palmisano et al. 2021* — Palmisano, A., Bevan, A., Kabelindde, A. et al. Long-Term Demographic Trends in Prehistoric Italy: Climate Impacts and Regionalised Socio-Ecological Trajectories. *J World Prehist* **34**, 381–432 (2021).
- Raineri & Debras 2020* — Raineri, S. and Debras, C. (2020). Corpora and representativeness: where to go from now? *CogniTextes*. 2019, n.p. <https://journals.openedition.org/cognitextes/1671>
- Raja et al. 2022* — Raja, N. B., Dunne, E. M., Matiwane, A., Khan, T. M., Nätischer, P. S., Ghilardi, A. M., & Chattopadhyay, D. (2022). Colonial history

and global economics distort our understanding of deep-time biodiversity. *Nature ecology & evolution*, 6 (2), 145-154.

Rei 2023 — Rei, Luis, et al. "Multimodal metadata assignment for cultural heritage artifacts.-*Multimedia Systems* 29.2 (2023): 847-869.

Sköld et al. 2023 — Sköld, O., Kaiser, J., Andersson, L., Huvila, I., & Liu, Y. H. (2023). Facilitating Data Re-use by Better Understanding Paradata, *DHNB2023*, Book of Abstracts, pp. 48-51.

Star & Griesemer 1989 — Star, Susan Leigh, and James R. Griesemer. "Institutional ecology, translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39.-*Social studies of science* 19.3 (1989): 387-420.

Tolfo et al. 2021 — Tolfo, Giorgia, et al. "Hunting for Treasure: Living with Machines and the British Library Newspaper Collection."-*Studies in Digital History and Hermeneutics* (2021): 25.

Vezina et al. 2022 — Vezina, Brigitte, et al. "Towards Better Sharing of Cultural Heritage—An Agenda for Copyright Reform: A Creative Commons Policy Paper."(2022).

Wallace & Euler 2020 — Wallace, A., Euler, E. Revisiting Access to Cultural Heritage in the Public Domain: EU and International Developments. *IIC* 51, 823–855 (2020).

Wallace & McCarthy — Wallace, A., McCarthy, D. Survey of GLAM open access policy and practice, 2018-2023, https://docs.google.com/spreadsheets/d/1WPS-KJptUJ-o8SXtg00llcxq0IKJu8eO6Ege_GrLaNc/edit#gid=1409426267

Wyatt 2022 — Wyatt, S. (2022). Critical (big) data studies. In *The Necessity of Critique: Andrew Feenberg and the Philosophy of Technology* (pp. 127-142). Cham: Springer International Publishing.

Zhao 2023 — Zhao, Fudie. "A systematic review of Wikidata in Digital Humanities projects.-*Digital Scholarship in the Humanities* 38.2 (2023): 852-874.

Zhitomirsky-Geffet 2019 — Zhitomirsky-Geffet, Maayan. "Towards a diversified knowledge organization system: An open network of inter-linked subsystems with multiple validity scopes.-*Journal of Documentation* 75.5 (2019): 1124-1138.

Zhitomirsky-Geffet 2023 — Zhitomirsky-Geffet, Maayan, and Sara Minster. "Cultural information bubbles: A new approach for automatic ethical evaluation of digital artwork collections based on Wikidata.-*Digital Scholarship in the Humanities* 38.2 (2023): 891-911.

ДИСКУССИЯ

Борис Орехов, Андрей Володин

DIGITAL HUMANITIES В РОССИИ И КОНЕЦ ИСТОРИИ

В статье в полемике с текстом Д. Скоринкина рисуется история отечественных цифровых исследований в литературоведении и исторической науке. Демонстрируется, что традиция применения точных методов глубока, не заканчивается и не может внезапно закончиться за пару лет. История применения цифровых методов в гуманитарном поле предстает как постепенное проникновение в эту область компьютерных технологий, которые, однажды закрепившись, остаются в этой сфере навсегда.

Ключевые слова: история digital humanities, digital humanities в России, цифровая история, цифровое литературоведение

В журнале *Canadian-American Slavic Studies* 57 (2023) вышла важная для русскоязычного цифрового гуманитарного сообщества статья Даниила Скоринкина «Digital Humanities in Russia Was Forever, Until It Was No More. The Story of Russian Digital Humanities in 2011–2022» («Цифровые гуманитарные науки в России были навсегда, пока не кончились. История цифровых гуманитарных наук в России в 2011–2022 годах») [Skorinkin 2023].

Автор представляет свою версию становления цифровых гуманитарных исследований в России последнего десятилетия, и это вторая после коллективной работы [Kizhner et al. 2022] попытка обобщить исторический опыт отечественного предметного поля. Примечательно, что оба текста опубликованы по-английски, то есть до некоторой степени представляют собой «историю на экспорт».

Андрей Юрьевич Володин

МГУ им. Ломоносова; Сибирский федеральный университет
volodin@hist.msu.ru

У нашего текста несколько задач, одна из которых — перенести дискуссию в русскоязычное пространство или хотя бы создать прецедент русскоязычного автоописания истории цифровых гуманитарных наук. Дополнительная мотивация настоящего отклика в том, чтобы предложить альтернативный способ упорядочивания исторического материала и скорректировать некоторые тезисы Д. А. Скоринкина, которые показались нам не вполне точными.

Сразу оговоримся: несмотря на скорее полемическую тональность нашего отклика, статья Д. А. Скоринкина представляется нам эмоционально цельным, интеллектуально и по-человечески честным высказыванием, заслуживающим внимания ученого читателя. Хотя в ней не всегда твердо можно отделить строгий научный стиль от облеченной в эссеистическую форму фиксации личных впечатлений, статья Д. А. Скоринкина претендует на то, чтобы в будущем занять важное место в своде источников, характеризующих современную нам историю гуманитарной науки. Личный опыт участника событий в таком контексте только добавляет тексту ценности, понятой в духе антропоцентрической *oral history*.

Статья разбита на четыре тематических блока (цифровые литературоведение, историю, культурологию, публичную репрезентацию) и пятый раздел, в котором подводятся неутешительный итог цифровым гуманитарным исследованиям в России, провоцируя воспринимать их как завершенный и завершившийся отрезок истории отечественной науки. К той же идее подводит и заглавие всей статьи.

Первый и самый объемный (три с половиной страницы) раздел посвящен цифровому литературоведению. У Д. А. Скоринкина его история выстраивается вокруг институций и больших академических проектов. Ключевыми новейшими событиями оказываются запуск Национального корпуса русского языка (НКРЯ) и поэтического подкорпуса в его экосистеме, создание Центра цифровых гуманитарных исследований в НИУ «Высшая школа экономики», проведенные Центром летние школы и выставка, участие российских исследователей в международных тематических конференциях, проекты Tolstoy.Digital и Dracor, открытие лаборатории цифровых исследований в Институте русской литературы РАН и Репозитория открытых данных под его эгидой. В связи с этими инициативами называются ключевые действующие лица заявленной истории: Е. В. Рахилина, А. А. Бонч-Осмоловская, Б. В. Орехов, Р. Г. Лейбов, К. А. Маслинский. Очевидно, скромность не позволила упомянуть в этом ряду и еще одно важное имя, имя автора

статьи — Д. А. Скоринкина, принимавшего самое активное участие в большей части перечисленного.

Такой взгляд на историю имеет право на жизнь, но некоторые позиции в нарисованной картине все же представляются нам проблематичными как с концептуальной, так и с фактической точки зрения.

Прежде всего, история и науки в целом, и отдельного поля может быть написана не только как история институций и проектов (упомянутые Д. А. Скоринкиным проекты настолько большие, что сами приобретают признаки институционализации), но как история исследований и идей. История институций, как и история дисциплинарного оформления различных областей научного знания [Дмитриев 2015], часто оказывается в сложных отношениях с историей конкретных персоналий и их идей [Алпатов 2023], потому что теории, методы и подходы противопоставляются человеческим характерам, личным качествами и пристрастиям. Можно заметить любопытную тенденцию: чем больше мы хронологически отдаляемся от ученого, тем меньше нас интересует, где и как он работал, а больше интересует то, что было им сформулировано.

Более того, внимательный взгляд обнаруживает в этом смысле некоторую непоследовательность и у самого автора статьи. Первые строки раздела возводят истоки цифрового литературоведения к 1910-м годам, за чем, очевидно, кроются стиховедческие подсчеты Андрея Белого, опубликованные в его книге «Символизм», а также, возможно, работы Н. А. Морозова на тему количественной атрибуции. Далее среди предшественников, обеспечивших историческую глубину применения количественных методов в отечественном литературоведении, называются Б. И. Ярхо, Р. О. Якобсон, М. Л. Гаспаров, Ю. М. Лотман. Существенно, что имена этих исследователей представлены в оптике Д. А. Скоринкина отдельно от институций, сообществ и иных объединяющих инстанций. В нарисованной ретроспективе они получили свое место именно благодаря своим научным работам и высказанным в них идеям, а не из-за принадлежности к определенным кафедрам или даже кружкам.

Вероятно, было бы полезно рассмотреть поле цифрового литературоведения как совокупность циркулирующих идей и методов.

Не претендуя на исчерпывающее описание и имея в виду предварительный, сугубо черновой характер такой попытки, мы представим свою версию новейшей истории этой области.

Надо сказать, что три направления количественных исследований на материале художественных текстов в отечественном лите-

ратуроведении имели свою историю и традиции задолго до описываемого Д. А. Скоринкиным периода. Речь идет о стилеметрии, авторской лексикографии и стиховедении. Именно стиховедческие прозрения Андрея Белого положили начало цифровой гуманитарной науке в России, а две группы работ, принадлежащих А. Н. Колмогорову и М. Л. Гаспарову, помогли ей приобрести окончательное оформление.

Ряд стиховедческих статей математика А. Н. Колмогорова, собранных не так давно в единую книгу [Колмогоров 2016], привили изучающим стих необходимую количественную выучку и строгость, а М. Л. Гаспаров в своих многочисленных работах, из которых самой важной стала обобщающая монография «Очерк истории русского стиха» [Гаспаров 1984], разработал сравнительно стройную систему понятий, соотносимую с традиционным литературоведческим знанием.

Несмотря на очевидность в этой истории вершинных авторов и сочинений, само поле стиховедения в кратком экскурсе обозреть затруднительно. Свои значительные работы в этой области создало несколько поколений выдающихся исследователей, любое перечисление которых будет случайным и неполным. Так как без списка обойтись нельзя, назовем имена Б. В. Томашевского, Г. А. Шенгели, К. Ф. Тарановского, В. Е. Холшевникова, П. А. Руднева, М. А. Красноперовой, Ю. Б. Орлицкого, М. И. Шапира, публиковавшихся в течение 1920—2000-х годов, но еще раз подчеркнем, что в зависимости от принадлежности к той или иной научной школе этот перечень может быть как расширен так и сужен.

В 2010-х годах целый ряд исследователей, использующих количественные методы, продолжает публиковать заметные работы в стиховедческом поле. Контекст этих публикаций и высказанных в них идей слишком сложен для краткой содержательной характеристики, поскольку соотносится с десятилетиями предшествующего научного поиска. Но если все же попытаться нащупать основные силовые линии, то это уточнение созданной Гаспаровым терминологической системы и case studies метрических особенностей отдельных поэтов и поэтических направлений. В 2010-х активно публикуются С. Е. Ляпин (ритмика четырехстопного ямба) [Larin 2020], В. А. Плунгян (классификация неклассического стиха) [Плунгян 2005], Ю. Б. Орлицкий (история верлибра), Е. В. Казарцев (ритмика четырехстопного ямба, случайные метры в прозе), Т. В. Скулачева (особенности синтаксиса в стихотворном тексте), В. В. Файер (количественная характеристика античного

гекзаметра), А. С. Белоусова (поэтика твердых форм) [Belousova & Páramo Rueda 2019], В. С. Полилова [Полилова 2017], К. М. Корчагин (история цезуры, особенности неклассического стиха).

Эти имена в очерке Д. А. Скоринкина отсутствуют, и, если отнестись к тексту внимательно, становится понятно, почему. Основным различием для Д. А. Скоринкина оказывается не применение количественных методов, а активное использование компьютерных технологий. Действительно, сама традиция квантитативного стиховедения не прерывалась с 1960-х, но парадоксальным образом ее наследники оказались традиционалистами, и до самого последнего времени производили подсчеты с помощью ручки и бумаги. Работы соавторов М. Л. Гаспарова [Гаспаров & Исаков 2003], М. А. Красноперовой и А. В. Козьмина [Козьмин 2006] оказались исключениями. В новейшее время известно несколько теоретических публикаций Пильщиков-Старостин [Пильщиков & Старостин 2010], но реальное прикладное значение в стиховедческих исследованиях компьютерные методы стали приобретать только в последние годы: Короткова для русского стиха, Файер для древнегреческого, Орехов для тюркского (башкирского, киргизского, узбекского).

Стилеметрия, а точнее ее подмножество — количественная атрибуция — также имеет сильную отечественную традицию. В 1910-х годах, уже после квантитативных опытов А. Белого, выходят работы Н. А. Морозова [Морозов 1916], указывающие на распределение служебных слов как на значимый признак при установлении авторства текста. Впоследствии высказанные им идеи были подхвачены В. П. Фоменко и Т. Г. Фоменко, обобщившими их в понятие авторского инварианта. Работа соавторов Фоменко стала особенно популярной благодаря концепции Новой хронологии, развиваемой их сыном, академиком А. Т. Фоменко, но все еще основывалась на ручных подсчетах. В рамках этой традиционалистской парадигмы производились поиски других факторов репрезентации стиля в художественном тексте, см., например, работы В. С. Андреева [Андреев 2021].

Подводящей промежуточные итоги в истории отечественной стилиметрии нужно признать книгу Г. Я. Мартыненко «Методы математической лингвистики в стилистических исследованиях» [Мартыненко 2019], в тематическом плане выходящую за рамки вопросов атрибуции, захватывающую в том числе и стиховедение.

Но и компьютер в этой области тоже применялся, ярким примером этому служит книга об атрибуции 2001-го года [Марусенко и

др. 2001]. Едва ли не более важным, чем сама эта монография, для истории digital humanities в России текстом является рецензия на нее А. Ю. Балакина [Балакин 2003], высвечивающая проблемные точки дискуссии между традиционными филологами и количественниками. О весомости этой рецензии говорит и последовавший полемический ответ автора монографии [Марусенко 2004].

Наиболее значимым событием в отечественной компьютерной стилиметрии 2010-х годов стало заимствование уже до этого широко применявшегося в зарубежных исследованиях метода Delta. Первенство в этом смысле принадлежит Д. А. Скоринкину и А. А. Бонч-Осмоловской [Скоринкин & Бонч-Осмоловская 2016]. Эта пионерская работа открыла дорогу новым изысканиям, как связанным с атрибуцией [Великанова & Орехов 2019], так и ставящим альтернативные теоретические вопросы в рамках переводоведения [Орехов 2020; Орехов 2021; Ковалев 2024]. В чем-то сходную собственную методику параллельно разрабатывают М. Ю. Михеев и Л. И. Эрлих [Михеев & Эрлих 2018].

Наконец, следует упомянуть и о таком сегменте авторской лексикографии, как частотные словари, которые активно составлялись отечественными учеными и в течение XX века (см., например, «Частотный словарь романа Л. Н. Толстого “Война и мир”» [Частотный 1978]), и продолжали составляться в 2010-х, несмотря на появление Национального корпуса русского языка. Особое место здесь занимает последовательная работа Н. Л. Васильева [Васильев 2013а; Васильев 2013б; Васильев 2016]. Сама идея, что количественное распределение лексики в текстах значимо и поддается интерпретации, лежит в основе и работы лексикографов, и цифровых литературоведов. Близка к этой области и количественная грамматика Ф. Н. Двинятина [Двинятин 2015], которая, однако, ограничивается сделанными в позитивистском духе наблюдениями, не получающими развития в форме выводов, соотносимых с гуманитарным знанием.

Прототипической для количественного литературоведения является методология Б. И. Ярхо, которая, в отличие от названных выше направлений, не нашла прямых последователей до публикации его трудов усилиями М. Л. Гаспарова, М. И. Шапира и М. В. Акимовой. Этот подход предполагает нахождение значимых для поэтики параметров, и подсчет их распределения в тексте. Такая исследовательская программа действительно может иметь серьезные перспективы только с развитием средств автоматиче-

ской обработки текста на русском языке и освоением этих средств гуманитариями. Эти процессы пришлось на 2010-е.

Тем не менее, и у компьютерного литературоведения в России есть своя история, которая начинается до описываемых Д. А. Скоринкиным событий. Прежде всего, упоминания требует книга [Бавевский 2001], где в компьютерной части описывается справочная система о жизни и творчестве Б. Л. Пастернака. Свои компьютерные разработки в то же время были и у фольклористов [Архипова и др. 2002].

В рамках этого поля в 2010-х вышло несколько заметных работ, транслирующих значимые для цифрового литературоведения идеи.

В 2013 году А. А. Бонч-Осмоловская и Б. В. Орехов своей статьей [Бонч-Осмоловская & Орехов 2013] открыли для количественного изучения стихотворный материал, размещаемый авторами на сайте stih.ru. Его особенность в том, что он представляет собой по-настоящему большие данные, что является редкостью для коллекций художественных произведений. Речь идет о миллионах стихотворений. Кроме того, в распоряжении исследователя оказывается дата публикации стихотворения, дающая простор для сопоставления текста и внетекстовой реальности.

Это исследование было представлено на ежегодной конференции Digital Humanities в Лозанне в 2014 году, и с опорой на него в дальнейшем был опубликован целый ряд научных трудов, развивающих применение количественных методов к обозначенному набору текстов: помещение их в корпусный контекст и извлечение акцентологических данных [Гришина и др. 2015], поиск новостного субстрата [Панасюк 2022], тематическое моделирование [Лейбов & Орехов 2022], семантический ореол метра [Бутенко 2022], уточнение границ наивной и профессиональной поэзии [Орехов 2022].

Важной работой, использующей теоретико-лингвистическое знание в анализе художественного текста, стала статья А. А. Бонч-Осмоловской и Д. А. Скоринкиной, в которой понятие семантической роли рассматривается в контексте создания портрета персонажа [Skorinkin & Bonch-Osmolovskaya 2017]. На внушительной выборке романа «Война и мир» показано, как языковые средства становятся материалом для художественного приема. Та же связь языка, частотности и сочетаемости его единиц с одной стороны и принципов построения художественного текста с другой актуализируется в статьях К. А. Маслинского и соавторов о детской литературе [Маслинский и др. 2018].

Мощным фундаментом для количественного исследования текстов стал НКРЯ. Такая сфера digital humanities, как культуромика, адаптирована с его помощью для русского языка А. А. Бонч-Осмоловской [Бонч-Осмоловская 2015; Бонч-Осмоловская 2018]. Поэтический подкорпус в составе НКРЯ стал основой для количественных работ о русской рифме Р. Г. Лейбова [Лейбов 2014], О. Н. Ляшевской и соавторов [Lyashevskaya et al. 2020]. Он же открыл перспективы для исследования семантического ореола метра [Shelya et al. 2020], длины текстов [Shelya & Sobchuk 2017] и других не собственно метрических характеристик стихотворного произведения.

Значимыми публикациями следует признать и тестирование на русскоязычном литературном материале методов анализа, ставших традиционными для digital humanities. Это работы Т. Ю. Шерстиновой и соавторов о тематическом моделировании в художественной прозе [Шерстинова и др. 2022], статьи Е. В. Ягуновой и Л. М. Пивоваровой о ключевых словах и коллокациях у Гоголя [Ягунова & Пивоварова 2014], диссертация Д. А. Скоринкина [Скоринкин 2018] и статья Б. В. Орехова и соавторов [Орехов и др. 2018], использующие преимущества сетевого анализа, а также статья Б. В. Орехова о межтекстовом расстоянии современных поэтических книг [Орехов 2019].

Если большинство перечисленных работ и направлений могут восприниматься как имеющие догоняющий характер, переносящие на русскоязычный материал привычные для западной науки методы и практики, то о работах Б. В. Орехова, осмысляющих исследовательский потенциал генеративных нейросетей, стоит говорить как о новаторских не только в отечественном, но и в мировом контексте [Орехов 2017; Orekhov & Fischer 2020].

В связи с компьютерными технологиями и художественными текстами не следует забывать и об области, самой по себе не предполагающей учета количественных показателей, но значимой для выстраивания на ее основе квантитативных исследований. Это область цифровых изданий. На эту тему за последнее десятилетие также было опубликовано заметное число работ как программного характера, так и описывающих конкретные инструменты создания таких публикаций. Внимания заслуживают статьи М. Б. Гронаса и Б. В. Орехова [Gronas & Orekhov 2018], Н. В. Перцова [Перцов 2015], А. А. Бонч-Осмоловской и Е. М. Севериной [Северина и др. 2022], М. Н. Виролайнен [Виролайнен 2021], А. А. Чевтаева [Чевтаев 2022]. Специально нужно сказать и о публикации,

представляющей электронный ресурс «Текстограф», предназначенный для создания цифровых представлений текстовых источников [Orekhov & Tolstoy 2017].

Из этого беглого обзора видно, что описанный Д. А. Скоринкиным период не так просто выделить из общего исторического контекста. Квантитативное (и даже компьютерное) литературоведение существовало в отечественной науке до 2010-х годов, и публикации на эту тему продолжают выходить [Маслинский 2022b]. Цифровое литературоведение в оптике Д. А. Скоринкина представляет собой довольно специфическое подмножество digital humanities, из которого заведомо исключены имеющие сильные для отечественной традиции направления. Смена угла зрения, перевод акцента с институций на исследования заставляет усомниться в конце истории отечественной цифровой гуманитарной науки. Несмотря на сложные внешние обстоятельства, многие из перечисленных авторов активны, и есть основания полагать, что будут продолжать публиковаться, развивая уже сформулированные и порождая новые идеи.

Отдельно стоит обсудить судьбу «цифрового поворота» в исторических исследованиях, который пошёл важные этапы количественной истории, становления компьютерного источниковедения, институционализации исторической информатики, а затем спорных взаимоотношений с цифровой историей и цифровыми гуманитарными науками. Во втором разделе своей статьи Д. А. Скоринкин справедливо отмечает, что начало советской квантитативной истории принято отсчитывать от 1968 года, когда под руководством И.Д.Ковальченко была создана Комиссия по применению математических методов и ЭВМ в исторических исследованиях при Отделении истории АН СССР. В этот период появляются «теоретические концепции работы с информацией исторических источников, разработки компьютерной реализации методов анализа этой информации, а также опыт создания коллекций оцифрованных (в 1960-е – 1970-е годы – машиночитаемых, machine readable) данных» [Гарскова 2018b]. В этот период появляется целая плеяда исследователей, которые не только сами начали применять количественные методы в исторических исследованиях [Количественные 1983], но и оказались весьма влиятельными педагогами, распространяющими свой опыт среди студентов и коллег [Количественные 1984]. Стоит отметить, что опыт преподавания количественных методов сохраняется, например, на историческом факультете МГУ до сих пор [Селунская и др. 2018].

Важным вкладом квантитативного подхода стала дискуссия о массовых исторических источниках. Понятие массовых источников возникло в связи с дискуссиями о доказательности исторического знания и репрезентативности научных наблюдений в истории. Столкнулись два подхода к определению этого типа источника. Б.Г.Литвак описывал массовые источники «от документа», выделяя такие черты массового источника как ординарность обстоятельств происхождения, однородность и повторяемость содержания, однотипность формы, тяготеющая к стандартизации, наличие установленного или сложившегося формуляра [Литвак 1979]. И.Д. Ковальченко признавал детерминирующим не свойства документов, а то, какие явления от отражают: «Массовыми являются источники, характеризующие такие объекты действительности, которые образуют определенные общественные системы с соответствующими структурами. Массовые источники отражают сущность и взаимодействие массовых объектов, составляющих эти системы, а следовательно, строение, свойства и состояние самих систем» [Массовые 1979]. Эта дискуссия оказалась весьма живой, потому что после десятилетий опыта изучения массовых источников в социальной и экономической истории, подобный подход распространился и на такие коллекции источников, как воспоминания. Можно привести диссертационное исследование о революционной мемуаристике Л. А. Колесниковой [Колесникова 2005] или серию публикаций И. М. Гарсковой и Е. М. Горецкой (Симонженковой) мемуарах репрессированных [Гарскова & Симонженкова 2019; Горецкая 2022]. И такое развитие исследовательской мысли приводит нас естественным образом к вопросам накопления цифровых архивов воспоминаний и дневников, где современным эталоном (не имеющим аналогов в мире) проект «Прожито»¹. Путь цифрового проекта «Прожито» отражает этапы освоения онлайн-платформ историками: от резкой критики и неприятия до включения в повседневный инструментарий своих исследований².

Развитие квантитативной истории в СССР стало импульсом к большим методологическим дискуссиям о путях развития исторического знания [Ковальченко 1969; Бессмертный 1972; Кахк & Ковальченко 1974; Хвостова 1975], доказательности исторических наблюдений, границах применимости математических методов. И если символом успехов советской количественной истории стали такие исследования, как например «Всероссийский аграрный рынок» И. Д. Ковальченко и Л. В. Милова [Ковальченко & Милова 1974]. То не менее, значимыми для осмысления количественных

подходов к истории оказались стали работы Б. Н. Миронова [Миронов 1976; Миронов 1991]. Любопытно, что уже в постсоветский период именно Миронов, обобщив многочисленные и разнообразные историко-статистические данные, стал создателем новой концепции развития Российской империи, существенно пересматривающий сложившиеся в историографии взгляды [Миронов 2018].

Прямым продолжением применения статистических методов в исторических исследованиях стали такие направления, как компьютерное источниковедение и историческая информатика. Термин «компьютерное источниковедение» впервые прозвучал в 1992 г. на конференции в Ужгороде, а уже на следующий год появился в работе, посвященной аграрной истории [Моисеенко & Свищев 1993]. Первоначально компьютерное источниковедение даже отождествлялось с исторической информатикой [Гарскова 2010]. Компьютерное источниковедение нацелено, во-первых, на повышение информационной отдачи исторических источников (классическая дискуссия об «информационном потенциале» конкретного типа или фонда исторических источников), во-вторых, на поиск возможностей комплексного подхода к историческим источникам (т.н. «комплексное источниковедение»³), в-третьих, на решение проблем семантической разметки исторических источников и, наконец, в-четвертых, на актуальные возможности связывания информации исторических источников в том числе в контексте открытых связанных данных (linked open data). Задачи компьютерного источниковедения состоят в изучении как традиционных проблем источниковедения (например, атрибуции) средствами новых компьютерных технологий, так и источниковедческие проблемы, порождаемые развитием информационных технологий (появление электронных документов и технотронных архивов и др.). Главный вклад теории информации в источниковедческую практику состоит в переходе от изучения «свидетельств исторических источников» к введению в научный арсенал понятия «информация исторических источников», и, как следствие, поиск методик извлечения скрытой (структурной) информации из исторических источников. Компьютерное источниковедение основывается на идее формирования комплекса источников, обогащая новыми технологиями традицию комплексного источниковедческого подхода Л. В. Черепнина, С. О. Шмидта, В. Л. Янина, что позволяет перейти от «иллюстративного» к «сплошному» изучению исторических источников. Алгоритм применения инструментария компьютерного источниковедения включает несколько этапов: перевод тек-

стов источников и метаинформации о них в электронную форму; разметка текстов и метаинформации; связывание информации в источник-ориентированных базах данных или семантических сетях; применение аналитического инструментария компьютерного источниковедения и генерация новых знаний на основе автоматического вывода при творческом участии исследователя. В России сегодня проблемы компьютерного источниковедения и источниковедческих исследований электронных ресурсов разрабатываются на кафедре исторической информатики МГУ (Л. И. Бородкин [Бородкин 2016], А. Ю. Володин [Володин 2020], И. М. Гарскова [Гарскова 2018a]). Проблемам цифровой трансформации традиционных архивов на примере аудиовизуальных архивов, переводу аудио-, фото-, кинодокументов в цифровой вид, созданию электронных ресурсов на основе созданных копий и их представлению в интернете посвящена фундаментальная монография Ю. Ю. Юмашевой [Юмашева 2020].

Становления исторической информатики как междисциплинарного направления в исторической науке происходило в конце 1980-х—начале 1990-х гг. на волне бурной информатизации общества. Историческая информатика сложилась как область исследований, в которой происходит апробация аналитических методов и компьютерных технологий в конкретно-исторических исследованиях. Формирование исторической информатики началось во многом благодаря микрокомпьютерной революции, когда историки обратились к созданию баз данных на основе информации исторических источников, а затем это направление начало расширяться и теперь включает такие современные методы и технологии, как ГИС, 3D-моделирование, сетевой и контент-анализ, одновременно продолжая традиции применения статистических методов и компьютерного моделирования исторических процессов. Определяя ориентиры развития исторической информатики Л. И. Бородкин отмечает: «Инвариантной на протяжении трех десятилетий является ее главная функция (миссия) – проводить апробацию (а иногда адаптацию) информационных/цифровых технологий и формализованных методов при решении аналитических задач исторического исследования, создании цифровых ресурсов» [Бородкин 2019].

В 1992 году была создана Ассоциация «История и компьютер» (АИК) как ветвь международной ассоциации “History and Computing” (АНС), созданной в 1986 году [Бородкин 2017a].

Опыт исторической информатики показывает, что технологические ориентиры могут меняться — компьютеризация, инфор-

матизация, цифровизация — а исследовательские цели остаются прежними: поиск нового доказанного знания.

Следует отметить, что в российском научном поле ни раз возникали трения по вопросам самоопределения и установления отношений между информатикой, цифровой историей и цифровыми гуманитарными науками [Бородкин 2017b]. И такие демаркационные споры возникают совсем не только в отечественной традиции, в пример можно привести статью директора Центра истории и новых медиа имени Роя Розенцвейга «Различия между цифровыми гуманитарными науками и цифровой историей» [Робертсон 2016]. Но самоопределение в отношении цифровой гуманитаристике, кажется, не таким существенным вопросом [Володин 2014], насколько важным оказывается вопрос методологического выбора: использования количественных методов, информационных технологий и цифровых подходов.

В исторических исследованиях «цифровой путь» (такое название кажется более точным в сравнении с привычным «цифровой поворот» или “digital turn”) существенно связан с технологическими возможностями компьютерных платформ. В период больших вычислительных машин историки довольствовались табличными данными и были рады возможности доверить расчеты машине. Существенные изменения произошли в ходе микрокомпьютерной революции, когда историк стал осваивать компьютер как рабочее место — стали появляться базы данных [Гарскова 1994]. Примечательной была дискуссия «Я и моя база данных», начатая П. Доорном, который предостерегал, что нельзя увлекаться самими инструментами, нужно ориентироваться на научные результаты и новые знания: «Наиболее процветающим направлением является применение баз данных, и я боюсь, что внимание к ним будет еще возрастать, сейчас, когда они содержат картины и звук, песни и танцы. Никогда опасность столкнуться с ситуацией “garbage in, garbage out” не была столь велика, как сегодня» [Доорн 1995]. Вслед за базами стали появляться коллекции оцифрованных документов, затем с ростом производительности компьютеров исторические геоинформационные системы, а с ещё большими процессорными возможностями — трехмерные реконструкции и создание исторических виртуальных миров. Таким образом, расширился спектр источников, которые оцифровывались и применялись для исторического исследования, а за источниками приходили и новые технологии. Палитра подходов, методов и исторических источников, с которыми работает сегодня «цифровые историки» можно увидеть, обратившись к повестке

дня журнала «Историческая информатика»⁴: базы данных, цифровые ресурсы, квантитативная история, моделирование исторических процессов, геоинформационные системы, 3D-реконструкции, компьютеризованный анализ исторических текстов, методы и технологии сетевого анализа, информационные технологии в архивах и музеях, технологии цифровой археологии, новые методы и технологии обработки исторических источников. Журналу уже больше 10 лет, за которые на его электронных страницах увидели свет сотни статей авторов, для которых технология исследования важна не меньше методологии.

Кроме того, нам кажется бессмысленным сделать несколько фактических уточнений к нарисованной Д. А. Скоринкиным картине.

В тексте Д. А. Скоринкина совершенно справедливо указывается на ключевую системообразующую роль поэтического корпуса НКРЯ в становлении отечественных цифровых гуманитарных исследований. С этим тезисом мы полностью солидаризируемся выше. Но у Д. А. Скоринкина среди основателей корпуса (стр. 210) называется один из авторов этих строк Б. В. Орехов, что, к сожалению, неверно. Настоящие создатели, оформившие общую концепцию, направлявшие первые технические шаги, вложившие много ручного труда в получившийся продукт, — это В. А. Плунгян, Д. В. Сичинава и Е. А. Гришина [Гришина и др. 2009]. Полтора десятилетия главным куратором и разметчиком корпуса был и остается К. М. Корчагин.

К команде активных разработчиков этого ресурса Б. В. Орехов присоединился на более поздних этапах, а если говорить о цифровых проектах, то в заслугу ему следует поставить непосредственно следующий во времени за поэтическим корпусом (конец 2006 года) Параллельный корпус переводов «Слова о полку Игореве» (начало 2007 года), то есть, вероятно, второй в России полноценный общедоступный digital humanities-инструмент в области литературоведения [Орехов 2009].

Наконец, И. А. Кижнер, о которой говорится как о покинувшей команду digital humanities-исследователей СФУ, в тот момент оставалась сотрудником Сибирского федерального университета.

Вывод Д. А. Скоринкина неутешителен. Сложившаяся ситуация в России не способствует развитию цифровых гуманитарных наук. Успевшие оформиться за последнее десятилетие формы их организации распадаются, центробежные силы берут верх над интеграционными. Хотя Д. А. Скоринкин как будто такого не утверждает,

но за него говорит название статьи: на повестке дня исчезновение отечественных digital humanities как научной дисциплины.

Несмотря на несомненную тревожность происходящих событий, такой прогноз все же представляется слишком радикальным. Можно закрыть или переформатировать институты, в рамках которых ведутся исследования, но гораздо труднее элиминировать перспективные научные темы и существовавшие десятилетиями направления изысканий. В рамках отечественной науки такие темы и направления есть, сомневаться в их потенциале нет причин.

Мы бы предложили скорректировать список возможных угроз. По нашему мнению, описанные Д. А. Скоринкиным факторы могут вести не к полному исчезновению цифровых гуманитарных наук в России, а к их маргинализации и отставанию от мировых тенденций. Молодые научные кадры, возможно, станут хуже ориентироваться в актуальных методах исследований, вроде сетевого анализа, упустят появление новых прорывных инструментов, сравнимых с векторными моделями, не воспримут важные для области стандарты аналогичные TEI. Накопительный эффект такого расхождения с state of the art западной науки вряд ли позитивно скажется на предметной области. Но всего этого еще не случилось, и в наших силах печального итога избежать.

Мы бы сказали, что текущее положение цифровых гуманитарных наук в России сравнимо не с положением советского гражданина в 1990-е годы, на которое намекает название статьи Д. А. Скоринкина благодаря цитированию книги А. Юрчака о конце советской цивилизации, а с положением Гэри Селдона из фантастического цикла А. Азимова «Foundation» («Основание» или «Академия»). Предвидя падение обеспечивающей стабильность государственности, персонаж писателя (одновременно математик, социолог и историк, каковое сочетание дисциплин должно быть особенно близко специалисту по digital humanities) озабочился созданием альтернативной научной институции⁵. Она должна была сохранить знания для будущих поколений, пронести их через века хаоса и варварства.

Такая деятельная ролевая модель представляется нам более конструктивной, чем произнесение надгробных речей. Одним из способов скрепления распадающегося сообщества нам видится издание настоящего журнала. Насколько это поможет отечественным цифровым гуманитариям, покажет будущее.

Примечания

- ¹ Центр «Прожито» ЕУСПб. URL: <https://prozhito.org/> ; Личные истории в электронном корпусе дневников и воспоминаний. URL: <https://corpus.prozhito.org/>
- ² Здравствуйте, Прожито! Корпус русских дневников он-лайн. Интервью Андрея Володина и Ирины Савельевой с Мишей Мельниченко // Public History Weekly. 23.12.2021. URL: <https://public-history-weekly.degruyter.com/9-2021-10/zdravstvujte-prozhito/>
- ³ Термин «комплексное источниковедение» ввел в научный оборот В.Л. Янин, предполагая, что такой подход позволит историкам выявлять новые, ранее неочевидные им взаимосвязи свойств и черт, присущих объекту познания и прийти к выводам, «которые никак не могут быть получены без такой интеграции» (см.: [Янин 1977]).
- ⁴ Историческая информатика. Журнал. URL: <https://e-notabene.ru/istinf/>
- ⁵ Орехов Б. «Основание» Азимова наоборот // Хабр. 24.09.2023. URL: <https://habr.com/ru/articles/761674/>

Литература

- Алпатов 2023* — Алпатов В. М. История идей науки и история людей науки // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 6: Языкознание. 2023. № 2. С 7–26.
- Андреев 2021* — Андреев В. С. Индивидуальный стиль как система соотношений разноразмерных признаков // Известия Смоленского государственного университета. 2021. № 3. С. 88-101.
- Архипова и др. 2002* — Архипова А. С., Козьмин А. В., Рафаева А. В. Текст–функция–мотив: программа "МЕДИАТОР"// Компьютерная лингвистика и интеллектуальные технологии: Труды Международного семинара Диалог'2002 (Протвино, 6-11 июня 2002 г.): в 2 т. Т. 2: Прикладные проблемы. М.: Наука, 2002. С. 276-282.
- Баевский 2001* — Баевский В. С. Лингвистические, математические, семиотические и компьютерные модели в истории и теории литературы. М.: Языки славянской культуры, 2001.
- Балакин 2003* — Балакин А. Ю. В плену «холодных чисел» // Русская литература. 2003. № 2. С. 208—214.
- Бессмертный 1972* — Бессмертный Ю. Л. Некоторые вопросы применения математических методов в исследованиях советских историков // Математические методы в исторических исследованиях. М.: Наука, 1972. С. 3–14
- Бонч-Осмоловская 2015* — 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

- Бонч-Осмоловская 2018* — Бонч-Осмоловская А. А. Имена времени: эпитеты десятилетий в Национальном корпусе русского языка как проекция культурной памяти // Шаги/Steps. 2018. № 4. С. 115-146.
- Бонч-Осмоловская & Орехов 2013* — Бонч-Осмоловская, А. А., Орехов Б. В. Некоторые применения корпусных методов к наивной поэзии // Статьи на случай: Сборник в честь 50-летия Р. Г. Лейбова, 2013. URL: <http://nevmenandr.net/pages/bo-rl.php>
- Бородкин 2016* — Бородкин Л. И. Моделирование исторических процессов: от реконструкции реальности к анализу альтернатив. СПб.: Алетейя, 2016.
- Бородкин 2017a* — Бородкин Л. И. Становление исторической информатики в России: первые шаги историков на пути «микрокомпьютерной революции» // Историческая информатика. 2017. № 3. С. 155-172.
- Бородкин 2017b* — Бородкин Л. И. Digital history и историческая информатика: конвергенция или дивергенция? // Цифровая гуманитаристика: ресурсы, методы, исследования: материалы Междунар. науч. конф. (г. Пермь, 16–18 мая 2017 г.): в 2 ч. / Перм. гос. нац. исслед. ун-т. Пермь, 2017. Ч. 1. С.15–19.
- Бородкин 2019* — Бородкин Л. И. Инварианты исторической информатики в изменяющемся мире // Историческая информатика. 2019. № 1. С. 1–7. DOI: 10.7256/25857797.2019.1.29508
- Бутенко 2022* — Бутенко З. А. Тематическое моделирование гомогенно метрически размеченных корпусов наивной поэзии // Квантитативная филология. 2022. № 2. С. 39-62
- Васильев 2013a* — Васильев Н. Л. Словарь поэтического языка Н. П. Огарева. Саранск: Издательство Мордовского университета, 2013.
- Васильев 2013b* — Васильев Н. Л., Жаткин Д. Н. Словарь Н. М. Языкова. М.: ФЛИНТА : Наука, 2013.
- Васильев 2016* — Васильев Н. Л., Жаткин Д. Н. Словарь поэтического языка Н. М. Карамзина : монография. М.: ФЛИНТА : Наука, 2016.
- Великанова & Орехов 2019* — Великанова Н. П., Орехов Б. В. Цифровая текстология: атрибуция текста на примере романа М. А. Шолохова «Тихий Дон» // Мир Шолохова. Научно-просветительский общенациональный журнал. 2019. № 1. С. 70—82.
- Виролайнен 2021* — Виролайнен М. Н. Проект «Pushkin-digital» в контексте цифрового развития Пушкинского Дома // Труды Отделения историко-филологических наук 2021. Ежегодник. Т. 11. М.: Наука, 2022. С. 95-100
- Володин 2014* — Володин А. Ю. Digital humanities (цифровые гуманитарные науки): в поисках самоопределения // Вестник Пермского университета. Серия История. 2014. Т. 26, № 3. С. 5–12.

- Володин 2020* — Володин А. Ю. Цифровая трансформация истории? Данные, стандарты, подходы // Электронный научно-образовательный журнал «История». 2020. Т. 11, № 3.
- Гарскова 1994* — Гарскова И. М. Базы и банки данных в исторических исследованиях. Гёттинген: Konrad Pachnicke Max-Planck-Institut für Geschichte, 1994.
- Гарскова 2010* — Гарскова И. М. Источниковедческие проблемы исторической информатики // Российская история. 2010. № 3. С. 155–156.
- Гарскова 2018a* — Гарскова И. М. Историческая информатика. Эволюция междисциплинарного направления. СПб.: Алетей, 2018.
- Гарскова 2018b* — Гарскова И. М. Квантитативная история 1960-х – 1980-х гг. в СССР и ее роль в становлении исторической информатики // Историческая информатика. 2018. № 3. С. 7-24.
- Гарскова 2019* — Гарскова И. М., Симонженкова Е. М. О формализованной методике анализа комплексов мемуарных источников // Историческая информатика. 2019. № 1. С. 169-188
- Гарскова & Симонженкова 2019* — Гарскова И. М., Симонженкова Е. М. О формализованной методике анализа комплексов мемуарных источников // Историческая информатика. 2019. № 1. С. 169-188
- Гаспаров 1984* — Гаспаров М. Л. Очерк истории русского стиха : Метрика, ритмика, рифма, строфика / М. Л. Гаспаров. М.: Наука, 1984
- Гаспаров & Исаков 2003* — Гаспаров М. Л., Исаков О. П. Стихование и ЭВМ // Известия Академии наук. ОЛЯ, 1989, № 2, С. 149-155.
- Горецкая 2022* — Горецкая Е. М. Сравнительный контент-анализ мемуаров репрессированных: гендерный аспект // Историческая информатика. 2022. № 1. С. 108-127.
- Гришина и др. 2009* — Гришина Е. А., Корчагин К. М., Плунгян В. А., Сичинава Д. В. Поэтический корпус в рамках НКРЯ: общая структура и перспективы использования // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009, 71–113.
- Гришина и др. 2015* — Гришина Е. А., Зеленков Ю. Г., Орехов Б. В. Наивная поэзия в акцентологическом корпусе // Труды Института русского языка им. В. В. Виноградова. 2015. Вып. 6. С. 257—271
- Двнятин 2015* — Двнятин Ф. Н. Количественная грамматика и поэтика личных форм глагола в «Гусли доброгласной» Симеона Полоцкого // Slověne. International Journal of Slavic Studies. 2015. Vol. 4. N 1. С. 159-169.
- Дмитриев 2015* — Дмитриев А. Н. Дисциплинарные порядки в гуманитарных и социальных науках // Науки о человеке: история дисциплин: коллект. моногр. / сост. и отв. ред. А.Н. Дмитриев, И.М. Савельева; НИУ ВШЭ. М.: Изд. дом ВШЭ, 2015. С. 7–38.

- Доорн 1995* — Доорн П. Я и моя база данных: движение к концу направления «История и компьютеринг» // Информационный бюллетень ассоциации «История и компьютер». 1995. №13. С. 48–77.
- Кахк & Ковальченко 1974* — Кахк Ю. Ю., Ковальченко И.Д. Методологические проблемы применения количественных методов в исторических исследованиях // История СССР. 1974. №5. С. 89–103
- Ковалев 2024* — Ковалев Б. В. О некоторых квантитативных подходах к периодизации поэтических текстов Хорхе Луиса Борхеса // Terra Linguistica. 2024. 15(1). С. 36-48. <https://doi.org/10.18721/JHSS.15103>
- Ковальченко 1969* — Ковальченко И. Д. О применении математико-статистических методов в исторических исследованиях // Источниковедение. Теоретические и методические проблемы. М.: Наука, 1969. С. 115–133
- Ковальченко & Милов 1974* — Ковальченко И. Д., Милов Л. В. Всероссийский аграрный рынок. XVIII - начало XX века. Опыт количественного анализа. М.: Наука, 1974.
- Козьмин 2006* — Козьмин А. В. Автоматический анализ стиха в системе Starling // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2006». М.: Издательский центр РГГУ, 2006. С. 265-268.
- Колесникова 2005* — Колесникова Л. А. Историко-революционная мемуаристика (1917–1935 гг.) как массовый источник по истории русских революций (Методика количественного анализа): Дис. ... д-ра ист. наук: 07.00.09. Москва, 2005.
- Количественные 1983* — Количественные методы в гуманитарных науках. М.: Изд-во Моск. ун-та, 1981; Количественные методы в советской и американской историографии. М.: Наука, 1983.
- Количественные 1984* — Количественные методы в исторических исследованиях: Учеб. пособие для студ. вузов, обуч-ся по спец. «История»/ Гарскова И. М., Измestьева Т. Ф., Милов Л. В. и др. Под ред. И. Д. Ковальченко. М.: Высш. шк., 1984.
- Колмогоров 2016* — Колмогоров А. Н. Труды по стиховедению. М.: МЦНМО, 2016
- Лейбов 2014* — Лейбов Р. Г. Город и рифма: К описанию морфологической динамики рифменного гнезда // Русско-французский разговорник, или/ou les causeries du 7 septembre: Сборник статей в честь Веры Аркадьевны Мильчиной. М.: Новое литературное обозрение, 2014. С. 521–534
- Лейбов & Орехов 2022* — Лейбов Р. Г., Орехов Б. В. Между политикой и поэтикой: топика Крыма в современной русскоязычной наивной лирике // Шаги/Steps. Т. 8. 2022. № 2. С. 205—232.
- Литвак 1979* — Литвак Б. Т. Очерки источниковедения массовой документации XIX — начала XX в. М., 1979.

Мартыненко 2019 — Мартыненко Г. Я. Методы математической лингвистики в стилистических исследованиях. СПб.: Нестор-История, 2019

Марусенко 2004 — Марусенко М. А. В редакцию журнала «Русская литература» // Вестник Санкт-Петербургского университета. 2004. Сер. 9. Вып. 1-2. С. 99—102.

Марусенко и др. 2001 — Марусенко М. А., Бессонов Б. Л., Богданова Л. М., Аникин М. А., Мясоедова Н. Е. В поисках потерянного автора: Этюды атрибуции. СПб., 2001.

Маслинский 2022a — Маслинский К. А. Сто лет счастья в детской литературе (1920–2020): сталинский канон и его долгосрочные последствия // Шаги / Steps. Т. 8. № 4. 2022. С. 226–247.

Маслинский 2022b — Маслинский К. А. Стилистика детской прозы Виктора Голявкина: синтаксический профиль // Квантитативная филология. 2022. №1(3). С. 32–42.

Маслинский и др. 2018 — Маслинский К. А., Видяева А. В., Додонова Е. Д., Кожевникова Ю. В., Никифоров Н. В. От анализа портрета к поэтике шаблона: о стереотипии и гендерных моделях в изображении детей в советской детской прозе // Детские чтения. 2018. №2 (14). С. 138–159

Массовые 1979 — Массовые источники по социально-экономической истории России периода капитализма. М., 1979

Миронов 1976 — Миронов Б. Н. Историк и математика. Ленинград: Наука, 1976

Миронов 1991 — Миронов Б. Н. Историк и математика. Ленинград: Наука, 1976; Миронов Б.Н. История в цифрах. Математика в исторических исследованиях. М.: Наука, 1991.

Миронов 2018 — Миронов Б. Н. Российская империя: от традиции к модерну. В 3 тт. СПб: Дм.Буланин, 2018.

Михеев & Эрлих 2018 — Михеев М. Ю., Эрлих Л. И. Идиостилевой профиль и определение авторства текста по частотам служебных слов // Научно-техническая информация. Серия 2: Информационные процессы и системы. 2018. № 2. С. 25–34.

Моисеенко & Свищев 1993 — Моисеенко Т., Свищев М. Изучение аграрной истории России последних десятилетий: перспективы «компьютерного источниковедения» // История и компьютер: новые информационные технологии в исторических исследованиях и образовании / под ред. Л. Бородкина и В. Леверманна. Gottingen, 1993. С. 152.

Морозов 1916 — Морозов Н. А. Лингвистические спектры : Средство для отличения плагиатов от истин. произведений того или др. известного авт. / Н.А. Морозов. Петроград : тип. Имп. Акад. наук, 1916.

- Орехов 2009* — Орехов Б. В. Параллельный корпус переводов «Слова о полку Игореве»: итоги и перспективы // Национальный корпус русского языка: 2006—2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 462—473.
- Орехов 2017* — Орехов Б. В. Искусственные нейронные сети как особый тип distant reading // Вестник Приамурского государственного университета им. Шолом-Алейхема. 2017. № 2. С. 32—43
- Орехов 2019* — Орехов Б. В. Лексика как классифицирующий признак современной поэзии // Slavica Revalensia. 2019. VI. С. 251—273
- Орехов 2020* — Орехов Б. В. «Илиада» Е.И. Кострова и «Илиада» А. И. Любжина: стилеметрический аспект // Аристей. 2020. Т. XXI. С. 282—296; Орехов Б. В. Текст и перевод Владимира Набокова через призму стилеметрии // Новый филологический вестник. 2021. № 3. С. 200—213.
- Орехов 2022* — Орехов Б. В. Рецепция классических поэтических формул у наивных авторов в акцентологическом корпусе НКРЯ // Труды Института русского языка им. В. В. Виноградова. 2022. № 2. С. 124—132
- Орехов и др. 2018* — Орехов Б. В., Успенский П. Ф., Файнберг В. В. Цифровые подходы к Камер-фурьерскому журналу В. Ф. Ходасевича // Русская литература. 2018. № 3. С. 19—53.
- Панасюк 2022* — Панасюк А. В. Автоматическое обнаружение тематической связности новостной повестки и наивной поэзии // Квантитативная филология. 2022. № 1 С. 52-70
- Перцов 2015* — Перцов Н. В. Об аутентичном факсимильно-транскрипционном представлении рукописей русских классиков // Филологические науки. 2015. № 1. С. 75-94.
- Пильщиков & Старостин 2010* — Пильщиков И. А., Старостин А. С. Автоматическое распознавание стихотворных размеров: теория и практика // Поэтика и фоностилистика: Бриковский сборник. М., 2010. Вып. 1: Мат-лы Междунар. науч. конф. «I-е Бриковские чтения: Поэтика и фоностилистика» (Москва, 10—12 февраля 2010 года). С. 41—49.
- Плунгян 2005* — Плунгян В. А. К эволюции русской метрики: немонотонная силлабо-тоника / В. А. Плунгян // Язык. Личность. Текст : сб. статей к 70-летию Т. М. Николаевой. М. : ЯСК, 2005. С. 857—869.
- Полилова 2017* — Полилова В. С. О неклассическом стихе Бальмонта: Ритмическая структура прерывистых строк // Zbornik matice srpske za slavistiku. 2017. Т. 92. С. 731—746
- Робертсон 2016* — Робертсон С. Различия между цифровыми гуманитарными науками и цифровой историей // Электронный научно-образовательный журнал «История». 2016. Т. 7. Выпуск 7 (51) [Электронный ресурс]. DOI: 10.18254/S0001648-1-1

Северина и др. 2022 — Северина Е. М., Бонч-Осмоловская А. А., Кудин А. М. Цифровые филологические практики: проект «Chekhov Digital» // Актуальные проблемы филологии и педагогической лингвистики. 2022. № 2. С. 153-165.

Селунская и др. 2018 — Селунская Н. Б., Петрова О. С., Карагодин А. В. Измерение прошлого. СПб: Алетейя, 2018.

Скоринкин 2018 — Скоринкин Д. А. Семантическая разметка художественных текстов для количественных исследований в филологии (на примере романа «Война и мир» Л.Н. Толстого). М., 2018.

Скоринкин & Бонч-Осмоловская 2016 — Скоринкин Д. А., Бонч-Осмоловская А. А. «Особые приметы» в речи художественных персонажей: количественный анализ диалогов в «Войне и мире» Л.Н. Толстого // Электронный научно-образовательный журнал «История». 2016. Т. 7. Вып. 7 (51).

Хвостова 1975 — Хвостова К. В. Методологические проблемы применения количественных методов в исторических исследованиях // Вопросы истории. 1975. №11.

Частотный 1978 — Частотный словарь романа Л. Н. Толстого "Война и мир" / М-во просвещения РСФСР. Тула: Б. и., 1978.

Чевтаев 2022 — Чевтаев А. А. О Применимости новых текстологических моделей в представлении рукописей Ф. М. Достоевского на сайте цифрового архива писателя // Новый филологический вестник. 2022. №3. 279–297

Шерстинова и др. 2022 — Шерстинова Т. Ю., Москвина А. Д., Кирина М. А., Карышева А. С., Колпащикова Е. О. Тематическое моделирование русского рассказа 1900–1930: наиболее частотные темы и их динамика // Компьютерная лингвистика и интеллектуальные технологии: по материалам международной конференции «Диалог 2022», Вып. 21. Изд-во РГГУ, 2022. С. 512–526.

Юмашева 2020 — Юмашева, Ю. Ю. Цифровая трансформация аудиовизуальных архивов. Аудиовизуальные архивы онлайн. М.; Берлин: Директ-Медиа, 2020.

Ягунова & Пивоварова 2014 — Ягунова Е. В., Пивоварова Л. М. Экспериментально-вычислительные исследования художественной прозы Н. В. Гоголя // Structures & Functions: Studies in Russian Linguistics. Volume 1, Issue 3. Tallinn (2014). pp. 83-104

Янин 1977 — Янин В. Л. Очерки комплексного источниковедения. М.: Высшая школа, 1977).

Belousova & Páramo Rueda 2019 — Belousova A., Páramo Rueda J. S. Macroanalysis of the Strophic Syntax and the History of the Italian Ottava Rima // Quantitative Approaches to Versification. Prague: Institute of Czech Literature of the Czech Academy of Sciences, 2019. P. 23–30.

- Gronas & Orekhov 2018* — Gronas M., Orekhov B. Что такое семантическое издание и почему в будущем все издания станут семантическими? // A/Z: Essays in honor of Alexander Zholkovsky. Boston: Academic Studies Press, 2018. С. 246—268.
- Kizhner et al. 2022* — Kizhner I., Terras M., Orekhov B., Manovich L., Kim I., Rummyantsev M., Bonch-Osmolovskaya A. The History and Context of the Digital Humanities in Russia // Global Debates in the Digital Humanities / ed. Domenico Fiormonte, Paola Ricaurte, Sukanta Chaudhuri. University of Minnesota Press, 2022. P. 55—70.
- Liapin 2020* — Liapin S. Russian Iambic Tetrameter: The Evolution of Its Rhythmic Structure // Studia Metrica et Poetica: Vol. 7 No. 2 (2020) 7–22
- Lyashevskaya et al. 2020* — Lyashevskaya O., Litvintseva K., Sechina E., Vlasova E. Verb rhyme in Russian Poetry: a quantitative analysis // Russian linguistics. 2020. No. 44(2). P. 145–156.
- Orekhov & Tolstoy 2017* — Orekhov B., Tolstoy F. Textograf: A Web Application for Manuscript Digitization // Digital Humanities 2017. Conference Abstracts, 2017. P. 772—773.
- Orekhov & Fischer 2020* — Orekhov B., Fischer F. Neural reading. Insights from the analysis of poetry generated by artificial neural networks // Orbis Litterarum. 2020. Vol. 75. No 5. P. 230—246
- Shelya & Sobchuk 2017* — Shelya A., Sobchuk O. The shortest species: how the length of Russian poetry changed (1750–1921) // Studia Metrica et Poetica: Vol. 4 No. 1 (2017) 66–84
- Shelya et al. 2020* — Šeļa A., Orekhov B., Leibov R. Weak Genres: Modeling Association Between Poetic Meter and Meaning in Russian Poetry // Proceedings of the Workshop on Computational Humanities Research (CHR 2020): Amsterdam, the Netherlands, November 18–20, 2020. CEUR-WS, P. 12—31.
- Skorinkin & Bonch-Osmolovskaya 2017* — Skorinkin D., Bonch-Osmolovskaya A. Text mining War and Peace: Automatic extraction of character traits from literary pieces // Digital Scholarship in the Humanities, Volume 32, Issue suppl_1, April 2017, Pages i17–i24
- Skorinkin 2023* — Skorinkin D. "Digital Humanities in Russia Was Forever, Until It Was No More: The Story of Russian Digital Humanities in 2011–2022 Canadian-American Slavic Studies 57, 1-2 (2023): 209-221. <https://doi.org/10.30965/22102396-05701025>

ХРОНИКА

Динара Гагарина

DIGITAL HUMANITIES В ЦЕНТРАЛЬНОЙ АЗИИ: ЗАМЕТКИ НА ПОЛЯХ КРУГЛОГО СТОЛА

В статье обсуждается круглый стол “Digital Humanities в Центральной Азии”, который состоялся 15 марта 2023 года в онлайн-формате, и стал первым мероприятием подобной тематики и формата в регионе. Основной фокус обсуждений касался трех стран Центральной Азии – Казахстана, Кыргызстана, Узбекистана. Анализируется состав участников, содержание дискуссий, а также результаты круглого стола и результаты проведенных в рамках мероприятия опросов.

Введение

Digital humanities (цифровые гуманитарные науки) как самостоятельная научная область начала формироваться в середине 2000-х годов и стала результатом объединения ряда других направлений – компьютерной лингвистики, цифровой истории (исторической информатики), цифрового литературоведения, цифровых исследований культуры и т.п. Методологические, организационные и даже маркетинговые причины и обоснования такого объединения многократно рассмотрены в литературе и обсуждались на конференциях¹.

Развитие направления в Европе, США, Австралии и Канаде достигло сегодня высокого уровня, научные центры digital humanities открыты в большом количестве университетов², издаются профильные журналы (среди них – Digital Humanities Quarterly (DHQ),

Динара Амировна Гагарина

Американский университет в Центральной Азии, Университет Эрлангена – Нюрнберга, DH CLOUD
dinara.gagarina@fau.de

Digital Scholarship in the Humanities (DSH), Digital Studies / Le champ numérique, Journal of Digital Humanities (JDH), International Journal of Humanities and Arts Computing), проводятся специализированные конференции, функционируют десятки образовательных программ³. В части вышеперечисленных критериев развитие digital humanities в странах Центральной Азии отстает от западных стран, при том, что сами гуманитарные исследования на основе цифровых методов проводятся достаточно активно в целом ряде университетов, то же верно и для цифровизации сферы культуры.

Значимую роль в развитии digital humanities в международном научном пространстве играют профессиональные сообщества, многие из них объединены в “Alliance of Digital Humanities Organizations” (ADHO)⁴, куда входят 11 национальных и наднациональных ассоциаций. Профессиональные сообщества на протяжении многих лет проводят конференции, обучающие и другие мероприятия.

Другой важный фактор согласованного развития digital humanities – исследовательские инфраструктуры, которые задают правила и рекомендации по работе с цифровыми гуманитарными данными. В Европе к таким относятся CLARIN-ERIC⁵ и DARIAH-EU⁶. Подобных структур и сообществ по цифровым гуманитарным исследованиям в странах Центральной Азии на сегодня нет.

Для обсуждения цифровой трансформации гуманитарных исследований и образования в университетах стран Центральной Азии, актуализация таких исследований и проектов, формирования сообщества для развития digital humanities в регионе в марте 2023 года был проведен Круглый стол “Digital Humanities в Центральной Азии”.

Круглый стол стал первым подобным мероприятием, объединившим исследователей и практиков, ученых из разных предметных областей, университетов и стран Центральной Азии.

Концепция круглого стола и состав участников

Мероприятие стало результатом инициативы нескольких исследователей (перечислены в алфавитном порядке):

- Шайыркул Джолдошевны Батырбаевой, профессора кафедры археологии, этнографии, источниковедения и историографии Кыргызского национального университета им. Ж. Баласагына;

- Динары Амировны Гагариной, сооснователя DH CLOUD, научного сотрудника Американского университета в Центральной Азии, Кыргызстан;
- Бауржана Олжабаевича Жангуттина, заведующего кафедрой истории Казахстана Казахского национального педагогического университета имени Абая;
- Замиры Раймовны Ишанходжаевой, профессора кафедры истории Узбекистана Национального университета Узбекистана имени Мирзо Улугбека;
- Ироды Махмудовны Шамсиевой, доцента кафедры истории Ташкентского международного университета Кимё, Узбекистан.

Ни один из перечисленных в аффилиациях инициаторов университетов формально не был организатором круглого стола. В программный и организационный комитет помимо указанных выше исследователей вошел также Андрей Юрьевич Володин, доцент кафедры исторической информатики МГУ и директор Центра цифровых гуманитарных исследований Сибирского федерального университета. Роль в проведении мероприятия сыграло DH CLOUD Community⁷, от его имени также выдавались сертификаты об участии.

Распространение информации о круглом столе шло по нескольким каналам. Были использованы площадки сообщества DH CLOUD в социальных сетях; размещение информации в профильных сообществах, в том числе “Inner and Central Asian Studies Dissertation Reviews” в Facebook⁸; почтовая рассылка по тем, кто ранее участвовала в мероприятиях DH CLOUD, рассылка по листу А Цифровые гуманитарные науки без институциональных и географических границ. Объединяем русскоязычное ЦТ ссоциации “История и компьютер”, рассылка по университетам и гуманитарным факультетам Казахстана, Кыргызстана, Таджикистана, Туркменистана, Узбекистана. Кроме того, был создан сайт⁹, где шла регистрации и размещалась вся необходимая информация. Распространение информации шло также по личным каналам инициаторов.

Изначально мероприятие планировалось как двуязычное – русский и английский, но в ходе подготовки было решено использовать преимущественно русский язык, являющийся своего рода lingua

franca в Центральной Азии. На английском было сделано только одно из выступлений. Как предварительные обсуждения, так и опрос аудитории во время мероприятия показали предпочтение к русскому языку.

При подготовке и проведении круглого стола мы исходили из широкого понимания digital humanities как междисциплинарного направления, которое включает исследования в различных областях гуманитарных наук на основе цифровых методов, теорию и практику создания цифровых гуманитарных ресурсов, цифровое сохранение и репрезентацию культурного наследия.

Связывающим признаком представленных проектов и обсуждаемых проблем был фокус на истории, языках и культурах стран Центральной Азии, а также науке и ее организации в регионе. Такое объединение по региональному признаку обосновано как с точки зрения существования общего культурного слоя, лингвистических и исторических пересечений, схожей системы организации образования, так и с точки зрения опыта региональных сообществ digital humanities – ассоциаций “Digital Humanities in the Nordic and Baltic Countries”, “Humanities im deutschsprachigen Raum” и других.

В соответствии с указанным выше пониманием digital humanities дискуссии были организованы в виде трех последовательных секций, где спикеры и слушатели рассказали о проектах, обменялись опытом исследований и образовательных программ по digital humanities, поделились существующими проблемами и путями их решения:

- “Цифровые исследования истории, культуры и языков народов и стран Центральной Азии”;
- “Цифровое сохранение и репрезентация культурного наследия и исторической памяти”;
- “Образовательные программы и инфраструктурное развитие Digital Humanities в Центральной Азии”.

Состав экспертов-спикеров формировался программным комитетом. Среди более 20 выступающих – лингвисты, историки, искусствоведы, специалисты по медиа и цифровому сохранению культурного наследия, разработчики цифровых гуманитарных проектов, ориентированных на Центральную Азию.

На круглый стол зарегистрировались более 200 человек, приняли участие – более 150 человек из Узбекистана, Казахстана,

Кыргызстана, Таджикистана, России и других стран. Доля историков на мероприятии была значительно больше, чем это типично для сообществ и конференций по digital humanities, что, по-видимому, связано со “скошенностью” в сторону истории состава программного комитета.

Ниже представлена характеристика секций-дискуссий круглого стола, все их записи можно найти на канале сообщества DH CLOUD¹⁰.

Секция “Цифровые исследования истории, культуры и языков народов и стран Центральной Азии”

Открыла мероприятие секция об исследовательском направлении digital humanities, ее модератором выступил А.Ю. Володин. Секция была выстроена через презентацию исследовательских проектов, представляющих широкое разнообразие объектов изучения и применяемых методов.

Секция началась с доклада Нилюфар Зайнобиддин кызы Абдурахмоновой, профессора кафедры узбекского языкознания Национального университета Узбекистана им. Мирзо Улугбека, содержащего обзор корпусных технологий для изучения узбекского языка. Внимание было уделено “O‘zbek tili korpusi”¹¹ – проекту создания электронного корпуса узбекского языка. Проект включает комплекс лингвистических ресурсов, среди которых лингвистические словари, морфологическая база данных узбекского языка, тексты различных жанров, включая научные, официальные, художественные и устные тексты, параллельные корпуса узбекско-английского и узбекско-русского языков, программное обеспечение, позволяющее управлять и использовать эти данные. Проект находится в разработке, Н. Абдурахмонова отметила, что коллектив открыт для сотрудничества.

Ландшафт разработки корпусов казахского языка представил Андрей Юрьевич Фильченко, профессор кафедры мировых языков, лингвистики и литературы Назарбаев Университета, Казахстан. Команда под руководством А.Ю. Фильченко занимается разработкой мультимедийного корпуса современного казахского языка¹², включающего устный разговорный материал, и исследовательской инфраструктуры для изучения казахского языка. На сайте хранятся записи речи людей в разных ситуациях для дальнейшего фонетического, морфологического и синтаксического анализа.

В лингвистической части первой дискуссии короткое сообщение о проекте ЛингвоДок¹³ также сделала Юлия Викторовна Норманская, ведущий научный сотрудник отдела урало-алтайских языков Института Языкознания. На этой платформе собраны аудиословари и корпуса более чем на 1000 исчезающих диалектах уральских и алтайских языках, в частности, 57 словарей и 6 корпусов казахского языка, в настоящее время разрабатывается обучающая платформа на казахском языке.

Данияр Карабаев, директор подразделения искусств, гуманитарных наук и коммуникаций Американского университета в Центральной Азии, Кыргызстан, презентовал портал “Manas Discovery”¹⁴. Проект посвящен Манасу – главному кыргызскому эпосу, его цифровому сохранению и изучению. На сайте проекта представлены записи чтения Манаса различными манасчи (рассказчиками), научные статьи, диссертации, документальные фильмы.

Научное поле цифрового стиховедения и результаты исследований тюркского стиха представил на круглом столе Борис Валерьевич Орехов, доцент школы лингвистики Высшей школы экономики, сооснователь сообщества DH CLOUD. С учетом тематики круглого стола отдельно было рассмотрено формальное и количественное изучение тюркской стиховедческой традиции, в том числе узбекского стиха. Подробнее о различиях различных тюркских стихов можно прочитать в монографии Б.В. Орехова “Башкирский стих XX века. Корпусное исследование”¹⁵.

Об пространственных технологиях и использовании ГИС в исследовании истории и этнокультурного пространства Кыргызстана рассказала Сайкал Чубаковна Жумалиева, доцент кафедры археологии, этнологии, источниковедения и историографии Кыргызского национального университета им. Ж. Баласагына. Было представлено несколько проектов и исследований в области исторической геоинформатики и исторической демографии, в том числе изучение истории Шелкового пути, проходящему по территории современного Кыргызстана, кочевой культуры и оседлости кыргызского народа, трансформации и инфраструктуре населенных пунктов.

Помимо собственно исследовательских проектов участники дискуссии также обсудили вовлечение студентов в их реализацию, как в части оцифровки и сбора материалов, так и анализа. Будучи деканом факультетов, в этой части ценным опытом поделились Д. Карабаев и А.Ю. Фильченко.

Секция “Цифровое сохранение и репрезентация культурного наследия и исторической памяти”

На второй дискуссии круглого стола обсуждались проблемы оцифровки исторического и культурного наследия и его включения в исследовательское и публичное поле через цифровые инструменты. Программа была сформирована так, чтобы было представлено видение отрасли с разных сторон – исследователями, аналитиками, учреждениями хранения, разработчиками. Модератором секции выступила И.М. Шамсиева.

Масштабы оцифровки в Казахстане были предметом выступления Кайрата Шакаримовича Алимгазинова, заместителя директора Архива Президента Республики Казахстан¹⁶. В этом архиве оцифрованы более миллиона документов, внедрена система интеллектуального поиска, удаленный доступ к каталогу и заказу документов, разработаны базы данных:

- “Номенклатурные кадры советского Казахстана”¹⁷, содержащая биографии руководителей партийных, комсомольских, административных, финансовых, профсоюзных органов, образования и здравоохранения, а также комиссаров военных комиссариатов Казахстана (1919-1991 гг.) начиная с уездного и районного уровней;
- “Военнопленные-казахстанцы Второй мировой войны. Республиканская база данных”¹⁸, где собраны более 700 тыс. документов и информация о 36 тыс. людей;
- “1930–1933 жж. Қазақстандағы ашаршылық”¹⁹ – единый банк данных жертв голода, где собрана информация о 1700 людях, погибших от голода в 1930-х годах.

Историческую тему продолжила З.Р. Ишанходжаева, представившая проект “Raқamli Tarix”²⁰ (“цифровая история”) – инициативу по сохранению в цифровом виде и изучению истории и культурного наследия Узбекистана, разработки и размещения учебных материалов по цифровой истории разного типа, проведению профильных мероприятий. В настоящее время на сайте инициативы размещен проект, посвященный изучению наследия и увековечению памяти о репрессированных, включающий в том числе мобильное приложение – книгу памяти о жертвах репрессий.

О том, что Узбекистан и Центральная Азия в целом находятся на начальной стадии цифровой трансформации сферы культурного наследия, в частности музеев, говорили Нигора Рахимовна Ахмедова, главный научный сотрудник Института искусствознания Академии наук и академик Академии художеств Республики Узбекистан, и Владимир Викторович Определенов, директор по цифровому развитию в ART Domains, также представляющий Узбекистан на момент проведения круглого стола. Оба эксперта отметили нехватку специалистов в Узбекистане.

Н.Р. Ахмедова говорила о цифровых методы работы с культурным наследием со стороны искусствоведа, отмечая значимость этого для культуры, науки и образования. По мнению эксперта в ряде институций процессы цифровой трансформации еще не начаты, но “картина не пессимистическая, а реалистическая”. Одним из значимых проектов в этом направлении является создание цифрового государственного каталога музейных коллекций Узбекистана.

В.В. Определенов продемонстрировал, как современные системы поиска, основанные на технологиях искусственного интеллекта и работающие в виде чата, не могут выдать существенной информации по истории и культуре Узбекистана. Это означает, что важно не только оцифровать культурное наследие, но и обеспечить их открытость, доступность и машиночитаемость, чтобы эта информация не выпадала от поисковых систем и соответственно из научного изучения и современного образования.

Среди прочего в выступлении В.В. Определенова была затронута тема цифрового искусства и генеративного искусственного интеллекта для создания изображений, что вызвало интерес аудитории, в дискуссию включилось сразу несколько участников. Со стороны художника цифровое искусство прокомментировал цифровой художник и исследователь-биолог Георгий Федорович Курякин, со стороны истории – И.М. Шамсиева, отметившая, что портреты Амира Темура и других исторических деятелей, созданные генеративным искусственным интеллектом, не соответствуют действительности. Это, как подчеркнул В.В. Определенов, еще раз доказывает и демонстрирует недоступность исторических и культурных данных для современных технологий. На тему возможностей использования искусственного интеллекта высказался также Сергей Иванович Корниенко, профессор кафедры гуманитарных дисциплин НИУ ВШЭ – Пермь, подчеркнув, что развитие технологий требуют проработки вопроса того, какие из задач на новом этапе будут решаться человеком, а какие компьютером. “Человек

теряет монополию на креативность”, – отметил Никита Александрович Клюев, приглашенный преподаватель той же кафедры, что вместе с тем позволяет использовать разную оптику восприятия тех или иных проблем и даже задач при обращении к культурному наследию через цифровую призму.

Тему проблем цифровизации Центральной Азии продолжил Дамир Сагынбаевич Саттаров, эксперт Central Asian Analytical Network²¹, представляющий Казахстан. Он привел статистику по покрытию региона Интернетом, скорости связи, электронному правительству и другим аспектам информатизации Казахстана, подчеркнув сложности в сфере цифровизации культуры и важность открытия профильных образовательных программ, в том числе по digital humanities.

Инновационный проект использования технологий дополненной реальности в музеях и туристических объектах презентовал и продемонстрировал Алишер Камолович Рахимов, генеральный директор ООО “SmartChain” и советник хокима Ферганской области по вопросам ИКТ, Узбекистан. Под руководством А.К. Рахимова создано приложение дополненной реальности NazzAR²², внедренное в музеях Самарканда, Ташкента, Хивы и Бухары. Эксперт подчеркнул важность государственной поддержки в этом вопросе и запрос на разработку подобных проектов со стороны государства, как это было в случае с приложением NazzAR. Интерес аудитории вызвала организация подобных команд и проектов, вовлечение в них разных специалистов, взаимодействие с музеями. В проекте работает 12 человек разных специальностей, привлекаются историки и представители музеев. Описывая цикл создания и внедрения приложения, А.К. Рахимов сказал о недостаточном качестве существующего контента в музеях, отсутствии описанных экскурсий, которые можно переложить в цифровой формат.

Тема образования и вовлечения студентов в digital humanities проекты прошла красной линией через все три секции круглого стола. Обсуждалась она и в связи с оцифровкой и цифровой репрезентацией культурного наследия. К ней так или иначе обратились практически все эксперты, рассказывая об опыте (З.Р. Ишанходжаева, К.Ш. Алимгазинов) или запросах в этой отрасли (Д.С. Саттаров, Н.Р. Ахмедова). Один из возможных способов совместной работы университетов и музеев представила независимый исследователь Анна Григорьевна Качкаева, автор проекта “Пространство соучастия”²³, в рамках которого около десяти вузов в разных регионах России работают с несколькими десятками музеев, архивов, театров

и НКО и совместно создают цифровые культурные проекты разного типа.

Секция “Образовательные программы и инфраструктурное развитие Digital Humanities в Центральной Азии”

Круглый стол завершился секцией про комплексное развитие digital humanities в регионе, важной составляющей которого является образование. Модератором секции выступила Д.А. Гагарина.

В первой части секции обсуждались сообщества и ассоциации в предметном поле digital humanities и их роль в развитии направления на опыте Европы и других регионов.

Здесь важным является опыт межрегиональной ассоциации “История и компьютер”²⁴, объединяющей уже 30 лет исследователей из постсоветских стран, в том числе Казахстана, Кыргызстана, Узбекистана. Этот опыт представил Леонид Иосифович Бородин, заведующий кафедрой исторической информатики МГУ, показав, как тематически развивалось использование цифровых технологий в исторических исследованиях, в том числе в России и странах Центральной Азии, и как это может быть масштабировано в структуре направления digital humanities.

Тему межрегиональных контактов историков, использующих математические и компьютерные методы, продолжила Сауле Ауке-новна Жакишева, профессор кафедры всемирной истории, историографии и источниковедения Казахского Национального университета им. аль-Фараби. Об исторической информатике в Кыргызстане говорила Ш.Д. Батырбаева, она представила серию проектов, связанных с математическими методами, 3D-моделированием, ГИС-технологиями в истории, реализованными в Кыргызском национальном университете им. Ж. Баласагына. Обе исследовательницы сделали акцент на разном восприятии проблематики digital humanities, исторической информатики и цифровой истории представителями разных коллективов и университетов.

Развитие digital humanities со стороны лингвистов представила Анастасия Александровна Бонч-Осмоловская, сооснователь DH CLOUD. Она показала на своем опыте и опыте международного сообщества, как исследователи-одиночки, применяющие цифровые методы в различных гуманитарных дисциплинах, обмениваются идеями и объединяются в междисциплинарные команды, и какую роль играют профессиональные сообщества и ассоциации в этом направлении.

Успешным опытом развития цифровой истории поделился Б.О. Жангуттин, на кафедре заведующим которой он является, действует профильные магистерская и бакалаврская программы. Он подчеркнул важность институциональной поддержки такого рода проектов и направления в целом со стороны администрации университетов и со стороны партнерских институций – архивов, институтов Академии наук, школьных структур. На примере проектов по устной истории, истории репрессий Б.О. Жангуттин показал значимость совместной работы и совместных мероприятий.

О роли сообществ digital humanities на примере партнерства с коллегами из Бейрута в рамках секции говорила Ангелина Попова, директор Центра преподавания, обучения и технологий Американского университета в Центральной Азии (в Бейруте работает университет той же сети – Американский университет Бейрута). Она подчеркнула важность междууниверситетской работы и рассказала о возможностях для коллабораций на разных уровнях.

Во второй части дискуссии было представлено несколько образовательных программ по digital humanities.

Опытom разработки и запуска двух программ поделилась Шолпан Турегельдиевна Нурман, доцент кафедрой истории Казахстана Казахского национального педагогического университета им. Абая, продемонстрировав путь от идеи и до выпуска и трудоустройства студентов, уделив внимание регламентной базе, административной поддержке, принципам реализации программ и роли партнерств, также была представлена структура учебного плана.

Ш.Д. Батырбаева рассказала об опыте запуска серии курсов по цифровым технологиям в истории, здесь среди прочего акцент был сделан на возникающих проблемах, что позволило сделать более объемную картину отрасли.

Завершая секцию, эксперты высказали свое видение развития направления в регионе, возможные совместные мероприятия и анонсировали научные конференции, семинары и школы в очном и онлайн формате в своих университетах.

Заключение

В ходе подготовки мероприятия и во время его проведения проводился опрос о потребностях сообщества и собраны предложения коллег о новых темах и форматах встреч. Наиболее востребованными оказались обучающие вебинары и мастер-классы – в ходе опроса их отметили 70 % респондентов. Среди актуальных тем ме-

роприятий – digital humanities в целом, цифровая история (включая количественные методы в истории), data science в гуманитарных науках, компьютерные методы анализа текстов.

За время, прошедшее после проведения круглого стола до написания этого обзора, уже состоялось несколько мероприятий, так или иначе связанных с digital humanities, в Казахстане, Кыргызстане, Узбекистане, России и онлайн. Круглый стол и сформированное им сообщество способствовало распространению информации и участию исследователей в этих мероприятиях. Digital humanities как научное и образовательное направление было представлено Д.А. Гагариной на пленарных заседаниях двух международных конференций – в Ташкентском международном университете Киме и Ташкентском университете прикладных наук. Запланированы дальнейшие мероприятия, идет обсуждение создания региональной ассоциации.

Внимание к круглому столу проявило издание Системный Блок²⁵, ими был подготовлен обзор всех выступлений мероприятия с подбором дополнительных ссылок на ресурсы и проекты²⁶.

В целом круглый стол показал значимую потребность исследователей в совместных тематических научных мероприятиях, а также в более детальном обсуждении и совместной работе над образовательными программами по digital humanities.

Примечания

- ¹ Gold, Matthew K. (ред.). "Debates in the Digital Humanities". University of Minnesota Press, 2012; McCarty, Willard. "Humanities Computing". Palgrave Macmillan, 2005; Moretti, Franco. "Graphs, Maps, Trees: Abstract Models for Literary History". Verso, 2007; Ramsay, Stephen. "Reading Machines: Toward an Algorithmic Criticism". University of Illinois Press, 2011; Schreibman, Susan, Siemens, Ray, и Unsworth, John. "A Companion to Digital Humanities". Blackwell, 2004; Berry, David M. "Understanding Digital Humanities". Palgrave Macmillan, 2012; Burdick, Anne, Drucker, Johanna, Lunenfeld, Peter, Presner, Todd, и Schnapp, Jeffrey. "Digital Humanities". MIT Press, 2012; Terras, Melissa, Nyhan, Julianne, и Vanhoutte, Edward. "Defining Digital Humanities: A Reader". Ashgate Publishing, Ltd., 2013.
- ² centerNet. <https://dhcenternet.org>.
- ³ Digital Humanities Course Registry. <https://dhcr.clarin-dariah.eu>; Гагарина Д.А., Корниенко С.И. Digital Humanities: образовательный ландшафт. Электронный научно-образовательный журнал "История". 2020. № 3 (89). С. 1-15.

- 4 Alliance of Digital Humanities Organizations (ADHO). <https://adho.org>.
- 5 CLARIN-ERIC. <https://www.clarin.eu>.
- 6 DARIAH-EU. <https://www.dariah.eu>.
- 7 DH CLOUD – профессиональное сообщество, объединяющее русскоязычных специалистов по цифровым гуманитарным наукам, не зависимо от институциональной принадлежности и страны работы или проживания; основано в 2022 году А.А. Бонч-Осмоловской, Д.А. Гагариной, Б.В. Ореховым. Сайт DH CLOUD: <https://dhcloud.org>.
- 8 Inner and Central Asian Studies Dissertation Reviews. <https://www.facebook.com/groups/1424>
- 9 Круглый стол Digital Humanities в Центральной Азии. <https://ca.dhcloud.org>.
- 10 Канал DH CLOUD на YouTube. <https://www.youtube.com/@dhcloud>.
- 11 O'zbek tili korpusi. <https://uzbekcorpus.uz>.
- 12 БҮГІНГІ АУЫЗЕКІ ҚАЗАҚ ТІЛІНІҢ МУЛЬТИМЕДИАЛЫҚ КОРПУСЫ. <https://project6817623.tilda.ws>
- 13 Lingvodoc 3.0. <http://lingvodoc.ispras.ru/>
- 14 Manas Discovery: всё о кыргызском эпосе. <https://manas-discovery.kg>
- 15 Орехов Б. В. Башкирский стих XX века. Корпусное исследование. СПб.: Алетейя, 2019. 344 с.
- 16 Қазақстан Республикасы Президентінің Архиві. <https://archive.president.kz>
- 17 Номенклатурные кадры советского Казахстана. <https://archive.president.kz/ru/bd-nomenklaturnye-kadry-sovetskogo-kazakhstan>.
- 18 Военнопленные-казахстанцы Второй мировой войны. Республиканская база данных. <http://tutkyn.kz>.
- 19 1930–1933 жж. Қазақстандағы ашаршылық. <http://asharshylyq.kz>.
- 20 Raqamli Tarix. <https://raqamlitarix.uz>.
- 21 Central Asian Analytical Network. <https://www.caa-network.org>.
- 22 NazzAR. <https://nazzar.uz>.
- 23 Пространство соучастия // Благотворительный фонд Владимира Потанина. <https://fondpotanin.ru/activity/specprjects/prostranstvo-souchastiya>.
- 24 Ассоциация "История и компьютер". <https://aik-hisc.ru>.
- 25 Системный Блок. <https://sysblok.ru>.
- 26 Подрядчикова М. От кыргызского эпоса до ChatGPT: что обсуждали на круглом столе "Digital Humanities в Центральной Азии" // Системный Блок. <https://sysblok.ru/philology/ot-kyrgyzskogo-jeposa-do-chatgpt-cto-obsuzhdali-na-kruglom-stole-digital-humanities-v-centralnoj-azii/>

Литература

Қазақстандағы — 1930–1933 жж. Қазақстандағы ашаршылық. <http://asharshylyq.kz>.

Ассоциация — Ассоциация «История и компьютер». <https://aik-hisc.ru>.

Бүгінгі — Бүгінгі ауызекі қазақ тілінің мультимедиалық корпусы. <https://project6817623.tilda.ws>.

Военнопленные — Военнопленные-казахстанцы Второй мировой войны. Республиканская база данных. <http://tutkyn.kz>.

Гагарина & Корниенко 2020 — Гагарина Д.А., Корниенко С.И. Digital Humanities: образовательный ландшафт. Электронный научно-образовательный журнал “История”. 2020. № 3 (89). С. 1-15.

DH CLOUD — Канал DH CLOUD на YouTube. <https://www.youtube.com/@dhcloud>.

Круглый — Круглый стол Digital Humanities в Центральной Азии. <https://ca.dhcloud.org>.

Қазақстан — Қазақстан Республикасы Президентінің Архиві. <https://archive.president.kz>.

Номенклатурные — Номенклатурные кадры советского Казахстана. <https://archive.president.kz/ru/bd-nomenklaturnye-kadry-sovetskogo-kazakhstan>.

Орехов 2019 — Орехов Б. В. Башкирский стих XX века. Корпусное исследование. СПб.: Алетейя, 2019. 344 с.

Подрядчикова — Подрядчикова М. От кыргызского эпоса до ChatGPT: что обсуждали на круглом столе “Digital Humanities в Центральной Азии” // Системный Блокъ. <https://sysblok.ru/philology/ot-kyrgyzskogo-jeposa-do-chatgpt-cto-obsuzhdali-na-kruglom-stole-digital-humanities-v-centralnoj-azii>.

Системный — Системный Блокъ. <https://sysblok.ru>.

Пространство — Пространство соучастия // Благотворительный фонд Владимира Потанина. <https://fondpotanin.ru/activity/specprjects/prostranstvo-souchastiya>.

Alliance — Alliance of Digital Humanities Organizations (ADHO). <https://adho.org>.

Berry 2012 — Berry, David M. "Understanding Digital Humanities". Palgrave Macmillan, 2012.

Burdick et al. 2012 — Burdick Anne, Drucker Johanna, Lunenfeld Peter, Presner Todd, и Schnapp Jeffrey. "Digital_Humanities". MIT Press, 2012.

centerNet — centerNet. <https://dhcenternet.org>.

Қазақстандағы — Central Asian Analytical Network. <https://www.caa-network.org>.

Central — CLARIN-ERIC. <https://www.clarin.eu>.

DARIAH — DARIAH-EU. <https://www.dariah.eu>.

Digital — Digital Humanities Course Registry. <https://dhcr.clarin-dariah.eu>.

DH CLOUD — DH CLOUD. <https://dhcloud.org>.

Gold 2012 — Gold Matthew K. (ред.). "Debates in the Digital Humanities". University of Minnesota Press, 2012.

Inner — Inner and Central Asian Studies Dissertation Reviews. <https://www.facebook.com/groups/1424533484431370>.

Lingvodoc — Lingvodoc 3.0. <http://lingvodoc.ispras.ru>.

Manas — Manas Discovery: всё о кыргызском эпосе. <https://manas-discovery.kg>.

McCarty 2005 — McCarty Willard. "Humanities Computing". Palgrave Macmillan, 2005.

Moretti 2007 — Moretti Franco. "Graphs, Maps, Trees: Abstract Models for Literary History". Verso, 2007.

NazzAR — NazzAR. <https://nazzar.uz>.

O'zbek — O'zbek tili korpusi. <https://uzbekcorpus.uz>.

Ramsay 2011 — Ramsay Stephen. "Reading Machines: Toward an Algorithmic Criticism". University of Illinois Press, 2011.

Raqamli — Raqamli Tarix. <https://raqamlitarix.uz>.

Schreibman et al. 2004 — Schreibman Susan, Siemens Ray, и Unsworth John. "A Companion to Digital Humanities". Blackwell, 2004.

Terras 2013 — Terras Melissa, Nyhan Julianne, и Vanhoutte Edward. "Defining Digital Humanities: A Reader". Ashgate Publishing, Ltd., 2013.

РЕЦЕНЗИЯ

Дарья Артемьева

ДЖУЛИЯ ТОМПСОН КЛЯЙН «МЕЖДИСЦИПЛИНАРНЫЕ ЦИФРОВЫЕ ГУМАНИТАРНЫЕ НАУКИ: РАБОТА С ГРАНИЦАМИ В РАЗВИВАЮЩЕЙСЯ СФЕРЕ»

Монография Джули Томпсон Кляйн «Междисциплинарные цифровые гуманитарные науки: работа с границами в развивающейся сфере» («Interdisciplining Digital Humanities: Boundary Work in an Emerging Field») рассматривает цифровые гуманитарные науки как междисциплинарную область. В том числе исследуется *работа с границами* (boundary work) в этой области, то есть создание, критика или укрепление границ между областями знания. Фокус на анализе цифровых гуманитарных наук как междисциплинарной области отличает эту работу от других похожих работ, таких как, например, *Companion to Digital Humanities*. Автор проводит комплексный анализ цифровых гуманитарных наук с использованием приемов социологического, историографического и риторического анализа. Помимо этого, отдельное внимание уделяется происхождению и развитию названия области *Digital Humanities*, поскольку названия являются некими историческими ориентирами, которые определяют лакуны, а также разграничивают, что являются касательным или ортогональным по отношению к области [Klein 2015. С. 7]; основная часть этого исследования сосредоточена в первой главе, однако автор на протяжении всей книги регулярно возвращается к этому вопросу, а также приводит в начале каждой главы набор ключевых слов для лучшего понимания коннотаций имени [Klein 2015. С. 8].

Дарья Сергеевна Артемьева
НИУ «Высшая школа экономики»
artemjevads@yandex.ru

Основная часть монографии состоит из шести глав: *междисциплинарный подход* (interdiscipling), *определение* (defining), *институционализация* (institutionalization), *профессионализация* (professionalizing), *обучение* (educating) и *коллаборация и вознаграждение* (collaboration and rewarding). Все главы также делятся на несколько разделов и содержат синопсис о развитии конкретного аспекта с зарождения цифровых гуманитарных наук до современного их состояния, поскольку монография предназначена для широкой аудитории с разным уровнем подготовки: как для профессионалов в области цифровых гуманитарных наук, так и для исследователей, преподавателей и студентов в области междисциплинарных исследований. Также, поскольку значительная часть дискуссии о цифровых гуманитарных науках происходит на онлайн-ресурсах, в конце каждой главы приводится список ссылки в порядке их упоминания. Заключительный раздел *снабжение источниками* (resourcing) написан Энди Энджелом (Andy Engel) и представляет собой список полезных ресурсов для исследователей и преподавателей в области цифровых гуманитарных наук.

Первая глава «Междисциплинарный подход» (*Interdiscipling*) делится на три раздела: в первом, «Концептуальный словарь междисциплинарности» (*A Conceptual Vocabulary of Interdisciplinarity*) рассматривается основная терминология. В первую очередь демонстрируется разница между междисциплинарностью (interdisciplinarity) и *мультидисциплинарностью* (multidisciplinarity): мультидисциплинарность – это сопоставление разных дисциплинарных вкладов (*disciplinary inputs*), не объединенных вокруг какого-то одного вопроса или проблемы, а, например, представленных в алфавитном порядке, в то время как междисциплинарность – это объединение методов, подходов и/или теорий и др. из двух и более дисциплин, что обеспечивает более полное понимание рассматриваемой проблемы [Klein 2015. С. 15].

Междисциплинарность, в свою очередь, делится на:

1. методологическую и теоретическую *Методологическая междисциплинарность* использует методы, концепции другой дисциплины для тестирования гипотезы или ответа на исследовательский вопрос, таким образом улучшая результаты, в то время как теоретическая разрабатывает более общие *концептуальные рамки* (conceptual frameworks). Цифровые гуманитарные науки обычно считаются методологическими по своей природе [Klein 2015. С. 17].

2. инструментальную и критическую *Инструментальная междисциплинарность* направлена на создание продукта или удовлетворение прагматических потребностей, *критическая* – подвергает сомнению доминирующую структуру знания и образования с целью их преобразования [Klein 2015. С. 18].

3. горизонтальную и вертикальную *Горизонтальная интердисциплинарность* является аддитивной, она связывает такие области, как история или литература, без фундаментального изменения их дисциплинарной структуры или логики. *Вертикальная интердисциплинарность*, напротив, бросает вызов дискурсивным категориям или формальным свойствам поля [Klein 2015. С. 20]

В конце раздела рассматриваются еще два важных термина: *трансдисциплинарность* (transdisciplinarity) и *интерпрофессиональный подход* (interprofessional). *Трансдисциплинарность* – это всеобъемлющий синтез, переходящий дисциплинарные границы; *трансдисциплинарность* демонстрирует разницу между двумя метафорами междисциплинарности, предложенными Фондом Наффилда: *строительство мостов* (bridge building) и *реструктуризация* (restructuring). Эти метафоры демонстрируют две точки зрения на предназначение цифровых гуманитарных наук: строительство мостов между гуманитарными науками и технологиям в интересах культурного и исторического наследия или полная реструктуризация гуманитарных наук. [там же]

Интерпрофессиональный подход – это сотрудничество представителей разных специальностей, таких как медицинских или социальных работников, педагогов, юристов и инженеров. Однако поскольку цифровые гуманитарные проекты обычно работают с традиционными академическими областями, даже если вовлечены представители разных профессий, *междисциплинарный* остается наиболее удобным термином. [Klein 2015. С. 22]

Во втором разделе, «Скрещивание» («Cross-hatching») исследуется природа и развитие междисциплинарности в гуманитарных науках. Базовые идеи о холизме развивались еще в древнегреческой философии и передавались впоследствии через историю либерального образования; однако интерес к синтетическим теориям, таким как структурализм и семиотика, а также к изучению медиа, начинает развиваться к концу 20-го века. Также, следует отметить тот факт, что распространенное мнение о том, что цифровые гуманитарные науки объединяют только гуманитарные науки и технологии, не учитывают увеличивающееся количество «пере-

сечения границ» между гуманитарными и социальными науками [Klein 2015. С. 23-25]

Далее для более полного понимания междисциплинарности в цифровых гуманитарных науках рассматривается междисциплинарность в смежных сферах. На примере американистики демонстрируется, как со временем в дисциплине может происходить *сдвиг идентичности* (identity shift) [Klein 2015. С. 26]; на примере медиаисследований – как увеличение сферы применения может увеличивать объем «скрещивания» между разными дисциплинами [Klein 2015. С. 31]. В то же время междисциплинарная область, такая как цифровые гуманитарные науки, может так же «застыть» до такой степени, что у нее будет свой собственный «канон», а также институты – журналы, организации, исследовательские центры. Сравнение со смежными областями, таким образом, демонстрирует, что цифровые гуманитарные науки не могут быть представлены одной конкретной практикой. Они не существуют в единственном пространстве, но расположены внутри разных дисциплин. Эта область является мультидисциплинарной по своему охвату, но междисциплинарной – по подходу к интегративной работе и практике сотрудничества [Klein 2015. С. 32].

Вторая глава «Определение» («Defining») так же делится на три раздела: «Объявление» (Declaring), «Вовлечение» (Engaging), и «Размещение» (Locating). В первом разделе предлагается рассматривать *определение* (definition) не только как высказывание, но и как сам акт определения: анализ шести высказываний, которые автор называет основными, дает более подробное представление о том, как определяется область цифровых гуманитарных наук [Klein 2015. С. 38].

В качестве первого высказывания приводится цитата из введения к «A Companion to Digital Humanities»¹; она используется в качестве предпосылки для анализа истории развития цифровых гуманитарных наук как сферы: так, цифровые гуманитарные науки исторически идентифицируются с лингвистикой и компьютерно-опосредованным изучением текста. Однако с появлением персональных компьютеров задача цифровых гуманитарных наук сменилась от простой обработки материала к сетевой работе: увеличившаяся скорость работы компьютеров позволила обрабатывать и обмениваться большим объемом данных. Цифровые гуманитарные науки также развиваются вместе с развитием киберпространства; этот сдвиг очевиден в таких областях исследования, как цифровая криминалистика, исследования видеоигр, а также исследований с

использованием большого объема данных и визуализаций [Klein 2015. С. 40-41].

В качестве второго утверждения, отражающего еще одну контрольную точку в истории цифровых гуманитарных наук, приводится цитата из вступительной статьи от редакции журнала *Digital Humanities Quarterly*, в которой предлагается рассматривать цифровые гуманитарные науки как гибридную сферу, пересекающую границы дисциплинарности и между теорией и практикой; таким образом, предлагается не давать определения цифровым гуманитарным наукам, а позволить этому определению «родиться» из практики [Klein 2015. С. 40-41].

В качестве третьего высказывания приводится цитата Бретта Бобли, директора отдела Цифровых гуманитарных наук (**Office of Digital Humanities**) Национального Фонда гуманитарных наук, в которой утверждается, что цифровые гуманитарные науки – это зонтичный термин для разных проектов, так или иначе связанных с гуманитарными науками или технологиями. Главная задача ОЦГ – поддерживать инновационные проекты в области цифровых гуманитарных наук; его появление рассматривалось как важная точка в «признании» цифровых гуманитарных наук как отдельной области. [Klein 2015. С. 43].

Четвертое высказывание принадлежит Джоанне Друкер: она предлагает отличать «цифровые гуманитарные науки», характеризующиеся философией математизации, от «спекулятивных вычислений» (*speculative computing*), характеризующиеся философией эстетики. Таким образом, «цифровые гуманитарные науки» приоритезируют принципы формальной логики и объективности, проявляющиеся в использовании количественных методов, автоматической обработке текста, дихотомии субъекта и объекта. «Спекулятивные вычисления» же приоритезируют субъективность и эстетические принципы; проявляется это в том числе в использовании инструментов визуализации [Klein 2015. С. 45-46].

Пятое высказывание – цитата из *Digital Humanities Manifesto 2.0*, в которой утверждается, что цифровые гуманитарные науки – это набор конвергентных практик, которые исследуют мир, в котором печать больше не является исключительным или нормативным средством производства или распространения знаний, а медиа и цифровые инструменты изменили процесс получения и распространения знаний в искусстве, гуманитарных и социальных науках. Однако цифровые гуманитарные науки являются скорее естествен-

ным развитием новой междисциплинарной области, нежели чем полноценной новой культурой [Klein 2015. С. 47-48].

Высказывания 1–3 таким образом отражают ключевые точки в институционализации новой области: *определяющая литература* (defining literature), специализированный журнал, институционализованная финансовая поддержка. Высказывания 4-5, в свою очередь, фиксируют дискуссию о том, как определять цифровые гуманитарные науки.

Шестому высказыванию, в качестве которого рассматривается типология Патрика Свенссона о пяти модусах взаимодействия (mode of engagement)² между гуманитарными науками и информационными технологиями, посвящен второй раздел главы, «Вовлечение» (Engaging). Свенссон выявляет пять основных модусов взаимодействия: информационные технологии как инструмент, как объект, как экспериментальная лаборатория, как средство выражения (expressive medium) и как место проведения акции (activist venue) [Klein 2015. С. 49]. В первом модусе взаимодействия – инструменте – цифровые гуманитарные науки проявляют эпистемическую вовлеченность в инструменты и методологию, а также фокусируются на анализе текста [там же]. Во втором модусе взаимодействия – объекте – цифровая культура и трансформативные эффекты новых технологий коммуникации становятся объектами изучения; разработка новой методологии изучения же не является приоритетной для этого модуса. В третьем модусе – экспериментальной лаборатории – исследовательские центры и лаборатории становятся местами тестирования инструментов и модификации датасетов. Цифровые платформы, такие как *Second Life*, также могут выступать в качестве виртуальной площадки для экспериментов [Klein, 2015, с 52]. В четвертом модусе – *средстве выражения* – цифровизация обеспечивает доступ к разным видам контента и медиа; инновации в технологиях же «взращивают» новые формы искусства и литературы. Тем не менее, как отмечает Свенссон, это взаимодействие обычно происходит на уровне объекта изучения, а не производства выразительных медиа [Klein 2015. С. 53]. В пятом модусе – месте проведения акции – технологии используются, чтобы призвать к решению какой-либо проблемы; так, вопросы социальной справедливости занимают видное место в исследованиях культуры и медиа [Klein 2015. С. 54].

В последнем разделе, «Размещение» (Locating), в контексте трех дисциплин – английского языка, истории и археологии – рассматри-

ваются изменения методологии исследования под влиянием новых технологий.

В третьей главе, «Институционализация» («Institutionalization»), рассматривается процесс институционализации, то есть процесс превращения каких-либо отношений в институты; категории знания так же можно рассматривать как институты в том смысле, что это набор некоторых созданных и поддерживаемых отметок в культурной среде, которые позволяют членам коллективов определять свое поведение и положение в пространстве [Klein 2015. С. 68].

В первом разделе, «Институционализация междисциплинарных областей» (Institutionalizing Interdisciplinary Fields), рассматривается концептуальная основа институционализации междисциплинарности. Одним из аргументов критики институционализации междисциплинарных областей является то, что институционализация области, которая ставит под вопрос дисциплинарные основы, противоречит сути этой области. В то же время, без институционализированной поддержки локальные междисциплинарные проекты приходят в упадок; либо выживают в качестве «выставочных образцов» внутри исторически детерминированной дисциплинарной структуры высшего образования [Klein 2015. С. 68-69].

Кляйн предлагает идею «критической массы»: количества ядерного топлива, необходимого, прежде чем начнется цепная реакция. Факторами критической массы в междисциплинарных областях могут являться, например, адекватное количество учебных программ и исследовательских центров, соответствующее финансирование обучения и исследования, доступ к библиотечным и вычислительным ресурсам и др. (хотя, как именно определять уровень адекватности, не очень понятно). Таким образом, в качестве институционализированных структур можно рассматривать в том числе медиа-лаборатории, курсы и т.д. [Klein 2015. С. 71-72]. Список подобных междисциплинарных структур, выделен в одноименный раздел, поскольку они достаточно разнообразны.

В следующем разделе, «Аффилиация» («Affiliating»), рассматриваются сферы, в которых могут работать исследователи цифровых гуманитарных наук: чаще всего это английский язык и история, германские языки и литература, искусствоведения, философия, антропология; из междисциплинарных областей упоминаются медиэвистика, исследования кино и медиа. Среди профессий чаще всего упоминаются образование и библиотечное дело. Кроме того, исследование рынка вакансий также может демонстрировать, в каких

сферах цифровые гуманитарные науки укореняются: так, чаще всего ДН-специалисты требуются в более «традиционных» областях, таких как например библиотечарское дело, для развития их цифрового потенциала [Klein 2015. С. 75-76].

Последний раздел – «Централизация цифровых гуманитарных наук» (Centering Digital Humanities) исследует феномен исследовательских центров; можно сказать, что это самый распространенный тип институциональной структуры в цифровых гуманитарных науках. Несмотря на название «центр», большинство из них имеют весьма скромные размеры и представляют некоторое «теневое пространство» внутри организации [Klein 2015. С. 79]. Программные заявления исследовательских центров часто совпадают с целями центров гуманитарных наук; таким образом, создается пространство для исследования и обмена опытом. Помимо этого, они обладают специфическими междисциплинарными целями, в том числе создание пространства для взаимодействия между представителями разных дисциплин, разработка новых педагогических и исследовательских инструментов [Klein 2015. С. 80-81]. Также они могут предоставлять возможности для профессионального развития (там же). Однако подобные центры за счет своей междисциплинарной природы могут быть уязвимы при сокращениях бюджета [Klein 2015. С. 86].

В четвертой главе, «Профессионализация», рассматривается вопрос профессионализации цифровых гуманитарных наук. Под профессионализацией в данном случае понимается процесс, в результате которого группа приобретает и поддерживает контроль над социальной реальностью [Klein 2015. С. 90]. В качестве примеров механизмов профессионализации академической среды можно привести *специализированный дискурс* (specialized discourse), *стандарты практической работы* (standards of practice), *документов об образовании* (educational credentials) и др. Рассматриваются два основных механизма профессионализации в цифровых гуманитарных науках: формирование *сообществ практики* (communities of practice) и *научные публикации* (scholarly publications) [Klein 2015. С. 91].

Сообщества практики рассматриваются в первом разделе главы. Подобные сообщества образуются в результате взаимодействия людей, объединенных общим интересом, ремеслом, или профессией. По мере того, как они учатся друг у друга, они создают репертуар *общих знаний и ресурсов, норм и методов* (shared norms and best practices). Этот репертуар затем становится основой формирова-

ния *платформы* новой области исследования. В информационных технологиях *платформой* называется аппаратная архитектура или программная среда, позволяющая работать программному обеспечению. В данном же случае социальная архитектура является ключевой. Процесс формирования сообществ практики является двунаправленным; он направлен как в сторону расширения – например, к ДН-проектам могут присоединяться работники других сфер: музеев, архивов, библиотек – так и в сторону сужения в направлении конкретной специализации [Klein 2015. С. 91-93].

Во втором разделе главы, «Tweet, Loc.Cit», рассматриваются форматы научных публикаций в сфере цифровых гуманитарных наук. Само название раздела одновременно является метафорой к тому, как меняются стандарты научной коммуникации в этой области: посты в твиттере, блоги, вебсайты являются объектами повышенного внимания. У цифровых форм публикации есть свои преимущества: быстрая скорость публикации, бóльшая аудитория и уровень цитирования; однако они сталкиваются с критикой и оппозицией со стороны традиционной системы рецензирования [Klein 2015. С. 97].

Цифровая периодика, в том числе в области цифровых гуманитарных наук, появляется в 90-х – 00-х годах; цифровые монографии начали появляться позже, однако к 10-м годам цифровые публикации начали активно развиваться [Klein 2015. С. 98-99]. Также, помимо журналов и монографий, одним из форматов научной публикации можно считать *тематические коллекции* (thematic research collections). Тематические коллекции – это цифровые агрегаторы первоисточников и тематических материалов, относящихся к одной и той же теме [Palmer 2004]. Тематика может быть сколько угодно широкой: например, целью *The Rosetti Archive* является размещение всех текстов и изобразительных материалов Данте Габриэля Россетти, дополненных письмами, библиографией, биографией и др. Такие коллекции создаются вследствие коллаборативной работы с библиотеками, музеями и издательствами; эти коллаборации являются одним из междисциплинарных аспектов цифровых гуманитарных наук. Кроме того, разработка инфраструктуры контента также требует совместной работы представителей разных сфер, как гуманитарных, так и технических. Авторская идентичность также меняется от индивидуального к «коллективному» голосу. Концепт авторства в цифровой среде сложнее, поскольку работа над проектами требует разнопрофильных навыков [Klein 2015. С. 100-101].

Социальная природа открытых сред приводит к еще одной форме пересечения границ: поскольку результаты и процесс исследования распространяются через медиа, граница между профессионалом и любителем стираются, что приводит к децентрализации авторитета и *распределенном знании* [Klein 2015. С. 104]. *Распределенное знание* – это такое тип знания, которым обладает некоторое сообщество, даже если отдельный член этого сообщества им не обладает [Ågotnes & Wáng 2017].

В пятой главе, «Образование» («Educating»), рассматривается вопрос образования в сфере цифровых гуманитарных наук. В первом разделе рассматриваются особенности учебного плана; на примерах различных образовательных программ рассматриваются такие особенности, как вводные курсы, баланс между гуманитарными науками и техническими навыками, а также роль теоретического и критического анализа.

Вводные подготовительные курсы являются необходимыми, поскольку помогают как студентам, так и профессорско-преподавательскому составу определить дисциплинарные и междисциплинарные границы области. В общем и целом, они имеют тенденцию к мультидисциплинарности, поскольку должны представить теоретический обзор, а также методологию и инструментарий [Klein 2015. С. 111]. Соблюдение баланса между гуманитарной и технологической составляющей также является весьма важным: с одной стороны, необходимость навыков, например, в программировании сложно недооценить; с другой стороны, технологии не должны «перевешивать» собственно глубину гуманитарной составляющей. В этом случае сложно выделить какое-то единое решение: некоторые образовательные программы требуют от студентов опыта работы с технологиями, другие – «открытый ум и стремление к изучению новых методик» [Klein 2015. С. 115-116]. Наконец, критический анализ также играет важную роль, поскольку наличие технических навыков само по себе не должно замещать понимание того, как проекты должны быть структурированы. Необходим навык «думать вместе и против компьютера», а не использовать его исключительно в качестве инструмента [Klein 2015. С. 118].

Второй раздел посвящен *цифровой педагогике* (digital pedagogy) и ее связи с междисциплинарностью. Опираясь на 10 принципов редизайна образовательных учреждений в цифровую эпоху, предложенных Кэти Дэвидсон (самообучение, горизонтальные структуры, коллективный авторитет, децентрализованная педагогика,

социально-информационное обучение, открытый доступ к образованию, интерактивность, «пожизненное» обучение, концепция института как мобилизующей сети, масштабируемость), Кляйн показывает, что цифровое обучение является по своей сути междисциплинарной практикой: например, *социально-информационное обучение* (networked learning) подчеркивает важность сотрудничества и взаимодействия, а «пожизненное» обучение признает, что скорость изменчивости цифрового мира требует умения работать с новыми концепциями и быстро адаптироваться к ним [Klein 2015. С. 120-122].

Цифровое и междисциплинарное обучение имеют также и другие общие черты, например, *проектное обучение* (project-based learning): работая над групповым проектом, студенты получают новые знания через исследование и реальное «занятие» какой-либо темой. Возможности, предоставляемые цифровыми платформами, также изменяют то, что квалифицируется как знание: так, гейм-дизайн проекты или онлайн выставка могут оцениваться наравне с более традиционными эссе и научными работами [Klein 2015. С. 122-123]. Цифровые гуманитарные науки и междисциплинарные исследования также развивают у студентов схожие навыки, такие как способность решать сложные вопросы, развитое критическое мышление; на более высоких уровнях обучения, студенты также приобретают навык, например, сравнивать и сопоставлять разные методы и подходы [Klein 2015. С. 124]. В совокупности все эти характеристики идентифицируют ДН-преподавание и обучение как междисциплинарную практику.

В последнем разделе рассматриваются стратегии взаимодействия образовательных программ с институциональными структурами. Выделяются две основных стратегии: ассимиляция и разграничение. Ассимиляция предполагает коллаборацию и распространение существующих исследований в области гуманитарных наук; разграничение же уделяет особое внимание новым методологиям и новым исследовательским вопросам. Однако, никакую стратегию нельзя назвать превосходящей другую. В зависимости от местной специфики, одна или другая стратегия может оказаться выигрышной; кроме того, стратегии могут смешиваться [Klein 2015. С. 127-130].

В **шестой главе** рассматриваются две последние важные для цифровых гуманитарных наук темы: коллаборация и *культура признания* (culture of recognition). Коллаборации посвящен первый раздел главы; несмотря на то, что это нетипичный для гуманитарных

наук формат работы, она используется в других областях, и многие считают, что сотрудничество является необходимым для цифровых гуманитарных наук. У организации такого формата работы есть свои проблемы, поэтому грамотный лидер является важным участником команды: лидер должен не только заниматься организационными и финансовыми задачами, но и быть посредником в передаче информационных потоков, а также справляться с конфликтующими рабочими культурами участников проекта из разных сфер. Кроме того, гетерогенность междисциплинарных команд является одновременно их сильной стороной, и потенциальным источником конфликта [Klein 2015. С. 136-138]. Помимо этого, важную роль в коллаборативной работе играет «межъязыковая» коммуникация, а также возможность разработки общего словаря [Klein 2015. С. 141]. «Глубокая междисциплинарность» требует использования новых способов и средств коммуникации, умение на новых гибридных языках, готовность учиться и уплощение иерархии в управлении проектами [Klein 2015. С. 127-130]. Коллаборативная работа также должна отвечать этическими принципам, к которым относятся *интеллектуальная щедрость* (intellectual generosity) – признание работы других людей, выраженное в явном для них виде; *интеллектуальная уверенность* (intellectual confidence) – понимание того, что каждый вносит важный вклад в рабочий процесс; *интеллектуальная скромность* (intellectual humility) – понимание того, что субъективные знания всегда частичны, и могут быть дополнены со стороны; *интеллектуальная гибкость* (intellectual flexibility) – способность менять свою точку зрения, основываясь на новых знаниях, полученных от других людей; а также *интеллектуальная целостность* (intellectual integrity) – ответственное участие, которое служит основой для развития доверия между членами команды. Из этих пяти принципов первые два лежат на уровне индивидуальной ответственности, а остальные три относятся к уровню ответственности команды [Klein 2015. С. 142-143].

Второй раздел главы посвящен культуре признания, поскольку цифровые гуманитарные науки, как и междисциплинарные исследования, выходят за рамки общепринятых критериев оценивания; более того, они не могут быть измерены одной моделью оценивания [Klein 2015. С. 145]. Также рассматриваются четыре основные проблемы, выделенные Аделин Ко: «обучить свою аудиторию»: необходимость не только объяснить суть работы, но и обосновать область исследования; «разнообразие цифровых проектов»: новые технологии и цифровая публикация для гуманитарных наук

все еще относительно незнакомы; «фиксация своей роли в цифровых проектах»: коллаборативная природа многих ДН-проектов часто создает проблемы в том, чтобы определить индивидуальные обязанности и вклад каждого; «объяснение изменяющихся форм рецензирования»: в случае развивающейся междисциплинарной области определить эксперта, чья компетентность соответствовала бы проблемному пространству работу может быть сложно; одним из решений этой проблемы является открытие возможности рецензирования для широкой аудитории [Klein 2015. С. 145].

В последнем разделе рассматривается потенциал развития цифровых гуманитарных наук в будущем. Современная информационная эпоха является эрой культурно-исторических преобразований, как, например, промышленная революция: сообщества становятся более подвижными, а исследуемые вопросы не могут быть сведены к одной дисциплине или институту. Поэтому, рассматривая концепцию Энн Балзамо о «двойной эффективности» и Патрика Свенссона о «двойной принадлежности» цифровых гуманитарных наук, Кляйн предлагает расширить их до понятия «тройной эффективности»: внутри и между дисциплинарными областями и профессиями, внутри и между институтами, внутри и между ДН-организациями [Klein 2015. С. 152-154]. Цифровые гуманитарные науки, таким образом, продолжают развиваться; однако работа по определению совпадений, пересечений и отклонений будет так же продолжаться [Klein 2015. С. 155].

В заключительной части приводится репрезентативная выборка ресурсов, полезных для исследований/преподавания в области цифровых гуманитарных наук. Ресурсы делятся на два типа: *агрегаторы* (Aggregators), которые собирают в одном месте большое количество ресурсов, и обновления (Keeping Up to Date); ресурсы, которые не подходят под эти два типа, приводятся отдельно со ссылкой, описанием и указанием уникальных возможностей. Также предлагается пять стратегий поиска источников [Klein 2015. С. 157-158], или *спектров* (SPECTRA), чтобы облегчить ориентацию в среди этих источников. К этим спектрам относятся: «таксономический vs. фолксономический» (организацию и менеджмент агрегатора; таксономический подход – это структурированный, централизованный, иерархичный подход, фолксономический – управляется сообществом пользователей, а организационные структуры построены на генерируемых пользователями метках), «уровень специализации» (дисциплинарная специализация), «глубина vs ширина» (количество ресурсов, с которыми взаимодействует сайт, и дис-

циплинарное разнообразие этих ресурсов), «оперативность» (насколько часто и быстро обновляется информация), «цель» (скорее напоминание о том, что надо принимать во внимание намерения владельцев ресурса, изучая источники, которые они используют) [Klein 2015. С. 158-163].

Таким образом, монография представляет собой разносторонний обзор разных идей и концепций, которые составляют современные цифровые гуманитарные науки. Наравне с научными публикациями используются также материалы, например, блогов и комментариев, представляющих мнение профессионального сообщества «цифровых гуманитариев»; использование такого рода материалов вместе с историографическим анализом помогает составить подробную картину диахронического развития и современного состояния цифровых гуманитарных наук, а также их связи с другими, не обязательно «родственными», дисциплинами. Благодаря этому, работа является ценным источником теоретического материала как о цифровых гуманитарных науках, так и о междисциплинарности. Финальная секция с ресурсами также представляется достаточно полезной, поскольку в ней приводятся в том числе ресурсы, которые могут с трудом индексироваться простым поиском в браузере. Однако, поскольку информация представлена достаточно сжато, по всей видимости предполагается некоторый уровень теоретических и/или профессиональных знаний у читающего, без которых полноценное понимание может быть несколько затруднено.

Литература

Ågotnes & Wáng 2017 — Ågotnes T., Wáng Y.N. Resolving distributed knowledge // *Artificial Intelligence*. 2017. Vol. 252. P. 1–21.

Klein 2015 — Klein J.T. *Interdisciplining Digital Humanities: Boundary Work in an Emerging Field*. University of Michigan Press, 2015.

Palmer 2004 — Palmer C.L. *Thematic research collections // A Companion to Digital Humanities* / ed. Schreibman S., Siemens R., Unsworth J. Oxford: Blackwell, 2004.

Адрес редакции «Цифровые гуманитарные исследования»:
nevmenandr@gmail.com
199034 Санкт-Петербург, наб. Макарова, д. 4
Оформление обложки И. Гурьянов

Подписано в печать 14.11.2024. Формат 60×84 1/16.
Бумага офсетная. Печать офсетная.
Тираж 200 экз.