



ФОРМИРОВАНИЕ НАВЫКОВ ИНТЕРПРЕТАЦИИ ДАННЫХ ЦИФРОВОГО АНАЛИЗА ТЕКСТА В ФИЛОЛОГИЧЕСКОМ ОБРАЗОВАНИИ

В.Е. Шахова

*аспирант кафедры литературы и русского языка ГИ
филиала Северного (Арктического) федерального университета им. М.В. Ломоносова
Северодвинск, Россия
v.shahova@narfu.ru*

DOI: <https://doi.org/10.5281/zenodo.18101390>

***Аннотация.** Вызовы цифровой революции обострили проблемы образования в системе гуманитарных наук, включая филологию. Ключевой вызов для цифрового филолога – не освоение алгоритма работы с компьютерной программы, а интерпретация полученных результатов. В статье представлен методический опыт интеграции цифровых инструментов в учебный процесс курса «Цифровые технологии в предметной области» (филологическое образование). Раскрываются особенности обучения студентов «чтению» и нарративизации данных, полученных в ходе работы с программным пакетом Stylo и сервисом Voyant Tools. Обсуждаются типичные ошибки интерпретации и методические приемы, позволяющие интегрировать компьютерные методы в традиционное филологическое знание.*

***Ключевые слова:** цифровое литературоведение, современные технологии преподавания литературы, Voyant Tools, Stylo, интерпретация данных, Data Storytelling.*

Стремительная цифровизация гуманитарных наук предоставляет филологу мощный аналитический инструментарий. Однако ключевым вызовом становится не просто получение данных, а их содержательная интерпретация. Студенты-филологи, осваивающие в рамках курса «Цифровые технологии в предметной области» такие инструменты цифрового анализа текста, как Stylo и Voyant Tools, часто сталкиваются с проблемой «немых графиков»: они могут получить визуализацию, но испытывают трудности с ее переводом в убедительный филологический нарратив, подменяя анализ простым описанием картинки. Следовательно, цель нашей работы – формирование у студентов компетенции дата-сторителлинга – умения «оживлять» полученные в ходе компьютерного анализа данные, извлекать из них скрытые смыслы и строить на их основе аргументированное исследовательское высказывание.

Как утверждают современные российские исследователи «...за последние 15-20 лет это направление (цифровая гуманитаристика – В.Ш.) прошло путь от спорного, кажущегося радикальным нового подхода до по-прежнему нового, но уже институционализированного и обретшего устойчивые контуры исследовательского и преподавательского поля» [Степанчук, с. 665]. При этом обнаруживается проблема внедрения методов цифровой гуманитаристики в традиционное преподавание гуманитарных дисциплин и проектирование



практической составляющей обучения. В связи с этим современный этап развития цифровой гуманитаристики характеризуется смещением фокуса с разработки инструментария к методологии интерпретации получаемых с помощью этого инструментария данных.

Если первоначальные исследования в области Digital Humanities, вслед за основополагающей работой Ф. Морретти о дальнем чтении, были сосредоточены преимущественно на возможностях квантитативного анализа, то современная научная дискуссия актуализирует проблему герменевтического перехода от данных к смыслам и подвергает критике «культ количественных методов» [3, с. 57]. Этот «интерпретативный поворот» ставит перед цифровым филологом-преподавателем новую сложную задачу: необходимо не только владеть техническими навыками работы с ПО, но и формировать у студентов особую «критическую оптику», т.е. способность задавать цифровым данным содержательные филологические вопросы, видеть за графиками и статистическими распределениями культурные и эстетические закономерности, а главное – встраивать полученные «инсайты» в традиционный нарратив филологического исследования, где количественные данные служат не заменой, а усилением качественного анализа и/или подтверждением сформулированной гипотезы.

Стоит отметить, что к числу типичных ошибок интерпретации цифровых данных относится абсолютизация количественных показателей, когда высокая частотность слова принимается за прямое доказательство смысловой значимости темы без последующего контекстуального анализа [5]. Для интеграции компьютерных методов в традиционное филологическое исследование продуктивным оказывается принцип «цифрового герменевтического круга» [1, с. 116], предполагающий не просто использование технологий как инструментов, а переосмысление самой природы филологического знания через взаимодействие человеческого понимания и машинной интерпретации [4]. В педагогическом процессе эффективен прием «триангуляции методов», когда перекрестная проверка результатов разными инструментами (в нашем примере это Stylo и Voyant Tools) учит студентов критически оценивать надежность данных и аргументировать свои выводы [6]. В рамках нашего курса для преодоления этой проблемы используется прием «интерпретационного комментария», развивающий принципы цифровой герменевтики [4; 6]. Его суть заключается в том, что студенты составляют развернутый комментарий к полученной визуализации (например, к облаку слов из Voyant Tools или дендрограмме из Stylo), структурированный по трем обязательным пунктам: 1) «Данные» (формальное описание того, что



изображено на графике), 2) «Гипотеза» (предположение, объясняющее выявленную закономерность), 3) «Верификация» (поиск подтверждений или опровержений гипотезы в самом тексте и историко-культурном контексте). Такой структурированный подход позволяет преодолеть простое описание «картинки» и целенаправленно формирует навык перехода от количественных данных к качественной интерпретации.

Подготовка современного филолога требует не только знакомства с цифровыми инструментами, но и формирования целостного понимания их места в исследовательском процессе. Этой задаче подчинен цикл лабораторных работ, выстроенных по принципу «от макроанализа к микроанализу» и от статистической верификации к филологической интерпретации. Ярким примером такой последовательности являются работы по стилометрическому анализу в Stylo и анализу текстовых структур в Voyant Tools. Stylo – это пакет (library/package) для языка статистического программирования R; это не самостоятельная программа, а специализированный набор скриптов и функций, который «учит» R проводить стилометрические исследования: рассчитывать меры стилистического сходства, строить дендрограммы, создавать сети и т.д.

На первом этапе студенты погружаются в логику макроанализа, где текст рассматривается как целостный стилистический объект. Цель данной лабораторной работы заключается в освоении методов количественной проверки гипотез о стилистическом сходстве и авторском стиле. Студенты заранее готовят корпус текстов (например, произведения разных авторов одного периода/жанра или одного автора разных периодов, или работают с предложенным преподавателем специальным корпусом текстов «арктической робинзонады» [7]), формулируя предварительную гипотезу. Пример исследовательской гипотезы: корпус текстов «арктической робинзонады», несмотря на внешнее сходство сюжета и тематики, является стилистически неоднородным. Мы предполагаем, что анонимные тексты статистически распределятся между текстами известных авторов, что может указывать на возможное авторство. Практическая часть включает строгий алгоритм: установка рабочей директории, загрузка корпуса текстов, настройка параметров анализа (MFW, метрика расстояния).

Ключевой исследовательский компонент в данном контексте – проведение серии экспериментов с варьированием параметров (количество анализируемых слов, метрика) и сравнение полученных дендрограмм (рис. 1). Такой подход настраивает студентов на критическую оценку валидности результатов, показывая, что вывод зависит от выбранных настроек.

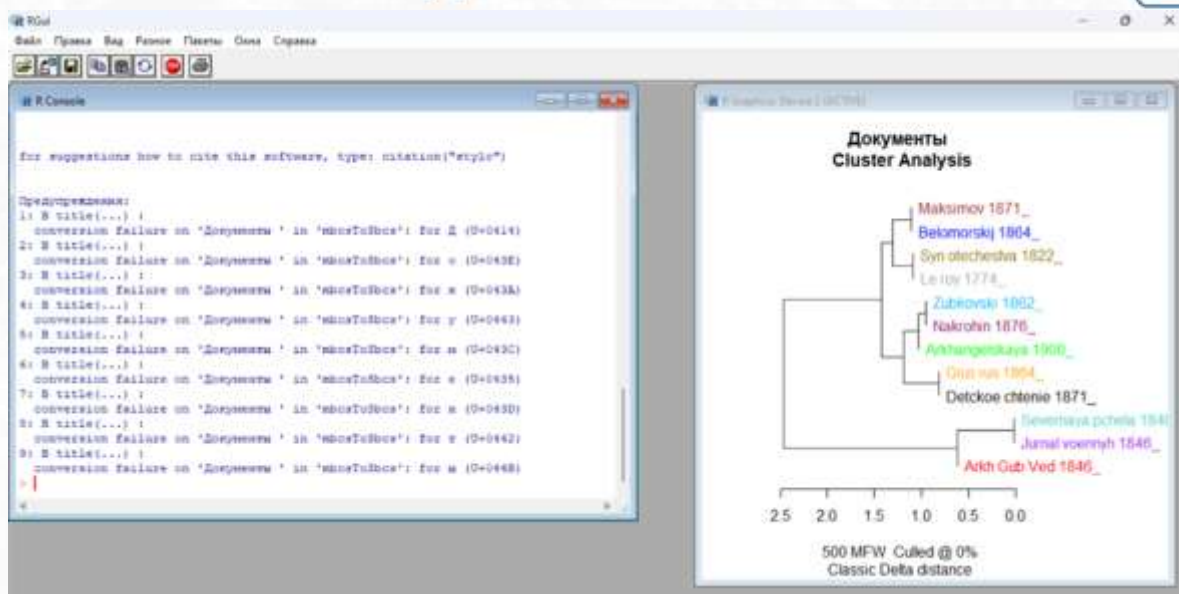


Рис. 1. – Пример кластеризации текстов «арктической робинзонады» XIX века с помощью пакета Stylo

Формируемая компетенция здесь заключается в понимании, что цифровые методы – это не «черный ящик», выдающий истину, а инструмент для верификации, требующий от исследователя методической рефлексии.

Второй этап переносит фокус на микроуровень текста, где количественные данные служат для выявления смысловых структур и закономерностей. Цель данного этапа – научиться «задавать вопросы» тексту с помощью визуализаций и работать с концептами, а не только с формальными единицами.

Работа строится как исследовательское расследование. Студенты работают с тем же или другим корпусом текстов, но теперь их задача – не классифицировать тексты, а понять их внутреннее устройство. Практическая часть посвящена последовательному освоению инструментов Voyant Tools: от создания облака слов и работы со стоп-словами до анализа трендов, коллокаций и визуализации связей («Мандала»).

Ключевой интерпретационный компонент – задание на составление таблицы коллокаций и анализ графиков «Тренды». Студенты должны не просто зафиксировать частотность, а предложить филологическое (литературоведческое) объяснение наблюдаемым картинам (рис. 2). Например, все тексты «арктической робинзонады» объединены единым историческим сюжетом (выживание русских промысловиков в Арктике в 1743-1749 гг.), однако смысловые доминанты этого сюжета в текстах разных авторов могут различаться. Мы предполагаем, что количественный анализ в Voyant Tools выявит статистически значимые различия в частотности лексики, относящейся



к разным тематическими блокам (например, «подготовка к выходу в море», «плавание», «выживание на острове», «починка избы», «охота» и т.п.).

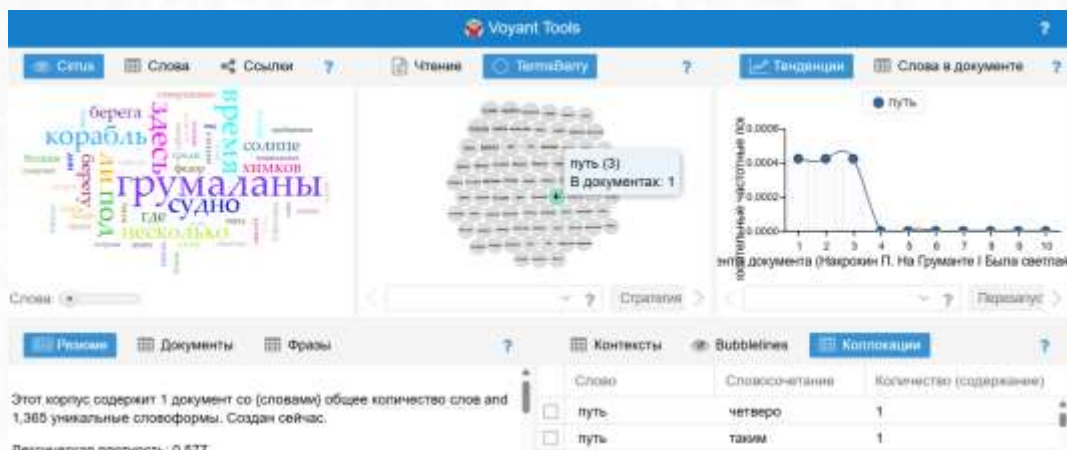


Рис. 2. – Пример цифрового анализа одного из произведений «арктической робинзонады» XIX века с помощью сервиса Voyant Tools

В ходе выполнения данной лабораторной работы формируется навык перехода от данных к интерпретации, то есть способность увидеть за графиком, диаграммой или списком слов – исторический, культурный или биографический контекст.

Например, цифровые паттерны, выявленные в Voyant Tools при анализе текста П. Накрохина «На Груманте» (1876), при их корректной интерпретации, служат ключом к пониманию духовной и традиционной основы поморского промысла.

Доминирование слова «Груманланы» в облаке слов указывает на определенный «статус» героев произведения, они не просто промысловики, а люди, прошедшие посвящение в сакральном пространстве Арктики. Номинацию «Груманлана» нельзя было получить по наследству или заработать в родной деревне, она обреталась и подтверждалась только одним способом – физическим и духовным пребыванием на Груманте (Шпицбергене). Само плавание на архипелаг было лишь подготовкой, «дорогой к храму», поэтому слово «путь» (на графике справа) концентрируется в начале и исчезает к финалу. Прибытие на остров и исчезновение темы пути означает начало процесса инициации. На Груманте обыденный промысел превращался в сакральный ритуал. Охота, борьба с холодом, выживание среди льдов – все это не просто работа, а форма диалога со стихией, которая является проявлением Божьей воли. Пройдя через это, человек становится частью избранного



братства, «посвященным», отмеченным особым знанием и опытом, недоступным тем, кто остался на «большой земле».

Представленный пример анализа показывает, как работа с Voyant Tools превращает статистические данные в литературоведческий нарратив, основанный на культурном контексте. Владение этим навыком позволяет выстроить целостную исследовательскую стратегию, в которой количественные методы и качественная интерпретация не противостоят, а взаимно обогащают друг друга.

Таким образом, представленные лабораторные работы являются не разрозненными модулями, а звеньями одной цепи. Так, стилометрический анализ дает «вид с высоты птичьего полета» – объективную, статистическую карту стилистических ландшафтов. А инструменты Voyant Tools предоставляют «лупу» для детального изучения особенностей каждого ландшафта.

Такой комплексный подход в цифровом анализе текста позволяет студенту на практике пройти полный цикл исследовательской работы: от выдвижения гипотезы и ее статистической проверки (Stylo) к углубленному качественному анализу и интерпретации выявленных закономерностей (Voyant Tools), что и составляет суть современной цифровой грамотности филолога.

Литература

1. Васюков В.Л., Шульга Е.Н. Цифровая онтология и цифровая герменевтика // Вопросы философии. – 2024. – № 4. – С. 116-125. – DOI 10.21146/0042-8744-2024-4-116-125.
2. Степанчук Ю.А., Кириллова Н.Б. Цифровые подходы в системе изучения гуманитарных дисциплин: проблемы интеграции // Перспективы Науки и Образования. – 2024. – № 6 (72). – С. 660–671. – DOI 10.32744/pse.2024.6.41.
3. Фененко А. Статистика против истории // Международные процессы. – 2018. – Т. 16, № 3(54). – С. 56-83. – DOI 10.17994/IT.2018.16.3.54.3.
4. Gonçalves M. The hermeneutic problem posed by digital humanities // Revista de Teoria da História. – 2022. – Т. 25. – № 1. – P. 136-150.
5. Jockers M.L. Macroanalysis: Digital methods and literary history. – University of Illinois Press, 2013. – 187 p.
6. Rockwell G., Sinclair S. Hermeneutica: Computer-assisted interpretation in the humanities. – mit Press, 2022. – 241 p.
7. <https://mezrob29.ru/category/korpus-tekstov/hudozhestvennyj-podkorpus/>