

ЦИФРОВЫЕ ГУМАНИТАРНЫЕ НАУКИ В ЧЕХИИ

Ричард Змелик

доктор философии, старший преподаватель,
Университет Палацкого в Оломоуце

Почтовый адрес: ул. Кршижковского, 10,
Оломоуц, 77900 Чехия

Электронный адрес: richard.zmelik@upol.cz

Аннотация

Статья посвящена проблеме цифровых гуманитарных наук (англ. *Digital Humanities*, далее – ДН) в чешском литературоведении. За последние десять лет различные методы, связанные главным образом с ДН, оказывали значительное влияние на развитие множества областей гуманитарных наук. В чешском литературоведении такого рода воздействие становится очевидным с 2017 г., когда среди ученых была инициирована официальная дискуссия по данному вопросу. С одной стороны, использование ДН в чешском литературоведении началось не так давно, поэтому они не занимают столь сильные позиции, как в зарубежной науке. С другой – в чешских научных кругах традиция применения так называемых точных методов в литературоведении существует и служит важным внутренним источником для более полного развития ДН. Эти точные методы исследуются в статье. Тем не менее основная цель работы – представить два современных чешских ДН-проекта в области литературоведения: первым из них является Корпус чешского стиха, а вторым – Корпус чешской прозы, который служит базой для тематического проекта «Вывымышленная топография Праги». Проекты могут рассматриваться в качестве характерных для положения и роли ДН в современном чешском литературоведении.

Ключевые слова

Чешское литературоведение, цифровые гуманитарные науки, цифровой корпус чешской поэзии, формализм, структурализм, топографические модели в литературоведении

Статья поступила в редакцию 25 февраля 2021 г.

Цитирование: *Змелик Р.* Цифровые гуманитарные науки в Чехии // Славянский мир в третьем тысячелетии. 2021. Т. 16. № 1–2. С. 81–98. <https://doi.org/10.31168/2412-6446.2021.16.1-2.04>

THE DIGITAL HUMANITIES IN THE CZECH REPUBLIC

Richard Změlík

Ph.D., Assistant Professor,

Palacký University in Olomouc

Postal address: Křížkovského 10, Olomouc, 77900
Czech Republic

E-mail: richard.zmelik@upol.cz

Abstract

This paper is devoted to the issue of Digital Humanities (DH) in Czech literary studies. Over the last decade, various methods primarily connected with DH have had a substantial impact on the development of numerous areas of the humanities. In Czech literary studies, this impact is visible only from 2017 onwards, when official discussion between Czech scholars about the topic was initiated. The recent development of this tradition is the main reason why DH have not gained as strong position in Czech literary studies as they have abroad. However, there exists a tradition of the use of so-called exact methods in literary studies in the Czech scholarly milieu, and this tradition serves as an important domestic source for the full development of DH in Czech literary studies. While these exact methods are examined in the paper, its main purpose is to introduce two contemporary Czech DH projects in literary studies – the first being the Corpus of Czech Verse and the second the Corpus of Czech Prose, the latter of which currently serves as a base for a subject-project on Prague Fictional Topography. These two projects can be viewed as representatives of the situation and role of DH in contemporary Czech literary studies.

Keywords

Czech literary studies, digital humanities, digital corpus of Czech verse, formalism, structuralism, topographical models in literary studies

Received 25 February 2021

How to cite: Změlík, R., 2021. Tsifroyve gumanitarnye nauki v Chekhii [The Digital Humanities in the Czech Republic]. *Slavic World in the Third Millennium*, vol. 16, no. 1–2, pp. 81–98. <https://doi.org/10.31168/2412-6446.2021.16.1-2.04>

Хотя зачатки цифровых гуманитарных наук (англ. *Digital Humanities*, далее – DH) можно обнаружить уже в конце XX в.¹, их значительное развитие и применение приходится на начало текущего тысячелетия, когда DH стали активно использоваться в иностранных гуманитарно ориентированных исследованиях, о чем свидетельствует не только ряд публикаций, уже ставших классическими, например «Справочник по цифровым гуманитарным наукам» (англ. *A Companion*

1 Авторы публикации о DH приводят данные о том, что зарождение этой дисциплины относится к 80-м гг. XX в. См.: Burdicková A., Druckerová J., Lunenfeld P., Presner T., Schnapp J. *Digital Humanities*. Praha: Academia, 2019. P. 20.

to *Digital Humanities*), но и большое количество специализированных монографий, отражающих роль ДН в ряде областей гуманитарных наук, к которым относится и литературоведение. Сегодня уже практически невозможно определить весь спектр проблематики ДН, поскольку она затрагивает многочисленные дисциплины (литературоведение, лингвистику, историю, археологию, музыковедение, музейное дело, педагогику, историю искусств и т. д.), а количество специализированных текстов, обсуждений или выступлений на конференциях по данной проблематике достигает внушительных масштабов. Однако задачей статьи не является охват всего многообразия ДН; как литературовед и богемист, я буду говорить лишь о некоторых сегментах данной проблематики. Прежде всего, о возможностях не только цифровой, но и количественной, статистической и компьютерной обработки художественных текстов, с чем очень тесно связан вопрос применения цифровых корпусов в литературоведении. Хотя бы частично мы попытаемся здесь наметить и другие ресурсы ДН, связанные не только с указанными подходами, но и, например, с пространственными моделями: сегодня они особенно распространены в современной картографии, пользующейся цифровыми геоинформационными системами. Названные аспекты мы намереваемся представить с оглядкой на чешский литературоведческий контекст. За рубежом исследования на эту тему широко распространены, о чем свидетельствует количество специализированных работ (назовем, например, статьи Мэтью Л. Джокерса или Франко Моретти), а также деятельность университетских центров, где группы литературоведов и программистов (иногда обеими специальностями владеет один человек, как показывает случай Мэтью Л. Джокерса) систематически занимаются цифровыми гуманитарными исследованиями². Ко всем этим областям и видам деятельности следует прибавить многочисленные конференции, в особенности устраиваемую «Европейской ассоциацией цифровых гуманитарных наук» (англ. *European Association for Digital Humanities*) ежегодную международную конференцию, которая должна в сентябре 2021 г. пройти в Красноярске³.

Если посмотреть на ситуацию в контексте литературоведения, можно констатировать, что и здесь в последние несколько лет ДН начинают всё сильнее проникать в общественное сознание. С другой стороны,

² Примером могут служить Stanford Literary Lab (URL: <https://litlab.stanford.edu/> (дата обращения: 06.04.2021)) или проект Ein literarischer Atlas Europas (URL: <http://www.literaturatlas.eu/> (дата обращения: 06.04.2021)).

³ Ссылка на страницу конференции: <https://eadh2020-2021.org/> (дата обращения: 06.04.2021).

этот факт нельзя никоим образом переоценивать. Определяя некий конкретный временной рубеж, когда ДН в чешском литературоведении становятся более или менее привычными, нужно назвать 2017 год. Тогда профессиональная общественность впервые обсуждала вопрос ДН в литературоведении – на базе Института чешской литературы Академии наук Чешской Республики⁴. Участниками дискуссии помимо автора данного текста были Петр Плевач, Карел Пиоретский, Михал Вогербауэр из ИЧЛ АН ЧР и Сильвия Цинкова из Института формальной и прикладной лингвистики в Карловом университете, модератором выступил Йозеф Шлерка, который работает на кафедре исследований новых медиа философского факультета Карлова университета. Главными темами являлись вопросы функционального использования методов ДН в литературоведении, были представлены и обсуждены первые планы и результаты соответствующих исследований. В 2017 г. в Чешской Республике существовало два результативных литературоведческих ДН-проекта. Первый – успешный в международном масштабе Корпус чешского стиха (далее – КЧС) Петра Плевача, второй – индивидуальный университетский проект Корпуса прозы Яна Чепы, разработанный автором данной статьи и реализованный в Университете Палацкого в Оломоуце. Об обоих проектах я выскажусь ниже.

В настоящий момент интенсивное развитие ДН в чешской профессиональной среде отражается главным образом в таких дисциплинах, как социология (Мартин Гаек), дискурсивный анализ медиапространства (Йозеф Шлерка) и лингвистика (Вацлав Цврчек)⁵. В 2017 г. была основана Чешская ассоциация цифровых гуманитарных наук⁶ – партнер Европейской ассоциации цифровых гуманитарных наук⁷, существующей с 1973 г. (первоначальное название – Ассоциация литературных и лингвистических вычислений). О долгой традиции, которую продолжают нынешние ДН в литературоведении (и, разумеется, в лингвистике), свидетельствует не только этот факт; в действительности история ДН

⁴ Запись дискуссии доступна здесь: <https://vimeo.com/244618425> (дата обращения: 06.04.2021).

⁵ Кроме того, в 2021 г. должна быть опубликована репрезентативная коллективная монография «Цифровой поворот в чешских гуманитарных и социальных науках» (чеш. – *Digitální obrat v českých humanitních a sociálních vědách*), в которой будут представлены статьи из исследований ДН в области лингвистики, литературоведения, социологии, истории, археологии, библиографии, кураторства и педагогики.

⁶ Интернет-страница ассоциации: Česká asociace pro digitální humanitní vědy (CzADH) [online]. URL: <https://www.czadh.cz/o-nas/> (дата обращения: 06.04.2021).

⁷ Интернет-страница ассоциации: European Association for Digital Humanities (EADH) [online]. URL: <https://eadh.org/about> (дата обращения: 06.04.2021).

начинается в XX в. благодаря работам русских формалистов и чехословацких структуралистов. Принципиальное влияние на лингвистику и литературоведение, однако, оказало развитие науки после Второй мировой войны, когда чем дальше, тем значительнее в разных научных дисциплинах стали продвигаться теория информации и кибернетика. Я не буду подробно останавливаться на общеизвестном репрезентативном международном контексте и сосредоточусь исключительно на ситуации, сложившейся в чехословацком литературоведении XX в. и предшествовавшей современным ДН-исследованиям.

Выше был упомянут русский формализм, проводивший параллели между точной литературной наукой (например, в работах Виктора Шкловского) и арифметикой, техническими науками⁸; чехословацкий межвоенный структурализм в свою очередь эту традицию естественным образом продолжает. Русских формалистов и чешских структуралистов связывал интерес к имманентным процессам литературной системы, исследование которой должно исходить из обстоятельного анализа языковых систем произведений. Разумеется, рано или поздно должен был встать вопрос, можно ли такой анализ расширить за счет неких точных математических методов. Ян Мукаржовский, литературный теоретик и эстетик, являвшийся вместе с Романом Якобсоном и рядом других значимых исследователей одним из ведущих представителей Пражского лингвистического кружка (1926), опубликовал в 1934 г. работу «Общие принципы и развитие чешского стиха Нового времени»⁹, где он использовал статистические диаграммы для изображения акцента, чтобы с их помощью моделировать характер ритма в выбранных текстах чешских поэтов. Кроме ритма он также изучал статистические отличия между мужским и женским стихом, рифмованным и нерифмованным. В том же году Мукаржовский опубликовал не менее обширный труд «“Возвышенность природы” Полака»¹⁰, в котором кроме метрических моделей привел и основные статистические таблицы слов с количеством слогов от одного до шести. В своих трудах он дополнил исследование Романа Якобсона, занимавшегося статистикой чешского стиха, например, в работе «Стих

⁸ Тропологией дискурса русского формализма занимался Петр Штайнер (Petr Steiner) в книге «Русский формализм: метапоэтика». См.: *Steiner P. Ruský formalismus: meta-poetika*. Brno: Host, 2011.

⁹ *Mukařovský J. Obecné zásady a vývoj novočeského verše // Studie z poetiky / red. H. Mukařovská. Praha: Odeon, 1982. S. 365–438.*

¹⁰ *Mukařovský J. Polákova Vznešenost přírody // Studie z poetiky / red. H. Mukařovská. Praha: Odeon, 1982. S. 439–513.*

древнечешский»¹¹, где его помимо прочего интересовали количественные показатели межсловных границ.

Традиция чехословацкого структурализма была прервана войной и послевоенным развитием Чехословакии. С 1950-х гг. в Европе стали интересоваться вопросами теории информации и кибернетики, которые затронули и эстетику с литературоведением. В этой связи упомянем так называемую штутгартскую школу¹², где с 1950-х гг. работал Макс Бензе, также повлиявший на лингволитературные исследования в Чехословакии 1960-х гг.¹³ Здесь следует привести несколько имен значимых чехословацких литературоведов. Любомир Долежел (1922–2017) в работе «Пражская школа и теория статистики поэтического языка»¹⁴ обдумывал возможности статистической дифференциации так называемых нормированных и актуализированных текстов. Для разграничения двух главных групп, которые можно рассматривать в различных коммуникационных сферах, а значит и в литературной, Долежел вводит несколько стилистических критериев, определяемых статистически¹⁵. Вместе с тем он переводит статистические показатели характерности/нехарактерности в рамках каждого критерия в бинарную систему (0–1) и с помощью нее кодирует информацию. Используя подход Долежела, можно очень хорошо отследить влияние обусловленных эпохой количественных методов исследования языка наряду с теорией информации, а также внедрение инновационного подхода к анализу поэтического языка, непосредственно опирающегося на данные методы. Это стало прорывом в использовании точных математических методов в литературоведческом контексте: они уже перестали носить лишь иллюстративный характер,

¹¹ *Jakobson R. Verš staročeský // Poetická funkce / přípr. M. Červenka. Jinočany: H&H, 1995. S. 275–318.*

¹² С 1950-х гг. в Штутгартском университете группа исследователей занималась экспериментальными творческими методами с использованием компьютера. Кроме Бензе, в нее входили Рул Гунценхойзер, Хельмар Франк и Георг Нис, экспериментировавший в области компьютерной графики. См.: *Klütsch Ch. Information Aesthetics and the Stuttgart School // Mainframe Experimentalims: Early Computing and the Foundations of the Digital Arts / ed. by H. B. Higgins, D. Kahn. Berkeley; Los Angeles; London: University of California press, 2012. P. 66–67. <https://doi.org/10.2307/j.ctv114c7b3>.*

¹³ *Doležel L. Úvod do informačně teoretické koncepce uměleckých textů. [Bense M. Theorie der Texte (Kolín n. R., 1962)] // Česká literatura. 1965. Roč. 13. Č. 3. S. 258–260.*

¹⁴ Во втором номере журнала «Чешская литература» за 1965 г. Долежел положительно оценил переведенную на чешский язык книгу Бензе «Теория текстов». См.: *Doležel L. Úvod do informačně teoretické koncepce uměleckých textů.*

¹⁵ Лишь кратко скажу, что речь идет о статистических показателях, связанных с величинами стандартизации/актуализации стиливых средств, а также степени субъективности/объективности текста и его стабильности во времени, то есть подвижности/неподвижности.

а их результаты начали активно влиять на интерпретацию¹⁶. Кроме Долежела необходимо отметить и других чешских литературоведов, в чьих трудах отразились современные знания в области теории информации и кибернетики: Иржи Левый¹⁷ рассматривал теорию информации в рамках теории перевода, Эдуард Петру также исследовал возможность экспериментирования в литературоведении при помощи перфокарт, Павел Вашак использовал статистику в текстологии, в особенности в вопросе определения авторства, Мирослав Червенка применял квантитативные методы при анализе чешского стиха. Следует сказать, что эта ситуация в чехословацком литературоведении в 1960–1980-х гг. была не только реакцией на развитие лингвистики и эстетики за границей, но и продолжением межвоенной структуралистской традиции. Долежел и Червенка были членами Пражского лингвистического кружка (ПЛК), а Левый в своей профессиональной деятельности впоследствии активно применял теоретические постулаты и принципы ПЛК.

В начале 1990-х гг. в ряде стран происходит формирование цифровых языковых корпусов¹⁸, применяющих для всё большего объема текстов квантитативные и статистические методы, позволяющие изучать языковую систему в макроаналитической перспективе. Развитие поисковых механизмов и появление возможности собственно машинного анализа текстов вывели корпусную лингвистику и ее методику за ее собственные профессиональные границы – в область ДН, что проявилось и в литературоведении. В чешской среде подтверждением тому является прежде всего Корпус чешского стиха¹⁹ Петра Плевача. Корпус формировался с 2011 г., и его материальную базу составляет чешская поэзия XIX – начала XX в., то есть это массив текстов, уже ранее оцифрованный и доступный в Чешской электронной

¹⁶ Это также одна из главных причин, почему в данной работе Долежел определяет межвоенный чехословацкий структурализм как достатистический (до использования статистических методов) период.

¹⁷ Чешский перевод Бензе «Теория текстов» Левый сопроводил вводной статьей, в которой он обобщает вклад Бензе в современную эстетику. См.: *Levý J. Úvodní poznámka // Benze M. Teorie textů. Praha: Odeon, 1967. S. 7–9.*

¹⁸ Первым таким корпусом был British National Corpus, который создавался с начала 1980-х гг. и запустился в 1990 г. Функционирование Чешского национального корпуса началось в 1994 г. В России, например, с 1990-х гг. создавался корпус публицистических текстов (см.: *Marková K. Korpusy ruského jazyka v Rusku // Slovo a slovesnost. 2003. Roč. 64. Č. 2. S. 148–155*), а с 2001 г. – репрезентативный корпус современного русского языка (Там же).

¹⁹ Доступно здесь: *Versologický tým. Ústav pro českou literaturu AV ČR, v.v.i. [online]. URL: http://versologie.cz/v2/web_content/index.php?lang=cz (дата обращения: 06.04.2021).*

библиотеке²⁰. КЧС содержит 76 699 стихотворений, 1689 поэтических сборников с общим числом слов 14 592 037. Отдельные стихи морфологически и фонетически отмечены специальными стиховедческими тэгами, которые расширяют набор основных операций с корпусом за счет возможности анализировать метр и рифму. В результате мы имеем в распоряжении цифровой корпус чешских поэтических текстов, позволяющий осуществлять поиск, статистический анализ и сравнение на уровне стиха, строфы, метра, рифмы и целого стихотворения. Кроме прочих важных для количественного анализа поэтических текстов с точки зрения стиховедения данных (например, доступно измерение тематически важных слов или так называемого эвфонометра, определяющего степень эвфонии текста) частью корпуса являются и коды источников отдельных инструментов, словари частотности или тезаурус чешского метра. Благодаря масштабу и степени проработанности мы можем назвать данный проект уникальным на международном уровне. Следует добавить, что общие результаты количественного анализа чешского стиха, полученные в этом проекте, представлены в «Главах из корпусного стиховедения» (чеш. *Kapitoly z korpusové versologie*) Петра Плекха и Роберта Колара²¹.

Если количественно-корпусный подход к чешской поэзии непрерывно развивается и расширяется, свидетельством чего может быть готовящийся корпус чешской поэзии XX в., первые полученные результаты которого в конце 2020 г. опубликовали Михал Шкрабал и Карел Пиорекский в шестом номере журнала «Словацкая литература» (словац. *Slovenská literatúra*)²², то аналогичная обработка чешской прозы в настоящее время находится в стадии зарождения и концептуальных проектов. Квантификация поэтических текстов имеет традицию не только в чешском литературоведении (само собой разумеется, сегодня благодаря цифровым инструментам она развивается по нарастающей), в то время как количественный анализ прозы у нас не может опереться на традицию, которая бы в прошлом применяла эти методы на нарративном литературном материале. Поэтому и результаты в данной области пока носят лишь предварительный характер.

Указанной проблематикой мы занимаемся с 2012 г., когда в рамках европейского гранта в Университете Палацкого в Оломоуце нами был

²⁰ Доступно здесь: Česká elektronická knihovna – Poezie 19. a počátku 20. století [online]. URL: <http://www.ceska-poezie.cz/cek/> (дата обращения: 06.04.2021).

²¹ Plecháč P., Kolár R. *Kapitoly z korpusové versologie*. Praha: Akropolis, 2017.

²² Piorecký K., Škrábal M. *Vícejazyčnost v současné české poezii. Několik úvodních postřehů z korpusové perspektivy // Slovenská literatúra*. 2020. Roč. 67. Č. 6. S. 568–583. <https://doi.org/10.31577/slovlit.2020.67.6.4>.

сформирован авторский корпус прозы чешского католического автора Яна Чепа (1902–1973)²³. Его главной целью было открыть возможности квантитативно-корпусного анализа для нарративных художественных текстов и тем самым подготовить будущую платформу для более комплексного проекта корпуса чешской прозы. Результаты данных усилий были опубликованы в двух монографиях: «Квантитативно-корпусный анализ и литературоведение»²⁴ (чеш. *Kvantitativně-korpusová analýza a literární věda*) и «Концептуализация цветов в художественном нарративе» (чеш. *Konceptualizace barev v narativní fikci*)²⁵, в последней мы исследовали частотность так называемых фундаментальных цветов в прозе Яна Чепа. С помощью статистических моделей, применявшихся для изучения не только самой частотности языковых наименований цвета, но и других аспектов квантитативного анализа (ими являются прежде всего коллокации и мотивный анализ, всегда проводимый на первых сотнях существительных), мы объяснили основные макроструктурные механизмы, которые определяют взаимное расположение цветов и цветовые спектры художественных миров прозы Чепа. Кроме того, отдельные эмпирические модели позволили отследить, как эти конфигурации меняются и развиваются в рамках хронологии всего творчества Чепа. В настоящее время мы занимаемся разработкой корпуса чешской прозы, которая в значительной степени тематизирует пражскую топографию²⁶. Здесь будут представлены основные параметры такого корпуса вместе с архитектурой одного из алгоритмов и с доступными на данный момент промежуточными результатами исследований²⁷. Основная цель этого проекта достигается в двух главных областях (рис. 1): в них

²³ Корпус доступен в Чешском национальном корпусе под названием «сер» (URL: <https://www.korpus.cz/kontext/query?corpname=ser> (дата обращения: 06.04.2021), в сокращенном поисковом режиме он был опубликован и здесь: *Autorský slovník Jana Čepa // KORBEL: Manažer beletristických korpusů* [online]. URL: <http://slovníkcep.g6.cz/index.php> (дата обращения: 06.04.2021).

²⁴ *Změlík R. Kvantitativně-korpusová analýza a literární věda*. Olomouc: Univerzita Palackého v Olomouci, 2015.

²⁵ *Změlík R. Konceptualizace barev v narativní fikci*. Olomouc: Univerzita Palackého v Olomouci, 2019.

²⁶ См.: *Literární topografie a kvantitativní analýza narativních textů: prostorové modely fikční topografie Prahy a kvantitativní modely vybraných segmentů literárních narativů v české proze 19. století* [online]. URL: <http://korpusprozy.cz/> (дата обращения: 06.04.2021).

²⁷ Подробно весь проект также описан в работе «Цифровая литературная картография и квантификация сегментов литературно-нарративного текста – методология, инструменты и интерпретационные стратегии (описание одного проекта)». См.: *Změlík R. Digitální literární kartografie a kvantifikace segmentů literárně narativního textu – metodologie, nástroje a interpretační strategie (popis jednoho projektu) // Slovenská literatúra*. 2020. Roč. 67. Č. 6. S. 523–551. <https://doi.org/10.31577/slovlit.2020.67.6.2>.

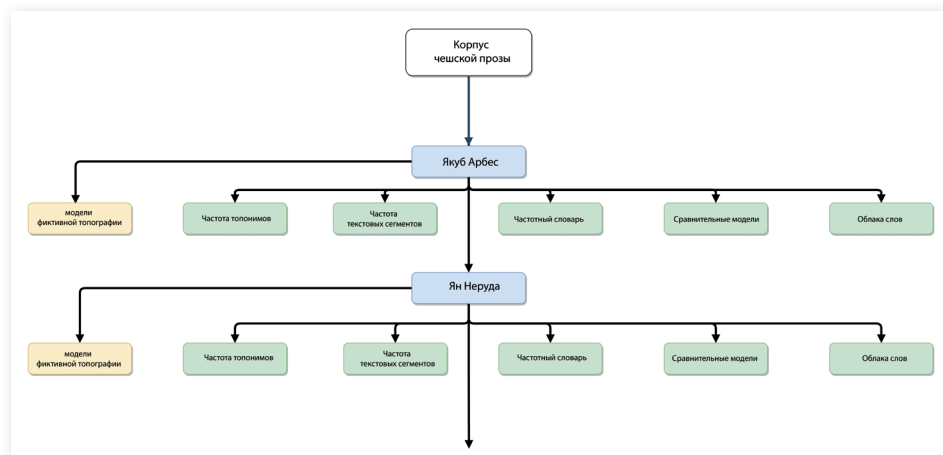


Рисунок 1. Основная модель корпуса чешской прозы

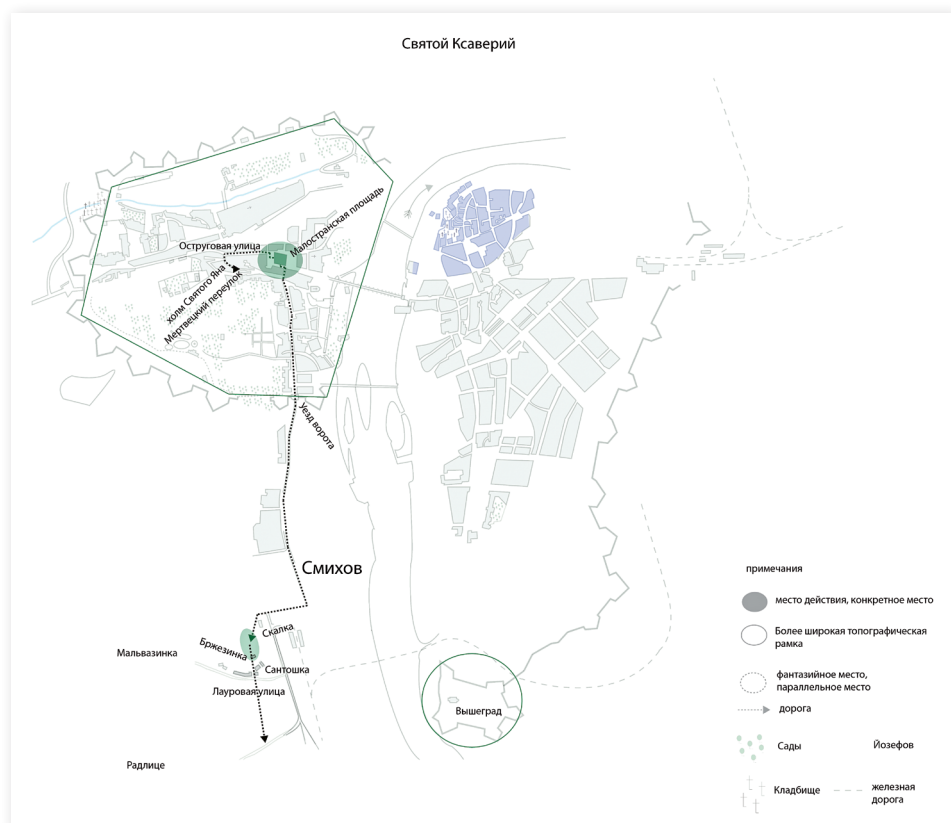


Рисунок 2. Пространственная модель романетто «Святой Ксаверий» (1873) Якуба Арбеса

используются одинаковые цифровые инструменты, техники моделирования и машинный квантитативный анализ. Первая область связана с созданием пространственных моделей пражской художественной топографии, призванных изобразить формальное распределение топографических сегментов вымышленного мира прозы вместе с траекториями движения их протагонистов. Как видно на модели (рис. 2), здесь фиксируются локусы главного места действия и эксплицитно названные более широкие топографические границы, которые не обязательно должны быть непосредственным местом развития сюжета. Конечно, смысл состоит не в том, чтобы найти аналогию между художественной и реальной городской топографией, а в том, чтобы путем сравнения моделей раскрыть анализ способов структурной проработки вымышленной топографии. Если мы посмотрим на представленную модель, то проследим две основные черты построения пространства выбранной прозы – романетто²⁸ «Святой Ксаверий» (1873) чешского писателя Якуба Арбеса. Первая черта заключается в ярко выраженной центрированности пространства. Хотя в указанном произведении это значение формируется и другими ее сегментами, в особенности нарративной стратегией, о чем будет идти речь ниже, на уровне художественной топографии оно организует топографические элементы, находящиеся внутри района Малой Страны. Главным местом действия является храм св. Николая, расположенный в центре воображаемой Малой Страны, чьи контуры составляют многоугольник. Вторая черта пространственного структурирования – положение на оси внешнего и внутреннего пространств. Оно топографически определено колебанием между храмовым внутренним устройством и внешним малостранским пейзажем и в особенности соотношением между внутренним обустройством города и его загородными окрестностями. Топографические отношения центрированности, с одной стороны, и полярности между внутренним и внешним пространством, с другой стороны (см. рис. 2), дополняются еще одним пространственным аспектом, возникающим в ситуации с рассказчиками (рис. 3) в упомянутом романетто Арбеса. Данная ситуация, создаваемая вложенными друг в друга нарративными высказываниями, создает пространственный эффект вертикальности. На этом примере мы видим, насколько продуктивным

²⁸ Речь идет о нарративном жанре среднего объема, создателем которого был Якуб Арбес (название жанру дал Ян Неруда). Типичным знаком романетто Арбеса в 1860-е и в особенности в 1870-е гг. является напряженный мистический сюжет с таинственной завязкой, но совершенно рациональной развязкой. В более поздний период (1880–1890-е гг.) фантастические элементы в романетто Арбеса отходят на второй план. Кроме Арбеса в чешской литературе романетто писал, например, Ладислав Клима.

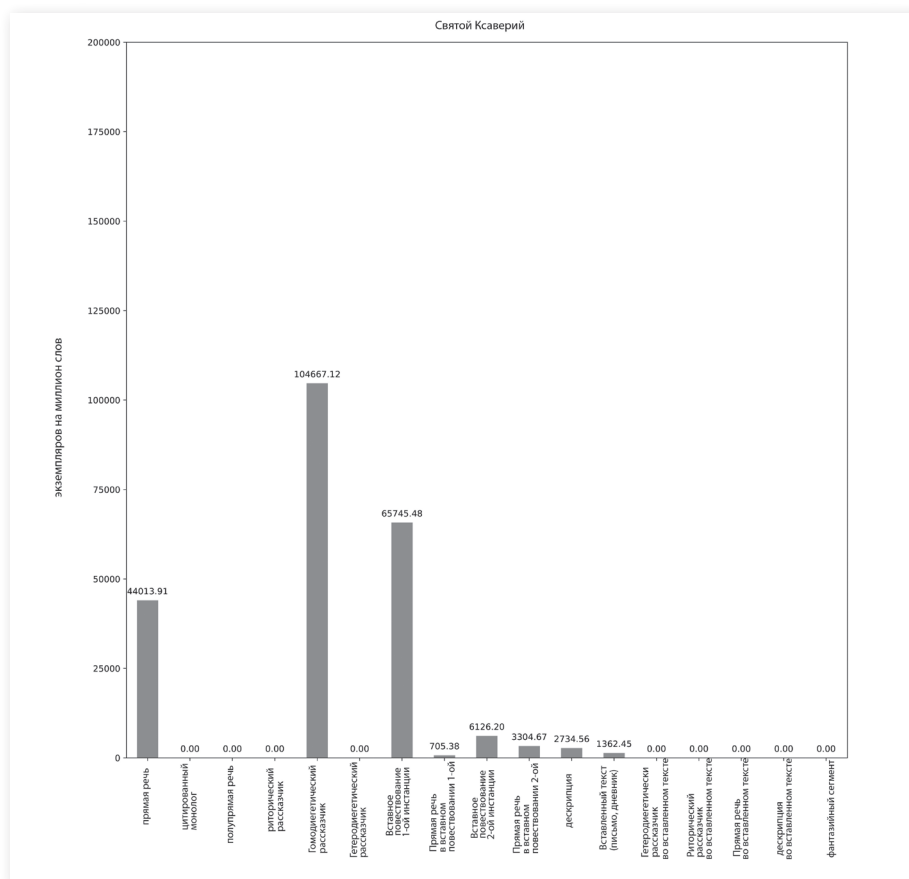


Рисунок 3. Квантитативная модель текстовых сегментов романетто «Святой Ксаверий»

может быть комбинирование разных моделей, в нашем случае – картографической модели с моделями квантитативного анализа высказываний в нарративе.

Именно такой тип моделей составляет вторую часть корпуса прозы. Его задача заключается в квантификации определенной области художественного произведения, каковыми являются различные типы нарративных стратегий, высказываний персонажей или описаний²⁹. Если мы

²⁹ К таким моделям присоединяются и другие модели, квантифицирующие топографические объекты, частотные словари, модели, сравнивающие тексты на уровне средней длины предложений, в том числе и отдельных сегментов текста, и других статистических показателей, например энтропии. На рис. 4 и 5 представлены элементы, с одной стороны, архитектуры алгоритма для обработки сегментации и квантификации нарратива по выделенным сегментам, а с другой – собственно кода, который представлен данной архитектурой.


```

split_precist_soubor_cely_korpus=precist_soubor_cely_korpus.split()
pocet_slov_cely_korpus=len(split_precist_soubor_cely_korpus)-(split_precist_soubor_cely_korpus.count('--'))

#pocet_slov_cely_korpus=len(slova_cely_soubor)-int(slova_cely_soubor.count('--'))
pocet_vet_cely_korpus=len(vety_cely_soubor)
prumerna_delka_vety_cely_korpus=round((pocet_slov_cely_korpus/pocet_vet_cely_korpus),2)

print('Pocet slov v celém korpusu: ',pocet_slov_cely_korpus)
print('Pocet vet v celém korpusu: ',pocet_vet_cely_korpus)
print('Průměrná délka věty v celém korpusu: ',prumerna_delka_vety_cely_korpus)

# Pocet slov a vet v daném textu
print(' ')
print('Pocet slov v celém textu: ',pocet_slov)
print('Pocet vet v celém textu: ',pocet_vet)
print('Průměrná délka věty v celém textu: ',prum_del_vety)
print(' ')

print('-----')
print('/% PROMĚLNÝVÝ PRŮBĚH TEXTU: ')
print('-----')
print(' a) PRŮBĚH SLOV: ')
print(' ')
if 'P' and 'R' in tokenizace1:
    prumilava=[]
    while 'P' and 'R' in tokenizace1:
        start_p=tokenizace1.index('P')
        end_r=tokenizace1.index('R')
        prumilava.extend(tokenizace1[start_p+1:end_r])
        del tokenizace1[start_p:end_r+1]
        pocet_vet_ppromilava.count('')*promilava.count('')*promilava.count('')
        pocet_slov_ppromilava=(promilava.count('')*promilava.count('')*promilava.count('')*promilava.count('')*promilava.count('')*promilava.count('')*promilava.count(''))
    print(' Pocet slov: ',pocet_slov_p)
    prumerna_delka_vety_ppromilava=(pocet_slov_p/pocet_vet_p,2)
    print(' Průměrná délka věty: ',prumerna_delka_vety_ppromilava)
    podil_p_promilava=(pocet_slov_p/pocet_slov)*100,2) # %
    podil_p_cely_korpus=round(((pocet_slov_p/pocet_slov_cely_korpus)*100000,2) # 4.p.p.m.
    print(' Podíl k celému korpusu: ',podil_p,'%')
    print(' Podíl k celému korpusu: ',podil_p_cely_korpus,'4.p.m.')
else:
    podil_p_cely_korpus=0
    print(' Podíl k celému textu: 0%')
print(' ')
print(' b) SLOVŮVNÝ MORFOLOG: ')
print(' ')
if 'CM X' and 'CM Y' in tokenizace2:
    cit_m=[]
    while 'CM X' and 'CM Y' in tokenizace2:
        start_cm=tokenizace2.index('CM X')
        end_cm=tokenizace2.index('CM Y')
        cit_m.extend(tokenizace2[start_cm+1:end_cm])
        del tokenizace2[start_cm:end_cm+1]
        pocet_vet_cm=cit_m.count('')*cit_m.count('')*cit_m.count('')*cit_m.count('')
        pocet_slov_cm=len(cit_m)+cit_m.count('')+cit_m.count('')+cit_m.count('')+cit_m.count('')+cit_m.count('')+cit_m.count('')+cit_m.count('')
    print(' Pocet vet: ',pocet_vet_cm)
    print(' Pocet slov: ',pocet_slov_cm)
    prumerna_delka_vety_cm=round((pocet_slov_cm/pocet_vet_cm,2)

```

Рисунок 5. Фрагмент алгоритма для делимитации и квантификации текстовых сегментов (отвечает модели его построения)

посмотрим на подобную модель в отношении упомянутого произведения Якуба Арбеса (см. рис. 2), то увидим, что при ее сравнении с пространственной моделью мы можем экстраполировать следующие данные. Пространственная вертикаль определяется структурированием нарративных уровней, представляющих собой вклинивающиеся друг в друга нарративные границы. Следует отметить, что в прозе присутствуют два типа интрадиегетического рассказчика (используя терминологию Жерара Женетта), причем их отношение является не эквивалентной, а структурной субординацией. Вся композиция нарративных зон в данной прозе выглядит в итоге следующим образом: исходная гомодиегетическая наррация обрамляет интрадиегетическую наррацию первой инстанции, которая обрамляет интрадиегетическую наррацию второй инстанции. Это движение в то же время отвечает за эвокацию эффекта пространственной (никак не топографической!) вертикали. Вертикаль (вместе с внешним/внутренним и границами переходов между этими зонами в прозе) обозначает колебание между сном и бодрствованием, галлюцинацией, миражом, с одной стороны, и реальностью – с другой.

Разумеется, мы можем задаться вопросом, действительно ли необходимы приведенные модели для подобных изысканий, нельзя ли было

к таким выводам прийти без них? Ответить, по нашему мнению, можно, учитывая два главных аспекта. Во-первых, даже если бы мы не использовали цифровые инструменты пространственного моделирования и алгоритмов для квантификации, нам бы пришлось всё равно прибегнуть к какой-либо форме схематизации, чтобы иметь возможность зафиксировать те факты, о которых идет речь. Во-вторых, цифровой подход не только обеспечивает более надежное и точное моделирование, в особенности в области квантификации, но и прежде всего позволяет обработать большое количество текстов и впоследствии их сравнить. Это очень важно, поскольку с помощью цифровой обработки появляется возможность анализировать проблематику художественной топографии – способы ее пространственной конфигурации вместе с количественными моделями.

Аналоговый подход к указанной проблематике при необходимости систематического решения предполагает линейную стратегию, чьим механизмом является интроспективный анализ, движущийся в рамках хронологии произведения. Де-факто же речь идет о так называемом близком чтении (*close reading*), обусловленном избирательностью субъективной памяти и линейно-рецептивной последовательностью. Представим себе такую ситуацию на примере литературоведа, перед которым как раз стоит задача проанализировать художественную топографию в творчестве выбранных авторов. Стандартные рабочие методы будут основываться на линейно-хронологическом продвижении и далее – на сегментации группы текстов по принадлежности к данному автору или периоду. Степень аналитической точности в таком случае будет уменьшаться пропорционально нарастающему объему текстов. Уже в случае нескольких текстов одного автора с их приращением будет связана и степень их взаимной детерминации и избирательности памяти в области исследовательской интроспекции. Под влиянием предшествующего читательского опыта, закономерно обуславливающего каждый последующий, будет с высокой вероятностью происходить, пусть и непреднамеренное, искажение литературоведом результатов своего анализа. С увеличением количества материала рано или поздно появится необходимость в его формальной фиксации, что является стадией, предшествующей созданию цифровых пространственных моделей. Их преимущество заключается в возможности построения комбинаций, при помощи которых можно сравнивать эти модели и которые определяются рядом переменных критериев (автор, период возникновения произведения, структурный характер пространственной модели и т. п.). Цифровой подход в целом позволяет

осуществлять сравнения без оглядки на линейную причинность, заменяемую причинностью комбинационной. Смысл таких моделей, конечно, не в самой литературоведческой интерпретации, а в картографической схематизации художественной топографии, а она уже приведет к литературоведческой интерпретации.

Вторую часть проекта составят основные количественные модели нарративов, определяемых по их сегментам (прямая речь, несобственно-прямая речь, описания или типы рассказчиков). Представленные модели будут сопровождаться словарями частотности, сравнительными моделями и другой информацией статистического анализа нарративных текстов. Обе главные составляющие корпуса, то есть как пространственные модели, так и количественные модели, предполагается еще и сравнивать. Небольшой пробный образец такого корпуса показал потенциал указанного подхода для литературоведческого исследования и изложения (см. прим. 20).

В настоящее время всё еще не очень многочисленное литературоведческое сообщество, активно применяющее методы ДН, разрабатывает в публикациях эти два основных направления³⁰ – создание картографических и количественных моделей, на воображаемой периферии которых появляются или будут постепенно появляться не только корпуса литературных текстов вместе с развитым инструментарием для их использования, но и интерпретационные сдвиги, переводящие формальную и техническую плоскость исследований ДН в конкретную литературоведческую практику.

Перевод с чешского А. В. Амелиной

Литература

Bense M. Teorie textů. Praha: Odeon, 1967.

Burdicková A., Druckerová J., Lunenfeld P., Presner T., Schnapp J. Digital Humanities. Praha: Academia, 2019.

Doležel L. Pražská škola a statistická teorie básnického jazyka. Rozšířené znění přednášky proslavené v kolokviu prof. Maxe Bense na Technische Hochschule, Stuttgart // Česká literatura. 1965. Roč. 13. Č. 2. S. 101–113.

Doležel L. Úvod do informačně teoretické koncepce uměleckých textů. [Bense Max. Theorie der Texte (Kolín n. R., 1962)] // Česká literatura. 1965. Roč. 13. Č. 3. S. 258–260.

³⁰ Сюда можно отнести также проект цифровой базы Чешской литературной библиографии, доступной здесь: Czech Literary Bibliography Research Infrastructure [online]. URL: <https://clb.ucl.cas.cz/> (дата обращения: 06.04.2021).

- Jakobson R.* Poetická funkce. Jinočany: H&H, 1995.
- Klütsch Ch.* Information Aesthetics and the Stuttgart School // *Mainframe Experimentalims: Early Computing and the Foundations of the Digital Arts* / ed. by H. B. Higgins, D. Kahn. Berkley; Los Angeles; London: University of California Press, 2012. P. 65–89.
- Marková K.* Korpusy ruského jazyka v Rusku // *Slovo a slovesnost*. 2003. Roč. 64. Č. 2. S. 148–155.
- Mukařovský J.* Studie z poetiky. Praha: Odeon, 1982.
- Piorecký K., Škrábal M.* Vicejazyčnost v současné české poezii. Několik úvodních postřehů z korpusové perspektivy // *Slovenská literatúra*. 2020. Roč. 67. Č. 6. S. 568–583. <https://doi.org/10.31577/slovlit.2020.67.6.4>.
- Plecháč P., Kolár R.* Kapitoly z korpusové versologie. Praha: Akropolis, 2017.
- Steiner P.* Ruský formalismus: metapoetiky. Brno: Host, 2011.
- Změlík R.* Digitální literární kartografie a kvantifikace segmentů literárně narativního textu – metodologie, nástroje a interpretační strategie (popis jednoho projektu) // *Slovenská literatúra*. 2020. Roč. 67. Č. 6. S. 523–551. <https://doi.org/10.31577/slovlit.2020.67.6.2>.
- Změlík R.* Konceptualizace barev v narativní fikci. Olomouc: Univerzita Palackého v Olomouci, 2019.
- Změlík R.* Kvantitativně-korpusová analýza a literární věda. Olomouc: Univerzita Palackého v Olomouci, 2015.

References

- Bense, M., 1967. *Teorie textů*. Prague: Odeon.
- Burdicková, A., Druckerová, J., Lunenfeld, P., Presner, T., Schnapp, J., 2019. *Digital Humanities*. Prague: Academia.
- Doležel, L., 1965. Pražská škola a statistická teorie básnického jazyka. Rozšířene znění přednášky proslovené v kolokviu prof. Maxe Bense na Technische Hochschule, Stuttgart. *Česká literatura*, 13 (2), pp. 101–113.
- Doležel, L., 1965. Úvod do informačně teoretické koncepce uměleckých textů. [Bense Max. *Theorie der Texte* (Kolín n. R., 1962)]. *Česká literatura*, 13 (3), pp. 258–260.
- Jakobson, R., 1995. *Poetická funkce*. Jinočany: H&H, pp. 275–318.
- Klütsch, Ch., 2012. Information Aesthetics and the Stuttgart School. In: Higgins, H.B., Kahn, D., eds., 2012. *Mainframe Experimentalims: Early Computing and the Foundations of the Digital Arts*. Berkley; Los Angeles; London: University of California Press, pp. 65–89.
- Marková, K., 2003. Korpusy ruského jazyka v Rusku. *Slovo a slovesnost*, 64 (2), pp. 148–155.
- Mukařovský, J., 1982. *Studie z poetiky*. Praha: Odeon.

- Piorecký, K., Škrábal, M., 2020. Vícejazyčnost v současné české poezii. Několik úvodních postřehů z korpusové perspektivy. *Slovenská literatúra*, 67 (6), pp. 568–583. <https://doi.org/10.31577/slovlit.2020.67.6.4>.
- Plecháč, P., Kolár, R., 2017. *Kapitoly z korpusové versologie*. Prague: Akropolis.
- Steiner, P., 2011. *Ruský formalismus: metapoetiky*. Brno: Host.
- Změlík, R., 2020. Digitální literární kartografie a kvantifikace segmentů literárně narativního textu – metodologie, nástroje a interpretační strategie (popis jednoho projektu). *Slovenská literatúra*, 67 (6), pp. 523–551. <https://doi.org/10.31577/slovlit.2020.67.6.2>.
- Změlík, R., 2019. *Konceptualizace barev v narativní fikci*. Olomouc: Univerzita Palackého v Olomouci.
- Změlík, R., 2015. *Kvantitativně-korpusová analýza a literární věda*. Olomouc: Univerzita Palackého v Olomouci.