

ФИЛОЛОГИЯ

(шифр научной специальности: 10.02.19)

Научная статья

УДК 81

doi: 10.18522/2070-1403-2021-88-5-121-129

ЦИФРОВЫЕ ГУМАНИТАРНЫЕ ПРОЕКТЫ: ПРАКТИКИ МЕЖДИСЦИПЛИНАРНОСТИ¹

© *Елена Михайловна Северина¹, Анастасия Александровна Бонч-Осмоловская²,
Юлия Васильевна Бец¹, Марина Валерьевна Флягина¹*

¹Южный федеральный университет. г. Ростов-на-Дону, Россия; ²Национальный исследовательский университет «Высшая школа экономики», г. Москва, Россия,

¹emkovalenko@srfedu.ru ²abonch@hse.ru

Аннотация. Рассматриваются междисциплинарные «цифровые практики» в гуманитарной сфере, использующие компьютерные модели и цифровые технологии в качестве научного инструментария и реализуемые в виде цифровых проектов. Описана работа междисциплинарных коллективов, реализующих цифровые проекты, в контексте основного принципа Digital Humanities – принципа открытых исследовательских данных (Open data), целью которого является не только размещение информации в доступных форматах и свободном доступе, но и создание научных сообществ вокруг данных. В качестве примера такого подхода представлен научный проект «Конвергенция языковых пластов русского языка в зеркале цифровых решений», цель которого состоит в создании новых лингвистических и филологических цифровых ресурсов, расширяющих стандартные типы языковых данных и дополненных наукоёмкой экспертной разметкой, работа над которой осуществляется Центром цифровых гуманитарных исследований ЮФУ совместно с Международной лабораторией языковой конвергенции НИУ ВШЭ. В рамках проекта предполагается создание трех ресурсов: корпуса донских диалектных текстов, Русского учебного корпуса "Восток" для носителей восточных языков (китайского, туркменского) и цифровое издание текстов Полного собрания сочинений и писем А. П. Чехова (1974-1983), снабженное семантической разметкой.

Ключевые слова: цифровые гуманитарные проекты, Digital Humanities, Chekhov Digital, Русский учебный корпус «Восток», корпус донских диалектов, междисциплинарность, открытые данные.

Для цитирования: Северина Е.М., Бонч-Осмоловская А.А., Бец Ю.В., Флягина М.В. Цифровые гуманитарные проекты: практики междисциплинарности // Гуманитарные и социальные науки. 2021. Т. 88. №5. С. 121-129. doi: 10.18522/2070-1403-2021-88-5-121-129

PHILOLOGY

(specialty: 10.02.19)

Original article

Digital projects in Humanities: interdisciplinary practices

© *Elena M. Severina¹, Anastasiya A. Bonch-Osmolovskaya², Yulia V. Bets¹, Marina V. Flyagina¹*

¹Southern Federal University. Rostov-on-Don, Russian Federation; ²National Research University Higher School of Economics. Moscow, Russian Federation

¹emkovalenko@srfedu.ru ²abonch@hse.ru

Abstract. The work examines interdisciplinary "digital practices" in the human sciences, which use computer models and digital technologies as research tools and are implemented as digital projects. The work of interdisciplinary teams running digital projects is described with the reference to the main principle of Digital Humanities – the principle of Open data, the purpose of which is not only to place information in accessible formats and free access, but also to create academic communities around the data. The presented research project "Convergence of linguistic layers of the Russian language in the reflection of digital solutions" can be an example of such an approach. Its purpose is to create new linguistic and philological digital resources expanding the standard types of linguistic data and supplementing them with

¹Работа выполнена в рамках соглашения о научном сотрудничестве № 6.13.1-02/250821-1 между Южным федеральным университетом (ЮФУ) и Национальным исследовательским университетом "Высшая школа экономики" (НИУ ВШЭ)

high-tech expert markup, the work on which is carried out by the Center for Digital Humanities Research of the SfedU together with the International Laboratory for Language Convergence of the National Research University of Higher School of Economics. Within the framework of the project, it is planned to create three resources: the Don dialects corpus, the Russian educational corpus "Vostok" for speakers of oriental languages (Chinese, Turkmen) and a digital edition of the Complete Chekhov's Works and Letters (published from 1974 to 1983), provided with semantic markup.

Key words: Digital Humanitarian Projects, Digital Humanities, Chekhov Digital, Russian Learner Corpus (RLC) «Vostok», Corpus of Don Dialects, Interdisciplinarity, Open data.

For citation: Elena M. Severina, Anastasiya A. Bonch-Osmolovskaya, Yulia V. Bets, Marina V. Flyagina Digital projects in Humanities: interdisciplinary practices. *The Humanities and Social Sciences*. 2021. Vol. 88. No 5. P.121-129. doi: 10.18522/2070-1403-2021-88-5-121-129

Введение

Цифровые гуманитарные науки (Digital Humanities, DH) – междисциплинарное научное направление, стремящееся объединить методологию традиционных гуманитарных наук и логику исследований в области информатики для изучения гуманитарной сферы с использованием цифровых технологий [17]. Многие исследователи рассматривают Digital Humanities в качестве «зонтичного» термина для целого ряда практик по разработке, использованию и интерпретации цифровых технологий в гуманитарных науках.

Ведущей тенденцией организации современной науки, стремящейся к синтезу знания, является формирование междисциплинарности, основанной на проблемном и проектном подходах к исследованиям. При этом понимание междисциплинарности неоднозначно, оценка ее эвристического потенциала и эффективности специалистами существенно различается.

Междисциплинарность как интегративный подход к научному познанию интерпретируется по-разному, но в целом выделяются два направления его понимания. В рамках одного из них междисциплинарность рассматривается в контексте организации проектных исследований по научной проблеме, а не в рамках определенных дисциплинарных направлений, т.е. «как взаимодействие научных дисциплин, каждая из которых имеет свой предмет, свою терминологию и методы исследования» [5, с. 264], что позволяет синтезировать знания об объекте, полученные в рамках различных дисциплин, сделать представление об объекте более целостным, сформулировать новые концепции, углубляющие научное познание. Другой подход ориентирован на те области знания, которые находятся как бы «между» существующими научными дисциплинами, не являясь объектом исследования ни одной из них, т. е. на стыке дисциплин появляется новая дисциплина с новым объектом исследования, которая заимствует существующие язык и методы научного исследования.

Методологически междисциплинарность позволяет использовать методы, специфичные для одной дисциплины, в других сферах знания, порождая новые междисциплинарные подходы. В этом контексте к междисциплинарным относятся методы компьютерного моделирования, которые будучи разработанными в информатике применяются в других областях знания, например, для моделирования языковых процессов. Однако в этом случае необходимо также владение лингвистическим знанием, т. е. применение этого инструментария возможно либо в рамках междисциплинарного сотрудничества представителей разных направлений, либо теми, кто владеет обоими типами знания, например, компьютерными лингвистами. При этом преимущества междисциплинарности в исследованиях требуют «определения четких принципов работы междисциплинарных коллективов и сфер ответственности входящих в него ученых, критического использования терминологии, выработки и применения эффективной методологии, позволяющей осуществить подлинный синтез знания» [5, с. 264].

Обсуждение

Современный этап развития гуманитарного знания характеризуется появлением «глубинных сетевых структур процесса производства знаний и продуктов научного труда» [2], возникновением «цифровых практик», использующих компьютерные модели и цифровые технологии в качестве научного инструментария и одновременно объекта изучения. Распространение сети Интернет, увеличение мощностей компьютеров, процессы цифровизации об-

разования, культуры, социально-общественных институций сыграли существенную роль в расширении области гуманитарных исследований и становлении Цифровых гуманитарных наук (ДН). Работа исследователей в этой области направлена на оцифровку различных видов гуманитарной информации, разработку цифровых инструментов для ее обработки [7]. Цифровые гуманитарные науки по своей сути являются областью пересечения различных дисциплинарных исследовательских практик и технологических реализаций, «несут в себе все методы, средства и перспективы познания, связанные с цифровыми технологиями в области гуманитарных наук» [17], при этом стремятся преодолеть дисциплинарные границы, разрабатывая свой понятийный аппарат и собственную методологию, позволяющую использовать специфичную дисциплинарную информацию и методы специалистами другой сферы знания.

По аналогии с традиционной моделью гуманитарного знания возникла классификация гуманитарной науки в цифровой среде: цифровая история, цифровая археология, цифровая лингвистика и т. п., общими чертами формирования которых стала ориентированность на цифровые данные и информацию, создание инфраструктуры и разработка методов исследования, институализация в научной и образовательной среде. Методология цифровых гуманитарных наук открывает новые возможности для гуманитарных исследований, предоставляя открытый доступ к цифровым ресурсам, разрабатывая цифровые инструменты и исследовательские инфраструктуры, развивая новые формы и исследовательские подходы в гуманитарной среде.

Сфера применения Digital Humanities постоянно расширяется, с развитием технологических средств появляются новые области исследования. При этом практические исследования в сфере ДН в основном реализуются в формате цифровых проектов, причем еще на этапе планирования должна быть выработана междисциплинарная коммуникативная стратегия, учтены приоритеты всех партнеров проекта, установлены сроки его реализации. Кроме того, разработка и планирование цифровых проектов должна включать еще и «контентную стратегию, визуальный дизайн, интерактивный дизайн, информационную архитектуру, функциональность», а также достаточно специализированные области ДН, к которым специалисты относят «открытые данные, большие данные, базы данных, сетевую модель данных, визуализацию данных» [7].

Важнейшим принципом Digital Humanities является принцип открытых исследовательских данных (Open data), цель которого «междисциплинарный обмен данными, размещение массивов информации в различных форматах, создание научных сообществ не вокруг научных школ, а вокруг данных» [7]. Такой подход способствует поддержке публикации данных в свободном доступе. Примером реализации такого подхода является ТЕІ (Text Encoding Initiative) [18] – инициатива, ориентированная на разработку методов декодирования текстовой информации, которые делают тексты машиночитаемыми и пригодным для автоматической обработки.

Современные цифровые лингвистические решения, как правило, строятся на наиболее доступных языковых данных, не требующих специальной экспертной разметки. Это так называемые “born digital” тексты, которые изначально создавались в цифровом формате (например, Википедия). Такого рода ресурсы, как правило, доступны для скачивания и обработки, и поэтому широко используются для различных исследований в социально-гуманитарной сфере, в области автоматического языкового моделирования и искусственного интеллекта, но они не имеют специализированной экспертной разметки, что становится их естественным ограничителем как для постановки исследовательских задач, так и для применения языковых моделей. На сегодняшний день достаточно остро стоит проблема создания таких цифровых текстовых ресурсов, которые имеют качественную экспертную разметку, с одной стороны, а с другой, представляют языковые сферы и пласты, которых нет в доступном виде для обработки с помощью цифровых методов.

В 2021 году Центр цифровых гуманитарных исследований ЮФУ совместно с Международной лабораторией языковой конвергенции НИУ ВШЭ начали работу над научным

проектом «Конвергенция языковых пластов русского языка в зеркале цифровых решений» (<https://www.hse.ru/news/science/484383688.html>), цель которого состоит в создании новых лингвистических и филологических цифровых ресурсов, расширяющих стандартные типы языковых данных и дополненных наукоемкой экспертной разметкой. Реализация проекта предполагается в трех научных направлениях цифровой диалектологии, цифрового инструментария преподавания русского языка как иностранного, семантической разметки русской классической литературы. Каждое из этих направлений имеет научный задел, реализованный в ЮФУ или в НИУ ВШЭ. Общий проект обеспечит контаминацию подходов и имеющихся наработок, выработку единых стандартов и форматов представления данных, совместимость разных цифровых решений.

В рамках проекта предполагается создание трех ресурсов: корпуса донских диалектных текстов, Русского учебного корпуса «Восток» для носителей восточных языков (китайского, туркменского) и цифровое издание текстов Полного собрания сочинений и писем А. П. Чехова (1974-1983), снабженное семантической разметкой. Корпус донских диалектных текстов и Русский учебный корпус «Восток» включают в себя новый языковой материал, не представленный на доступных ресурсах. Языковые данные и глубокая экспертная разметка востребована как для научных количественных исследований, так и технологий машинного обучения, разработки образовательных продуктов. Цифровое издание текстов А. П. Чехова должно стать основой для исследования произведений писателя методами Digital Humanities [11], которые дополняют и расширяют возможности традиционных филологических подходов.

Корпус донских диалектных текстов. Работа над корпусом ведется по модели, разработанной в Международной лаборатории языковой конвергенции НИУ ВШЭ.

Изучение разговорной речи, особенно ее диалектной формы, во многом зависит от наличия и доступности соответствующего языкового материала. В настоящее время ведется активная работа по созданию диалектных электронных ресурсов, которые представляют собой результат работы исследователей-диалектологов, команды разработчиков и создателей корпуса, хотя доступ к диалектным текстам все еще ограничен для исследователей. Очевидно, что это связано как с проблемой сбора языковых данных, так и с их представлением в цифровом формате, в частности с решением вопросов автоматической разметки разговорных текстов, а также с их размещением для свободного доступа в сети Интернет.

Однако при создании такого рода электронных ресурсов возникают проблемы репрезентации фонетических, лексических, грамматических особенностей диалектных языковых единиц, которые по-разному решаются представителями различных научных центров. В состав Национального Корпуса Русского языка (НКРЯ) входит Диалектный корпус, включающий диалектные тексты, записанные на территории 22 областей Российской Федерации исконного проживания русского населения, раннего и позднего заселения, а также миграций. Объем диалектного подкорпуса составляет 390 тыс. словоупотреблений со снятой омонимией. В январе 2021 года он пополнился текстами разных фольклорных и бытовых жанров и тематик из различных регионов и диалектных зон, включая южнорусские (смоленские, тамбовские) говоры. Причем некоторые тексты сопровождаются аудио- и видеозаписями [6]. Однако донские говоры территории Ростовской области в корпусе не представлены, хотя историко-культурный и языковой ландшафт Дона особое явление, обусловленное условиями заселения территории и особенностями формирования донских говоров как говоров вторичного образования. Полагаем, что совместная работа над проектом позволит подготовить репрезентативную выборку текстов регионального диалекта для включения в НКРЯ.

Создание диалектного корпуса как полноценного научно-исследовательского и образовательного ресурса требует решения специфических задач, связанных с особенностями самого диалектного материала и его представления в корпусе – от единиц хранения, видов лингвистической разметки, формы представления метатекстовой информации, типов представленного материала до параметров выдачи текстовых/звуковых фрагментов по поисковым запросам. При этом моно-/политематическое и полижанровое представление мате-

риала определяется организацией единиц хранения в корпусе. В НКРЯ единицами хранения являются монотематические отрезки диалектной речи, в то время как в диалектных корпусах, разработанных при поддержке Лаборатории международной конвергенции (http://lingconlab.ru/index_ru.html), в качестве единиц хранения выступает устный текст, выделенный на основе зафиксированного непрерывного общения интервьюера и информанта, записанный в аудиоформате, расшифрованный и отгlossированный. Такое представление диалектного текста существенно расширяет исследовательские возможности корпуса, в том числе в коммуникативно-когнитивном аспекте.

Свободный доступ к диалектному материалу, их «грамматическая, семантическая и метатекстовая характеристика» позволяет исследователям обращаться к диалектным текстам для решения разнообразных задач – от поиска примеров до изучения и «демонстрации этнографических, этнокультурных традиций, особенностей русского менталитета» [3, с. 81], а доступ к исходным аудиозаписям обеспечивает возможность создать свою транскрипцию, не полагаясь на работу стандартных парсеров, с помощью которых создается разметка текстов для поиска. Для диалектных корпусов также важное значение имеет структура метаразметки, т. к. разговорные тексты требуют воссоздания социокультурного контекста бытования говора, поэтому значимой оказывается как информация о социокультурных характеристиках информанта, так и сведения о конкретной ситуации записи текста, адресах сообщения, времени описываемых событий. Наличие лингвистических метаданных о говорящих, включая информацию о возрасте, поле, образовании, месте жительства, владении другими языками, является отличительной особенностью разговорных корпусов Лаборатории (http://lingconlab.ru/resources_ru.html#m1). Кроме того, для расширения культурологического контекста может быть включена и разнообразная лингвистическая информация – сведения «исторического, социокультурного характера, демографические, этнографические, географические данные» [4, с. 137].

Изучением донских говоров занимаются специалисты кафедры общего и сравнительного языкознания ИФЖиМКК ЮФУ и лаборатории филологии ЮНЦ РАН, важнейшим направлением работы которых являются исследования историко-культурного наследия Юга России. В рамках постоянно действующей с 1995 г. этнолингвистической экспедиции проводится систематическое полевое обследование территории проживания донских казаков, родственных и контактных этносоциальных групп. На сегодняшний день накоплен значительный материал по языку, устной истории и традиционной культуре населения Юга России. В Корпус донских диалектных текстов войдут материалы, записанные на территории Ростовской области.

Создание корпуса донских диалектных текстов даст возможность провести систематическое изучение собранного лингвистического материала с помощью современных цифровых методов, исследовать и сохранить самобытную культуру региона. В корпус войдут полевые записи, аудиорасшифровки, представленные пользователю в виде текстовых фрагментов вместе с согласованными звуковыми сегментами. Текстовые фрагменты представляются в орфографической (стандартизированной) записи с элементами фонетической транскрипции. Включение в корпус аудиоматериалов расширит возможности для изучения «живой» диалектной речи. На материале корпуса предполагается создание Частотного словаря донских говоров.

Русский учебный корпус “Восток” (RLC). Ресурс разрабатывается по модели Русского учебного корпуса RLC [10], который позволяет проводить исследования в области усвоения иностранных языков (Second Language Acquisition), преподавания иностранных языков (Second Language teaching), языковой интерференции и т. п. Основная задача корпуса представить исследователю новый тип данных для выявления типичных речевых «ошибок» и языковых стратегий нестандартных носителей русского языка, которыми они пользуются, «чтобы преодолеть ощущаемое ими «расстояние» до слушающего» [8, с. 4]. Системное изучение речевых «ошибок», которые «дают представление о действующих языковых правилах», маркируя отклонения от нормы [9, с. 21], представляет особый интерес для теоретической лингвистики.

Важным методологическим аспектом использования учебного корпуса как современного лингвистического инструментария является возможность сформулировать «новые задачи, существенно уточнив и расширив традиционное описание русского языка» [9, с. 25].

Русский учебный корпус RLC (Russian Learner Corpus) [10] разрабатывается с 2013 г. под руководством Е. В. Рахилиной и А. С. Выренковой в Лингвистической лаборатории по корпусным технологиям НИУ ВШЭ [1, с. 368]. В корпусе представлены 9123 текста, созданных нестандартными говорящими – носителями 25 доминантных языков. Основу корпуса составляют устные и письменные тексты на русском языке двух категорий нестандартных говорящих – тех, кто изучает русский язык как иностранный, и так называемых эритажных носителей русского языка («унаследованный» язык), для которых русский язык является родным, но в качестве основного средства общения используется другой язык (чаще всего из-за эмиграции). Корпусные технологии позволяют проводить поиск как по лексико-грамматическим категориям, так и различным типам отклонений от стандартной русской речи: орфографические ошибки, различные нарушения речевых норм, неправильное употребление грамматических конструкций и т. п. Устные тексты эритажных носителей были собраны в результате лингвистических экспериментов. В тексте размечены различные типы ненормативного употребления языковых средств.

Основная часть текстов для корпуса RLC была собрана и обработана преподавателями русского языка как иностранного за рубежом. Тексты для корпуса «Восток» предоставляют преподаватели русского языка как иностранного в Южном федеральном университете. Основу корпуса составляют тексты на русском языке носителей туркменского языка, которые в корпусе RLC ранее не были представлены, и тексты носителей китайского языка, представленные в корпусе в недостаточной мере.

Важнейшим результатом проекта должна стать разработка методических рекомендаций для преподавателей русского языка как иностранного с целью предупреждения и преодоления типичных ошибок туркменских и китайских обучающихся на основе анализа материалов создаваемого корпуса. В контексте обучения русскому языку существенным представляется не только исследование и выявление типичных ошибок, но и изучение речевых стратегий и тактик нестандартных носителей языка, которые влияют на выбор говорящего. Такой подход позволит описать «причину нарушения правильности в данной точке языковой системы» [8, с. 7], а также предложить способ устранения речевой ошибки. При этом исследования на материале речевых ошибок носителей европейских языков ведутся достаточно активно, но нарушения норм русского литературного языка носителями восточных языков исследованы в недостаточной мере. Представляется важным отметить, что изучение русского языка носителями туркменского языка имеет свою специфику, т. к. в Туркменистане русский язык как иностранный изучается в рамках школьной программы.

Digital Chekhov. В рамках проекта ведется работа по созданию семантической разметки собрания произведений А. П. Чехова с опорой на стандарты цифровой публикации Text Encoding Initiative (TEI) [18], создание цифровой инфраструктуры сохранения и исследования текстов. Основной источник – академическое Полное собрание сочинений и писем А. П. Чехова в 30 томах (ПССиП) [14]. В процессе реализации проекта используются инструменты семантической разметки Полного собрания сочинений Л. Н. Толстого [12], разработанные проектной группой Центра Digital Humanities НИУ ВШЭ [16].

В данный момент разработана структура разметки ПССиП А. П. Чехова с опорой на работы исследователей-литературоведов, проводится ее уточнение и корректировка с использованием цифровых методов (см. [11]). К работе над проектом привлечены студенты магистратуры «Цифровые технологии в филологии. Компьютерная лингвистика» Института филологии журналистики и межкультурной коммуникации ЮФУ: проведена первоначальная разметка текстов 1-3 тома ПССиП А.П. Чехова в TEI-формате с учетом имен/названий, дат, формата представления, проводятся исследования для уточнения разметки текстов писем, реализована тестовая версия сайта с семантическим поиском (1-3 тт. ПССиП) [13].

Полное собрание сочинений и писем А. П. Чехова состоит из томов как основной структурной единицы, в которые включены тексты писателя за определенный период. Однако проект реализуется с опорой на структуру ЭНИ «Чехов» (ФЭБ) [15], в котором представлены тома, сегментированные по отдельным текстам. Структура аннотации корпуса текстов писателя с помощью элементов TEI включает как структурную разметку, так и разметку именованных сущностей, а также категории «время», «цвет», «природные явления», «социальный статус», «профессия». Примечания и комментарии к текстам включены в разметку каждого произведения, общие комментарии также будут размечены и представлены в отдельном описании каждого тома.

Создание ресурса включает в себя создание базы данных размеченных текстов ПС-СиП А. П. Чехова в TEI-формате с учетом имен/названий, дат, формата представления справочных материалов, текстологических, библиографических, историко-литературных комментариев, а также дополнительной разметки, определенной экспертами, в том числе разметки метафор; полной базы данных именованных сущностей, упомянутых в текстах или комментариях/примечаниях; словаря метафор языка писателя с возможностью их семантического поиска по текстам и предоставлением контекстов их употребления (открытый доступ); специализированного сайта с гибкими возможностями поиска, инструментами извлечения статистической информации и визуализации по БД размеченных текстов ПС-СиП и БД именованных сущностей.

Выводы

Реализация научного проекта «Конвергенция языковых пластов русского языка в зеркале цифровых решений» требует междисциплинарного сотрудничества представителей разных направлений, поэтому привлекаются специалисты в области лингвистики, филологии, компьютерных наук, студенты и магистранты соответствующих направлений обучения, проектная и научно-исследовательская деятельность которых ориентирована на проведение исследований по темам заявленных цифровых гуманитарных проектов. Такого рода проекты являются формами институализации Digital Humanities в научной и образовательной среде, примером создания «научных сообществ не вокруг научных школ, а вокруг данных» [7].

Список источников

1. *Власова Е. А.* Изменения в предложно-падежном управлении при финско-русском двуязычии: корпусный анализ // *Acta Linguistica Petropolitana. Труды института лингвистических исследований.* 2020. Т. 16. № 2. С. 366-396.
2. *Журавлева Е. Ю.* Современные модели развития гуманитарных наук в цифровой среде // *Вопросы философии.* 2011. № 5. С. 91-98.
3. *Качинская И. Б., Сичинава Д. В.* О корпусе диалектных текстов в национальном корпусе русского языка // *Вопросы лексикографии.* 2017. №11. С. 71-85.
4. *Крючкова О. Ю., Гольдин В. Е.* Коммуникативные свойства диалектной речи и специфика диалектного текстового корпуса // *Известия ВГПУ.* 2016. № 9-10 (113). С. 133-138.
5. *Лысак И. В.* Междисциплинарность: преимущества и проблемы применения // *Современные проблемы науки и образования.* 2016. № 5. URL: <http://www.science-education.ru/ru/article/view?id=25376> (дата обращения: 05.06.2021).
6. *Национальный корпус русского языка (НКРЯ).* URL: <https://ruscorpora.ru/new/> (дата обращения: 08.09.2021).
7. *Погорский Э. К.* Особенности цифровых гуманитарных наук // Информационный гуманитарный портал «Знание. Понимание. Умение». 2014. № 5 (сентябрь-октябрь). URL: http://www.zpu-journal.ru/e-zpu/2014/5/Pogorskiy_Digital-Humanities/ (дата обращения: 02.06.2021).

8. *Полинская М., Рахилина Е. В., Выренкова А. С.* Грамматика ошибок и грамматика конструкций: «эритажный» («унаследованный») русский язык // Вопросы языкознания. 2014. № 3. С. 3-19.
9. *Рахилина Е. В.* О новых инструментах описания русской грамматики: корпус ошибок // Русский язык за рубежом. 2016. № 3. С. 20-25.
10. *Русский учебный корпус (RLC)*. URL: <http://web-corpora.net/RLC/> (дата обращения: 08.09.2021).
11. *Северина Е. М., Ларионова М. Ч.* Новые филологические практики: семантическое издание текстов А. П. Чехова // Филология: научные исследования. 2020. № 10. С.13-21.
12. *Толстой Л. Н.* Полное собрание сочинений в 90 томах. М.; Л.: Изд-во «Художественная литература», 1928-1964.
13. Цифровой проект Chekhov Digital. URL: <http://chekhov-digital.sfedu.ru/> (дата обращения: 08.09.2021).
14. *Чехов А. П.* Полное собрание сочинений и писем: В 30 т. / АН СССР. Ин-т мировой лит. им. А. М. Горького. М.: Наука, 1974-1983. URL: <http://feb-web.ru/feb/chekhov/default.asp?feb/chekhov/texts/che-te02.html> (дата обращения: 08.09.2021).
15. Электронное научное издание «ЧЕХОВ» (ЭНИ «ЧЕХОВ»). URL: <http://feb-web.ru/feb/chekhov/default.asp> (дата обращения: 08.09.2021).
16. *Bonch-Osmolovskaya A. A., Skorinkin D., Orekhov B., Pavlova I. S., Kolbasov M. G.* Tolstoy semanticized: Constructing a digital edition for knowledge discovery // Web Semantics. 2019. No. 59. P. 100483.
17. *Dakos M.* Manifesto for the Digital Humanities. 2011 URL: <https://tcp.hypotheses.org/category/manifeste> (дата обращения: 01.09.2021).
18. Text Encoding Initiative. TEI. URL: <https://tei-c.org/> (дата обращения: 08.09.2021).

References

1. *Vlasova E.A.* Changes in prepositional-case management in Finnish-Russian bilingualism: corpus analysis // Acta Linguistica Petropolitana. Proceedings of the Institute for Linguistic Research. 2020.Vol. 16. No. 2.P. 366-396.
2. *Zhuravleva E.Yu.* Modern models of the development of the humanities in the digital environment // Problems of Philosophy. 2011. No. 5. P. 91-98.
3. *Kachinskaya IB, Sichinava DV* About the corpus of dialect texts in the national corpus of the Russian language // Questions of lexicography. 2017. No. 11. P. 71-85.
4. *Kryuchkova O. Yu., Goldin V.E.* Communicative properties of dialect speech and the specifics of the dialect text corpus // Bulletin of VSPU. 2016. No. 9-10 (113). P. 133-138.
5. *Lysak I.V.* Interdisciplinarity: advantages and problems of application // Modern problems of science and education. 2016. No. 5. URL: <http://www.science-education.ru/ru/article/view?id=25376> (accessed: 05.06.2021).
6. National Corpus of the Russian Language (NCRL). URL: <https://ruscorpora.ru/new/> (accessed: 08.09.2021).
7. *Pogorskiy E.K.* Features of digital humanities // Informational humanitarian portal "Knowledge. Understanding. Skill". 2014. No. 5 (September-October). URL: http://www.zpu-journal.ru/e-zpu/2014/5/Pogorskiy_Digital-Humanities/ (accessed: 02.06.2021).
8. *Polinskaya M., Rakhilina E. V., Vyrenkova A. S.* Grammar of errors and grammar of constructions: "eritazhniy" ("inherited") Russian language // Problems of linguistics. 2014. No. 3. P. 3-19.
9. *Rakhilina E.V.* On new tools for describing Russian grammar: the corpus of errors // Russian language abroad. 2016. No. 3. P. 20-25.

10. Russian academic corpus (RLC). URL: <http://web-corpora.net/RLC/> (accessed: 08.09.2021).
11. *Severina E.M., Larionova M. Ch.* New philological practices: semantic publication of texts by A. P. Chekhov // *Philology: scientific research*. 2020. No. 10. P.13-21.
12. *Tolstoy L.N.* Complete works in 90 volumes. M.; L.: Publishing house "Artistic Literature", 1928-1964.
13. Digital project Chekhov Digital. URL: <http://chekhov-digital.sfedu.ru/> (accessed: 08.09.2021).
14. *Chekhov A. P.* Complete works and letters: In 30 volumes / Academy of Sciences of the USSR. Institute of World Literature them. A. M. Gorky. Moscow: Nauka, 1974-1983. URL: <http://feb-web.ru/feb/chekhov/default.asp?feb/chekhov/texts/che-te02.html> (accessed: 08.09.2021).
15. Electronic scientific publication "CHEKHOV" (ENI "CHEKHOV"). URL: <http://feb-web.ru/feb/chekhov/default.asp> (accessed: 08.09.2021).
16. *Bonch-Osmolovskaya A. A., Skorinkin D., Orekhov B., Pavlova I. S., Kolbasov M. G.* Tolstoy semanticized: Constructing a digital edition for knowledge discovery // *Web Semantics*. 2019. No. 59. P. 100483.
17. *Dakos M.* Manifesto for the Digital Humanities. 2011 URL: <https://tcp.hypotheses.org/category/manifeste> (accessed: 01.09.2021).
18. Text Encoding Initiative. TEI. URL: <https://tei-c.org/> (accessed: 08.09.2021).

Статья поступила в редакцию 20.09.2021; одобрена после рецензирования 10.10.2021; принята к публикации 15.10.2021.

The article was submitted 20.09.2021; approved after reviewing 10.10.2021; accepted for publication 15.10.2021.