

МУЛЬТИМЕДИЙНЫЙ КОРПУС РЕГИОНАЛЬНЫХ ТЕКСТОВ «ЖИЗНЕННЫЙ КРУГ»: К ПРОБЛЕМЕ АРХИТЕКТониКИ ОБОЛОЧКИ

Статья подготовлена при финансовой поддержке Российского государственного научного фонда (конкурс поддержки молодых ученых 2014 года, проект № 14-34-01263 «Мультимедийный корпус вологодских текстов “Жизненный круг”»)

Проблемы построения базы данных электронного корпуса «Жизненный круг» связаны с тематическим делением материала на два крупных блока (*ТРУД* и *ИГРА*), необходимостью учета функциональных и содержательных характеристик текстов (устные / письменные; бытовое общение / художественные тексты / публицистика), введением научного комментария к текстам. Мультимедийная составляющая корпуса заключается в представлении аудио- и видеофайлов, фотографий и иллюстраций и также требует своего места в общей архитектонике базы данных.

Русская диалектология, вологодские говоры, корпусная лингвистика, электронный корпус, база данных.

Problems of a database building of the *Circle of Life* electronic corpus are associated with thematic material dividing into two major sections (*LABOUR* and *GAME*), the need to consider the functional and substantial characteristics of texts (oral / written form; social interaction / literary texts / media texts), the introduction of scientific commentary to the texts. Multimedia component of the corpus is aimed to provide audio and video files, photographs and illustrations, and also requires its place in the general database architectonics.

Russian dialectology, Vologda dialects, corpus linguistics, electronic corpus, database.

Введение.

Современная корпусная лингвистика в настоящее время переживает бум своего развития. Создание электронных корпусов не только способствует развитию и упорядочиванию различных направлений филологического знания, но и позволяет отразить историю языка, его грамматику и функционально-типологические характеристики. На протяжении нескольких последних десятилетий активно развивается диалектное направление корпусной лингвистики, причем как в России, так и за рубежом существуют несколько крупных проектов как полных корпусов национальных языков, так и собственно диалектных корпусов.

К электронным корпусам общего типа (или общего назначения) относятся национальные корпусы, которые отражают все многообразие и вариативность национального языка, все возможные его подсистемы, формы существования и разновидности, например, ряд ресурсов, посвященных английскому языку: Британский национальный корпус (*British National Corpus (BNC)*), Корпус современного американского английского языка (*Corpus of Contemporary American English (COCA)*), Национальный корпус американского английского (*American National Corpus (ANC)*) и т. д.; также известны франкоязычные корпусы *American and French Research on the Treasury of the French Language (ARTFL-FRANTEXT)*, *Lexique* и пр., разрабатываются корпусы немецкоязычных (например, *Das Deutsche Referenzkorpus (DeReKo)*), испаноязычных текстов (например, *Corpus de Referencia del Español Actual (CREA)* и др.

Часто корпус имеет ограничения на материалы, включаемые в базу данных. Так, в «Корпусе русского литературного языка» представлены тексты на

русском литературном языке; широко известны корпусы медиатекстов: например, немецкоязычный корпус *NEGR@* включает в себя газетные тексты, ресурс *TIME Magazine Corpus* составляют материалы американского журнала «Тайм» (*Time*) и пр. Ряд корпусов включает в себя такие функциональные разновидности естественных языков, как устная и письменная речь (например, Международный корпус английского языка – *International Corpus of English (ICE)*) или видеофрагменты (Национальный корпус русского языка), однако далеко не все содержат записи устной речи, ограничиваясь той или иной формой представления расшифрованного материала: запись орфографическая, транскрипция, различные комбинации орфографической записи и записи, отражающей основные фонетические особенности произношения, а также другие комплексные способы расшифровки устных текстов.

Диалектные же корпусы могут быть включены в качестве подкорпусов в национальные корпусы или же существовать автономно в виде отдельных электронных ресурсов: ср. корпус разговорных употреблений северо-германских языков *Nordic Dialect Corpus*, корпус английских диалектов *Freiburg English Dialect Corpus*, корпус эстонских диалектов *Estonian Dialect Corpus* и пр. В русскоязычном пространстве представлены оба варианта: существуют корпусы региональной речи как диалектные подкорпусы (ср. подкорпусы Национального корпуса русского языка [6]), так и проекты, направленные на описание именно диалектной речи (например, Мультимедийный Саратовский диалектологический текстовый корпус [8], другие региональные диалектные корпусы: томских говоров [12], коми диалектов [10] и др.). Типология диалектных корпусов включает в себя также

лингвокультурологические диалектные корпуса (см.: [7], [11]).

Предметом анализа данной статьи является структура лингвокультурологического мультимедийного корпуса, основной же проблемой, требующей обсуждения, становится построение электронной оболочки корпуса таким образом, чтобы получившийся электронный продукт смог стать эффективным инструментом филологических изысканий.

Основная часть.

Функциональность электронного корпуса связана с его потенциальной наполняемостью и возможностью верификации материала, поскольку необходимо представить в базе данных корпуса достаточное количество записей для объективных выводов. Поскольку в основе корпуса как информационной системы лежит та или иная вербализованная единица, все корпуса объединяются прежде всего тем, что основной единицей является *текст* на одном или нескольких языках. Данный текст как основная единица системы определяет структуру всего корпуса. В основе же типологии корпусов лежит функциональный принцип, который определяется целью создания электронного продукта. Таким образом, на этапе концептуального проектирования базы данных предполагается учет прежде всего двух факторов: характеристики текстовой единицы базы данных и назначения корпуса.

База данных «Мультимедийный корпус региональных текстов «Жизненный круг» в качестве основной единицы использует *региональный текст*, который позволяет объединить несколько подходов к изучению языка в его региональной специфике.

Во-первых, это собственно диалектологический подход, рассматривающий региональные варианты языка. С точки зрения этого подхода, региональный текст представляет собой образец диалектной речи. Данная единица представлена в рассматриваемом корпусе двояким образом: как аудио- и видеозаписи фрагментов бесед с информантами – жителями Вологодской области и носителями вологодских говоров, и также как расшифрованный текст этих записей.

Во-вторых, подход с точки зрения научного направления с не вполне устоявшимся названием (см. анализ в [9]), например, лингворегиология [1], которое изучает не «язык в регионе», а «регион в языке», т. е. «особенности самого конкретного региона как индивидуальной дискурсивной сущности, зафиксированные в тех или иных языковых фактах» [1, с. 20]. С этой точки зрения региональный текст представлен в корпусе публицистическими и художественными текстами, репрезентирующими лингвистические и культурные особенности региона. Отбор текстов производился по территориальному признаку: примеры публицистических текстов представляют собой статьи в региональной прессе – районных (например, «Северная Новь», «Звезда») и областных (например, «Премьер», «Красный Север») газетах. Выбор местной прессы обусловлен тем, что в ней наиболее последовательно и объективно отражается

культура региона, в то время как «внешние» средства массовой информации преимущественно обращаются к стереотипизированным представлениям о регионе [3]. Выборка художественных произведений определяется тематическим принципом: через критерий отражения в них индивидуально-авторского осмысления региональной картины мира.

Лингвокультурологическая направленность рассматриваемого корпуса эксплицируется в его наименовании – «Жизненный круг», – которое восходит к описанию жизни севернорусского крестьянина и его художественно-философскому осмыслению в книге В. И. Белова «Лад». Исследовательская часть корпуса строится на сопоставлении лексико-семантической репрезентации концептов <труд> и <игра> в различных дискурсивных практиках, поэтому корпус «Жизненный круг» основан на противопоставлении двух макрокомпонентов – разделов *ТРУД* и *ИГРА* [5].

Перечисленные выше составляющие концепции корпуса определяют архитектуру оболочки электронной базы данных как ключевого логического решения при ее проектировании. С точки зрения пользовательского интерфейса, целесообразно разделение электронной базы на два крупных тематических блока, внутри которых представлен текстовый материал различной дискурсивной направленности: устные тексты, публицистические и художественные. В оболочке корпуса также следует учитывать наличие поисковых инструментов, метаразметки, возможности обращения к мультимедийным материалам.

Концептуальная модель базы данных как логической системы включает в себя прежде всего «описание информационных объектов или понятий предметной области и связей между ними» [2, с. 386]. Модель электронной базы данных «Жизненный круг» ограничивается представлением тематических областей, в которых реализуются описываемые концепты (например, концепт <игра> представлен тематическими блоками «Детские игры», «Календарные праздники» и пр.). Внутри каждого тематического модуля региональные тексты сгруппированы по типам дискурсов: народно-бытовая речь, медиатексты, художественные тексты (подробнее см.: [4]).

Электронная база данных строится как реляционная модель данных, в основу которой положена совокупность взаимосвязанных таблиц. Для реализации мультимедийного корпуса используется технология объектного (объектно-ориентированного) моделирования. Она не противоречит, а, наоборот, дополняет и развивает реляционную модель данных за счет своей открытости и возможности совмещения неоднородных ресурсов. Так в мультимедийном корпусе «Жизненный круг» производится одновременная работа с такими типами данных, как текст, графика, звук, видео.

Для внешнего представления в мультимедийной оболочке выбран иерархический принцип структурирования материала (см. рисунок).

Стартовая страница	
Концепт ТРУД	Концепт ИГРА
Сенокос	Детские игры
Полевые работы	Развлечения молодежи
Огородничество	Свадьба
Обработка льна	Календарные праздники
Работа по дому	
Разведение скота, птицы	
Приготовление пищи	
Рукоделия	
Ремесла	

Рисунок. Основные тематические модули корпуса «Жизненный круг»

Для удобства работы с материалом и его поиска со стартовой страницы организована возможность перехода на элемент поиска по географическому признаку, автору, тематике текстов и их типу. Таким образом, для индексации материала внутри корпуса используются следующие признаки: концепт, тема текста, тип текста, район, автор, ключевые слова.

Для программной реализации мультимедийного корпуса была выбрана интегрированная среда программирования "Turbo Delphi Explorer", которая имеет инструменты для работы с базами данных, различными типами данных и большое количество визуальных компонентов, для организации дружественного интерфейса.

В мультимедийной оболочке корпуса используются следующие визуальные компоненты: TImage (изображение, используется для включения фото и основных изображений); TPanel (панель, для визуального оформления страниц и объединения других компонентов страницы); TMemo (большие фрагменты текста, используется для хранения объемных устных текстов корпуса); TWebBrowser (веб-страница для организации визуального поиска по географическому принципу); TMediaPlayer (встроенный медиаплеер, используется в корпусе для добавления управляющих кнопок по работе со звуковыми файлами); TTreeView (дерево данных, позволяет просматривать логическую иерархическую структуру данных корпуса); TLabel (надписи, компонент для добавления небольших текстовых фрагментов); TButton (кнопки, используются для добавления кнопок перехода между страницами корпуса).

Выводы.

Таким образом, при создании оболочки корпуса вологодских текстов «Жизненный круг» учитывались следующие факторы: лингвокультурологическая направленность базы данных, отразившаяся в иерархической структуре тематических модулей; региональный характер корпуса, реализованный включением в базу данных подкорпусы локально и тематически ограниченных текстов разных дискурсов; мультимедийный способ представления информации, иллюстрирующей и верифицирующей текстовые материалы.

Литература

1. *Ворожбитова, А. А.* Дискурсивное пространство Сочинского региона как объект лингвориторического исследования / А. А. Ворожбитова, А. С. Зубцов. – М., 2014.
2. *Дейт, К. Дж.* Введение в системы баз данных / К. Дж. Дейт; пер. с англ. – М., 2005.
3. *Драчева, Ю. Н.* Партворк «Куклы в народных костюмах»: коммуникативные особенности, лингвистическая основа, дидактический потенциал / Ю. Н. Драчева, Е. Н. Ильина // Вестник Томского государственного педагогического университета. – 2014. – №10. – С. 140–144.
4. *Драчева, Ю. Н.* Электронный корпус диалектных текстов в аспекте изучения динамики культурных концептов (на примере мультимедийного корпуса вологодских текстов) / Ю. Н. Драчева, П. Н. Задумина // Современная русская лексикология, лексикография и лингвогеография / отв. ред. О. Н. Крылова. – СПб., 2014. – С. 114–121.
5. *Драчева, Ю. Н.* Электронный корпус диалектных текстов как источник изучения региональной концептосферы / Ю. Н. Драчева, П. Н. Задумина // Славянская диалектная лексикография / отв. ред. О. Н. Крылова. – СПб., 2014. – С. 48–49.
6. *Качинская, И. Б.* Корпус диалектных текстов в Национальном корпусе русского языка: состояние и перспективы / И. Б. Качинская // Лексический атлас русских народных говоров (Материалы и исследования): 2009. – СПб., 2009. – С. 57–68.
7. *Крючкова, О. Ю.* Диалектологический корпус как источник лингвокультурологического изучения русских народных говоров / О. Ю. Крючкова // Славянский альманах. – 2010. – Т. 2009. – С. 67–84.
8. *Крючкова, О. Ю.* Корпус русской диалектной речи: концепция и параметры оценки / О. Ю. Крючкова, В. Е. Гольдин // Компьютерная лингвистика и интеллектуальные технологии. – Вып. 10 (17). – М., 2011. – С. 359–367.
9. *Мишанкина, Н. А.* Лингвистический корпус «Томский региональный текст»: теоретико-методологическое обоснование проекта / Н. А. Мишанкина // Вестник Томского государственного университета. – 2014. – № 389. – С. 28–37.
10. *Некрасова, Г. А.* Электронный диалектный корпус как ресурс сохранения и изучения коми диалектов / Г. А. Некрасова // Финно-угорский мир. – 2010. – № 1. – С. 13–16.
11. *Финько, О. С.* Возможности «Электронного корпуса диалектной культуры Кубани» для проведения этнолингвистического исследования / О. С. Финько // Современ-

ные вопросы языкознания и переводоведения: сборник научных статей / отв. ред. Н. В. Кормилина, Н. Ю. Шугаева. – Чебоксары, 2014. – С. 219–222.

12. Юрина, Е. А. Томский диалектный корпус: в начале пути / Е. А. Юрина // Вестник Томского государственного университета. Филология. – 2011. – № 2. – С. 58–63.

УДК 811.161.1

Е. Н. Ильина, Н. Н. Карачева
Вологодский государственный университет

ДИАЛЕКТНАЯ ЯЗЫКОВАЯ ЛИЧНОСТЬ И ЕЕ РЕПРЕЗЕНТАЦИЯ В ПРОЗЕ АЛЕКСАНДРА ЯШИНА

*Работа выполнена при финансовой поддержке гранта РГНФ № 15-04-00364
«Вологодский текст в русской словесности»*

Статья посвящена проблеме изучения текстов художественной прозы Александра Яшина в лингвоперсонологическом аспекте. Исследуются когнитивные, вербально-семантические и прагматические параметры диалектной языковой личности. В результате анализа явлений сюжетно-тематического, композиционного, персонажного и образного ряда делается вывод о специфике художественной картины мира писателя-вологжанина.

Диалектная языковая личность, региональная проза, А.Я. Яшин.

The article is devoted to the problem of studying the texts of Alexander Yashin's art prose in lingvo-person aspect. Cognitive, verbal and semantic and pragmatical parameters of the dialect language personality are investigated. As a result of the analysis of a subject and thematic, composite, personal and figurative row, the authors made a conclusion about specifics of art world of the writer-resident of Vologda.

Dialect language personality, regional prose, Alexander Yashin.

Введение.

Изучение феномена языковой личности на материале текстов художественной литературы относится к числу актуальных проблем лингвоперсонологии. Работа с материалом художественного наследия автора позволяет, с одной стороны, описать его самого как языковую личность [2] [5], а с другой стороны, определить авторский эстетический идеал, опираясь на речевые портреты персонажей [9]. В нашей работе на основе исследования художественной прозы Александра Яшина делается попытка изучения феномена диалектной языковой личности, черты которой, по нашему мнению, реализуются как в речи повествователя, так и в речи персонажей его текстов.

Основная часть.

Диалектной языковой личностью ученые называют совокупность проявляемых в языке свойств человека – жителя определенной местности [3]. Основу для их описания обычно составляют значительные по объему записи бытовой речи [7], [8]. На основании их анализа делается вывод о характерных особенностях диалектной языковой личности на всех уровнях ее репрезентации: когнитивном, вербально-семантическом и прагматическом [4]. На когнитивном уровне у диалектной языковой личности обнаруживаются локальные особенности восприятия действительности, картины мира. Вербально-семантический уровень определяется тем, что диалектная языковая личность использует локальный вариант национального языка – территориальный

диалект во всей совокупности его лингвистических черт. На прагматический уровень диалектной языковой личности влияет реализация ее коммуникативных задач преимущественно в сфере устноречевого бытового общения.

Личность и творчество Александра Яковлевича Яшина (1913–1968) неразрывно связаны с Вологодским краем. Здесь началась трудовая деятельность и поэтическая биография писателя. Александр Яшин довольно поздно осознал, что его талант – это, прежде всего, талант прозаика. Все, что было свойственно поэзии А. Яшина, – лиричность, драматизм, юмор, эпическое начало, народность и стройность языка, естественность диалогов, – стало характерным и для его прозы [1]. Исследуя язык повествователя в прозе Александра Яшина, можно заметить, что наряду с многими признаками креативной языковой личности русского интеллигента здесь отчетливо проявляются локальные черты. На когнитивном уровне об этом свидетельствует репрезентация традиционных ценностей крестьянской культуры, оценка всего происходящего с точки зрения нравственных устоев деревенского социума. На вербально-семантическом уровне обнаруживает себя словесная культура вологодских говоров – местная лексика и фразеология, вкрапления в авторский текст произведений устного народного творчества. На прагматическом уровне обращает на себя позиция повествователя, который наблюдает за происходящим не с превосходством цивилизованного городского наблюдателя, а с деятельным сочувствием земляка.