

УДК 811.161.1; 81-25; 81-39; 81'322
DOI 10.17223/19986645/33/6

Е.Н. Трегубова

МНОГОУРОВНЕВАЯ ТЕМАТИЧЕСКАЯ РАЗМЕТКА КАК ИНСТРУМЕНТ ЭТНОЛИНГВИСТИЧЕСКОЙ РЕПРЕЗЕНТАЦИИ ДИАЛЕКТНОГО ДИСКУРСА В ЭЛЕКТРОННОМ ТЕКСТОВОМ КОРПУСЕ¹

В статье рассматриваются принципы многоуровневой тематической разметки диалектного дискурса, используемые при создании диалектного текстового корпуса с целью моделирования локального варианта традиционной культуры. Обсуждаются возможности символического представления коммуникативных, содержательных и концептуальных параметров диалектного дискурса. Представлены уровни корпуса – подкорпусы, макродискурсы, дискурсы и микродискурсы, соотносимые с понятиями глобальной темы, макротемы, темы и микротемы.

Ключевые слова: электронный текстовый корпус, диалектный дискурс, традиционная локальная культура, говоры Кубани, концепт, гипертема, микротема, многоуровневая разметка.

Создание электронных диалектных ресурсов – одна из важнейших задач русской диалектологии, развитие которой на современном этапе определяется ростом интереса к антропологической проблематике. Дискурс как механизм и способ репрезентации концептуальной картины мира носителей региональных вариантов традиционной культуры в трудах диалектологов последнего десятилетия является одним из ведущих объектов исследования (В.Е. Гольдин, О.Ю. Крючкова, Н.В. Большакова, Е.А. Юрина и др.). Обширные фонды диалектных текстов требуют специальных форм хранения и аннотирования, таковыми являются электронные библиотеки и электронные корпуса.

Электронный текстовый корпус решает многие задачи традиционной диалектологии – создание и хранение источниковой базы, формирование картотек, получение словников, создание словарей разного типа. Диалектолог получает возможность работать с электронной библиотекой текстов, с приложениями автоматического и полуавтоматического построения словников и словарей, решается также задача включения справочных модулей. Преимуществом текстового корпуса является и его многофункциональность. Корпус – это система описания, открытая для внесения новых материалов, экспериментально-исследовательская площадка, оснащенная программным обеспечением для лингвиста-диалектолога, поисковая система, ориентированная на нужды пользователя. Кроме того, в настоящий момент представляется важным рассматривать корпус не столько как результат, сколько как

¹ Исследование выполнено при финансовой поддержке РГНФ и администрации Краснодарского края в рамках научно-исследовательского проекта «Создание электронного корпуса диалектной культуры Кубани», проект № 14-14-23007.

процесс, так как уже на этапе разработки модулей корпуса очевидной становится необходимость решения принципиально новых задач как в области анализа, систематизации данных, так и их интерпретации. Тематически размеченный диалектный корпус «позволяет получить принципиально новые сведения о диалекте, открывает перед исследователями возможности изучения диалекта не только как специфической языковой системы, но и как особого культурно-коммуникативного образования. В частности, широкая и узкая тематическая рубрикация диалектных текстов, анализ дискурсивных различий, наблюдающихся в пределах различных тематических блоков текста, записанного от одного информанта, дают возможность определить место и роль той или иной предметной области в когнитивно-дискурсивном поле индивида» [1. С. 74].

Для исследователей особенный интерес представляют моделирующие свойства текстового корпуса, позволяющие на основе заданных параметров обнаруживать особенности диалектного идиома. Инструментом такого моделирования является многоуровневая разметка, обеспечивающая возможность извлекать из корпуса скрытые «качества» диалектного текста – грамматические, семантические, коммуникативные, культурологические. Саратовский корпус, разрабатываемый В.Е. Гольдиным и О.Ю. Крючковой, описывает диалектную речь как модель сельского общения [2]. Авторы проекта считают, что «текстовая база корпуса отдельного говора должна включать не любой набор текстовых фрагментов, а стремиться к моделированию коммуникации в конкретном говоре, т.е. к моделированию говора как целостного культурно-коммуникативного образования. В связи с этим выделяются основные принципы формирования диалектологического корпуса как научно-исследовательского источника [3. С. 253]. В концепции разработки Томского диалектного корпуса используется принцип сочетания семантических и дискурсивных параметров. Список маркеров тематической метаразметки формируется на основе «ядерных концептов, выявленных в процессе анализа диалектного дискурса». В список жанровых параметров включены биографический рассказ, сюжетный рассказ, описание, рассуждение, интервью, сказка, песня, частушка, пословица. Такой принцип аннотирования диалектного дискурса, по мнению авторов, позволит в полной мере отразить «ключевые понятия крестьянской культуры, актуализированные в дискурсивной практике сибирских старожилов» [4. С. 61].

Электронный корпус диалектной культуры Кубани разрабатывается на основе лингвокультурологической концепции репрезентации диалектного дискурса [5]. Объектом описания в Корпусе является локальная традиция северо-западной части Кубани, формировавшаяся на материнской основе украинской и южнорусской культур в непосредственном контакте с линейными казаками (восточная Кубань) в среде русскоязычного населения.

В качестве единицы экспликации лингвокультурных концептов рассматривается дискурс. Нам близка точка зрения А.А. Кибрика, определяющего дискурс как «единственный заведомо реальный лингвистический объект» на том основании, что «люди разговаривают между собой дискурсами, а не предложениями и тем более не морфемами или фонемами» [6]. В настоящей статье анализу подвергается многоуровневая разметка дискурса как инстру-

мент моделирования диалектной культуры и как средство, определяющее аспекты репрезентации материалов в корпусе.

Текстовую базу электронного корпуса диалектной культуры Кубани составляют оцифрованные речевые произведения, которые представляют собой большей частью тексты-монологи, тексты-рассказы, инициированные исследователем-собирателем в ходе целенаправленной беседы, ограниченной определенной тематикой. Диалектоноситель оказывается в сложной ситуации, так как вступает в процесс коммуникативного и культурного взаимодействия с человеком другой среды, что порождает сложный дискурс – нарративный (рассказ о чем-либо), рефлексивный (попутно дается оценка явлениям старого и нового времени), пояснительный (говорящий учитывает, что слушателю может быть не все понятно). Несмотря на искусственность ситуации для говорящего, которому присваивается в этом случае особая роль – выступать в качестве репрезентанта культуры, войти в процесс воспроизведения особенностей традиции, носителем которой он является, полученные тексты могут быть рассмотрены как особая группа: тексты, в которых традиция фиксируется не в момент ее свершения, а в ситуации воспроизведения ее в процессе воспоминания, извлечения из ментальной картины мира говорящего. Ценность воспоминаний отдельного члена этнокультурного сообщества состоит в том, что события в процессе порождения текста интерпретируются на основе общей (коллективной) концептуальной картины мира и объективируются в знаках семиотических и языковых кодов, бытующих в локальной культуре. В мировидении отдельной языковой диалектной личности отражается совокупность установок социума, определяющая закономерности восприятия окружающей действительности [7]. Преобразование «сцен реального мира» в процессе речемыслительной деятельности происходит в соответствии со стандартизированными представлениями, соответствующим типовым ситуациям, типовому набору характеристик объектов (к ним относятся сценарии, фреймы, схемы) [8. С. 28].

В концепции Корпуса снимается противопоставление дискурс vs текст, поскольку отрезки речи, зафиксированные и представленные в графическом варианте, близки большей частью по своей структуре к текстам и строятся по схеме инварианта, жанровому стандарту, принятому в культуре. Отметим высокую речевую способность говорящих: информанты не затрудняются при порождении текста, как правило, изложение или описание строится в хорошем темпе. Не всегда сегодня в среде представителей письменной культуры можно наблюдать такую свободную речевую манеру. Думаем, что это и обусловлено наличием комплекса стандартных моделей речепорождения для каждой сферы деятельности человека. Человек культуры традиционной в осмыслении мира исходит из точки мировосприятия «мы», «я» понимается как неотъемлемая часть социума, встроенная в него в соответствии с правилами жизни, принятыми в нем, и сложившимися на протяжении длительного времени как результат коллективного опыта. Таким образом, материнская культура сотворяет человека ментального, помещая его в логос традиции, отличительным свойством которого является априорность опыта старших, безусловное принятие принципов жизни, существующих в культуре. Говорящий, как правило, является наследником традиции, потомком, идущим по следам

своих предков, воспроизводя ее элементы. Способность репродуцировать культурные ментальные модели, схемы закрепляется в процессе коммуникации и становится частью языковой личности представителя культуры. Исходя из этого, можно утверждать, что зафиксированный дискурс содержит речевые фрагменты, которые представляют собой текст, так как реализуются по готовым жанровым стандартам. Спонтанно порождаемый текст генерируется сознанием на базе соответствующих коммуникативной ситуации и теме речи когнитивных схем, усвоенных им в процессе разнообразной коммуникативной деятельности. В частности, при анализе диалектного дискурса *домовой* был вычленен инвариантный культурный текст, который сохраняется в сознании членов традиционного социума и является базовым в ситуации продуцирования информантами рассказа о данном мифологическом персонаже [9]. Таким образом, тексты Корпуса в процессе сегментации дискурса распределяются на группы в соответствии с жанровыми и структурно-семантическими особенностями.

В основу концепции Корпуса положен принцип тематического членения, который уже был использован при создании текстовой базы диалектного лексикографического фонда локальной культуры Кубани [10; 11]. В ряде работ были предложены принципы лексикографического описания тематического диалектного дискурса на основе информационных технологий [12; 13; 14]. В гипертекстовом словаре дискурса свадебного обряда применялся опыт фреймовой репрезентации слова и его грамматической парадигмы на материале текстов диалектоносителей станиц и хуторов северо-западной части Кубани [15].

Макроуровень диалектного корпуса представлен подкорпусами, репрезентирующими разные сферы традиционной культуры посредством тематического дискурса: «Обрядовая культура», «Традиционные верования», «Промысловая культура», «Бытовая культура» и т.д. Следующий уровень членения представлен тематическими дискурсами, входящими в определенный подкорпус. В подкорпус «Обрядовая культура» включены дискурсы «Свадебный обряд», «Рекрутский обряд», «Родинный обряд», «Крестинный обряд». Подкорпус «Традиционные верования» образуют дискурсы «Мифология» и «Народное православие».

Диалектный дискурс темы рассматривается как единица и одновременно как инструмент описания локального варианта культуры. Совокупность заданных параметров позволяет системе «описывать самое себя» в разных аспектах, выводя на уровень гипертекста глубинные свойства системы – грамматические, семантические, концептуальные, коммуникативные. Современные исследователи дискурса приходят к выводу о существовании семантических *макроструктур* и процедур их выделения на основании того, что содержание большого текста может быть выражено одним предложением [16]. В ряде современных работ отечественных лингвистов, диалектологов в том числе, также выделяется гипертематический уровень текста [17. С. 52].

Всякий дискурс, согласно точке зрения В.З. Демьянова, формально «представляет собой объединение предложений или их фрагментов», в то время как содержание дискурса концентрируется вокруг некоторого «опорного» концепта, который называется «топиком дискурса», или дискурсным

топиком [18]. Дискурсный топик может быть соотнесен с понятием микротемы. И в том и в другом случае речь идет о сегменте дискурса, характеризующемся структурно-семантической целостностью. Под дискурсом понимается «вербализованная деятельность, понимаемая как совокупность процесса и результата и обладающая как собственно лингвистическим, так и экстралингвистическим планами» [19. С. 113]. Эта обращенность дискурса одним вектором к говорящему как лицу, преобразующему ментальное в вербальное, и другим вектором – к культуре, которая располагает готовыми стандартными лексико-фразеологическими и жанровыми единицами отображения реальности (особенно в традиционной культуре), составляет предмет наблюдений лингвистов и лежит в основе сегментации дискурса определенной темы в Корпусе. Иными словами, под дискурсом мы понимаем фиксированный текст, в котором репрезентирована лингвокультурная (этнокультурная) модель отражения одной из сфер реального мира. Дискурс в рамках лингвокультурной концепции Корпуса исследуется прежде всего как сложная когнитивная структура.

Таким образом, макроединицей Корпуса является дискурс подкорпуса, единство которого обеспечивается макротемой (гипертемой), а единицей минимального членения – микродискурс (топос), сегмент дискурса, вычлененный на основе микротемы. Тема микродискурса задавалась на этапе сбора материала, и опрос информантов велся целенаправленно по блокам Корпуса (подкорпусам) – обрядовая, промысловая культура, традиционные верования и др. В подкорпус «Свадебный обряд» вошли тексты, которые были получены в ответ на вопросы «Как раньше выдавали замуж?», «Как вы выходили замуж?», «Как проходил свадебный обряд в вашей станице (вашем хуторе)?». В подкорпус «Традиционные верования» (дискурс «Мифология») включены рассказы-ответы на вопросы «Знаете ли вы что-нибудь о домовом?», «Приходилось ли вам видеть домового?», «Есть ли в Вашем селе ведьмы?», «Верите ли Вы, что сон может предсказывать будущее?», «Как раньше гадали?», «Случались ли с вами необычные истории?».

При параметризации дискурса исследователь-этнолингвист вычленяет фрагмент текста, присваивая ему символ микротемы. Так, макрособытие свадьба репрезентируется в дискурсе как ряд микрособытий, происходящих в определенных локусах и в строго определенной временной последовательности, следствием которых является трансформация в сфере социальных отношений – создается семья, меняется семейный статус целого ряда лиц. Дискурс о свадебном обряде реализуется как рассказ о комплексе событий, происходящих в линейной временной последовательности: сватовство – договор – приглашение – выпекание ритуальной пищи – приданое – вечеринки – первый день (выкупы) – венчание – застолье – дарение – брачная ночь, *садовыть сад – просыть мзду – нэсты завтрак* – второй день (куры) – ряженные – катание батькив – нарядить батькив – танцевать на одияле – заметить двор. Сценарная организация каждого события предполагает конкретных участников и роли, которые они исполняют. Рассказ о свадьбе строится на основе последовательности, известной каждому члену этнокультурного социума. Поэтому дискурс членится на соответствующие темы *сватовство, договор, приглашение, приданое* и т.д.

Микротему можно рассматривать как текстовую экспликацию коммуникативной интенции. Например, микротема *невеста* вводится в поле разметки дискурса маркером звездочка в том случае, если в рассказе есть описание невесты или ее действий. Присвоение параметра обеспечит возможность получить пользователю в Корпусе дискурс о невесте, т.е. все суждения о невесте будут представлены конкордансом микротемы.

Имена культурных концептов включаются в состав параметров, образуя отдельный уровень разметки. Наличие концепта фиксируется специальным символом – лексемой, именующей ядерный смысл концепта, например *воспитание*, *обязанность*. Список концептов формируется экспериментально в ходе анализа дискурса в программном приложении «Рабочее место лингвиста». Такой принцип параметризации обеспечит возможность концептуально-го исследования дискурса в этнолингвистическом аспекте.

Приведем пример тематической разметки. В знаках решетки размещается название подкорпуса, в фигурных скобках – тематический дискурс, в угловых скобках – тема, знак звездочки символизирует микротему (топос), концепт, элемент семиотического кода.

**#Обрядовая культура# {Свадебный обряд} <Приданое> *оценка*
четыре *дом* *время* *дружки* *деятель* *жених* *суббота*
воскресенье *символическое число***

Дальшэ... э... в суботу идуть... сейчас то колы прывызуть тэ прыданэ нызвесно, а туди обязательно ихалы в суботу за прыданым, потому шо в ныдилью начинаеця свадьба. От в суботу идуть за прыданым, за прыданым идуть до невесты: жыных, дружка – ёго друх и сват... той шо... и сващка. Все. Четыри чёловика, бильшэ за прыданым ны идуть. Значить прыхалы за прыданым, тут сразу их на воротах встричают, встритылы, повылы в хату, в хати стоить невестынэ прыданэ, ну раньшэ якэ було прыданэ: кровать, на кровати обязательно пырына, одеяло, дви подушкы обязательно було, цэ однэ... скрыня... О! В скрыни там ши одэжа була. Дальшэ шо... стил, стулья, и всё... отако... цэ сийчас всяких понавядумувалы грыцивни, а жыть нэ жывуть.

Знаком маркера звездочка помечается наличие в дискурсе элементов разных уровней – макроконцептов и концептов. К макроконцептам относятся *оценка* *время* *локус* *символическое число* *деятель*; концепты микротемы представлены единицами *дом* *ворота* *дружки* *жених* *суббота* *воскресенье* *четыре*.

**<Выкуп, приданое> *деятель* *деятели* *огласка (извещение)*
ритуальный акт *транспорт* *выставить на обозрение* *жилье*
угощение *дети* *песня* *бражка* *невеста* *жених* *соседи*
количество *дом* *свекровь* *водка* *конфеты* *переспать на кровати* *постель** *Ну и шо... на каждой вещи стоять... от, например, возли кровати стоить ши... ужэ ши ротственики вси, стоить там за кровать дэржыця, за прыголовыч дэржыця одын чёловик, манэнькэ дитё схватыло подушэчку маненьку, дэржэ, бильшэ взяло другу подушэчку – дэржэ, там... э... возли сундука стоять двое чи трое. И шо воны потрэбують. Шо воны потрэбують, ныхто нычёго ны зна. Раньшэ шо там, канхвэты, всэ там трэбувалы, малэнькым дитям канхвэты, побильшым там по рюмки водкы дава-*

лы... О!...Выкупылы оцэ всэ... и... оци, хто стояв у прыданого, хто брав выкуп, воны должни сами погрузыць ёго на... машыны, чи машыны, яки там машыны, на гарбі. На гарбу должни погрузыць ёго сами, сами прывызты. Ихать, як ихать, так должни ш ны мовчыкі ихать, а должни ш спивать писні, шоб чулы, шо прыданэ вызуть. А тут ужэ людэ... вси сосиды пособыраюця. Дывляця якэ ш прыданэ прывызты... невестынэ. И оци ж людэ, шо грузылы прыданэ, дэ там им хатку вжэ отдылылы, комнатку, воны должни занысты цэ прыданэ, поставыць ёго всэ на мисто дэ матэ-свыкруха скажэ, дэ воно должно стоять цэ прыданэ. Поставыць, скластэ постель, заправыць ш, и туди их просяць до стола. По рюмочки далы, угостылы и всэ... и воны поухалы назат до дому. А жыных должэн обязательно цю ничь на цей кровати сам пырыспаць, в цю ничь ужэ до невесты вин ны ходыв... О!

Метаразметку следует прочитать следующим образом:

<Выкуп, приданое> – текст о том, как проходил выкуп приданого. Уровень макроконцептов дискурса вычленен маркерами *деятель* *деятели* *огласка (извещение)* *ритуальный акт* *транспорт* *жилье* *угощение*. Уровень концептов микротемы представлен маркерами *дети* *песня* *выставить на обозрение* *бражка* *невеста* *жених* *соседи* *количество* *дом* *свекровь* *водка* *переспать на кровати* *постель* *конфеты*.

Маркирование концептов микротемы при условии наличия достаточного количества текстов позволяет внести в список концептов дискурса дополнения. Таким образом, лингвист, работающий с разметкой Корпуса, редактирует списки концептов дискурса и микротем.

В процессе анализа мифологического дискурса адекватным представляется привлечение понятия концепт-фрейм. Рассмотрим сегменты дискурса с общей темой *встреча мифического персонажа (МП) с человеком*. В основе рассказов о водяном, ведьме, домовом и под. лежит когнитивная схема, которая может быть представлена как список микротем (топиков). В результате тематической параметризации обнаруживаем, что дискурс о водяном организован как фрейм событие-ситуация-портрет-сакральное время-сакральное место-действие МП-действие / реакция человека-семья МП-место встречи МП с человеком. В отдельных рассказах репрезентируется часть элементов этой схемы. Параметризация дискурса позволяет выявить все элементы фрейма конкретного МП, так как в процессе разметки программа собирает слоты фрейма в «ячейку» МП. В результате можно утверждать по данным программы, что все МП в лингвокультуре описываются системой слотов, образующих статичный фрейм. Происходящее представляется в дискурсе как событие ... *було...*, но строится как ситуация конкретной «встречи» человека с МП. Событие помещается в рамки хронотопа: *Сплю якось ничью и чую, шось в ногах у меня в шэрсти*. Портрет МП образует, по-видимому, отдельный слот, так как представлен в регистре гениритивном – *маленький человечек, кошэчка, гадюк бувае*. Таким образом, концепт *домовой* строится из фреймов разной структуры – фрейма-ситуации и фрейма-описания. Сценарий позволяет излагать впечатление от происходящего как лично пережитое, фрейм-описание порождает высказывания генеритивного характера. Такой фрейм строится на основе мифологического архаического

дискурса, в котором запечатлены верования старших поколений и представления их о домовом-предке в образе ужа.

Приведем пример размеченного дискурса:

#Традиционные верования# {Мифология} <Апелляция к чужому опыту> *Частный* *<Домовой (домовиха)>* Портрет* *Действие мифического персонажа* <Домашние животные> *Действие персонажа* А соседка рассказывала, шо спыть ночью, хтось прых на кровать и начаяв ластыця. Вона кажэ: «Васька, шо ты нэ спыш? Який тяжолый, крысу сожрав чи шо? А ну брысь отсель». И кит пэрэстав ластыця. Но вона нэ чула, шоб кит на пол спрыгнув, потом фспомныла, шо кита на нич выгнала и хворткы уси позакрыла. Фстала, свит включила, ныкого нэма, домовый, мабудь, був. А кажут, шо колы кит у хати есть, то вин выгоняе усю нэчисть, а цэ кот нэ було, так домовый и ластыся.

#Традиционные верования# {Мифология} <Домовой (домовиха)> *Семья мифического персонажа* <Домовой (домовиха)> *Ситуация* *Действие человека* <Ритуальный акт>* *Речевое действие* *Физическое действие* У домового жынка есть, клычиця вона домовыхой, або хозяйкой. Як уежаеш ты со своей хаты, то зовэи ёго с собой: «Домовой, домовой, пидэи ты домой». Развязуеш мишок, кыдаеш в нёго пирижок и оставляиш на нич. Колы домовый захоче йихать з вама, вин залазэ в мишок.

#Традиционные верования# {Мифология} <Чудо> Утром просыпаются. Два мужыка муку схватыть схватылы сыби на спыну, а одишлы писят митров от брычкы, як сталы и всэ, як укупани. Ны взад, ны впэрэд. Ото ж вышлы люды – глянулы. «О! Шо за чудеса?» Начялы между собой балакать – можэ ны добиглы домой, чи шо. А цэй жэ ш мужык выходэ, хозяин, и кажэ: «Та воны и так сёдня получили. Фсю ночь стоять з мишком муки на спыни!».

#Традиционные верования# {Мифология} <Нечистая сила(нечисть)> <Ведьма> <Ведьмак> <Оборотень> *Действие мифического персонажа* *Действие человека* *Место встречи мифического персонажа с человеком* *Домашнее животное* А раньшэ на хуторах часто... Ну, в общем, просыпаются хозяйка и мужык утром ранэнько доить корову. А корова ужэ здоена. Шо за дила? Ни каждый день, ну повторялося в определенные промижутки времени. А тут хто-то посоветовал подслэдить. Бывают, шо оборотны оборощаются в это, в яку-ныбуть жывотину, видьма чи видьмак. И доют корову. Ну вот, мужык собрал сосидей. Мужыки и решылы подслэдить. Ну, бачют малую кишку. Обрубалы ей ногу пэрэднюю и отпустылы. А черэз врэмя дальшэ, чэрэз годину, черэз дви нидили побачылы – бапка вышшла с хаты, рука забинтована. Чуть ли не по локоть обрубана.

Параметризация дискурса позволяет вычленить отдельную микротему, получить конкорданс микротемы и проанализировать коммуникативные стратегии ее воплощения, установить типологические особенности фрагментов, способы ее языковой экспликации. В частности, в дискурсе микротемы ***Портрет мифического персонажа*** наблюдаем комбинацию генеритивного и репродуктивного регистров либо информативного и генеритивного. При этом используются коммуникативные тактики апелляции к чужому опыту, констатации характерных признаков мифического персонажа (информатив-

ный регистр), констатации собственного мнения, но с опорой на устойчивый культурный образ, что выражается в сочетании генеритивного и репродуктивного регистров. Приведем примеры:

<Домовой> Домовый – цэ малый чёловик, нэвидимка. Я колы у цю хату пэрэхала, стали рэмонт робыць, стала на стул бilyць, а у другой комнати «шур-шур», и ушов хтось, як вроди чёловик малэнький.

<Кикимора> и выроста кикимора, мала та нэвзрачна, як соломынка.

<Леший> Казала, шо лишый нэряшлывый, вэсь у шэрсти, левый лапаты на правой наги. Вин кривый, клышаногый, со сквэрнячим голосом, хужэ нэма нэ в кого.

<Водяной> нам казали, шо воны голи, уси в жабурэни, шо по им жабы та лягушкы прыгають, а у борода милка рыба, та головастькы путаюця.

<Водяной> Кажуть, шо у водяного е жинка – водяныха, баба с большой грудю.

<Русалка> ще у рички е русалкы, цэ полудифка, полурыба с бilyми волосымы.

<Полудница> Похожа вона була на дивку молоду, красыву с золотымы волосымы, ну мы-то сами нэ бачили, так воно е чи ни, бох ёго знае.

<Ведьма> видьму узнаты – цэ трудно. Тико колы спыть, кажуть, хвист выглядае. Колы опытна, то волосатый, колы молода, то голый. А ще и в церкви левой рукой хрэстяця да взад вид цэрквы пятаця, колы на колени станэ, так ногы обязательно скрэстэ.

<Смерть (мифический персонаж)> постучали. Я открываю – стоить большая жэницина отут на двэрях, чёрная вся, у чёрном, и сама смутная, чёрная, ага. «Так, – говорю, – а чё ты прышла?» А она ужэ пэрэт смэртью явилась. Я закрыла дверь и заципнула, а вона добиваеца, вона добиваеца. Ну дней можэ за пять, за шэсть, пэрэд тем, шо мама умерла. Ужэ явилось мне, шо мама умрёт. Ужэ я знала.

Таким образом, использование многоуровневой тематической разметки позволяет представить диалектный дискурс как лингвокультурный феномен, извлекая на уровень гипертекста скрытые коммуникативные и культурологические свойства описываемой системы, представляя их в совокупности параметров, коррелируемых с разными модулями корпуса. В списках полученных экспериментальным путем параметров, будет отражена лингвокультурная модель диалектной традиции.

Литература

1. Гольдин В.Е., Крючкова О.Ю. Тематическая разметка и тематический анализ диалектного текстового корпуса // Языковая личность – текст – дискурс: теоретические и прикладные аспекты исследования: материалы междунар. науч. конф.: в 2 ч. Ч. 1. Самара, 2006. С. 71–80.
2. Крючкова О.Ю., Гольдин В.Е., Сдобнова А.П. Электронный диалектный корпус как новый источник изучения русских народных говоров // Язык и культура в России: состояние и эволюционные процессы: материалы междунар. науч. конф. Самара, 2007. URL: http://sarteorlingv.narod.ru/dialekt/isto4nik_izu4enia.html
3. Крючкова О.Ю., Гольдин В.Е. Проблемы создания электронного диалектологического корпуса // Русская устная речь: материалы междунар. науч. конф. «Баранниковские чтения. Устная речь: русская диалектная и разговорно-просторечная культура общения» и межвузов-

ского совещания «Проблемы создания и использования диалектологических корпусов», Саратов, 15–17 ноября 2010 г. Саратов, 2011. С. 248–259.

4. Юрина Е.А. Томский диалектный корпус: В начале пути // Вестн. Том. гос. ун-та. Филология. 2011. № 2 (14). С. 59–63.

5. Трегубова Е.Н., Емельянова М.В. Лингвокультурный дискурс и специфика его репрезентации в диалектном корпусе // Русская устная речь: материалы междунар. науч. конф. «Бараниковские чтения. Устная речь: русская диалектная и разговорно-просторечная культура общения» и межвуз. совещания «Проблемы создания и использования диалектологических корпусов», Саратов, 15–17 ноября 2010 г. Саратов, 2011. С. 264–269.

6. Кибрик А.А. Модус, жанр и другие параметры. Классификации дискурсов. Институт языкознания РАН. URL: kibrik@comtv.ru

7. Иванцова Е.В. Мировидение языковой личности в традиционной русской народно-речевой культуре // Вестн. Том. гос. ун-та. Филология. 2014. №4 (30). С. 27–42.

8. Темнова Е.В. Современные подходы к изучению дискурса // Язык, сознание, коммуникация: сб. ст. / отв. ред. В.В. Красных, А.И. Изотов. М., 2004. Вып. 26. С. 24–32.

9. Трегубова Е.Н., Резец Д.В. Кубанский культурный дискурс домового // Антропоцентрическая парадигма в филологии: Материалы Междунар. науч. конф., Ставрополь, 14–15 мая, 2003 г. Ч. 2. Лингвистика. Ставрополь, 2003. С. 300–307.

10. Трегубова Е.Н., Емельянова М.В. Лингвокультурный региональный дискурс и современные способы его лексикографического описания // Конференция грантодержателей регионального конкурса Российского гуманитарного научного фонда и администрации Краснодарского края «Северный Кавказ: традиции и современность»: сб. тез. Краснодар, 2007. С. 48–51.

11. Трегубова Е.Н. Кубанский лингвокультурный дискурс «Свадьба» и его лексикографическое описание: к вопросу о научной версии современного диалектного словаря // Региональное своеобразие языка и культурных традиций кубанского фольклора. Славянск-на-Кубани, 2005. С. 9–12.

12. Емельянова М.В., Трегубова Е.Н. Применение информационных технологий при обработке лингвокультурного дискурса: материалы 33-й Междунар. филол. конф. Вып. 25, секция прикладной и математической лингвистики, т. 2 15–20 марта 2004 г., филологический факультет Санкт-Петербургского государственного университета. СПб., 2004. С. 38–44.

13. Емельянова М.В. Принципы создания компьютерного словаря кубанских говоров // Филология на рубеже тысячелетий: материалы Междунар. науч. конф. Вып. 2: Язык как функционирующая система. Ростов н/Д, 2000. С. 23–25.

14. Трегубова Е.Н. Диалектный дискурс: новые аспекты изучения // Вестн. Славянского на- Кубани гос. пед. ин-та. 2005. № 1 (1). С. 17–22.

15. Емельянова М.В., Трегубова Е.Н. Гипертекстовый мультимедийный словарь лингвокультурного дискурса // Русский язык: исторические судьбы и современность: 2-й Междунар. конгресс исследователей русского языка, Москва, МГУ им. М.В. Ломоносова, филологический факультет, 18–21 марта 2004 г.: Труды и материалы / сост. М.Л. Ремнёва, О.В. Дедова, А.А. Поликарпов. 2004. С. 472–473.

16. Дейк Т.А. ван. Язык. Познание. Коммуникация: пер. с англ. / сост. В.В. Петрова; под ред. В.И. Герасимова; вступ. ст. Ю.Н. Караулова, В.В. Петрова. М.: Прогресс, 1989. 312 с.

17. Косицина Ю.В. Статико-динамическая модель тематической организации диалектного монологического текста. Ст. 1 / Ю.В. Косицина, С.П. Петрунина // Вестн. Челябин. гос. ун-та. 2013. № 4. С. 52–54.

18. Демьянков В.З. Политический дискурс как предмет политологической филологии // Политическая наука. Политический дискурс: История и современные исследования. № 3. М., 2002. С. 32–43.

19. Красных В.В. «Свой» среди «чужих»: миф или реальность? М.: Гнозис, 2003. 375 с.

MULTILEVEL THEMATIC MARKING AS AN ETHNOLINGUISTIC TOOL OF DIALECTAL DISCOURSE REPRESENTATION IN DIGITAL TEXT CORPORA.

Tomsk State University Journal of Philology, 2015, 1(33), pp. 66–77. DOI 10.17223/19986645/33/6
Tregubova Yelena N., Kuban State University, Branch in Slavyansk-on-Kuban (Slavyansk-on-Kuban, Russian Federation). E-mail: tregubova.el@yandex.ru

Keywords: digital text corpora, dialectal discourse, traditional local culture, Kuban dialects, concept, macro theme, micro theme, multilevel marking.

The elaboration of dialectal corpora is one of the up-to-date problems in modern dialectology, for digital resources allow to solve problems either in traditional dialectology or in cognitive and discourse linguistics. The most important tool to build up corpora is multilevel thematic marking. Its structure is determined by the conception of the corpora authors.

Accordingly, the communicative parameters are set in Saratov corpus describing the dialectal system as a type of rural communication. Digital Kuban dialectal corpus is being elaborated on the basis of a linguocultural approach to represent dialectal discourse. The object described in the Corpus is a local tradition of the north-west of Kuban territory. This tradition is based on Ukrainian and Southern Russian cultures in immediate contact with Linear Cossacks (Eastern part of Kuban) among Russian-speaking population. Discourse is discussed as an explication unit of linguocultural concepts. The opposition discourse vs text is removed in the Corpus conception because discourse pieces recorded and represented graphically are structurally close to texts and are built up according to the scheme of the invariant which is a genre standard accepted in culture.

The minimal unit of division is micro discourse (topos) or discourse segment based on a micro theme. The multilevel thematic marking mechanism allows to build up dialectal discourse as a linguocultural phenomenon. Also it helps to reveal latent communicative and cultural properties of the described system and represent these properties within the parameters correlated with various corpus modules. Cultural concept nominations are included in the parameters to form a new level of marking. Concept is represented by a special sign – a lexeme nominating the nuclear meaning of the concept. Segment marking of thematic discourse in sub-corpora is as follows:

#Ritual culture#{Wedding rite}<Dowry>*mark**four* *house* *time* * bridesmen* *doer* *bridegroom* *Saturday*Sunday* *symbolic number

#Spiritual culture#{Mythology}<Hobgoblin (female hobgoblin)>*Family of a mythological character*Situation**Man's act*<Ritual act>* *Speech act* *Physical action*. In text marking the following signs are used: # (octothorpe) sub corpus, {} (curly braces) thematic discourse, < > (angle brackets) theme, * (asterisk) micro theme (topos), concept, an element of a semiotic code.

References

1. Gol'din V.E., Kryuchkova O.Yu. [Theme marks and thematic analysis of dialect text corpus]. *Yazykovaya lichnost' – tekst – diskurs: teoreticheskie i prikladnye aspekty issledovaniya: materialy mezhdunar. nauch. konf.: v 2 ch.* [Language personality – text – discourse: theoretical and applied aspects of research: Proceedings of the International Scientific Conference. In 2 parts]. Samara: Samarskiy Universitet Publ., 2006. Pt. 1, pp. 71–80. (In Russian).
2. Kryuchkova O.Yu., Gol'din V.E., Sdobnova A.P. [Electronic dialect corpus as a new source of studying Russian folk dialects]. *Yazyk i kul'tura v Rossii: sostoyanie i evolyutsionnye protsessy: materialy mezhdunarodnoy nauchnoy konferentsii* [Language and Culture in Russia: state and evolutionary processes: Proceedings of the International Conference]. Samara: Samarskiy Universitet Publ., 2007. Available from: http://sarteorlingv.narod.ru/dialekt/isto4nik_izu4enia.html. (In Russian).
3. Kryuchkova O.Yu., Gol'din V.E. Problemy sozdaniya elektronnoy dialektologicheskogo korpusa [Problems of creating an electronic dialectological corpus]. *Russkaya ustnaya rech'. Materialy mezhdunar. nauch. konf. "Barannikovskie chteniya. Ustnaya rech': russkaya dialektnaya i razgovorno-prostorechnaya kul'tura obshcheniya"* [Russian speaking. Proceedings of the International Scientific Conference "Barannikovskie readings. Speaking: Russian dialect and colloquial vernacular culture of communication"]. Saratov: Nauka Publ., 2011, pp. 248–259. (In Russian).
4. Yurina E.A. Tomsk dialectal corpora: the starting point. *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology*, 2011, no. 2 (14), pp. 59–63. (In Russian).
5. Tregubova E.N., Emel'yanova M.V. [Linguo-cultural discourse and specificity of its representation in the dialect corpus]. *Russkaya ustnaya rech'. Materialy mezhdunar. nauch. konf. "Barannikovskie chteniya. Ustnaya rech': russkaya dialektnaya i razgovorno-prostorechnaya kul'tura obshcheniya"* [Russian speaking. Proceedings of the International Scientific Conference "Barannikovskie readings. Speaking: Russian dialect and colloquial vernacular culture of communication"]. Saratov: Nauka Publ., 2011, pp. 264–269. (In Russian).
6. Kibrik A.A. *Modus, zhanr i drugie parametry. Klassifikatsii diskursov* [Modus, genre and other parameters. Classifications of discourse]. Available from: http://iling-ran.ru/kibrik/Discourse_classification@VJa_2009.pdf.

7. Ivantsova E.V. Worldview of the language personality in the traditional Russian folk-speech culture. *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology*, 2014, no. 4 (30), pp. 27–42. (In Russian).

8. Temnova E.V. *Sovremennye podkhody k izucheniyu diskursa* [Modern approaches to the study of discourse]. In: Krasnykh V.V., Izotov A.I. (eds.) *Yazyk, soznanie, kommunikatsiya* [Language, consciousness, communication]. Moscow: MAKSS Press Publ., 2004. Issue 26, pp. 24–32.

9. Tregubova E.N., Rezets D.V. [Cuban cultural house discourse]. *Antropotsentricheskaya paradigma v filologii: Materialy Mezhdunar. nauch. konf.* [Anthropocentric paradigm in Philology: Proceedings of the International Scientific Conference]. Stavropol, 2003. Pt. 2, pp. 300–307. (In Russian).

10. Tregubova E.N., Emel'yanova M.V. [Linguo-cultural regional discourse and modern methods of its lexicographic description]. *Konferentsiya grantoderzhateley regional'nogo konkursa Rossiyskogo gumanitarnogo nauchnogo fonda i administratsii Krasnodarskogo kraia "Severnyy Kavkaz: traditsii i sovremennost'"* [Conference of the grantees regional competition of the Russian Humanitarian Science Foundation and the Administration of Krasnodar Region "North Caucasus: tradition and modernity"]. Krasnodar: Prosveshchenie-Yug Publ., 2007, pp. 48–51. (In Russian).

11. Tregubova E.N. *Kubanskiy lingvokul'turnyy diskurs "Svad'ba" i ego leksikograficheskoe opisanie: k voprosu o nauchnoy versii sovremennogo dialektного slovary* [Kuban linguo-cultural discourse "Wedding" and its lexicographical description: On the scientific version of the modern dialect dictionary]. In: *Regional'noe svoeobrazie yazyka i kul'turnykh traditsiy kubanskogo fol'klora* [Regional originality of language and cultural traditions of the Kuban folklore]. Slavyansk-on-Kuban: SGPI Publ., 2005, pp. 9–12.

12. Emel'yanova M.V., Tregubova E.N. [Application of information technology in the processing of linguocultural discourse materials]. *Materialy XXXIII Mezhdunar. filologicheskoy konf.* [Proceedings of XXXIII International Philological Conference]. St. Petersburg: St. Petersburg State University Publ., 2004. Issue 25, vol. 2, pp. 38–44. (In Russian).

13. Emel'yanova M.V. [Principles of creation of a computer dictionary of Kuban dialects]. *Filologiya na rubezhe tysyacheletiy: materialy Mezhdunar. nauch. konf.* [Philology at the turn of the Millennia. Proceedings of International Scientific Conference]. Rostov-on-Don: Donskoy izdatel'skiy dom Publ., 2000. Issue 2, pp. 23–25. (In Russian).

14. Tregubova E.N. *Dialektnyy diskurs: novye aspekty izucheniya* [Dialect discourse: new aspects of the study]. *Vestnik Slavyanskogo-na-Kubani gosudarstvennogo pedagogicheskogo instituta*, 2005, no. 1 (1), pp. 17–22.

15. Emel'yanova M.V., Tregubova E.N. [Multimedia hypertext dictionary of linguocultural discourse] *Russkiy yazyk: istoricheskie sud'by i sovremennost': II Mezhdunar. kongress issledovateley russkogo yazyka: Trudy i materialy* [Russian language: historical destiny and present: II International Congress of Russian researchers: Proceedings and materials]. Moscow: Moscow State University Publ., 2004, pp. 472–473. (In Russian).

16. Dijk T.A. van. *Yazyk. Poznanie. Kommunikatsiya* [Language. Cognition. Communication]. Translated from English. Moscow: Progress, 1989. 312 p.

17. Kositsina Yu.V., Petrunina S.P. Statiko-dinamicheskaya model' tematicheskoy organizatsii dialektного monologicheskogo teksta. Stat'ya I [Static-dynamic model of thematic organization of dialect monologic text. Article One]. *Vestnik Chelyabinskogo gosudarstvennogo universiteta*, 2013, no. 4, pp. 52–54.

18. Dem'yankov V.Z. *Politicheskyy diskurs kak predmet politologicheskoy filologii* [Political discourse as an object of political philology]. In: Gerasimov V.I., Il'in M.V. (eds.) *Politicheskaya nauka. Politicheskyy diskurs: Istoriya i sovremennyye issledovaniya* [Political science. Political discourse: History and current research]. Moscow, 2002, no. 3. pp. 32–43.

19. Krasnykh V.V. *"Svoy" sredi "chuzhikh": mif ili real'nost'?* ["Proxy ally": myth or reality?] Moscow: Gnozis Publ., 2003. 375 p.