



НАУЧНАЯ СТАТЬЯ

УДК 81'44

Дата поступления: 17.07.2024
рецензирования: 19.09.2024
принятия: 15.11.2024

Тематическая разметка антропологического корпуса: методика классификации шахтерских нарративов

Л.Л. Мазитова

Пермский филиал Национального исследовательского университета
«Высшая школа экономики», г. Пермь, Российская Федерация
E-mail: mazitova-larisa@rambler.ru. ORCID: <http://orcid.org/0000-0002-6775-8233>

Л.М. Пантелеева

Пермский филиал Национального исследовательского университета
«Высшая школа экономики», г. Пермь, Российская Федерация
E-mail: liliya_pant@mail.ru. ORCID: <https://orcid.org/0000-0002-5815-1288>

Аннотация: В статье излагается методика создания антропологического корпуса текстов, объединенных принадлежностью к шахтерской профессии. Содержание работы коррелирует с тремя исследовательскими задачами: разработкой тематической классификации, введением условных обозначений для выделения нарратива в тексте, определением принципов организации корпуса по тематике нарративов. Тематическая классификация нарративов стала результатом анализа нескольких «контрольных» текстов. Она представляет собой ступенчатую систематизацию культурно-профессиональной тематики, поскольку, помимо основных (базовых) тем, может быть проведена их внутренняя детализация, что приводит к появлению микротем. Количество таких микротем может быть различным и определяется спецификой самой темы, т. е. возможностью характеризовать соответствующее явление действительности с различных сторон. Фрагменты текста, в которых реализуется та или иная тема или подтема, выделяются квадратными скобками, по обе стороны от которых указываются цифровые обозначения соответствующей темы/подтемы из приведенной ранее тематической классификации. При тематической разметке выдерживаются принципы соответствия нарратива основной теме корпуса, незавершенности тематической классификации, цельности нарратива, нежесткой разметки, учета «нулевых» тем. В статье описываются частные проблемы разметки, когда та или иная тема не разворачивается информантом, притом что контекст допускает различную трактовку темы нарратива. В подобных ситуациях в ходе процесса разметки отделяется целенаправленно репрезентируемая информантом тематика от фактов, которые могут иметь лишь косвенное отношение к какой-либо теме, но могут и не иметь такового. Таким образом, описанная методика представляет собой первый подход к разработке стандарта метаразметки. Применяемый тип разметки по специфике предметной информации можно назвать экстралингвистическим, по способу выявления – аналитическим, по объекту – нарративным, по глубине классификации – многоуровневым, по способу присвоения меток – ручным, по соотношению с определенной теорией – нейтральным.

Ключевые слова: антропологический корпус; компьютерная лингвистика; экстралингвистическая разметка; тематическая классификация; нарратив.

Благодарности. Статья подготовлена в рамках проекта «Зеркальные лаборатории» НИУ ВШЭ.

Цитирование. Мазитова Л.Л., Пантелеева Л.М. Тематическая разметка антропологического корпуса: методика классификации шахтерских нарративов // Вестник Самарского университета. История, педагогика, филология Vestnik of Samara University. History, pedagogics, philology. 2024. Т. 30, № 4. С. 156–164. DOI: <http://doi.org/10.18287/2542-0445-2024-30-4-156-164>.

Информация о конфликте интересов: авторы заявляют об отсутствии конфликта интересов.

© Мазитова Л.Л., Пантелеева Л.М., 2024

Лариса Львовна Мазитова – кандидат филологических наук, Пермский филиал Национального исследовательского университета «Высшая школа экономики», 614070, Российская Федерация, г. Пермь, ул. Студенческая, 38.

Лилия Михайловна Пантелеева – кандидат филологических наук, научный сотрудник научно-учебной лаборатории междисциплинарных эмпирических исследований, Пермский филиал Национального исследовательского университета «Высшая школа экономики», 614070, Российская Федерация, г. Пермь, ул. Студенческая, 38.

SCIENTIFIC ARTICLE

Submitted: 17.07.2024
Revised: 19.09.2024
Accepted: 15.11.2024

Thematic marking of the anthropological corpus: a methodology for classifying miners' narratives

L.L. Mazitova

HSE Campus in Perm, Perm, Russian Federation
E-mail: mazitova-larisa@rambler.ru. ORCID: <http://orcid.org/0000-0002-6775-8233>

L.M. Panteleeva

HSE Campus in Perm, Perm, Russian Federation

E-mail: liliya_pant@mail.ru. ORCID: <https://orcid.org/0000-0002-5815-1288>

Abstract: The article describes the methodology for creating an anthropological corpus of texts that are united by belonging to the mining profession. The content of the work correlates with three research tasks: development of a thematic classification, introduction of conventions for highlighting narratives in the text, 3) determination of principles for organizing the corpus according to the themes of narratives. Thematic classification of narratives was the result of the analysis of several «control» texts. It represents a multi-level systematization of topics of cultural and professional nature: when the main (basic) topics can have internal detail, which leads to the emergence of micro-topics. The number of such microtopics can be different and is determined by the specifics of the topic itself, i.e. the ability to characterize the corresponding phenomenon of reality in various aspects. Fragments of text in which a particular topic or subtopic is implemented are highlighted with square brackets. On both sides of the brackets, the numerical designations of the corresponding topic/subtopic from the previously given thematic classification are indicated. During thematic marking, the principles of compliance of the narrative with the main theme of the corpus, incompleteness of the thematic classification, integrity of the narrative, non-rigid marking, taking into account «zero» topics are maintained. The article describes particular problems of marking, when a particular topic is not developed by the informant, despite the fact that the context allows for different interpretations of the topic of the narrative. In such situations, during the marking process, the topic purposefully represented by the informant is separated from facts that may have only an indirect relation to any topic, but may not have any. Thus, the described methodology represents the first approach to developing a meta-markup standard. The type of marking used can be called extra-linguistic based on the specifics of the subject information, analytical based on the method of identification, narrative based on the object, multi-level based on the depth of classification, manual based on the method of assigning labels, and neutral in relation to a certain theory.

Key words: anthropological corpus; computational linguistics; extralinguistic markup; thematic classification; narrative.

Acknowledgements. The article was prepared within the framework of the project «Mirror Laboratories» of the HSE University, RF.

Citation. Mazitova L.L., Panteleeva L.M. Thematic marking of the anthropological corpus: a methodology for classifying miners' narratives. *Vestnik Samarskogo universiteta. Istorii, pedagogika, filologiya Vestnik of Samara University. History, pedagogics, philology*, 2024, vol. 30, no. 4, pp. 156–164. DOI: <http://doi.org/10.18287/2542-0445-2024-30-4-156-164>. (In Russ.)

Information on the conflict of interests: the authors declared no conflicts of interest.

© Mazitova L.L., Panteleeva L.M., 2024

Larisa L. Mazitova – Candidate of Philological Sciences, HSE Campus in Perm, 38, Studencheskaya Street, Perm, 614070, Russian Federation.

Liliya M. Panteleeva – Candidate of Philological Sciences, researcher at the Scientific and Educational Laboratory of Interdisciplinary Empirical Research, HSE Campus in Perm, 38, Studencheskaya Street, Perm, 614070, Russian Federation.

Введение

Разметка – это процесс или результат приписывания текстам и их компонентам служебных пометок – тегов. Лингво-антропологическая разметка – одно из основных понятий корпусной антропологии. Разметка дает возможность идентифицировать тексты по различным параметрам, позволяя осуществлять осмысленный поиск по корпусу [Atkins, Clear, Ostler 1992, p. 8].

Для решения задач корпусной антропологии релевантна дискурсная тематическая разметка, которая позволяет группировать материал в соответствии с различными аспектами антропологического исследования. Однако среди современных корпусов тематика отнюдь не частотный способ разметки материала (в отличие от внешних параметров), поскольку требует значительных затрат на анализ содержания текста.

Нам известно только о трех общедоступных антропологических корпусах, где тематическое содержание текстов выступает основным или одним из основных принципов разметки: «Народная история России» [Народная история], «Тематическая классификация и распределение фольклорно-мифологических мотивов по ареалам» [Тематическая классификация...], «Электронный текстовый корпус лингвокультуры Северного Приангарья» [Электронный текстовый корпус]. В первом из

них объектом классификации выступают рассказы о городском пространстве, легендах и персонажах, во втором – мифологический фольклор разных народов, в третьем – устная речь носителей традиционной культуры.

Тематическая разметка применяется и в корпусной лингвистике, ее применение осуществляется как в области литературного языка [Савчук 2005], так и нелитературных подсистем, в частности диалектов [Летучий 2005; Буранова 2012; Трегубова 2015; Лавров, Харламова, Костюшина 2018; Земичева, Иванцова 2020]. (Отметим, что темники, лежащие в основе классификаций, имеют разный состав и структуру, что нередко осмысливается составителями как проблема, препятствующая анализу материала в межкорпусном аспекте).

Создавая антропологический корпус текстов, объединенных принадлежностью к шахтерской профессии, мы не могли миновать этап тематической разметки, поскольку содержание нарративов с одинаковыми темами высвечивает предметную специфику профессиональной картины мира. В ходе этой работы перед нами встали следующие задачи: 1) разработать тематическую классификацию; 2) ввести условные обозначения для выделения нарратива в тексте; 3) определить принципы организации корпуса по тематике нарративов. Методика их выполнения последовательно изложена в настоящей статье.

1. Описание внешних параметров текста

Тексты, привлекаемые для изучения профессиональной культуры методами антропологического анализа в рамках корпусной антропологии в ходе данного исследования, представляют собой письменные расшифровки устных нарративов представителей шахтерской профессии. Внешне тексты, подвергающиеся разметке, представляют собой чередование вопросов, задаваемых интервьюером (И), и ответов (как кратких, так и более или менее развернутых), даваемых респондентом (Р). Нарративы получены в ходе устных бесед с работниками отрасли различных возрастов и различной общей культуры. Информанты неодинаковы по степени освоения тонкостей производственного процесса, по мере вовлеченности в социальные отношения и той профессионально-социальной роли, носителями которой они являются. С этим связаны заметные отличия в характере нарративов, включая большую или меньшую степень подробности изложения, связности, отчетливости выражения мысли. Очевидно, что нарративы неодинаковы по своей информативности. К тому же они обнаруживают явную асимметричность в репрезентации различных сторон профессиональной культуры.

2. Тематическая разметка

Для электронной систематизации имеющихся нарративов проведена их тематическая разметка, которая представляет собой выделение фрагментов текста, содержащих ту или иную информацию о различных проявлениях шахтерской профессиональной культуры, как материальной, так и духовной. Это сведения и о шахтерской повседневности в длительном историко-культурном контексте, и о функциональных состояниях работников и стадиях их профессиональной адаптации, и о формальных и неформальных отношениях в производственной среде. В качестве составляющих профессиональной культуры шахтеров рассматриваются также и различные социально-культурные практики, включая семейные традиции, что способствует выявлению мировоззренческих и аксиологических установок респондентов.

Для решения антропологических задач по разработке корпуса шахтерских нарративов необходима предварительная разработка классификации для тематической разметки. В качестве исходной базы для тематической систематизации целесообразно рассматривать вопросник (программу опроса), служивший ориентиром при построении беседы с информантами. Именно он задает основные направления характеристики профессионально-культурной среды. Однако, как показывает практика работы с конкретными текстами, эта рамка требует детализации, дополнения и корректировки. Такой подход обусловлен тем, что в ходе реализации полевых исследований живые контакты с информантами позволили расширить и углубить исходные установки планируемого антропологического описания. Таким образом, тематическая классификация нарративов стала резуль-

татом анализа нескольких «контрольных» текстов. Важно подчеркнуть, что классификация составлялась таким образом, чтобы ее можно было использовать для тематической разметки нарративов представителей не только шахтерской, но и ряда других производственных отраслей.

Тематическая классификация представляет собой ступенчатую систематизацию культурно-профессиональной тематики, поскольку помимо основных (базовых) тем может быть проведена их внутренняя детализация, что приводит к появлению микротем. Количество таких микротем может быть различным и определяется спецификой самой темы, т. е. возможностью характеризовать соответствующее явление действительности с различных сторон.

3.1. Тематическая классификация нарративов

1. Выбор профессии

- 1.1. причины
- 1.2. отношение родных
- 1.3. профессиональная династичность

2. Вход в профессию

- 2.1. первый рабочий день
- 2.2. обучение и адаптация

3. Посттрудова жизнь

4. Объект и результаты труда

- 4.1. характеристика объекта труда
- 4.2. применение продукта труда в экономике

5. Специфика работы

- 5.1. трудовые действия и операции
- 5.2. оборудование, материалы, инструменты
- 5.3. развлечения на рабочем месте

6. Условия труда

- 6.1. график работы
- 6.2. специальные помещения
- 6.3. форма одежды и экипировка
- 6.4. рабочее место
- 6.5. оценка условий труда
- 6.6. нештатные ситуации
- 6.7. профессиональные болезни
- 6.8. отдых на рабочем месте (перерывы на обед, перекур и т. п.)
- 6.9. взаимодействие с живой природой на рабочем месте

7. Трудовой коллектив

- 7.1. члены коллектива
- 7.2. достижения коллектива и личные достижения
- 7.3. конкуренция в трудовой среде
- 7.4. отношения в коллективе

8. Руководство

- 8.1. отношения с администрацией
- 8.2. руководитель, характеристика стиля руководства

9. Культурное пространство

- 9.1. Этнокультурная характеристика трудовой сферы
- 9.2. Массовая культура, фольклор и творчество в профессионально-бытовой среде

10. Корреляция профессии с инфраструктурой

11. Достаток

- 11.1. оценка доходов по сложности работы
- 11.2. оценка доходов по сравнению с городским окружением
- 11.3. оценка доходов по покупательской способности
- 11.4. оценка доходов в ретроспективе

11.5. оценка доходов по сравнению с другими профессиями, регионами

12. Девиантные события в трудовом процессе

12.1. Забастовки

12.2. Нарушения ТБ и правил поведения

13. Трудовые традиции, ритуалы и праздники

13.1. профессиональный праздник

13.2. приход нового работника

13.3. выход на пенсию

13.4. похороны члена коллектива

13.5. уход в отпуск

13.6. другие торжества и события

14. Досуг

14.1. отношения коллектива в бытовой повседневности

14.2. организация досуга руководством

14.3. проведение выходных и отпусков

15. Сны о трудовой деятельности

16. Аксиология

16.1. привлекательность профессии в общественном сознании

16.2. качества истинного представителя профессии

16.3. профессиональные авторитеты в коллективе

16.4. знаменитые представители профессии

16.5. приверженность профессии

16.6. альтернативные варианты труда

17. Семья

17.1. профессиональная символика в домашней обстановке

17.2. отношение семьи к профессии

17.3. трансляция профессиональной культуры

17.4. семейный уклад представителя профессии

18. Вера и суеверия

18.1. святой покровитель

18.2. иная мифология

18.3. суеверия, приметы, ритуалы

18.4. необъяснимые случаи

19. Профессиональные СМИ

19.1. официальные медиа

19.2. средства неофициальной коммуникации

20. Деградация отрасли

3.2. Условные обозначения

Фрагменты текста, в которых реализуется та или иная тема или подтема, выделяются квадратными скобками, по обе стороны от которых указываются цифровые обозначения соответствующей темы/подтемы из приведенной ранее тематической классификации. Например, в приведенном ниже фрагменте реализуется тема «Оценка условий труда»:

6.5[И: *А вам как у него работалось?* // **Р:** *Ну / как-то работалось // Вот например крепь / вот эта которая / она весит одна / их три штуки получается / и по середине / одна весит сорок килограмм* // **И:** *Кошмар* // **Р:** *Я сам весил в то время сорок килограмм* // **И:** *И как вы ее?* // **Р:** *Поначалу / вот вдвоем ее поднимут / и раз ее / и побежал / с ней* // **И:** *Кошмар* // **Р:** *Не хочешь ты бегать / значит все / вперед / на другую / это самое / участок переходит* // **И:** *А на другом че / полегче?* // **Р:** *Герои там только делались с помощью других* // **16.5** (Поманисточка 2022).

Фрагмент интервью и цифровые обозначения реализующейся в нем темы/подтемы разделяются пробелом.

3.3. Принципы тематической разметки

При тематической разметке считаем необходимым следовать нескольким принципам. Эти принципы определены на основе сравнительного тестового анализа интервью пятерых респондентов: А.И. Поманисточки, В.Н. Федосова, В.А. Бородин, В.С. Артамонова, N1 (информант не дал согласие на упоминание имени) (N1 2022).

3.3.1. Принцип соответствия нарратива основной теме корпуса. В ходе беседы с респондентами периодически и неизбежно происходят отступления от запланированного направления диалога. Несмотря на то что подобные фрагменты часто небезынтересны сами по себе, их тематика не соответствует цели исследования. Для разметки мы отбираем только те нарративы, которые либо имеют то или иное отношение к характеристике профессии, производственных действий и межличностных отношений в ее рамках, либо отражают образ жизни и мироощущение представителей шахтерской профессии.

3.3.2. Принцип незавершенности тематической классификации. Работа с текстами интервью пяти разных представителей профессии показала, что Вопросник, по которому проводились беседы, может быть лишь базовым ориентиром для составления Темника. В ходе интервьюирования любого респондента могут появляться такие аспекты, которые невозможно заранее предусмотреть. В связи с этим мы не только не исключаем, но и можем прогнозировать расширение и дальнейшую детализацию тематики. Таким образом, если при работе с текстами интервью других респондентов встретятся новые темы, связанные с исследуемым аспектом, то в Темник будут вноситься соответствующие коррективы.

3.3.3. Принцип цельности нарратива. В связи с тем что в ходе устной беседы ожидаемо возникают отступления от намеченной темы разговора (отвлечения на окружающие обстоятельства, полупутные повествования и рассуждения, связанные с разнообразными ассоциациями и т. п.) в нарратив, связанный одной темой (например, развернутый ответ на какой-либо вопрос направленной беседы), иногда «вклиниваются» сторонние отступления. В подобных случаях единый нарратив не разрывается, несмотря на наличие внутри его инородных фрагментов, поскольку изъятие этих фрагментов привело бы к разрыву целостности нарратива, необходимой для развертывания темы.

Примером такой ситуации может служить следующий отрывок беседы с В.Н. Федосовым:

1.1[И: *А как вы вообще попали / ну / на шахту?* // **Вы ... Р:** *Ну на шахту вот / я приехал / тетки привезли меня // В Томари я слесарем устроился учеником* // *А потом я иду / а в гостинице объявление висит* // *А я страсть любил и хотел работать машинистом электровоза* // *А я читаю объявление / набирают в ГЭПЭШа-три / это*

в Быкове ГэПэШа-три было // щас там нету / вот //
Я прочитал / машинист электровоза // Я прихожу /
там / ну с училища был преподаватель / вот // Я гово-
рю / “На / это самое / машинист электровоза” / да //
А у меня образования-то не было // У меня же в дет-
стве / прошло как в концлагере / тяжелое было дет-
ство / поэтому // Мать умерла / батька ногой под зад
/ и иди / из дому выгнал меня // В шес[т]на[д]цать лет
вот я на Сахалин приехал // Тетки привезли // Вот я
зашел... // И: А тетки здесь жили ? // Или ... Р: А? //
Материны сестры // Нет / они уже давно умерли // Вот
// И: Они с вами поехали или просто вас отправили? //
Р: Они поехали к сестре / на кладбище сходили // Вот я
их по-... / упротил / и они меня забрали / сюда // Вот я /
это / э-э-э месяц-полтора наверно / это / слесарем от-
работал / это / в Томари // А потом вот объявление / я /
вот машинистом электровоза // А у меня образование-
то пять классов // Он г[ово]рит / “Нет / не пойдет
/ надо семь классов // Но вот ты понимаешь / г[ово]рит
/ вот есть еще одна специальность // машинист
шахтных машин и механизмов / понимаешь / г[ово]рит
/ звучи-и-и-т” // Я помню его слова до сих пор // Ну хо-
рошо / и я поехал в это // Вот в Тельновске пятнадца-
тая шахта вот / три года там я / отработал слесарем
/ вот // [1.1 (Федосов 2022).

Здесь внутри нарратива на тему «Причины выбора профессии» содержится попутная информация о некоторых семейных обстоятельствах и событиях респондента (выделено подчеркиванием).

Во фрагменте интервью В.С. Артамонова:

5.1[Р: Это вот например / проработали определен-
ный участок / выбрали / а сзади осталась пустота //
Это называется отработанное пространство // По-
том это отработанное пространство ограничива-
ют / пробивают стойки одна к одной / чтоб по это-
му / по этому / по этой линии обрезало / то что сзади
завалило / а то что здесь / осталось целое // И вот /
траченная лава готовит посадки // (звонок телефона)
Мой или ваи? // Мой? // Откуда у меня / он э-э / у меня
эмчэсовец / они телефоны редко дают // И: Да / да //
Р: Вообще редко дают // И вот он вырвался / а то деду
не звонил / давно // Мне говорит стыдно уже стало /
надо деду позвонить (смеется) // Не договорил я по по-
воду / вот органку пробиваем да / вот органка (неразб.
0.36) // [5.1 (Артамонов 2022).

Рассказ на тему «Трудовые действия и операции» (5.1) прерывается телефонным звонком и сопровождающими это событие репликами. А затем респондент завершает разговор на исходную тему.

В некоторых случаях вставки, не относящиеся к теме, могут быть весьма объемными. Например, во фрагменте разговора с тем же В.С. Артамоновым:

4.1[4.2[И: А это / машина черная / это угольная
пыль? // Или / тут уголь возят / так открыто по горо-
ду? // Р: Где черная? // Вот щас проехала? // Это щас
развозят уголь населению // Щас начинается отопи-
тельный сезон / и людям возят его // Ну кто в частном
секторе живет / тем развозят уголь // Девять тонн
дают на сезон // И: Ааа // И: Девять тонн / и хвата-
ет этого? // Р: Нет / оно не хватает // Смотря какой
дом // Потом докупают // Это просто по льготной
цене дают по девять тонн // [4.2 И: Сколько вообще
стоит тонна угля? // Р: Ну вообще / в этом году у нас

случилось непонятное вообще // Мы сами не можем...
// В шоке // Не знаю / это правда нет / но по крайней
мере начали отпускать // Вообще стоила эта тонна в
пределах трех тысяч // Даже для населения / две с по-
ловиной // Живем на угле / и такая цена / представьте
// В этом году / ни с того ни с сего почему-то / кто про-
писан в частном секторе / по сто рублей за тонну // И: Мы
в газете тоже видели // Р: Че-то произошло непон-
ятно что блядь // До сих пор все не могут врубиться /
почему именно так / сто рублей за тонну / щас платят
в частном секторе // Но уголь говно / конечно / я вам
скажу // И: Поэтому? (смеются) // Р: Нет / он про-
сто такой // Он уголь с открытых / он как для костра
// Пламя есть от него / есть небольшое тепло / но нет
калорийности // Он не дает // Он тепло не держит //
Вот подземный уголь / он спекающийся / он коксуется
/ он же как бы собирается в ком / и он долго делает
жар / держит долго // Например / с вечера в доме вы-
сыпал ведро / на утро ты встал / тебе не надо печку
растопливать // Немного пошевелил / зола просыпалась
/ туду сверху ба-бах еще ведро / и все / и все нормально
// И: Понятно // (Голос некоего мужчины, не И¹ и не Р (разговор, видимо, о появившемся человеке в форме):
Это что за сопля такая? Прокурорская какая-то // И: Ка-
кая-то форма интересная // Р: Прокуроры / проку-
ратура // Прокурорская // он вишь весь зализанный //
И: Важный (смеется) // Р: Зализанный / и прическа / и
все // (С издевкой) Я в прокуратуре работаю // (Слыш-
ишум проезжающей машины. Возвращаются к обсуж-
дению этой машины) Р: А чего она черная / поехала?
// Обыкновенная черная машина // И: А это она воз-
ит? // Р: Да / вон она видите // Здесь склад не склад /
здесь ЖэКэХа в ЖэКа отпускает / там / через весы //
И вот она основная часть / нижняя часть населения //
И: А в советские годы / если весь уголь шел в Японию..
// Р: Перебью вас // В советские годы / годы / шахтерам
всем / у кого частный сектор / или у кого дома сто-
ял титан ... // И: Титан это / это печка угольная? //
Р: Это не печка / это титан / сейчас электрический /
воду греть // Раньше титаны такие стояли / ну их на
дровах топили / а это / и давали уголь бесплатно // Во-
обще бесплатно // И: Ух ты // И: Шахтерам? // Р: Да
/ всем / всем / кто на шахте работал // Всем бесплатно
уголь шел // И: А остальные люди как? // Р: Покупали
// Но все равно скидки были / льготы были большие /
не то что щас // А щас вообще // Вроде бы частники
да олигархи // Да называются олигархами // Никто их
олигархами-то никогда и не считал / тем более я во-
обще // И: А развозили в сельской местности точно
так же / на машине? // Р: Да / на машине // Просто
адрес дают / адреса / и по адресам развозят // И: Ин-
тересно // И: Ну список... // Р: Да / по списку // [4.2 И: А в Японию отправляют на танкере? // Р: Пароходы
приходили / да // Трех- пятидесятичники (неразб. 5.38)
приходили // Грузили там на пароходы // Японцы у нас
брали хорошо // У нас раньше здесь в порту японцы /
филиппинцы / греки / американцы приходили за углем //
Тогда подземный уголь был хорошего качества // А щас
это говно в основном китайцам тудят и все // И: По-
чему (неразб. 5.52) // Р: Ну там у них / свое там кино //
Получается не получается // Японцы с удовольствием /
они одно время брали / потом бросили / не стали брать
с того разреза уголь // Не уголь / а говно // Уголь подзем-
ный это так / вот это уголь // [4.1 (Артамонов 2022).

Реализуются две основные темы, относящиеся к предмету данного исследования, — «Характеристика объекта труда (угля)» и «Применение продукта труда (угля) в экономике». При этом есть и ситуативные, и контекстуальные отступления (здесь они также выделены подчеркиванием).

3.3.4. Принцип нежесткой разметки. Известно, что особенностями устной коммуникации является ее неформальный характер и спонтанность, которая влечет за собой отсутствие строгой последовательности изложения мысли, соскальзывание с одной темы на другую с последующим возвращением к исходной теме. Важно учитывать также, что некоторые темы и подтемы могут быть смежными, а потому легко перетекают друг в друга. В связи с этим один и тот же фрагмент может реализовывать различные темы и подтемы, которые пересекаются и накладываются полностью или частично. Это можно проиллюстрировать фрагментом интервью с В.С. Артамоновым:

2.2[И: Начало пятидесятых где-то / наверное // **Р:** Ну да / пятьдесят / начало шестидесятых / пятьдесят третий наверное там / пятьдесят пятый / в это время примерно // **И:** Вы тогда такелажником пришли / вы говорили / работать / к таким крепким мужикам уже // **7.1[Р:** Я пришел / я вообще пришел / *з[ово]рю* / джаз-бригада это называлось / джаз бригада // А это бригада / кто с морпехи / кто с морфлота / кто с десанта // Там такие дяди все // И я там среди них // Я с детства не маленький был // Я и в школе первый все время стоял // Я занимался все время // **7.1 7.4[И:** А они как вас приняли? // **Р:** Молодой / как / нормально восприняли / все // Че / бери потяжелее / неси подальше // Так приняли (смеются) // Тренировали / как // Потом ниче / нормально все // **6.5[Я** попал сразу в жесткую бригаду / честно говоря / в жесткую // Сразу в жесткие рамки попал / у-у-у // С первых дней / можно сказать / в шахте начал работать / в жесткие рамки попал // Не было мне никакой раскочки / ниче // Мне надо было просто / бжжжж / и попер // **7.4** Мне даже мать / мать обратила внимание // Хотя она не знала / что я / эт самое / первое время один месяц / она не знала / что я в шахте работаю // “То раньше / *з[ово]рит* / по вечерам / *з[ово]рит* / я заметила / дома не найдешь / шарахается где-то там / где-то тыры-пыры / придет там // А тут спит ест / спит ест / никуда не ходит / типа эт самое // Постоянно спит // Ниче дома / так типа сделает / но в основном это // Че случилось?” // А потом уже когда я принес первую зарплату / шампанское конфеты / она / “А-а / вон в чем дело / теперь *з[ово]рит* ясно” // **6.5 12.2** (Артамонов 2022).

В этом фрагменте внутри темы «Вход в профессию: обучение и адаптация» (2.2) содержатся также фрагменты, реализующие другие темы: «Трудовой коллектив: члены коллектива» (7.1), «Трудовой коллектив: отношения в коллективе» (7.4) и «Оценка условий труда» (6.5).

3.3.5. Принцип учета «нулевых» тем. Для полноты и объективности картины различных сторон профессиональной культуры мы считаем целесообразным включать в тематическую разметку в том числе и те фрагменты, которые указывают на отсутствие предполагаемого (потенциально воз-

можного) факта, явления, представления. По сути это фрагменты, в которых на заданный вопрос получен отрицательный ответ. Например:

13.6[13.5[И: А вот скажите / а-а-а вы еще какие-то праздники вместе отмечали например / рождение ребенка у кого-то из коллег / или отпуск / свадьбу там? // **Р:** Нет // После отпуска денег нет отмечать // **И:** (смех) А уход в отпуск? // **Р:** Уход некогда // Заранее билет покупаешь / отрабатываешь / и тут же садишься на поезд или на автобус и едешь / куда уже купил билет // **13.5 13.6** (Федосов 2022).

3.3.6. Частные проблемы разметки. Речь идет о случаях, когда та или иная тема не развивается информантом, при том что контекст допускает различную трактовку темы нарратива. В подобных ситуациях в ходе процесса разметки необходимо, на наш взгляд, отделять целенаправленно репрезентируемую информантом тематику от фактов, которые могут иметь лишь косвенное отношение к какой-либо теме, но могут и не иметь такового. В подобных случаях выделение или не выделение темы выражает мнение не информанта, а разметчика. В связи с этим принято решение такую тему не выделять. В качестве примера приведем фрагмент из интервью с В.Н. Федосовым:

8.1[И: А вот вы сказали / что в девяносто первом была заваруха по стране / а у вас тихо было // **Р:** Тихо // **12.1[11.4[И:** А не бастовали / э-э-э / шахтеры / на счет того что зарплату не платили? // **Р:** Это шахтеры / мы уже бастовали в девяносто седьмом году / в девяносто восьмом // **И:** А как уже бастовали? // **Р:** Ну как // Пришли на работу / вот так вот как мы с вами сидели / и все // Три часа / собрались / пошли домой // **И:** Просто приходили и не работали? // **Р:** Обязательно приходили // А не придешь / сразу прогул // А тогда же у нас льготы были / терять-то неохота было // Поэтому приходили // Да и народу было очень мало уже / на шахте // **И:** И как-то / подействовало это / на администрацию? // **Р:** А че подействовало? // Долги долги // И тут все ...// Тяжелый вопрос // Вот // Потом же еще / можно сказать / ну конечно не для записи // Эти все начальники / не надо записывать / вот / себе в карман / себе в карман // И уезжали // Им до лампочки были все это / и уголь / и люди / и прочее // Они вот все что-то / приходило там / они себе карман понабили и поехали все // Им до лампочки // **12.1 И:** Ну тогда шахта уже частная стала / ее купил какой-то олигарх? // **Р:** Нет / никто ее не покупал // **И:** Она государственная была? // **Р:** Да-а // Просто Сахуголь закрылся и все // Большие потери зарплат // Никому не платили // **И:** А вам не выдавали вместо зарплат продукты? // **Р:** Вот то[ль]ко что хотел сказать // **И:** Едой платили? // **Р:** Они там купят / допустим / килограмм картошки / ну я к примеру / за десять рублей / а нам продают уже за двадцать рублей // А десять рублей себе в карман // А это двадцать рублей с зарплат вычитывали // Вот так было // **8.1 11.4** Ну я к примеру рассказал // Да / продукты там все привозили / потому что у них у нас база в Южном // Я сам ездил в Южный на базу // Я и щас езжу / там дешевле / чем в магазине // Ну которые не портятся продукты / взял / и они лежат // Чтoб не бегать в магазин лишний раз // Удобно так // Я все вре-

мя езжу / весной и осенью / поэтому // И вот они вот так нажились // А потом с этими квартирами там махинации были / разборки были // Ну / короче г[ово]ря / американцы хорошо поработали / в девяносто первом году // (Федосов 2022).

Приведенный нарратив в принципе допускает его маркировки как фрагмента, где констатируется факт деградации шахтерской отрасли, однако, поскольку речь идет о глобальном кризисе в стране в 90-е годы XX века, сам информант не оценивает описанную ситуацию как кризисную именно в шахтерской отрасли, а скорее характеризует состояние производственной и бытовой сферы в целом в определенный период времени. Исходя из этого тема «Деградация отрасли» здесь при разметке не выделяется. При этом констатация неблагополучия реализуется через другие темы: 8.1. – «Отношения с администрацией», 11.4. – «Оценка доходов в ретроспективе», 12.1. – «Забастовки».

3.3.6. Разграничение смежных тем. В реальном нарративе некоторые темы настолько близки, что возникает необходимость установления критериев для их разграничения.

1) В частности, такие темы как «Трудовые действия и операции» (5.1) и «Оценка условий труда» (6.5) являются смежными. При этом фактором для их разграничения является наличие во фрагменте оценочной лексики типа «тяжело», «трудно», «опасно» или каких-либо аналогов на лексико-фразеологическом уровне, в которых те же характеристики даются описательно. В частности, фрагмент из интервью В.С. Артамонова заключает в себе информацию о тяжелых условиях труда шахтеров:

7.4[И: А они как вас приняли? // **Р:** Молодой / как / нормально восприняли / все // Че / бери потяжелее / неси подальше // Так приняли (смеются) // Тренировали / как // Потом ниче / нормально все // **6.5[Я** попал сразу в жесткую бригаду / честно говоря / в жесткую // Сразу в жесткие рамки попал / у-у-у // С первых дней / можно сказать / в шахте начал работать / в жесткие рамки попал // Не было мне никакой раскочки / ниче // Мне надо было просто / бжжжж / и попер // **7.4** Мне даже мать / мать обратила внимание // Хотя она не знала / что я / эт самое / первое время один месяц / она не знала / что я в шахте работаю // “То раньше / г[ово]рит / по вечерам / г[ово]рит / я заметила / дома не найдешь / шарахается где-то там / где-то тыры-пыры / придет там // А тут спит ест / спит ест / никуда не ходит / типа эт самое // Постоянно спит // Ниче дома / так типа сделает / но в основном это // Че случилось?” // А потом уже когда я принес первую зарплату / шампанское конфеты / она / “А-а / вон в чем дело / теперь г[ово]рит ясно” // **6.5** (Артамонов 2022).

А в нарративе В.А. Бородина условия труда характеризуются как опасные:

6.5[И вот / бегали / скажем так / на карачках / чик-чик-чик // Но там мощность была же разная // Где-то придавит / там / скажем так / нарушение / либо что-то / то есть / ну порода / когда либо снизу вылазит большие / либо сверху с кровли да // И эта мощность пласта / она естественно менялась / допустим с девяноста бывает до семидесяти / до шестидесяти / то есть плюс секция (смеется) // Даже я / скажем так / со

своей худобой / бывало довольно проблематично было проскочить / вот // Ну и вот так там работали // **6.5** (Бородин 2022).

В интервью шахтеров встречаются отрывки, описывающие те или иные трудовые действия и операции, не оценивая их трудоемкость или опасность, например:

5.1[И: А когда вот эту рудстойку ставили / вы наблюдали / где комель / где вершина / вот это учитывалось? // **Р:** Конечно / да // Старались ставить наоборот / чтобы / ну как говорили / влагу не тянуло / вот // Чтобы влагу не тянуло // Вроде как / если влаги натянет / быстрее сгниет / вроде того / вот // **И:** Ну то есть вы ставили как / сверху вниз получается / вершинкой вниз? // **Р:** Да / ну как бы вверх ногами // Обязательно // **5.1** (Бородин 2022).

Тематически могут пересекаться повествования о ритуалах «посвящения в профессию» в рамках темы «Суеверия, приметы, ритуалы» (18.3) и контексты на тему «Вход в профессию» (2) с подтемами «Первый рабочий день» (2.1) и «Обучение и адаптация» (2.2). Различия между ними, на наш взгляд, заключаются в том, что во втором случае предполагается описание конкретной ситуации без указания на ее закреплённость в профессионально-культурной среде и регулярную повторяемость. Наличие же определенного ритуала (в том числе и в процессе входа в профессию) очевидно предполагает его воспроизводимость как некоего правила, обычая. Об этом может свидетельствовать наличие вербальных маркеров типа: «обычно было так», «была такая традиция», «каждый раз так делали» и под.

Темы «Посттрудовая жизнь» (3) и «Альтернативные варианты труда» (16.6) разводятся на основе того, что в первом случае речь идет об образе жизни бывшего шахтера после завершения его трудовой деятельности в связи с выходом на заслуженный отдых по возрасту (или, возможно, из-за приобретения инвалидности), а во втором случае предполагается целенаправленный переход из шахтерской отрасли в иную сферу деятельности в период активного трудового возраста.

Заключение

Тематическая разметка корпуса шахтерских нарративов предпринята впервые, поэтому описанная работа представляет собой первый подход к разработке стандарта метаразметки (см. [Leech 1993]) текстов по антропологии труда. Соответственно, данная публикация о принципах и категориях аннотирования может послужить будущей согласованности исследовательских процедур с совпадающими целями, что обеспечит возможность единообразного поиска по разным ресурсам [Burnard 2004].

Применяемый тип разметки по специфике предметной информации можно назвать экстралингвистическим (ср. [Захаров 2016, с. 153]), по способу выявления – аналитическим [Burnard 2004], по объекту – нарративным, по глубине клас-

сификации – многоуровневым [Трегубова 2015], по способу присвоения меток – ручным, по соотношению с определенной теорией – нейтральным [Leech 2004].

Этот тип аннотирования может быть применен и на материале текстов других профессий, относящихся к типу производственно-технических. Применение части описанных тематических категорий к материалу о профессиях социально-гуманитарной сферы будет неадекватным, поскольку категоризация мира в разных культурах неодинаковы.

Возможности использования информации по метаданным корпуса представляются как многоаспектными, так и многоцелевыми. «Людам, создающим корпуса, знакома идея, что никто в здравом уме не станет предсказывать будущее использование корпуса – будущее использование всегда более изменчиво, чем мог себе представить создатель корпуса! То же самое верно и в отношении аннотированного корпуса: сами аннотации открывают совершенно новый диапазон использования, который был бы невозможен, если бы корпус не был аннотирован» [Leech 2004].

Материалы исследования

№1 2022 – *Интервью №1*, Сахалинская обл., 2022.

Артамонов 2022 – *Интервью В.С. Артамонова*, Сахалинская обл., 2022.

Бородин 2022 – *Интервью В.А. Бородина*, Сахалинская обл., 2022.

Поманисточка 2022 – *Интервью А.И. Поманисточки*, Сахалинская обл., 2022.

Федосов 2022 – *Интервью В.Н. Федосова*, Сахалинская обл., 2022.

Библиографический список

Atkins, Clear, Ostler 1992 – *Atkins S., Clear J., Ostler N.* Corpus Design Criteria // *Literary and Linguistic Computing*. 1992. Vol. 7, issue 1. P. 1–16. <http://doi.org/10.1093/lc/7.1.1>.

Burnard 2004 – *Burnard L.* Metadata for corpus work // *Developing Linguistic Corpora: a Guide to Good Practice*. URL: <https://users.ox.ac.uk/~martinw/dlc/chapter3.htm>.

Leech 1993 – *Leech G.* Corpus Annotation Schemes // *Literary and Linguistic Computing*. Vol. 8, issue 4. P. 275–281. DOI: <http://doi.org/10.1093/LLC%2F8.4.275>.

Leech 2004 – *Leech G.* Adding Linguistic Annotation // *Developing Linguistic Corpora: a Guide to Good Practice*. URL: <https://users.ox.ac.uk/~martinw/dlc/chapter2.htm>.

Буранова 2012 – *Буранова А.И.* Тематическая организация диалектной речи: квантитативный анализ // *Известия Саратовского университета. Новая серия. Серия Филология. Журналистика*. 2012. Т. 12, вып. 3. С. 35–38. URL: <https://cyberleninka.ru/article/n/tematicheskaya-organizatsiya-dialektnoy-rechi-kvantitativnyy-analiz/viewer>; <https://www.elibrary.ru/item.asp?id=18413341>. EDN: <https://www.elibrary.ru/pmyksf>.

Захаров 2016 – *Захаров В.П.* Прологомены к корпусной лингвистике // *Вопросы психолингвистики*. 2016. № 2 (28). С. 151–161. URL: <https://cyberleninka.ru/article/n/prolegomeny-k-korpusnoy-lingvistike/viewer>.

Земичева, Иванцова 2020 – *Земичева С.С., Иванцова Е.В.* Тематическая разметка диалектного корпуса: опыт томских диалектологов // *Вестник Томского государственного университета. Филология*. 2020. № 67. С. 45–61. DOI: <http://doi.org/10.17223/19986645/67/3>.

Лавров, Харламова, Костюшина 2018 – *Лавров Д.Н., Харламова М.А., Костюшина Е.А.* Представление разметки корпуса народной речи Среднего Прииртышья // *Математические структуры и моделирование*. 2018. № 4 (48). С. 85–91. URL: <https://www.elibrary.ru/item.asp?id=37083776>. EDN: <https://www.elibrary.ru/yzfhgp>.

Летучий 2005 – *Летучий А.Б.* Корпус диалектных текстов: задачи и проблемы // *Национальный корпус русского языка: 2003–2005. Результаты и перспективы*. Москва, 2005. С. 215–232. URL: <https://www.elibrary.ru/item.asp?id=19092154>. EDN: <https://www.elibrary.ru/qbuqid>.

Народная история – *Народная история России*. Общедоступный портал устных рассказов о городах России. URL: <https://pastandnow.ru>.

Савчук 2005 – *Савчук С.О.* Метатекстовая разметка в Национальном корпусе русского языка: базовые принципы и основные функции // *Национальный корпус русского языка: 2003–2005. Результаты и перспективы*. Москва, 2005. С. 62–88. URL: <https://www.elibrary.ru/item.asp?id=19090691>. EDN: <https://www.elibrary.ru/qbturl>.

Тематическая классификация... – *Тематическая классификация и распределение фольклорно-мифологических мотивов по ареалам*. Аналитический каталог. URL: <http://www.ruthenia.ru/folklore/berezkin/>.

Трегубова 2015 – *Трегубова Е.Н.* Многоуровневая тематическая разметка как инструмент этнолингвистической репрезентации диалектного дискурса в электронном текстовом корпусе // *Вестник Томского государственного университета. Филология*. 2015. № 1 (33). С. 66–77. DOI: <http://doi.org/10.17223/19986645/33/6>.

Электронный текстовый корпус... – *Электронный текстовый корпус лингвокультуры Северного Приангарья*.
URL: <http://angara.sfu-kras.ru>.

References

- Atkins, Clear, Ostler 1992 – *Atkins S., Clear J., Ostler N.* (1992) Corpus Design Criteria. *Literary and Linguistic Computing*, vol. 7, issue 1, pp. 1–16. DOI: <https://doi.org/10.1093/lc/7.1.1>.
- Burnard 2004 – *Burnard L.* (2004) Metadata for corpus work. In: *Developing Linguistic Corpora: a Guide to Good Practice*. Available at: <https://users.ox.ac.uk/~martinw/dlc/chapter3.htm>.
- Leech 1993 – *Leech G.* (1993) Corpus Annotation Schemes. *Literary and Linguistic Computing*, vol. 8, issue 4, pp. 275–281. DOI: <https://doi.org/10.1093/LLC%2F8.4.275>.
- Leech 2004 – *Leech G.* (2004) Adding linguistic annotation. *Developing Linguistic Corpora: a Guide to Good Practice*. Available at: <https://users.ox.ac.uk/~martinw/dlc/chapter2.htm>.
- Buranova 2012 – *Buranova A.I.* (2012) Thematic Organization of Dialect Speech: Quantitative Analysis. *Izvestiya of Saratov University. Philology. Journalism*, vol. 12, issue 3, pp. 35–38. Available at: <https://cyberleninka.ru/article/n/tematicheskaya-organizatsiya-dialektnoy-rechi-kvantitativnyy-analiz/viewer>; <https://www.elibrary.ru/item.asp?id=18413341>. EDN: <https://www.elibrary.ru/pmyksf>. (In Russ.)
- Zakharov 2016 – *Zakharov V.P.* (2016) Prolegomena to corpus linguistics. *Journal of psycholinguistics*, no. 2 (28), pp. 151–161. Available at: <https://cyberleninka.ru/article/n/prolegomeny-k-korpusnoy-lingvistike/viewer>. (In Russ.)
- Zemicheva, Ivantsova 2020 – *Zemicheva S.S., Ivantsova E.V.* (2020) Dialect Corpus Thematic Markup: The Experience of Tomsk Dialectologists. *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya = Tomsk State University Journal of Philology*, no. 67, pp. 45–61. DOI: <http://doi.org/10.17223/19986645/67/3> (In Russ.)
- Lavrov, Kharlamova, Kostushina 2018 – *Lavrov D.N., Kharlamova M.A., Kostushina E.A.* (2018) Representation of the corpus of Medium Irtysh folk dialect. *Mathematical Structures and Modeling*, no. 4 (48), pp. 85–91. Available at: <https://www.elibrary.ru/item.asp?id=37083776>. EDN: <https://www.elibrary.ru/yzfhgp>. (In Russ.)
- Letuchy 2005 – *Letuchy A.B.* (2005) Corpus of dialect texts: tasks and problems. In: *National corpus of the Russian Language: 2003–2005. Results and prospects*. Moscow, pp. 215–232. Available at: <https://www.elibrary.ru/item.asp?id=19092154>. EDN: <https://www.elibrary.ru/qbuqid>. (In Russ.)
- People's history – *People's history of Russia*. A public portal of oral stories about Russian cities. Available at: <https://pastandnow.ru/>. (In Russ.)
- Savchuk 2005 – *Savchuk S.O.* (2005) Metatext markup in the National Corpus of the Russian Language: basic principles and main functions. In: *National corpus of the Russian Language: 2003–2005. Results and prospects*. Moscow, pp. 62–88. Available at: <https://www.elibrary.ru/item.asp?id=19090691>. EDN: <https://www.elibrary.ru/qbturl>. (In Russ.)
- Thematic classification... – *Thematic classification and distribution of folklore and mythological motifs by area*. Analytical catalogue. Available at: <http://www.ruthenia.ru/folklore/berezkin/>. (In Russ.)
- Tregubova 2015 – *Tregubova E.N.* (2015) Multilevel Thematic Marking as an Ethnolinguistic Tool of Dialectal Discourse Representation in Digital Text Corpora. *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya = Tomsk State University Journal of Philology*, no. 1 (33), pp. 66–77. DOI: <http://doi.org/10.17223/19986645/33/6>. (In Russ.)
- Electronic text corpus – *Electronic text corpus of the linguistic culture of the Northern Angara region*. Available at: <http://angara.sfu-kras.ru>. (In Russ.)