

Онтология хантыйского корпуса текстов для электронного портала
«Фольклор народов Сибири»¹

Аннотация. В статье освещается опыт составления онтологии для электронного корпуса текстов, разработанного для хантыйского обрядового фольклора. Статья написана в рамках нового научного направления – корпусной фольклористики.

This paper highlights the experience of making the ontology for web corpus of texts developed for Khanty ritual folklore. The paper is written within the new scientific trend – corpus folkloristics.

Ключевые слова: онтология, обрядовый фольклор, функциональный метод.

Ontology, ritual folklore, functional method.

УДК: 398.3(=511.142)(082)

Контактная информация: 630090, г. Новосибирск, ул. Николаева, д. 8. Институт филологии СО РАН, сектор фольклора народов Сибири. Тел./факс: (383) 330-14-52. E-mail: annazor@mail.ru

Корпусная фольклористика – новое направление, которое начинает формироваться на базе 60-томной серии «Памятники фольклора народов Сибири и Дальнего Востока». Наряду с традиционным научным изданием фольклора Серия выходит в новое электронное пространство, представляя фольклорное произведение не просто как линейный текст, а как интерактивное единство печатного слова, аудио- и нотной записи этого же фрагмента, видеозаписи, фотоиллюстрации и комментария. Многомерная запись фольклорного произведения приближает издание к его живому бытованию. Заимствуя методы корпусной лингвистики, текст в электронном портале «Фольклор народов Сибири» подается в виде корпуса текстов, где устойчивым элементам фольклорного нарратива – формулам, типическим местам, устойчивым сюжетным темам – присваивается определенное значение. Размеченный таким образом текст позволяет осуществлять навигацию, поиск по выделенным элементам. Фольклористика не занимается индивидуальными феноменами. Фольклорный корпус текстов, снабженный метаданными, откроет новые возможности для статистического и автоматизированного анализа. Систематизация и структуризация данных по хантыйскому обрядовому фольклору выполнена на основе онтологии, вся информация на портале представлена в виде сети взаимосвязанных информационных объектов. Доступ к этим знаниям и данным осуществляется путем навигации по дереву понятий онтологии.

Разработка онтологии основана на всемирном опыте составления указателей фольклорных текстов. К настоящему времени мировая фольклористика накопила богатый опыт составления указателей. Работы, в которых обрабатывается сказочный материал, опираются в первую очередь на ставший общепризнанным указатель фольклорной сказки, созданный для систематизации сказочных типов [Thompson 1961]. Такие указатели стремятся, с одной стороны, найти общие основы для сравнения мирового сказочного фольклора, с другой же – учитывают национальные особенности фольклорного материала. Важным критерием для составления указателей в настоящее время является принцип максимального отражения внутренних категорий фольклорного нарратива. Универсальные категории фольклористики, хотя и дают базу для сравнения, зачастую чужды конкретному национальному фольклору. Выявить и описать эти культурные феномены важная задача фольклористики: «*Folklore texts and their performance have a different type of subjectivity – their own subjectivity. They constitute a reality that is culturally regulated. Like language, their repeated patterns, their regularities and rules, are cultural, not scholarly, constructs. We have the opportunity, duty, privilege to find the meaning of theses cultural creations, categories, and orders*» «Фольклорные тексты и их представление имеют иной тип субъективности – их собственную субъективность. Они порождают культурно регулируемую реальность. Подобно языку, их повторяющиеся структуры, их закономерности и правила являются культурными, а не научными конструктами. У нас есть возможность, обязанность, привилегия найти значения этих культурных творений, категорий и упорядоченностей» [Ben Amos 1992: 21] (здесь и далее в статье перевод цитат мой. – А.Г.). Даже общепризнанный источник сравнения сказочных типов, указатель Аарне–Томпсона разрабатывался, в первую очередь, для систематизации сказочного материала Северо-Запада Европы и только позднее распространился на весь индоевропейский материал. «*It would be a mistake to think that it could be extended to tales of such areas as central Africa, the North American Indi-*

¹ Статья подготовлена при поддержке РГНФ, проект № 14-04-12010/14 «Электронный ресурс “Фольклор народов Сибири”: текстовое, семантическое и мультимедийное представление».

ans, or Oceania. Each of those would need an index strictly upon its own traditions 'Было бы ошибкой думать, что он [указатель] может быть применен к сказкам таких областей, как центральная Африка, северо-американские индейцы или Океания. Каждая из них требует указатель, основанный строго на ее собственных традициях'» [Thompson 1961: 8]. Поэтому применять систему Аарне–Томпсона для фольклора других народов не всегда корректно. Действительно, составители указателей, систематизирующие сказочные сюжеты и мотивы народов, населяющих другие, отдаленные территории, сталкиваются с определенными проблемами, если они избирают в качестве основы для своей классификации систему Аарне–Томпсона. Как, например, столкнулась Патриция Ватерман, автор указателя австралийских сказочных типов: «*The tenor of the Australian narratives with their wandering creator heroes, animals which are neither domesticated or wild but rather transformations of people-who-behave-like-animals or animals-who-behave-like-people, their cycles of ancestral extraordinary beings, is quite foreign to European lore. While the major divisions of the Australian materials, the concept of type was found to serve, with modification, as a useful model* 'Темы австралийского нарратива с его странствующим героем-создателем, с животными, которые не являются ни домашними, ни дикими, а, скорее, представляют собой трансформацию людей-которые-ведут-себя-как-животные или животных-которые-ведут-себя-как-люди, с их родовыми циклами о сверхъестественных существах, в значительной степени чужд европейскому знанию. В то же время в большинстве делений австралийского материала понятие "тип" с оговорками было найдено подходящим как полезная модель'» [Waterman 1987: 11]. С подобной проблемой столкнулся автор классификации африканских сказок, который, хотя и соотнес африканские сказочные типы с системой Аарне–Томпсона, тем не менее разработал собственную классификацию, где «...*the scheme reflects native categories of folk narratives. '...схема отражает родные категории фольклорного нарратива'*» [Agewa 1961: 8]. Структурно-морфологические критерии в систематизации оказались неприменимыми к африканскому материалу, поскольку «... *if my typology were based on a rigid structural frame, thousands of tales could not come into my frame. The road to typology through structural and morphological means, an ideal in itself, is not an easily attainable as one might think. '...если бы моя типология была основана на жестких структурных основаниях, тысячи сказок не вошли бы в мою структуру. Путь к типологии через структурные и морфологические средства, идеальный сам по себе, не так легко достижим, как кажется'*» [Там же: 6].

Для «несказочных» указателей тем более требуется особенный подход. Авторы таких указателей стремятся максимально отразить композиционные и сюжетно-структурные особенности анализируемых произведений. Ярким примером является работа «A List of Motifs in the Lives of the Early Irish Saints», предпринятая Dorothy Ann Bray [Bray 1992]. В этом указателе задокументированы жития 40 ирландских святых. Список мотивов призван выделить и классифицировать большинство житийных мотивов, а структура указателя отражает композиционные особенности ирландских житий. «*The classifications are meant to reflect the main preoccupations of the hagiographers with regard to their saints and the principal types of miracles and motifs by which the saints manifested their holiness. 'Предполагалось, что классификация будет отражать главные заботы агиографов относительно их святых и основные типы чудес и мотивов, через которые святые проявляли свою святость'*» [Там же: 21]. За композиционными сюжетными элементами в построении указателя следует и Е.Н. Кузьмина, в чьем указателе систематизирован массив типических мест эпоса монголоязычных и тюркоязычных народов Сибири: «В сибирском эпосоведении наступил момент более пристального внимания к вопросам структуры, мотивики сказаний, их функционально-семантической природы. Одной из важных задач является решение проблем сюжетостроения, сюжетного контекста и этнопоэтической константности сибирского эпоса» [Кузьмина 2005: 4].

Материалом для нашей работы стали песни, записанные сотрудниками Сектора фольклора народов Сибири Института филологии СО РАН во время комплексной фольклорной экспедиции в Ханты-Мансийский автономный округ – Югра, д. Казым в 2002 г. Записанные песни были расшифрованы Т. Молдановым и готовятся к изданию в рамках 60-томной академической Серии «Памятники фольклора народов Сибири и Дальнего Востока». Это вторая попытка систематизации типических мест обрядовых песен медвежьего праздника. В качестве минимальной значимой повторяющейся единицы обрядового стиха мы приняли формулу. Формула – поэтическая строка, характеризующаяся типическим грамматико-синтаксическим строением и устойчивым лексическим наполнением, обладающая ритмическим единством, призванная передавать определенный образ. Например, *Ссам мөвләп ай ям пушах* 'Грудь сосущий маленький славный сыночек' – формула, построенная по типу конструкции характеристики, имеющая грамматико-синтаксическую структуру [N+Adv]+[Adj+Adj]+N. По сравнению с первой классификацией [Гриневич 2012] при выделении классов новой онтологии мы опирались на уже названный принцип использования только внутренних категорий фольклорного нарратива

ва. Это необходимо для того, чтобы не навязывать тексту чуждые ему категории. В первой редакции указателя перед нами стояла практическая задача, а именно – выборка и систематизация «пучка вариантов» каждого формульного выражения (по В.М. Гацаку), из которых состоят обрядовые медвежьи песни. Индекс поэтических строк был создан для того, чтобы осуществить удобную навигацию по рукописи тома «Медвежьи песни казымских хантов», который на 100 % состоял из типических мест и повторяющихся устойчивых формульных выражений и был объемом в 21 печатный лист (подробнее о целях и принципах систематизации Указателя см.: [Гриневиц 2014]). Поскольку новый указатель существует в электронном виде и привязан непосредственно к текстам, необходимости распределять пучки вариантов в отдельные классы нет. В создании онтологии, таким образом, первоочередной задачей было свести к минимуму приписываемые исследователем концепты и значения, чем грешила предыдущая классификация. Все категории должны были быть заимствованы из текста. Поэтому, например, в новой редакции нет таких универсальных категорий, как Время или Пространство.

В основу новой классификации лег функциональный принцип, который отражает внутренние текстуральные связи выделяемых классов. Типические места разделены на категории согласно той роли, которую они играют в тексте. На первом уровне классификации онтология разделена на четыре класса: мир медведя «Медведь», мир человека «Человек», мир божеств «Божества», а также «Обряд». Эти классы противопоставлены друг другу по ряду характеристик, часть из которых является общей для всех трех онтологических сфер, например, «Тело», «Жилище», «Еда». Другие же характеристики описывают только одну категорию, например, класс формул «Психофизиологические состояния» характеризует только медведя.

Таблица 1

Список классов онтологии хантыйского корпуса текстов

ВНУТРИТЕКСТОВЫЕ КЛАССЫ				
МЕДВЕДЬ	ЧЕЛОВЕК	БОЖЕСТВА	ОБРЯД	НЕЗАВИСИМЫЕ КЛАССЫ
Тело	Тело	Тело	Действующие лица	Композиционные формулы
Еда	Одежда	Жилище	Место проведения	
Передвижения медведя	Место проживания		Время проведения	
Место проживания	Дом человека			
Деструктивные действия	Деятельность человека			
Психофизиологические состояния	Обряд			

Вопреки логике функционального принципа классификации, в качестве четвертой онтологической сферы был выделен «Обряд». Этот класс должен был бы входить в онтологическую сферу «Человек», характеризуя его деятельность. Тем не менее класс «Обряд» претендует на статус четвертой онтологической сферы, поскольку он занимает в тексте особое положение. Особенный статус ему придает то, что именно благодаря проведению обряда стало возможным соприкосновение трех перечисленных онтологических сфер. Это точка пересечения разных миров. Кроме того, класс «Обряд» характеризуется тем же набором признаков, что и другие онтологические сферы: действующие лица, еда, деятельность, что сопоставляет этот класс с классами более высокого онтологического уровня.

Объем и количество экземпляров разных классов показательны. Так, для класса «Человек» наиболее объемным подклассом является «Жилище» и «Деятельность». Класс «Медведь» описан более подробно с точки зрения телесности: тело медведя описано очень подробно. Только у медведя есть психофизиологическая характеристика «Психофизиологические состояния». Элементы пространства (лес, тундра, реки) также относятся к онтологической сфере медведя, характеризуя его передвижения. Эти подклассы сближают мир медведя с миром богов, поскольку в своих лесных путешествиях медведь посещает священные места, что и является главной целью обрядовых песен жанра кайёен

(*кайөөн*): бродя по священным местам, медведь приглашает божеств на готовящийся обряд медвежьего праздника.

Пример1 : Подкласс I.1. «Тело».

I. МЕДВЕДЬ

I.1. Тело:

I.1.a. Голова, лицо

I.1.b. Горло, язык, голос

I.1.v. Глаза

I.1.g. Нос

I.1.d. Уши

I.1.e. Внешняя оболочка, облик

I.1.j. Душа *ис*

I.1.z. Сокровенное внутреннее пространство (*пухэл*)

I.1.i. Живот, желудок, кишки

I.1.k. Сердце

I.1.l. Табуированные названия частей тела медведя {город, звезды, пеньки, лодка, набирушка, весла, крючки, шуба}²

Как уже говорилось, в онтологии нет такого класса, как «Пространство». Логически выделенная категория, она складывается из названий и характеристик конкретных местностей и рек. Лес, тундра, реки и т.п. являются экземплярами класса «Медведь» и характеризуют его передвижения, описывая то, что он видит. Самостоятельная категория времени также отсутствует. Время в данной онтологии также является одной из характеристик передвижений медведя.

Пример 2: Подкласс I.4. «Передвижения медведя».

I. МЕДВЕДЬ

I.4. Передвижения медведя, то, что он видит, слышит, встречает:

I.4.1. Местность {лес, реки, земли и т.п. }

I.4.2. Время {время суток, времена года }

I.4.3. Звуки

I.4.4. Существа

Определенные сложности вызвало отнесение категории времени к конкретной онтологической сфере. Поскольку хантыйским текстам свойственна конкретность и наглядность в описаниях, выделять отдельный класс «Время» было бы неверным с точки зрения текстов. Временные характеристики появляются при описании существ – это их возраст. Таким образом, класс «Возраст» стал одним из атрибутов, противопоставляющих медведя, человека и божеств друг другу. Кроме того, время – это одна из характеристик передвижения медведя, который кружит по лесотундре, бывая на всех священных и значимых местах хантов, при этом день сменяется днем, времена года следуют друг за другом. Смена времен года – очень важная тема для медвежьего праздника, это, кроме прочего, структурная характеристика, влияющая на последовательность номеров на медвеьем празднике. Согласно же верованиям хантов, время сотворено верховным божеством Торумом, который управляет или является временем и погодой. Для демонстрации этих представлений в системе используется инструмент «связь». Связь устанавливается между подклассами «Времена года» и «Время суток» класса «Медведь» и подклассом «Торум» класса «Божества», название типа связи «управляет».

Благодаря инструменту «связь» в электронной онтологии хантыйского корпуса текстов возможно установление связей между разными классами внутри классификации. Во-первых, при функциональном подходе классификации разбиваются «пучки вариантов»: варианты одной и той же формулы могут характеризовать разных существ или разные ситуации. Чтобы показать, в каких контекстах встречаются варианты одной формулы, мы ввели связь, которую назвали «Пучок вариантов».

Инструмент «связь» позволяет воссоздавать микросюжеты в типологической фабуле фольклорного произведения. Медвежий праздник – это энциклопедия хантыйской жизни. В нем есть все – и бы-

² Фигурные скобки обозначают домены – области неразрывных значений.

товой, повседневный мир человека, и мифологический уровень, рассказывающий о сложной системе родства между богами, а значит, и людьми.

Онтология фольклорного текста включает в себя как внутритекстовые, так и экстратекстуальные данные. Каждому тексту обязательно дается жанровая характеристика, которая опирается на традиционные определения. «*Clearer and more careful analysis of terms in the languages of the world could yield us a better understanding of human thought, verbal art, and its performance.* ‘Ясный и более точный анализ терминов в языках мира может привести нас к лучшему пониманию человеческой мысли, вербального искусства и их представления’» [Amos 1992: 18]. Медвежьи песни, являясь частью обрядового фольклора, согласно традиции подразделяются на *кайөөн ар* (песня, посвященная медведю), *миш ар* (песня божества миш, приносящего удачу), *пойэкты ар* (песня-молитва). Каждый из жанров имеет строго заданное место в обряде, свои сюжетные характеристики в соответствии с традиционным представлением о жанровой природе исполняемого текста. Эта классификация и использована в электронной системе.

Пример 3: Жанровая характеристика фольклорного документа.

I. Хантыйский корпус текстов

I.1. Обрядовый фольклор

I.1.a. Медвежьи песни

I.1.a.1. Песни кайөөн (*кайөөн ар*)

I.1.a.2. Песни богов миш (*миш ар*)

I.1.a.3. Песни-молитвы (*пойэкты ар*)

I.1.a.4. Юмористические сценки (*ньӕхӕнты ар*)

Кроме жанровой характеристики экстратекстуальные данные содержат информацию об исполнителе фольклорного произведения, месте и дате записи, экспедиции и т.п., что является подробным паспортом текста. Эти данные сопровождают любой ресурс, который публикуется на портале: фотографию, видеозапись, аудиозапись, нотную расшифровку.

Произведение каждого жанра в медвежьем празднике состоит из набора устойчивых сюжетных тем. Для разметки константных единиц на уровне фабулы произведения разработана особая часть онтологии. Например, *кайөөн ар*, или хвалебные песни медведю, исполняются в начале каждого дня праздника. Основу этого жанра составляют песни доставки медведя (*вой тотты ар* или *вой тотты пант* – песни доставки зверя или песни пути). Это песни, имитирующие доставку медведя в стойбище. В них зверя символически «знакомят» с сакральными территориями. Благодаря этому путешествию исполнители вспоминают божеств – хозяев территорий. Таким образом божества приглашаются к участию в игрищах. Принцип построения таких песен базируется на географии. Другими разновидностями *кайөөн ар* являются песни родов и поселений, песни-мифы, песни-наставления, песни наказания медведя охотником и охотника медведем, песни поднятия души на небо. Жанровая принадлежность определяется по содержанию песни. Некоторые из них сложно отнести к конкретной разновидности. Их просто называют *вой ар* ‘медвежья песня’, как, например, «*Тал улам вой* ‘Зиму проспавший зверь’». В структуре песен жанра *кайөөн* выделяются следующие типические сюжеты:

Пример 4: Структура песен жанра *кайөөн*.

I. Жизнь медведя в лесу, описание природного мира глазами животного;

- a. Описание берлоги, зверь строит (роет) свой дом;
- b. Медведь ходит по описываемым землям (в этой части песни подробно описывается природа, леса, по которым бродит медведь);
- c. Медведь охотится, ищет пищу;
- d. Медведь спасается от комаров, подыскивая прохладное место;
- e. Медведь набредает на священные места, видит божеств хантыйского пантеона;

II. Появление человека;

- a. Эпитет человека;
- b. Зверь наблюдает за человеком;

III. Добыча медведя, начало ритуала медвежьего праздника;

- a. Ритуальное разделывание туши (расстегивание застёжек), угощение божеств, молитва;

- b. Транспортировка медведя в стойбище;
- c. Приглашение гостей;
- d. Жертвоприношение;
- e. Веселье, исполнение песен, танцев, игра на музыкальных инструментах;
- f. Приход божеств;

IV. Исполнение назначения обряда медвежьего праздника;

- a. Поднятие одной души медведя на Небо;
- b. Превращение другой души в духа-покровителя дома.

Кроме классификации устойчивых поэтических формул и типических сюжетов онтология на портале «Фольклор народов Сибири» содержит гипертекстовые классы, которые характеризуют текст с точки зрения его внешних связей: информация о месте записи, исполнителе, экспедиции. Благодаря тому, что «Исполнитель» произведения, «Собиратель», специалист, расшифровавший его, и другие являются отдельными классами, в системе можно будет установить связи, которые покажут, кто и в ходе каких экспедиций записал то или иное произведение и т.п. Каждый класс имеет свой набор атрибутов, характеризующих его.

Таблица 2

Гипертекстовые классы, паспорт текста

ГИПЕРТЕКСТОВЫЕ КЛАССЫ (ПАСПОРТ ТЕКСТА)			
ПРОИЗВЕДЕНИЕ	ИСПОЛНИТЕЛЬ	ЭКСПЕДИЦИЯ	ИЗДАТЕЛИ
Название	ФИО	Экспедиция (Дата)	Расшифровщик (ФИО)
Язык	Пол	Экспедиция (Место)	Переводчик (ФИО)
Диалект	Возраст	Собиратель (ФИО)	Музыковед (ФИО)
Жанр	Национальность		
Опубликовано	Род		
Варианты	Место проживания		
Ареал распространения	Наличие детей		
	Образование		
	Репертуар		
	Награды		

Электронный портал «Фольклор народов Сибири» позволит получить представление о поэтике, о внутренних взаимосвязях, о сюжетной характеристике публикуемого фольклорного произведения, получить данные об исполнителе и месте исполнения, узнать, кто и когда провел запись текста и расшифровал его. Благодаря тому, что перечисленные элементы являются частью общей онтологии, по размеченным элементам можно осуществлять специализированный поиск. Это позволит получить доступ к разного рода статистической информации и узнать, например, какие произведения исполнены тем или иным носителем традиции, какие произведения были исполнены в выбранный пользователем портала год, какие еще произведения относятся к интересующему пользователя жанру и т.п., что открывает простор для исследователя.

Гриневиц А.А. Опыт составления «Указателя поэтических формул песен медвежьего праздника казымских хантов» // Языки и фольклор коренных народов Сибири. 2014. № 1 (вып. 26). С. 75–79.

Гриневиц А.А. Указатель поэтических формул песен медвежьего праздника казымских хантов: Приложение к диссертации на соискание ученой степени канд. филол. наук. Новосибирск, 2012. 256 с.

Кузьмина Е.Н. Указатель типических мест героического эпоса народов Сибири (алтайцев, бурят, тувинцев, хакасов, шорцев, якутов): Экспериментальное изд. Новосибирск: Изд-во СО РАН, 2005. 1383 с.

Amos B. Do We Need Ideal Types (in Folklore)? An Address to Lauri Honko. Turku: Nordic Institute of Folklore, 1992. 35 p.

Bray D.A. A List of Motifs in the Lives of the Early Irish Saints // FF Communications № 252 (Vol. CIX₂) / Edited for the Folklore Fellows by Alan Dundes, Bengt Holbek, Lauri Honko, Matti Kuusi, Anna Birgitta Rooth. Helsinki: Suomalainen Tiedeakatemia. Academia Scientiarum Fennica, 1992. 589 p.

Arewa E.O. A Classification of the Folktales of the Northern East African Cattle Area by Types: Dissertation for the degree of Doctor of Philosophy in Anthropology. Berkeley: University of California, 1961. 298 p.

Waterman P. A Tale-Type Index of Australian Aboriginal Oral Narratives // FF Communications № 238, Helsinki: Suomalainen Tiedeakatemia. Academia Scientiarum Fennica, 1987. 173 p.

Thompson S. The Types of the Folktale. A Classification and Bibliography // FF Communications № 184 / Antti Aarne's Verzeichnis der Maerchentypen (FF Communications № 3) Translated and Enlarged by Stith Thompson. Helsinki: Suomalainen Tiedeakatemia. Academia Scientiarum Fennica, 1961. 588 p.