

# НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ  
ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 10

Москва 2015

## ИНФОРМАЦИОННЫЙ АНАЛИЗ

УДК 004.94 : 001.102

В.В. Грибова, А.С. Клещев, Ф.М. Москаленко, В.А. Тимченко

### Двухуровневая модель сложноструктурированных информационных единиц, соответствующая метафоре анкетирования\*

*Описывается двухуровневая модель информационных единиц, которая может быть положена в основу компьютерных средств редактирования инвариантной по отношению к технологическим пространствам (предметным областям) сложноструктурированной информации различных уровней абстракции в привычной для ее носителей системе понятий – без специального обучения последних или участия в этом процессе специалистов-посредников.*

**Ключевые слова:** *двухуровневая модель представления информации, орграфовая модель представления информации, иерархические орграфы, метаинформация*

#### ВВЕДЕНИЕ

Одним из классов задач, решаемых в информационном обществе, являются задачи «заказа информации»: **заказчик** информации хочет получить у **носителей** информации такую **информацию**, которая

соответствует его потребности, как по содержанию, так и по форме (структуре). Примеры таких задач – перепись населения, социологические опросы, заполнение историй болезни, чтение лекций, создание учебников, другие многочисленные формы документов, которые заполняются в банках, социальных, государственных и коммерческих учреждениях (оформление кредита, пособия, пенсии, покупка би-

\* Работа выполнена при частичной поддержке РФФИ (проект 13-07-00024, проект 14-07-00299 и проект 15-07-03193).

лета и т.п.). Для решения таких задач заказчик может сформировать **запрос**, в соответствии с которым носитель должен передать заказчику нужную ему информацию – **ответ** на запрос. Примерами запросов являются опросные листы и анкеты, образовательные стандарты и программы и т.п.

Если заказчик и носитель – это люди, которые могут взаимодействовать непосредственно, то запрос и ответ на него могут формироваться в процессе диалога между ними. Если же заказчик и носитель не могут непосредственно взаимодействовать, то для того, чтобы ответ на запрос содержал релевантную информацию, необходима некоторая «технология» решения такого рода задач. Примером такой «технологии» может служить анкетирование, когда запрос представляется некоторой анкетой (метаинформацией), а ответ – результатом заполнения этой анкеты (информацией, соответствующей метаинформации). Простейшая анкета может быть совокупностью вопросов, на которые необходимо получить ответы. Но в более сложных случаях анкета содержит структуру информации, которая: облегчает понимание запроса за счет того, что вопросы формулируются в контексте этой структуры; неявно руководит действиями носителя по заполнению анкеты (поскольку структура анкеты как запроса одновременно является и структурой ответа на запрос).

Примерами задач «заказа информации», когда заказчиком является компьютерная программа, а носителем – человек, может служить формирование баз данных и баз знаний. При этом компьютерная программа должна иметь такой интерфейс, чтобы, пользуясь им, носитель мог сам, без посторонней помощи передать необходимую информацию. Если передаваемая информация имеет большой объем и сложную структуру, то в качестве подходящей метафоры для разработки таких интерфейсов предпочтительно использовать одну из таких «технологий» решения задачи заказа информации, которая знакома носителю этой информации (поскольку последний, как правило, не станет тратить свое «драгоценное» время на получение специальной подготовки, чтобы научиться передавать компьютерной системе свои знания).

В основе многих систем извлечения экспертных знаний [1–7] лежит «технология» решения задачи заказа информации, базирующаяся на парадигме объектно-ориентированных онтологий [8–11], которая незнакома большинству реальных и потенциальных носителей этих знаний, что вызывает в их работе значительные трудности, а иногда делает ее практически невозможной.

В настоящей работе рассматривается двухуровневая модель сложноструктурированных информационных единиц<sup>1</sup>, соответствующая метафоре анкетирования. В соответствии с данной моделью любая информация (результат заполнения анкеты) связана с метаинформацией (анкетой), по которой она формируется или модифицируется.

<sup>1</sup> Под сложноструктурированными информационными единицами в работе понимаются онтологии, базы знаний, базы данных и вообще любая сложным образом организованная информация.

## 1. ДВУХУРОВНЕВАЯ МОДЕЛЬ АНКЕТЫ

В модели анкеты естественно различать два уровня: абстрактный (метаинформация), соответствующий запросу, и конкретный (информация), соответствующий ответу на запрос. Рассмотрим двухуровневую модель анкеты, в которой:

- абстрактный уровень (метаинформация) представлен конечным множеством понятий, а конкретный уровень (информация) – конечным множеством экземпляров этих понятий;
- каждое понятие либо имеет конечное множество атрибутов, либо представляет собой конечное множество других понятий<sup>2</sup>, либо представляет собой множество констант некоторого сорта (т.е. объем этого понятия есть множество констант этого сорта);
- каждый атрибут понятия имеет область возможных значений – либо множество констант некоторого сорта (конечное или бесконечное), либо конечное множество понятий;
- атрибуты понятия могут быть обязательными и необязательными;
- каждый экземпляр некоторого понятия имеет те же атрибуты (обязательные и необязательные), что и данное понятие, а каждый обязательный атрибут этого экземпляра понятия и некоторые его необязательные атрибуты имеют значение из области значений этого атрибута данного понятия – либо константу некоторого сорта, либо экземпляр некоторого понятия;
- если понятие есть множество других понятий, то множество экземпляров этого понятия есть объединение множеств экземпляров этих других понятий;
- если понятие есть множество констант некоторого сорта, то множество экземпляров этого понятия есть некоторое подмножество констант этого сорта.

Легко видеть, что такая модель анкеты является частным (упрощенным) случаем модели, лежащей в основе редактора WebProtégé [7].

Пример 1. Приведем модель анкеты, в которой могут быть заданы такие информационные единицы, как базы наблюдений в различных разделах медицины [12]. В этом описании:

- термины берутся в кавычки;
- термины, обозначающие понятия, начинаются с заглавной буквы, а обозначающие атрибуты – со строчной;
- определение понятия заключено в угловые скобки;
- фигурные скобки обозначают множества;
- вещественный или целый интервал представляется его нижней и верхней границами, разделенными дефисом;
- необязательные атрибуты заключены в квадратные скобки;
- названия сортов (множеств констант) пишутся жирным шрифтом;

<sup>2</sup> Объем этого понятия совпадает с объединением объемов множества других понятий.

- для экземпляров понятий в круглые скобки заключаются названия этих понятий; константы сортов пишутся курсивом;

- если понятие есть множество констант некоторого сорта, то его определение представлено этим сортом.

**Метаинформация** [13] описывается следующим образом.

- «Вещественные значения» <вещественные интервалы>;

- «Целые значения» <целые интервалы>;

- «Тип значения» <{«Целые значения», «Вещественные значения»}>;

- «Качественные значения» <{строки}>;

- «Составные значения» <{«характеристика» «Тип области возможных значений»}>;

- «Числовые значения» <«характеристика значения» «Тип значения», [«единица измерения» строки]>;

- «Тип области возможных значений» <{«Качественные значения», «Числовые значения», «Составные значения»}>;

- «Наблюдение» <«область возможных значений» «Тип области возможных значений»>;

- «Группа» <[«группы» {«Группа»}], [«наблюдения» {«Наблюдение»}]>;

- «База наблюдений» <[«группы» {«Группа»}], [«наблюдения» {«Наблюдение»}]>.

**Информация** [14] описывается следующим образом.

- «Описывается значениями слабая, умеренная, резкая» (экземпляр «Качественные значения») <{«слабая», «умеренная», «резкая»}>;

- «Описывается значениями постоянно, периодически» (экземпляр «Качественные значения») <{«постоянно», «периодически»}>;

- «Описывается значениями острое, постепенное» (экземпляр «Качественные значения») <{«острое», «постепенное»}>;

- «Описывается значениями справа, слева» (экземпляр «Качественные значения») <{«справа», «слева»}>;

- «Описывается значениями отсутствует, имеется» (экземпляр «Качественные значения») <{«отсутствует», «имеется»}>;

- «Диапазон 34.0-42.0» (экземпляр «Вещественные значения») <34,0-42,0>

- «Значения температуры» (экземпляр «Числовые значения») <«характеристика значения» «Диапазон 34.0-42.0», «единица измерения» «°C»>;

- «Описывается характеристиками присутствие, характер начала, степень проявления, периодичность, длительность» (экземпляр «Составные значения») <«присутствие» «Описывается значениями отсутствует, имеется», «характер начала» «Описывается значениями острое, постепенное», «степень проявления» «Описывается значениями слабая, умеренная, резкая», «периодичность» «Описывается значениями постоянно, периодически», «длительность» «Описывается значениями кратковременная, продолжительная»>;

- «Описывается характеристиками присутствие, сторона, характер начала, периодичность» (экземпляр «Составные значения») <«присутствие» «Описывается значениями отсутствует, имеется», «сторона» «Описывается значениями справа, слева», «характер начала» «Описывается значениями острое, постепенное», «периодичность» «Описывается значениями постоянно, периодически»>;

- «Температура» (экземпляр «Наблюдение») <«область возможных значений» «Значения температуры»>;

- «Головная боль» (экземпляр «Наблюдение») <«область возможных значений» «Описывается значениями отсутствует, имеется»>;

- «Светобоязнь» (экземпляр «Наблюдение») <«область возможных значений» «Описывается характеристиками присутствие, характер начала, степень проявления, периодичность, длительность»>;

- «Ощущение инородного тела в глазу» (экземпляр «Наблюдение») <«область возможных значений» «Описывается характеристиками присутствие, сторона, характер начала, периодичность»>;

- «Дополнительные жалобы» (экземпляр «Группа») <«наблюдения» {«Головная боль», «Повышение температуры», ...}>;

- «Главные жалобы» (экземпляр «Группа») <«наблюдения» {«Ощущение инородного тела в глазу», «Светобоязнь», ...}>;

- «База наблюдений в области офтальмологии» (экземпляр «База наблюдений») <«группы» {«Главные жалобы», «Дополнительные жалобы», ...}>.

Из этого примера видно, что, следуя парадигме, лежащей в основе объектно-ориентированного подхода, реализованного в редакторе WebProtégé и ряде других программных средств [1–6], носитель информации должен каждому экземпляру каждого понятия сопоставить имя этого экземпляра. Часть этих имен является терминами предметной области: «База наблюдений в области офтальмологии», «Главные жалобы», «Дополнительные жалобы», «Ощущение инородного тела в глазу», «Светобоязнь», «Головная боль», «Повышение температуры». Другие же имена экземпляров понятий в предметной области не используются: «Описывается характеристиками присутствие, сторона, характер начала, периодичность», «Описывается характеристиками присутствие, характер начала, степень проявления, периодичность, длительность», «Описывается значениями отсутствует, имеется», «Описывается значениями справа, слева», «Описывается значениями острое, постепенное», «Описывается значениями постоянно, периодически», «Описывается значениями слабая, умеренная, резкая». Они возникают как следствие принятого порядка «снизу-вверх» определения понятий, при котором сначала вводятся понятия, которые определяются лишь через сорта и константы, затем понятия, которые определяются через уже определенные понятия и т.д. Вспомогательные имена необходимы для связи через атрибуты понятий одних экземпляров понятий с другими, ранее определенными. Из-за этого носитель информации, во-первых, испытывает значительный дискомфорт при необходимости при-

думывать эти искусственные имена, во-вторых, ему приходится изначально определять весь необходимый и достаточный набор экземпляров понятий, требуемый для задания целостной информационной единицы, в-третьих, процесс описания информационной единицы от понятий нижнего уровня к понятиям верхнего уровня является неестественным для человека, в-четвертых, при определении косвенно рекурсивных понятий порядок определения понятий «снизу-вверх» не может быть реализован (в примере 1 такой порядок вынужденно нарушается при определении понятий «Составные значения» и «Тип области возможных значений»). Для преодоления этой трудности в объектно-ориентированной парадигме помимо понятий и их атрибутов вводятся бинарные отношения между понятиями; атрибут понятия также рассматривается как бинарное отношение между этим понятием и областью возможных значений этого атрибута.

## 2. ОРГРАФОВАЯ СВЯЗНАЯ ДВУХУРОВНЕВАЯ МОДЕЛЬ АНКЕТЫ

Каждый атрибут понятия можно считать бинарным отношением между множеством экземпляров этого понятия и множеством возможных значений этого атрибута – либо множеством констант некоторого сорта, либо множеством экземпляров конечного множества понятий. Имя атрибута является именем этого отношения. Практика формализации систем понятий разных предметных областей показывает, что если множество возможных значений атрибута однозначно идентифицирует этот атрибут, то в ряде случаев имя этого атрибута является излишним. Таким образом, некоторые атрибуты понятия удобно обозначать именами, другие же оставлять безымянными. Не умаляя общности, введем орграфовую связную двухуровневую модель анкеты, в которой и абстрактный (метаинформация), и конкретный (информация) уровни представлены орграфами с размеченными вершинами. В орграфе, представляющем абстрактный уровень модели (метаинформации):

- каждому понятию и каждому именованному атрибуту соответствует вершина, меткой которой является имя понятия или атрибута;
- каждому значению (константе некоторого сорта) соответствует терминальная вершина, меткой которой является это значение.

Кроме того, введем следующие ограничения:

- в орграфе, представляющем абстрактный уровень, (метаинформации) выделяется одна вершина, которую будем называть начальной;
- в метаинформации определена такая ориентация дуг, что для любой вершины этого графа существует путь в нее из начальной вершины с учетом ориентации дуг.

В орграфе, представляющем конкретный уровень модели (информации):

- каждому экземпляру каждого понятия и каждому именованному атрибуту соответствует вершина;

- каждому значению соответствует терминальная вершина;
- из каждой нетерминальной вершины, соответствующей экземпляру понятия, выходит набор дуг, соответствующих атрибутам этого понятия;
- из каждой нетерминальной вершины, соответствующей именованному атрибуту, выходит дуга, входящая в вершину, представляющую значение этого атрибута у этого экземпляра понятия.

В орграфе, представляющем конкретный уровень модели (информации), также существует начальная вершина – она является вершиной, соответствующей экземпляру того понятия, которому соответствует начальная вершина в орграфе, представляющем абстрактный уровень модели. Аналогичным (орграфу метаинформации) образом вводится и ориентация дуг в орграфе информации. Далее будут рассматриваться только орграфовые связные двухуровневые модели анкеты.

Орграфовая связная двухуровневая модель анкеты является не только более экономной в некоторых отношениях, но и представляет как анкету (метаинформацию), так и заполненную анкету (информацию) «целиком», каждую в виде единой информационной единицы, представленной орграфом, а не в виде неестественным образом частично упорядоченного множества вопросов и ответов на них. Кроме того, в орграфе информации «искусственные» имена экземпляров понятий отсутствуют, поскольку обозначаемые ими экземпляры понятий идентифицируются своим входением в граф и своим содержанием.

В орграфовой связной двухуровневой модели анкеты каждый вопрос, представленный меткой вершины орграфа метаинформации, оказывается в контексте, который представляет собой последовательность вопросов, лежащих на пути из начальной вершины в вершину, соответствующую данному вопросу. Каждый ответ на вопрос, представленный меткой вершины орграфа информации, также оказывается в соответствующем контексте.

Пример 2. На рис. 1 приведен орграф, представляющий метаинформацию, приведенную в примере 1.

Начальная вершина, имеющая метку «База наблюдений», закрашена серым цветом. Названия сортов представлены на рис. 1 курсивным начертанием. Спецификаторы дуг (см. далее – раздел 5) имеют следующие символьные обозначения: «копия» – «=», «возможное отсутствие» – «[=]», «в точности один» – «!», «ноль или один» – «[!]», «непустое упорядоченное множество» – «+». Ограничители дуг (см. далее – раздел 5) имеют следующие символьные обозначения: «порождение» – «gen», «порождение или ссылка» – «all». Если из вершины выходит лишь одна дуга, то тип такого набора исходящих дуг (см. далее – раздел 5) – «список» – на рисунке не указывается.

Орграф, представляющий информацию, приведенную в примере 1, приведен на рис. 2.

Начальная вершина, имеющая метку «База наблюдений в области офтальмологии», закрашена серым цветом, а в квадратных скобках у вершин указаны метки соответствующих вершин из метаинформации.

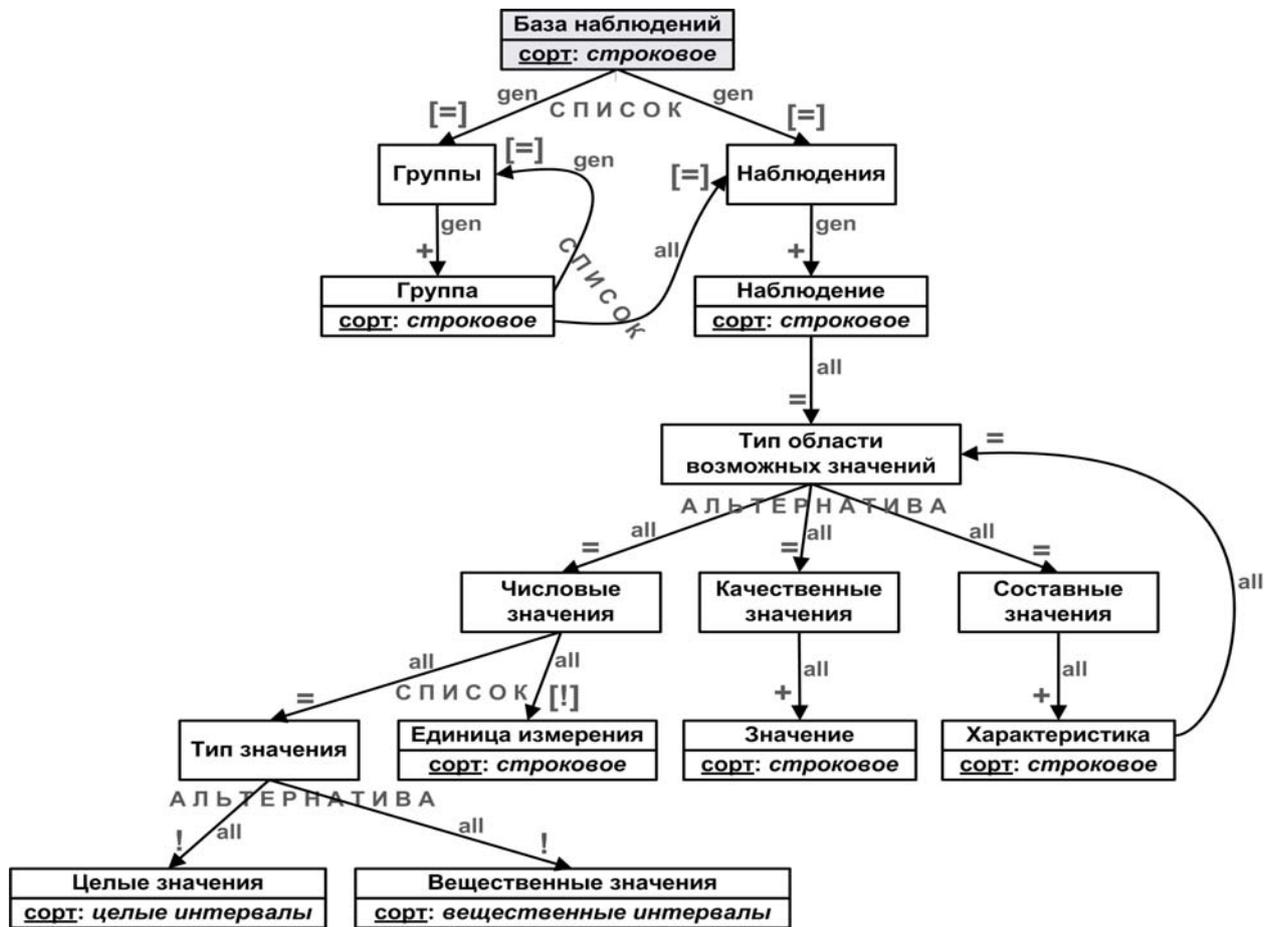


Рис.1. Орграф метаинформации для баз наблюдений в различных разделах медицины

### 3. ВЕРШИНЫ ОРГРАФОВ МЕТАИНФОРМАЦИИ И ИНФОРМАЦИИ

К вершинам орграфов метаинформации и информации может быть присоединена дополнительная информация различных видов. Будем говорить, что вершина имеет несколько «полей», каждое из которых имеет название.

В орграфе метаинформации каждая вершина имеет поле «комментарий», значением которого является строка (возможно пустая). Назначение информации в этом поле – пояснение вопроса анкеты.

Все вершины орграфа метаинформации можно разбить на два непересекающихся подмножества:

- вершины, имеющие поля «имя» со значением строковая константа (название понятия или атрибута понятия) и «сорт» со значением название сорта (область возможных значений атрибута понятия или множество возможных имен собственных для экземпляров понятия);
- вершины, имеющие поле «значение», значением которого может быть константа некоторого сорта (название понятия или атрибута понятия, в роли которого выступает не обязательно строка).

**Пример 3.** На рис. 1 к первому подмножеству относятся вершины «База наблюдений», «Группа», «Наблюдение», «Единица измерения», «Значение», «Характеристика», «Целые значения», «Вещественные значения», которые представлены прямоуголь-

никами с горизонтальной разделительной линией. В верхней половине таких прямоугольников отображено значение поля «имя», а в нижней (курсивным начертанием) – значение поля «сорт» соответствующей вершины орграфа. Название сорта, множество констант которого представляют собой строки конечной длины, есть «строковое» (см. далее – в соответствии с разделом 4). Ко второму подмножеству относятся вершины «Группы», «Наблюдения», «Тип области возможных значений», «Числовые значения», «Качественные значения», «Составные значения», «Тип значения», которые представлены прямоугольниками без разделительной линии. В таких прямоугольниках отображено значение поля «значение» соответствующей вершины орграфа.

Начальная вершина орграфа метаинформации имеет поля «имя», значением которого является название системы понятий, представляемой орграфом метаинформации, и «сорт» со значением «строковое».

**Пример 4.** На рис. 1 значение поля «имя» у начальной вершины есть «База наблюдений».

В орграфе информации каждая вершина имеет поле «комментарий», значением которого является строка (возможно пустая). Назначение этого поля – пояснение ответа на вопрос анкеты. Кроме того, каждая вершина имеет поле «значение», значением которого может быть константа некоторого сорта (имя собственное экземпляра понятия или название атрибута понятия, экземпляр которого представляет вершина).

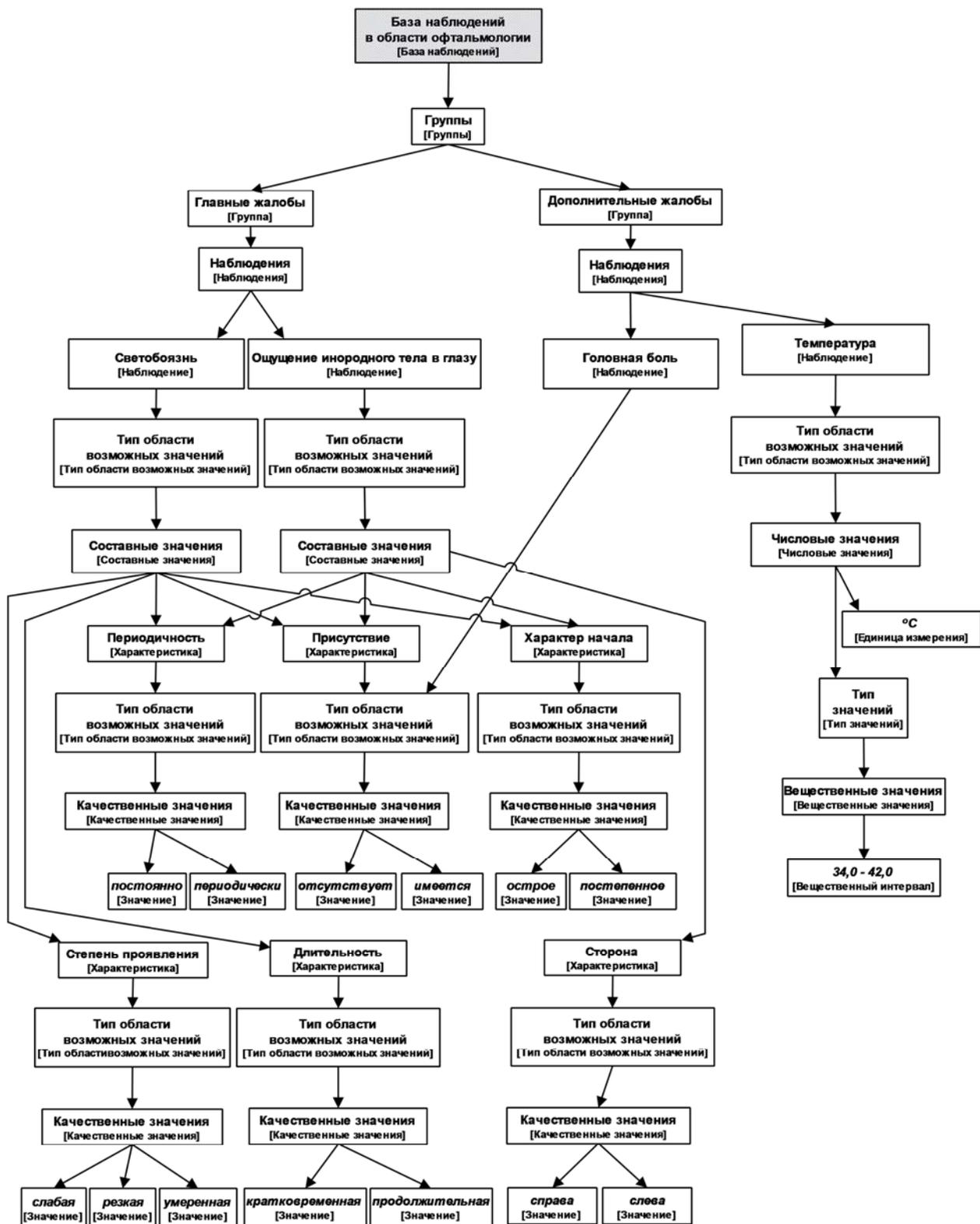


Рис. 2. Орграф информации для базы наблюдений в области офтальмологии

**Пример 5.** В примере 2 все вершины орграфа информации имеют поле «значение». Их значения приведены на рис. 2.

Начальная вершина орграфа информации имеет поле «значение» со значением сорта «строковое», являющимся названием системы экземпляров понятий, представляемой орграфом информации.

**Пример 6.** В информации из примера 2 у начальной вершины поле «значение» имеет значение «База наблюдений в области офтальмологии».

#### 4. СОРТА И КОНСТАНТЫ ОРГРАФОВ МЕТАИНФОРМАЦИИ И ИНФОРМАЦИИ

Каждый сорт орграфов метаинформации и информации имеет название и множество констант этого сорта. Сорта могут быть базовыми и библиотечными. Базовые сорта имеют названия:

- «строковое», константами которого являются строки конечной длины;
- «целое», константами которого являются целые числа;
- «вещественное», константами которого являются вещественные числа;
- «логическое», константами которого являются «истина» и «ложь»;
- «бинарные данные», константами которого являются массивы байтов конечной длины.

Каждый библиотечный сорт имеет название, отличное от названий базовых сортов и других библиотечных сортов. Для каждого библиотечного сорта определяется свое множество констант этого сорта. Множество библиотечных сортов может быть произвольным. Множество констант сорта служит для представления множества возможных значений атрибута понятия или имен собственных экземпляров понятий.

**Пример 7.** На рис. 1 у вершин «База наблюдений», «Группа», «Наблюдение», «Единица измерения», «Значение», «Характеристика» значением поля «сорт» является название одного из базовых сортов («строковое»), у вершины «Целые значения» – название одного из библиотечных сортов – «целые интервалы», а у вершины «Вещественные значения» – «вещественные интервалы».

#### 5. ДУГИ ОРГРАФОВ МЕТАИНФОРМАЦИИ И ИНФОРМАЦИИ

Каждая дуга в орграфе на рис. 1 (метаинформации или информации) соединяет две вершины. Вершину, из которой выходит дуга, назовем *началом*, а вершину, в которую она входит, – ее *концом*.

В орграфе метаинформации для каждой нетерминальной вершины (из которой выходит хотя бы одна дуга) определен тип набора исходящих дуг: «список» или «альтернатива». Если тип набора дуг, выходящих из вершины, есть «список», то концы этих дуг, которые являются нетерминальными вершинами орграфа, задают упорядоченное множество атрибутов понятия или значений безымянных атрибутов, а концы, которые являются терминальными вершинами, – значения атрибутов понятия. Тип набора дуг «альтернатива» определяет возможные значения атрибута понятия.

**Пример 8.** На рис. 1 для вершин «База наблюдений», «Группы», «Наблюдения», «Группа», «Наблюдение», «Числовые значения», «Качественные значе-

ния», «Составные значения», «Характеристика» типом наборов исходящих дуг является «список», а для вершин «Тип области возможных значений» и «Тип значения» – «альтернатива».

Вершины, являющиеся соответственно началом и концом дуги, могут принадлежать как одному, так и разным орграфам (метаинформации или информации). Вершины, принадлежащие орграфу, отличному от текущего редактируемого орграфа, но связанные дугами с последним, будем называть *вершинами, не принадлежащими редактируемому орграфу*.

Каждая дуга в орграфе метаинформации имеет ограничитель и спецификатор, которые определяют возможные способы ответов на вопросы анкеты. Значениями ограничителя могут быть «порождение», «ссылка», «порождение или ссылка». Если значение ограничителя есть:

- «порождение», то ответом на вопрос анкеты является новый экземпляр понятия (в орграфе информации дуга входит в новую вершину);
- «ссылка», то ответом на вопрос анкеты является подходящий экземпляр понятия, который уже создан при ответе на другой из вопросов анкеты (в орграфе информации дуга входит в уже существующую вершину);
- «порождение или ссылка», то возможен любой из этих вариантов.

Значениями спецификатора могут быть «копия», «возможное отсутствие», «пропуск», «в точности один», «непустой список», «непустое упорядоченное множество», «непустая последовательность», «ноль или один», «возможно пустой список», «возможно пустое упорядоченное множество» и «возможно пустая последовательность».

Хотя структура информации (ответа на анкету) соответствует структуре метаинформации (анкеты), между ними имеются и некоторые различия: могут отсутствовать ответы на необязательные вопросы анкеты; ответ может выбираться из предложенного списка; ответом на некоторые вопросы может быть множество экземпляров понятия или значений сорта. Анкета (метаинформация) и ответ на нее (информация) являются разными информационными единицами, поэтому ответ содержит структуру, которая формируется по структуре анкеты и ответам на ее вопросы. Если в метаинформации (анкете) спецификатор дуги  $D$ , выходящей из вершины с типом набора дуг «список», имеет значение «копия», то в информации (ответе на анкету) создается дуга  $d$ , поле «значение» конца которой является копией значения поля «значение» конца дуги  $D$ ; если у дуги  $D$  значение спецификатора есть «в точности один», то в информации создается дуга  $d$ , а значение поля «значение» ее конца задается в ответе; если у дуги  $D$  значение спецификатора есть «непустой список», то на основе ответа в информации создается множество дуг  $\{d_1, \dots, d_n\}$ , у каждой из которых значение поля «значение» ее конца задается в ответе, причем эти значения у концов разных дуг могут совпадать; если у дуги  $D$  значение спецификатора есть «непустое упорядоченное множество», то на основе ответа в информации создается множество дуг  $\{d_1, \dots, d_n\}$ , у каждой из которых значение поля «значение» ее конца задается в ответе, причем эти значения у концов разных дуг должны быть попарно различны; если у дуги  $D$  зна-

чение спецификатора есть «непустая последовательность», то на основе ответа в информации создается множество дуг  $\{d_1, \dots, d_n\}$ , у каждой из которых значение поля «значение» ее конца формируется как порядковый номер этой дуги; если у дуги  $D$  значение спецификатора есть «возможное отсутствие», «ноль или один», «возможно пустой список», «возможно пустое упорядоченное множество» или «возможно пустая последовательность», то новые дуги в информации могут и не создаваться; если у дуги  $D$  значение спецификатора есть «пропуск», то для таких дуг в информации новые дуги не создаются.

Возможные значения спецификатора для дуги, выходящей из вершины с типом набора дуг «список», определяются следующим образом:

- если дуга входит в нетерминальную вершину, имеющую поле «значение», то значением спецификатора могут быть «копия», «возможное отсутствие» и «пропуск»<sup>3</sup>;
- если дуга входит в терминальную вершину, имеющую поле «значение», то значением спецификатора могут быть «копия» и «возможное отсутствие»;
- если дуга входит в вершину, имеющую поле «сорт», то значением спецификатора могут быть «в

точности один», «непустой список», «непустое упорядоченное множество», «непустая последовательность»<sup>4</sup>, «ноль или один», «возможно пустой список», «возможно пустое упорядоченное множество» и «возможно пустая последовательность»<sup>4</sup>.

Если в метаинформации из вершины выходит набор дуг с типом «альтернатива», то при построении ответа (информации) выбирается одна из дуг этого набора и для нее ответ формируется в соответствии со значением ее спецификатора, как описано выше.

Возможные значения спецификатора для дуги, выходящей из вершины с типом набора дуг «альтернатива», определяются следующим образом:

- если дуга входит в нетерминальную вершину, имеющую поле «значение», то значением спецификатора могут быть «копия» и «пропуск»;
- если дуга входит в терминальную вершину, имеющую поле «значение», то значением спецификатора может быть только «копия»;
- если дуга входит в вершину, имеющую поле «сорт», то значением спецификатора могут быть «в точности один», «непустое упорядоченное множество», «непустая последовательность».

Пример 9. Значения спецификатора и ограничителя дуг на рис. 1 приведены в Таблице.

**Значения спецификатора и ограничителя дуг на рис. 1**

№	Дуга	Значение спецификатора	Значение ограничителя
1	База наблюдений $\rightarrow^{[=]_{gen}}$ Группы	возможное отсутствие	порождение
2	База наблюдений $\rightarrow^{[=]_{gen}}$ Наблюдения	возможное отсутствие	порождение
3	Группы $\rightarrow^{+}_{gen}$ Группа	непустое упорядоченное множество	порождение
4	Группа $\rightarrow^{[=]_{gen}}$ Группы	возможное отсутствие	порождение
5	Группа $\rightarrow^{[=]_{all}}$ Наблюдения	возможное отсутствие	порождение или ссылка
6	Наблюдения $\rightarrow^{+}_{gen}$ Наблюдение	непустое упорядоченное множество	порождение
7	Наблюдение $\rightarrow^{=}_{all}$ Тип области возможных значений	копия	порождение или ссылка
8	Тип области возможных значений $\rightarrow^{=}_{all}$ Составные значения	копия	порождение или ссылка
9	Тип области возможных значений $\rightarrow^{=}_{all}$ Качественные значения	копия	порождение или ссылка
10	Тип области возможных значений $\rightarrow^{=}_{all}$ Числовые значения	копия	порождение или ссылка
11	Составные значения $\rightarrow^{+}_{all}$ Характеристика	непустое упорядоченное множество	порождение или ссылка
12	Характеристика $\rightarrow^{=}_{all}$ Тип области возможных значений	копия	порождение или ссылка
13	Качественные значения $\rightarrow^{+}_{all}$ Значение	непустое упорядоченное множество	порождение или ссылка
14	Числовые значения $\rightarrow^{=}_{all}$ Тип значений	копия	порождение или ссылка
15	Числовые значения $\rightarrow^{[!]}_{all}$ Единица измерения	ноль или один	порождение или ссылка
16	Тип значения $\rightarrow^{!}_{all}$ Целые значения	в точности один	порождение или ссылка
17	Тип значения $\rightarrow^{!}_{all}$ Вещественные значения	в точности один	порождение или ссылка

<sup>3</sup> Значение ограничителя в этом случае не задается.

<sup>4</sup> Значением поля «сорт» у вершины, куда входит дуга, у которой значение спецификатора есть «непустая последовательность» или «возможно пустая последовательность», может быть только «целое», а значение ограничителя – только «порождение».

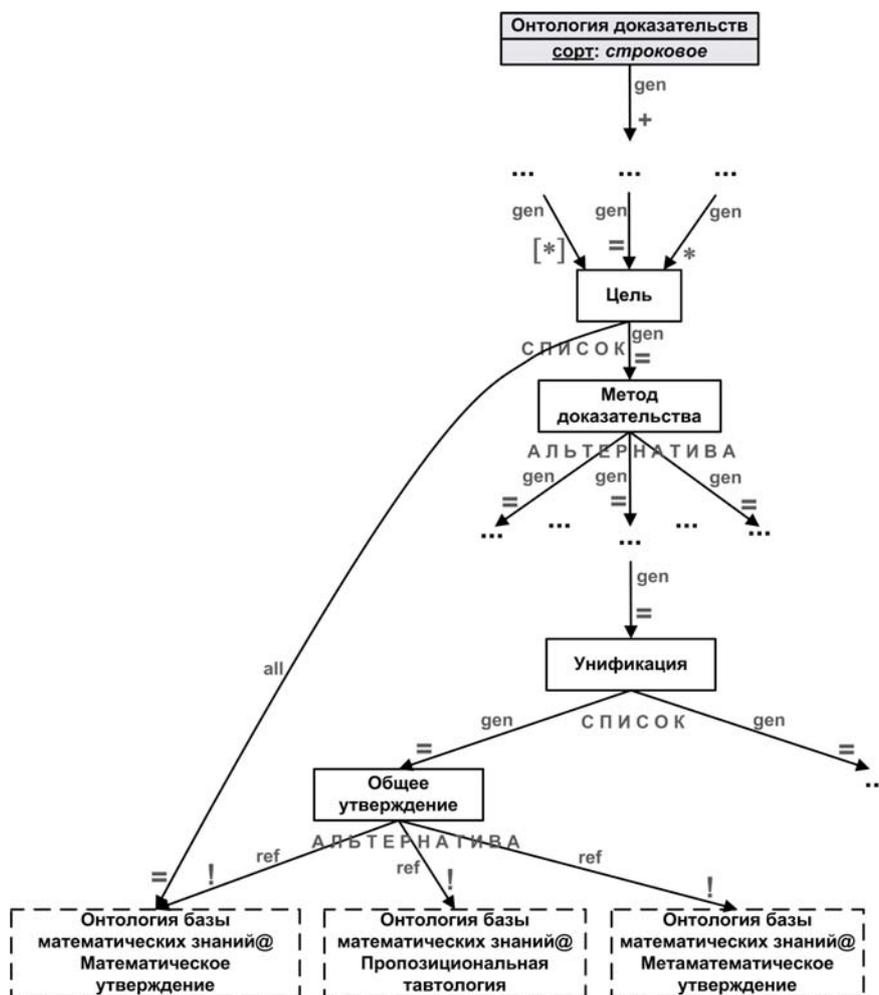


Рис.3. Фрагмент орграфа метаинформации, описывающего онтологию доказательств математических теорем

**Пример 10.** Одним из методов доказательства текущей цели (математического утверждения) является ее унификация с утверждением из базы математических знаний. Суть данного метода состоит в том, чтобы показать, что цель есть частный случай некоторого (более общего) математического утверждения, пропозициональной тавтологии или метаматематического утверждения [15], принадлежащего базе математических знаний. На рис. 3 представлен фрагмент орграфа онтологии доказательств теорем.

На рис. 3 вершины, представленные пунктирными прямоугольниками, принадлежат орграфу, представляющему метаинформацию для базы математических знаний. Символ «@» в метках этих вершин разделяет значение поля «имя» начальной вершины соответствующего орграфа и значение поля «имя» вершины этого же орграфа (не являющейся начальной), экземпляр которой является корневой вершиной повторно используемого фрагмента орграфа информации. Поскольку общее утверждение для унификации содержится в базе математических знаний, то концом дуги, началом которой является вершина-экземпляр вершины «Общее утверждение», должна быть корневая

вершина выбранного пользователем утверждения из базы математических знаний.

Спецификатор дуги «непустой список» имеет символическое обозначение «\*», а спецификатор дуги «возможно пустой список» – «[\*]». Значение ограничителя у дуг «Общее утверждение  $\rightarrow_{ref}^!$  Математическое утверждение», «Общее утверждение  $\rightarrow_{ref}^!$  Пропозициональная тавтология», «Общее утверждение  $\rightarrow_{ref}^!$  Метаматематическое утверждение» на рис. 3 есть «ссылка» (соответствующие дуги на рисунке имеют символическую метку «ref»).

В орграфе информации дуги не имеют ограничителей и спецификаторов.

## ЗАКЛЮЧЕНИЕ

В настоящей статье рассмотрены двухуровневая модель информационных единиц, основанная на широко распространенной в настоящее время парадигме объектно-ориентированного подхода к представлению информации, и орграфовая связанная двухуровневая модель информационных единиц, соответствующая метафоре анкетирования. Показано, как первая практически без умаления общности может быть

сведена ко второй, в которой, в свою очередь, разделены модель орграфов для представления метаинформации и модель орграфов для представления информации.

В соответствии с двухуровневой моделью различаются два типа сложноструктурированных информационных единиц – по уровню абстракции представляемой ими информации: информационные единицы, представляющие метаинформацию (абстрактный уровень), и информационные единицы, представляющие информацию (конкретный уровень). Назначение орграфа метаинформации (который можно рассматривать как порождающую орграфовую грамматику – грамматику, порождающую орграфы информации) состоит в описании языка (онтологии), в терминах которого формируется множество орграфов информации.

## СПИСОК ЛИТЕРАТУРЫ

1. Neologism: Easy Vocabulary Publishing. – URL: <http://ceur-ws.org/Vol-368/paper10.pdf> (дата обращения: 10.07.2015).
2. Dragoni M., Ghidini C., Bosca A. Multilingual MoKi: How to Manage Multilingual Ontologies in a Wiki / ed. Philipp Cimiano, Miriam Fernández, Vanessa Lopez, Stefan Schlobach, Johanna Völker // ESWC (Satellite Events). – Springer. – 2013. – P. 162-166.
3. Guoxing Fu, Weinian Rao. An introduction of Knoodl. – URL: <http://tinman.cs.gsu.edu/~raj/8711/sp11/presentations/knoodlReport.pdf> (дата обращения 10.07.2015).
4. OntoWiki: a tool providing support for agile, distributed knowledge engineering scenarios. – URL: <http://aksw.org/Projects/OntoWiki.html> (дата обращения 10.07.2015).
5. SOBOLEO – Editor and Repository for Living Ontologies. – URL: <http://knowledge-maturing.com/services-and-tools/tools/> (дата обращения 10.07.2015).
6. PoolParty – thesaurus management system and a SKOS editor for the Semantic Web. – URL: <http://www.poolparty.biz/> (дата обращения 10.07.2015).
7. The Protégé Ontology Editor and Knowledge Acquisition System. – URL: <http://protege.stanford.edu/> (дата обращения 07.10.2015).
8. Kleppe A., Warmer S., Bast W. MDA Explained. The Model Driven Architecture: Practice and Promise. – Addison-Wesley, 2003. – 192 p.
9. Varro D., Pataricza A. VPM: A visual, precise and multilevel metamodeling framework for describing mathematical domains and UML // Journal of Software and Systems Modeling. – 2003. – Vol. 2(3). – P. 187-210.
10. Muller P.-A., Fleurey F., Jezequel J.-M. Weaving Executability into Object-Oriented Meta-languages / ed. L. S.K. Briand // MoDELS 2005. LNCS. – Springer. – 2005. – Vol. 3713. – P. 264 – 278.
11. MDA web site. – URL: <http://www.omg.org/mda> (дата обращения: 18.02.2015).
12. Черняховская М.Ю., Никифорова Н.Ю. Информационная база для интеллектуальной, дистанционной поддержки обследования больных // Искусственный интеллект. – 2006. – № 2. – С. 510-519.
13. Клещев А.С., Москаленко Ф.М., Черняховская М.Ю. Модель онтологии предметной области «Медицинская диагностика». Часть 1. Неформальное описание и определение базовых терминов // Научно-техническая информация. Сер. 2. – 2005. – №12. – С.1-7.
14. Черняховская М.Ю., Москаленко Ф.М. База наблюдений в области офтальмологии для банка медицинских знаний // Информатика и системы управления. – 2009. – № 2(20). – С. 40-49.
15. Клещев А.С., Тимченко В.А. Алгоритм унификации для расширяемой модели математического диалекта // Информатика и системы управления. – 2012. – № 1(31). – С. 155-165.

*Материал поступил в редакцию 23.07.15.*

## Сведения об авторах

**ГРИБОВА Валерия Викторовна** – доктор технических наук, заместитель директора по научной работе, старший научный сотрудник Института автоматизации и процессов управления Дальневосточного отделения Российской академии наук (ИАПУ ДВО РАН), г. Владивосток  
e-mail: [gribova@iacp.dvo.ru](mailto:gribova@iacp.dvo.ru)

**КЛЕЩЕВ Александр Сергеевич** – доктор физико-математических наук, профессор, главный научный сотрудник ИАПУ ДВО РАН, г. Владивосток  
e-mail: [kleshev@iacp.dvo.ru](mailto:kleshev@iacp.dvo.ru)

**МОСКАЛЕНКО Филипп Михайлович** – кандидат технических наук, научный сотрудник ИАПУ ДВО РАН, г. Владивосток  
e-mail: [philipmm@iacp.dvo.ru](mailto:philipmm@iacp.dvo.ru)

**ТИМЧЕНКО Вадим Андреевич** – кандидат технических наук, научный сотрудник ИАПУ ДВО РАН, г. Владивосток  
e-mail: [vadim@dvo.ru](mailto:vadim@dvo.ru)

# АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

УДК 81'322'367.3

Е.В. Падучева

## Коммуникативная структура предложения\*

*Исследуется коммуникативная структура предложения. Планом выражения для коммуникативной структуры является линейно-акцентная структура. Принят своего рода трансформационный подход к коммуникативной структуре: описываются линейно-акцентные преобразования предложения и тот вклад, который они вносят в его семантику и прагматику.*

**Ключевые слова:** коммуникативная структура, линейно-акцентная структура, линейно-акцентные преобразования

**Коммуникативная структура** (иначе – **актуальное членение**) выделяет в предложении компоненты, существенные с прагматической точки зрения (а именно, с точки зрения важности, новизны информации, порядка ее восприятия адресатом), и устанавливает отношения между ними. Важнейшими из этих компонентов являются **тема** (= о чем идет речь) и **рема** (= что сообщается); ср. английские термины – *topic* и *focus*. Например, во фразе *По железной дороге мчался курьерский поезд* речь идет о том, что происходило на железной дороге, а во фразе *Курьерский поезд мчался по железной дороге* – о том, что происходило с курьерским поездом. Коммуникативная структура неутвердительного предложения (в частности, вопросительного) требует отдельного рассмотрения.

### 1. ПЛАН СОДЕРЖАНИЯ И ПЛАН ВЫРАЖЕНИЯ КОММУНИКАТИВНОЙ СТРУКТУРЫ. ЛИНЕЙНО- АКЦЕНТНАЯ СТРУКТУРА

Коммуникативная структура предложения имеет план выражения и план содержания. Планом выражения для коммуникативной структуры (**К-структуры**) является в русском языке **линейно-интонационная** структура предложения [1, с. 108], она же – **линейно-акцентная** структура [2]. В некоторых языках есть специальные грамматикализованные лексические показатели К-структуры, см., например, о К-показателях японского языка в [3].

Линейно-акцентная структура (**ЛА-структура**) задается порядком слов и фразовой просодией, *взятыми в их взаимодействии* [4, с. 31–35; 5, с. 10 ff]. В [6, с. 8] было введено, в том же значении, понятие

**линейно-динамическая структура**: «Определенный порядок слов и определенное место ударения, вместе взятые, образуют так называемую линейно-динамическую структуру предложения».

Планом содержания для К-структуры является членение предложения на коммуникативно релевантные (иначе – **коммуникативно охарактеризованные**) компоненты и отношения между ними. До недавних пор единственными коммуникативно охарактеризованными компонентами предложения считались **тема** и **рема**. А единственным семантическим ресурсом при описании семантического вклада коммуникативной структуры в семантику высказывания были понятия **логический субъект** и **логический предикат** (см., например, [7]), ранее – **психологический субъект** и **предикат** (см. [8]). (Об истории вопроса см. подробно [9, с. 441–468].) Иными словами, в классическом репертуаре было только одно отношение между компонентами коммуникативного членения – субъектно-предикатное. Моделируя семантический вклад противопоставления темы и ремы, исследователи пытались опираться на синтаксис, исходя из того, что тема – это подлежащее, иначе – субъект, а рема – сказуемое, т.е. предикат. Современная теория К-структуры располагает более развитым концептуальным аппаратом.

Связи между синтаксической и коммуникативной структурой достаточно свободные. Например, логический субъект (и тема) может состоять из группы имени и группы глагола, а роль логического предиката (ремы) при этом играет адвербиал. Так, в предложении (1) при определенной интонации усматривается тема *скорая помощь приехала*, т.е. событие «приезд скорой помощи», и рема *удивительно быстро*:

(1) Скорая помощь приехала удивительно быстро.

\* Данная работа выполнена при финансовой поддержке РФФИ, грант № 14-04-00604а

ЛА-структура – это не что иное, как последовательность тоновых (т.е. акцентных) групп; акценты распознаются говорящими и объективно идентифицируются с помощью современной аппаратуры, см. [10–12]. К-структура – это конструкт. Она задает коммуникативно релевантные компоненты предложения и отношения между ними, принимая во внимание не только ЛА-структуру предложения, но и его лексико-синтаксические характеристики, включая референциальную семантику именных групп, лексическую семантику предикатов и коммуникативную семантику частиц, которые способны дублировать или даже замещать фразовую просодию. Так, частица *это*, как в предложениях *Это вы МЕНЯ должны спросить*, *Это мне БРАТ посоветовал лечь в больницу*, служит маркером экспрессивной препозитивной ремы. Частица *ли* в *да-нет*-вопросе является единственным средством выражения иллюкутивной силы вопроса. При этом просодическая структура, свойственная общему вопросу без *ли*, – подъем частоты основного тона на вопросительном компоненте вопроса – не используется. Предложение *Нашел ли Ваня работу?* просодически ближе к *Ваня нашел работу*, чем к *Ваня нашел работу?*

ЛА-структура сугубо линейна. Между тем К-структура может быть иерархической: и тема, и рема могут, в свою очередь, члениться на тему и рему, см. [9, с. 445].

В генеративной грамматике порядок слов рассматривается в отрыве от просодии. Это обстоятельство подвергается справедливой критике в [13, с. 293]; так, в предложениях *Пирог испекла МАША* и *МАША испекла пирог* разный порядок слов, но К-структуры практически равнозначны.

## 2. ЛИНЕЙНО-АКЦЕНТНОЕ ПРЕОБРАЗОВАНИЕ И КОММУНИКАТИВНАЯ ПАРАДИГМА

Коммуникативная структура предложения опирается на его лексико-синтаксическую структуру, т.е. на его синтаксис и лексический состав. Но она не предопределена ею однозначно: на базе одной и той же лексико-синтаксической структуры можно построить целую **парадигму** коммуникативных структур [5, с. 34 ff; 1, с. 108ff]). Опыты с варьированием порядка слов и акцентов в предложении проводились уже в [3] и даже в [8]. На этой основе возник своего рода «трансформационный» подход к описанию того вклада, который К-структура вносит в семантику предложения. Именно ему мы и следуем.

Коммуникативная парадигма порождается набором **линейно-акцентных преобразований** коммуникативных структур (см. [5, с. 34 ff; 1, с. 118; 14]). Среди коммуникативных структур, образующих К-парадигму, имеется **исходная**, или **первичная** (не обязательно единственная, см. [5, с. 181–183]), и **вторичные**, производные. Вторичные К-структуры описываются с помощью линейно-акцентных преобразований (ЛА-преобразований) исходных структур. ЛА-преобразование – это средство описания формально-смысловых соотношений между членами К-парадигмы. Члены К-парадигмы – это **разные** предложения.

ЛА-преобразования нужны не только для того, чтобы **исчислить** набор возможных К-структур (как в трансформационной грамматике), но и для того, чтобы описать их **функционально**: для каждого линейно-акцентного преобразования указывается изменение коммуникативной семантики предложения, которым оно сопровождается.

Исходные К-структуры являются **нейтральными** – в том смысле, что они находятся в наиболее естественном соответствии с синтаксической структурой и лексическим составом предложения: при нейтральной К-структуре вклад К-структуры в семантику предложения минимален или даже равен нулю.

Список исходных К-структур приводится в разделе 4, их возможные ЛА-преобразования обсуждаются в разделе 5.

## 3. КОММУНИКАТИВНО РЕЛЕВАНТНЫЕ КОМПОНЕНТЫ ПРЕДЛОЖЕНИЯ

Традиционно выделяемые компоненты коммуникативной структуры – это **тема** и **рема**. Иногда проводится различие между темой (которая обычно является субъектом-подлежащим при предикатесказуемом) и **кулисой** (обычно это детерминант предложения), см. [6, с. 6]. Оба эти компонента могут быть названы **началом** [1, с. 232], так что в предложении может быть несколько начал. Так, в предложении (2) два начала – *рано утром* (кулиса) и *пионер Петя* (тема):

(2) Рано утром пионер Петя открыл калитку.

ПРИМЕЧАНИЕ. Понятие «начала» имеет применение в теории коммуникативной структуры. Например, есть важное понятие **нерасчлененное** (= тетическое) предложение, в котором нет ни темы, ни кулисы – т.е. нет начала. В [2, с. 228] про один из коммуникативных типов предложений говорится, что у них «нет привычного начала». У Вилема Матезиуса, родоначальника теории актуального членения, не было термина «тема», а был только **«исходный пункт»**, т.е. то же начало. Термин исходный пункт используется в [9, с. 447].

У разных синтаксических составляющих, таких как подлежащее и детерминант, набор возможных коммуникативных ролей не в точности совпадает. Термин «начало» можно использовать как обобщение для этих двух разных синтаксических типов коммуникативных составляющих, так что тема – это одно из возможных начал. В [15, с. 69] сказано, что в предложении *На холмах Грузии лежит ночная мгла* трудно выделить тему как то, о чем говорится, и рему как то, что говорится. Впрочем, в книге [16], посвященной бытийным предложениям, Локализатор (в данном случае – *на холмах Грузии*) называется темой, а Бытующий объект (*ночная мгла*) входит в рему, и с этим можно согласиться.

В современных работах термин «тема» часто употребляется в чисто просодическом смысле, так что слова «тема» и «начало» выступают как синонимы; тогда во фразе *Однажды я отдыхал в Египте* две темы – как и в (2). Но термин «тема» может иметь и другие употребления, апеллирующие к более тонким аспектам коммуникативной функции слова

или составляющей (т.е. группы слов). Если необходимо четко обозначить именно **смысловую тему**, т.е. то, о чем говорится (см. термин *aboutness* в [17]; о смысловой теме – в [1, с. 115]), можно употребить термин **топик**, ср. эффект топикализации в п. 5.10 и топикальность в п. 5.11.

Термин **детерминант** используется в значении «адвербиал, который относится к высказыванию в целом». Обычно это адвербиал с семантикой времени или места в начальной позиции в предложении. Адвербиалы со значением образа действия не употребляются в качестве естественного начала: \**Быстро Иван оседлал коня*. Но в конечной позиции может оказаться любой адвербиал, и он может быть ремой, в том числе для темы, образованной синтаксическим комплексом «подлежащее + сказуемое», как в примере (1).

Дальнейшее изложение во многом ориентируется на книги [2, 18], которые позволяют а) уточнить арсенал акцентных средств, используемых в русских линейно-акцентных структурах, б) расширить набор коммуникативно релевантных компонентов предложения и коммуникативных структур.

Линейно-акцентная структура характеризует как порядок слов, так и просодию предложения. В основе акцентов, которые играют смысловозначительную роль в русских ЛА-структурах, лежат **интонационные конструкции** (ИК) Е.А. Брызгуновой, см. [19, с. 99 ff]. Следуя за [2], мы принимаем следующие обозначения.

ИК-1 – это обычный понижающийся тон, как правило, конечный; обозначается  $\searrow$ .

ИК-2 – усиленный (он же контрастный) понижающийся тон, часто не конечный, обозначается  $\searrow$ ; ИК-2 – это интенсивный и совершающийся с большим, чем при ИК-1, перепадом частот вариант ИК-1.

ИК-3 – обычный повышающийся тон, обозначается  $\nearrow$ ; усиленный (он же контрастный) повышающийся тон; обозначается  $\nearrow$ .

ИК-4 – понижение с последующим повышением; обозначается  $\cup$ : в дальнейшем рассматривается в роли контекстного варианта ИК-3 (хотя может иметь и другие функции). Так, ниже, в примере (4) акцент ИК-4 имеет составляющая *рано утром*.

ИК-6 – специальный повышающийся тон, обозначается  $\rightarrow$ .

Конструкции ИК-5 и ИК-7 по [19] в дальнейшем изложении не используются.

К этому списку следует добавить обозначение для понижения тона на конечном слове конечной составляющей, которое структурно нерелевантно, т.е. эквивалентно безударности; в [20] для него принято обозначение — (в [14] в этой функции используется подстрочный индекс 0). Например:

(3) Совесьь  $\searrow$  надо иметь — .

Такие аспекты интонации, как долгота, наличие гортанной смычки и др. (см. [12, с. 382–397; 19, с. 107]), связаны с более частными аспектами коммуникативного значения и здесь не рассматриваются.

В любом предложении есть **главное фразовое ударение** (= **фразовый акцент**); в повествовательном предложении оно нисходящее –  $\searrow$  или  $\searrow$ , в во-

просительном (а именно, в общем вопросе) – восходящее  $\nearrow$ . В расчлененном предложении есть также **второстепенное фразовое ударение**, обычно восходящее –  $\nearrow$ .

Теперь можно сказать, что **ремой** предложения является, в каноническом случае, составляющая, в которую входит слово – носитель главного фразового акцента (в случае множественной ремы, см. п. 5.12, главный фразовый акцент несет только последняя из рематических составляющих). **Темой** является составляющая, которая отмечена второстепенным фразовым акцентом (см., однако, в п. 5.12 об атоначеской теме). Так, в предложении (4) есть рема – *открыл калитку* – и две темы, о которых уже шла речь. В коротком предложении второстепенное фразовое ударение на теме факультативно – можно произнести фразу *Петя открыл калитку* так, что тема *Петя* будет выражена только лексико-синтаксической структурой.

На каждый коммуникативно охарактеризованный компонент ЛА-структуры приходится один фразовый акцент. Слово, на которое падает акцент, – это **акцентоноситель**, см. [2]. В русском языке в нейтральной К-структуре акцентоносителем является обычно конечное слово составляющей. Так, в предложении (2), здесь повторяемом под номером (4), главное фразовое ударение падает на *калитку* и второстепенное – на *утром* и *Петя*:

(4) Рано утром  $\cup$  пионер Петя  $\nearrow$  открыл калитку  $\searrow$  .

В простом случае акцентоноситель в рематической составляющей – это то же, что **собственно рема**, см. [1, с. 112]; аналогично, в тематической составляющей может быть **собственно тема** – акцентоноситель темы. Но собственно ремой может быть не только слово, но и составляющая – меньшей протяженности, чем рема в целом, см. п. 5.1.

Особым объектом в ЛА-структуре предложения является то, что в [1, с. 113] называется **заударный компонент** (в [21] для составляющей, расположенной после рематического акцентоносителя, принят термин **хвост**; он может не быть никаким отдельным компонентом ЛА-структуры), см. именную группу (ИГ) *свой стул* в примере (5). В [5, с. 177] в контекстах такого рода усматривается вторая тема.

(5) Ипполит Матвеевич  $\nearrow$  сразу узнал  $\searrow$  *свой стул*.

Предложение может быть **нерасчлененным**, т.е. состоять из одной ремы, см. (6), (7); но состоять из одной темы предложение не может: рема должна быть обязательно.

(6) Послышалось ворчание старой служанки  $\searrow$  в темноте [М.А. Булгаков. Мастер и Маргарита (1929–1940)].

(7) Начались скандалы на новую тему  $\searrow$ .  
[В. Токарева. Своя правда (2002)].

Связь нисходящего тона с главным фразовым ударением и восходящего – со второстепенным принципиальна для предложений с относительно простой структурой. При синтаксическом усложнении предложения происходят автоматические акцентные сдвиги. Так, нисходящий рематический акцент должен быть заменен на восходящий, если в ходе синтаксической деривации составляющая попадает в контекст

сочинительной конструкции, как в (8), или условной конструкции, как в (9):

(8) Пионер Петя открыл калитку↗ и вышел на большую зеленую лужайку↘.

(9) Если Петя открыл калитку↗, в сад забегают собаки↘.

Считается неслучайной связь восходящего ударения, с одной стороны, с незавершенностью, как в (8), (9), а с другой – с вопросом, как в (10):

(10) Петя открыл↗ калитку?

Термин «интонация незавершенности» употребляется в [22]: в немецком оригинале (с. 199) различаются «weiterweisende» Funktion, **интонация незавершенности**, и «abschließende» Funktion, **интонация завершенности** (в русском переводе А.А. Холодовича, см. с. 248, термина «интонация незавершенности» нет; англ. термин – nonterminal intonation).

Дать представление о коммуникативной структуре русского предложения – значит иметь возможность указать для любого предложения его коммуникативную структуру, охарактеризовав при этом тот вклад, который она вносит в его семантику, в дополнение к лексико-синтаксической структуре. В дальнейшем изложении принят последовательно **интерпретирующий** подход к К-структуре. Речь идет о том, как выявить в предложении ремю, тему или какой-то другой компонент К-структуры, дать интерпретацию исходным К-структурам и описать изменение коммуникативного значения предложения при изменении его К-структуры.

ПРИМЕЧАНИЕ. В [23] речь о коммуникативной структуре идет в связи с принципиально иной задачей – задачей построения предложения на базе его семантического представления в рамках теории «Смысл–Текст». Т.е. речь идет о построении предложения по его структуре, а не о построении структуры по предложению, как в нашем случае.

Ниже следует описание коммуникативных структур русских предложений. Оно состоит из двух частей:

а) перечня исходных коммуникативных структур, где для каждой из них приведена соответствующая ей линейно-акцентная структура и семантическая интерпретация (см. раздел 4);

б) набора линейно-акцентных преобразований, которые меняют исходную ЛА-структуру и, соответственно, интерпретацию, вводя тем самым новые коммуникативные структуры с приписанной им семантической интерпретацией (см. раздел 5).

В качестве иллюстративных примеров ниже частично используется материал Основного корпуса русского языка, и это, соответственно, только письменные тексты; акцентная разметка предложений, как и в искусственных примерах, авторская.

Примеры, в которых указание на источник дано в квадратных скобках, взяты из Национального корпуса русского языка, <http://ruscorpora.ru>.

#### 4. К-СТРУКТУРЫ С ИСХОДНЫМ СООТНОШЕНИЕМ МЕЖДУ СИНТАКСИЧЕСКИМ И КОММУНИКАТИВНЫМ ЧЛЕНЕНИЕМ

В [5, с. 65] перечислены коммуникативные структуры, про которые можно сказать, что они порожда-

ются синтаксическим членением предложения. Выделены основные типы высказываний, «в которых актуальное членение не нарушает иерархию синтаксического членения», так что ЛА-структура не добавляет нового содержания по сравнению с тем, которое заключено в лексико-синтаксической структуре предложения.

##### Тип 1. Субъектно-предикатное расчленение

Это классическое расчленение на логический субъект и логический предикат, см. раздел 3 (вертикальная черта в примерах обозначает границу между темой и ремой).

(11) Шаляпин↗ | уехал в Италию↘.

(12) Пожилые люди↗ | легко простужаются↘.

##### Тип 2. Детерминантно-событийное расчленение

Рематический событийный компонент (т.е. субъектно-предикатный комплекс) может быть нерасчленным, как в (13), или, в свою очередь, потенциально членимым на тему и ремю, как в (14):

(13) 25 сентября 1970 года↗ | умер Эрих Мария Ремарк↘.

(14) После тяжелой и продолжительной болезни↗ | писатель умер↘.

##### Тип 3. Нерасчленная структура (тетическое предложение)

(15) Наступила весна↘.

В нерасчленном предложении, таком как (15), нормой является **субъектно-предикатная инверсия**, т.е. порядок слов, при котором группа сказуемого стоит перед группой подлежащего.

В предложении (13), с детерминантно-событийным расчленением, такая же инверсия осуществляется в пределах рематической составляющей. В [24] эта инверсия объяснена тем, что препозитивное подлежащее имеет тенденцию расчленять синтаксический фрагмент на тему и ремю. Об инверсии как способе консолидации рематического субъектно-предикатного комплекса см. в п. 5.9.

Односоставные предложения не образуют (вопреки [5, с. 65]) отдельных коммуникативных типов: у предложения *На улице быстро темнело* К-структура типа 2, в предложении *Становилось темно* – типа 3.

Все три типа К-структур могут усложняться при усложнении синтаксиса. Так, при одной реме может быть несколько тем, см. пример (4). Аналогично, при одной теме может быть несколько рем – так, в (16) их три:

(16) Умер↗ Ремарк | 25 сентября 1970 года↗ | в возрасте 72-х лет↗ | в городе Локарно↘.

Варьировать К-структуру может наличие зависимых слов при глаголе. О К-структурах синтаксически сложных предложений (в частности, о том, что восходящий акцент может выражать ремю) см. [25].

Все три типа К-структур являются нейтральными – в том смысле, что К-структура предложения не добавляет ничего существенного в его семантическую интерпретацию. В разделе 5 мы обратимся к К-структурам с неисходным соотношением между синтаксическим и коммуникативным членением: вклад коммуникативной структуры в семантику предложения описывается как результат ЛА-преобразования.

## 5. ЛИНЕЙНО-АКЦЕНТНЫЕ ПРЕОБРАЗОВАНИЯ В К-СТРУКТУРАХ

Ниже следует перечень ЛА-преобразований, которые описывают К-структуры с неисходным соотношением между синтаксическим и коммуникативным членением. В отличие от синтаксических трансформаций в поздней генеративной грамматике, ЛА-преобразования меняют смысл предложения: их назначение в том и состоит, чтобы описать вклад ЛА-структуры в коммуникативную семантику предложения.

### 5.1. Экспрессивная препозиция собственно ремы

Это ЛА-преобразование (см. [5, с. 98]) переносит собственно рему (например, *последние огни* в (17а)) в начало предложения, сохраняя неизменным свойственный ей нисходящий фразовый акцент. Остаток ремы и бывшая тема (*потухли* и *в хатах* в (17а)) соединяются в одну составляющую, причем безударную: второстепенное ударение утрачивается – понижение тона на последнем слове не имеет смысловозначительного статуса.

(17) а. *Последние огни* \> потухли в хатах — ⇐ В хатах \> | потухли *последние огни* \> .

б. Вдруг *странный треск* \> послышался в лесу — ⇐ Вдруг в лесу \> | послышался *странный треск* \> .

в. *Серебристый туман* \> стоял над станцией — ⇐ (Л. Толстой) ⇐ Над станцией \> | стоял *серебристый туман* \> .

У предложений с препозицией собственно ремы (в отличие от предложений с суперпозицией ремы, см. п. 5.10) конечная составляющая не несет нормально-го акцента ИК-1, а является в функциональном отношении безударной: предложение теряет одно из фразовых ударений. Предложение, которое членилось на тему и рему, становится нерасчлененным.

Фронтированию может подвергаться группа подлежащего – если в исходной К-структуре порядок слов «сказуемое – подлежащее»:

(18) *Дикая ярость* \> охватила Якова — ⇐ Якова \> охватила *дикая ярость* \> .

Экспрессивная препозиция собственно ремы не дает никакого специального коммуникативного отличия от предложений с конечной ремой – кроме трудноуловимой экспрессии. См. примеры – частично из [5]:

(19) а. *Любви и заботы* \> требует ребенок — ⇐ Ребенок \> требует *любви и заботы* \> .

б. *Запахом осени* \> наполнилась комната — ⇐ Комната \> наполнилась *запахом осени* \> .

в. *Посредственный* \> поэт был Шаховской — ⇐ Шаховской \> был *посредственный* \> поэт.

г. *На двести пунктов* \> упал вчера курс доллара — ⇐ Курс доллара \> упал вчера *на двести пунктов* \> .

д. *Не однажды* \> вставал передо мною этот вопрос — ⇐ Этот вопрос \> вставал передо мною *не однажды* \> .

Предложение (20) можно рассматривать как результат экспрессивной препозиции ремы целиком – включая заударный остаток:

(20) *Старили* \> Аксинью горячие ночи — ⇐ Горячие ночи \> *старили* \> Аксинью. (М. Шолохов. Тихий Дон).

Экспрессивная препозиция часто применяется в предложениях **введения в рассмотрение**, (17б), и в **бытийных**, (17в). (В словаре [26] предложения введения в рассмотрение включены в состав бытийных.) Дело в том, что имеется тесная связь и иногда трудно прочерчиваемая граница между существованием и восприятием [27; 20, с. 443 ff]. Бытийные предложения занимают важное место в русском синтаксисе, см. [28, 16]. Например, они играют важную роль в морфосинтаксисе отрицания, см. [29, 30]. Бытийным предложениям, равно как и предложениям введения в рассмотрение, свойственна субъектно-предикатная инверсия в рематической части.

Экспрессивная препозиция сопровождается субъектно-предикатной инверсией, см. (19). В предложениях введения в рассмотрение и бытийных, где инверсия изначальна, как в (17б), (17в), при экспрессивной препозиции восстанавливается порядок слов «подлежащее – сказуемое».

### 5.2. Дислокация ремы с рецессией темы

В [5, с. 120] описаны два ЛА-преобразования, которые характеризуются как **инверсия с дислокацией ремы**: рема, состоящая из глагола и зависимой от него синтаксической группы, расчленяется на две части, и глагол выносится в начало предложения (в позицию перед подлежащим), а оставшаяся синтаксическая группа – в конец. При этом бывшая тема оказывается внутри расчлененной ремы. Глагол получает акцент ИК-6, а оставшаяся синтаксическая группа сохраняет свой обычный главный фразовый акцент.

Коммуникативный эффект дислокации существенно зависит от того, было ли темой исходного предложения подлежащее или детерминант. Рассмотрим преобразование, при котором тема исходного предложения – подлежащее. Преобразование, при котором тема исходного предложения – детерминант, обсуждается ниже, в п. 5.3.

Предложение с темой-подлежащим при дислокации ремы превращается, согласно [2, с. 201], в структуру с **рецессивной** (т.е. подавленной) **темой**; предложение становится в каком-то смысле нерасчлененным – оно обозначает ситуацию в целом, а не свойство или действие субъекта:

(21) Посадил \> *дед* репку \> ⇐ *Дед* \> посадил репку \> .

Другие примеры:

(22) Возвращается \> *муж* из командировки \> ...

(23) Оглянулся \> *Жилин* на свою лошадь \> (пример из [5, с. 121]).

(24) Пустился \> *мой доктор* в разговоры \> (пример из [5, с. 171]).

(25) <Выйду \> *я* за Ганса замуж \> . Родится \> *у нас сын* \> > Пойдет \> он в погреб за пивом \> ... (Сказка «Умная Эльза»).

Основное коммуникативное свойство этих предложений – они неавтономны: типичная позиция такого предложения – в начале рассказа или анекдота; оно предполагает продолжение. Фраза *Плеханов ушел в кусты* (из «Краткого курса истории ВКП(б)», известного своей коммуникативной безупречностью) после такого преобразования оказалась бы неуместной в своем контексте: фраза однозначно интродуктивная, т.е. требует продолжения – что же произошло там дальше?

(26) Ушел ↗ Плеханов в кусты ↘ .

Контекстная связанность предложений с дислокацией ремы очевидна в примерах из [5, с. 122]:

(27) Запустил ↗ Иван руку в борозду ↘ , ощупал – мягкое (Л. Толстой).

(28) Сел ↗ Никишка на корточки ↘ , смотрит во все глаза (Ю. Казаков).

Частица *раз* делает рецессию темы и интродуктивный интонационный контур практически обязательным:

(29) а. Раз пришел ↗ я к ним на семинар ↘ ; б. <sup>?</sup>Раз пришел ↗ я к ним на семинар ↘ .

В [2, с. 214] у предложений этого рода усматривается «семантика умозрительного наблюдения». В самом деле, они создают «эффект развертывания ситуации перед мысленным взором наблюдателя» – в том смысле, что введенная в рассмотрение ситуация служит ареной для развертывающихся далее событий.

Дислокация темы – это специально нарративная структура, когда данное предложение предполагает следующее предложение. Она не используется в репортаже, где каждое предложение – это отдельный речевой акт.

Не следует думать, что ЛА-структура с акцентом ИК-6 свойственна одним только предложениям с дислокацией ремы. Тот же интонационный рисунок имеют предложения с препозитивным адвербиалом, см. (30). ЛА-структуру с акцентом ИК-6 может иметь вопросительное предложение с вопросительным словом (ср. о связывании акцентов в [12]), см. (31):

(30) а. Раненько ↗ задумал ты жениться ↘ . <Но невеста твоя мне нравится.>

б. Напрасно ↗ ты так старался ↘ . <Не выйдет ничего.>

(31) Куда ↗ ты положил Булгакова ↘ ?

### 5.3. Дислокация ремы в интродуктивном предложении

В п. 5.2 обсуждается один из случаев **инверсии с дислокацией ремы** по [5, с. 120] – с темой-подлежащим в исходном предложении. Если же в исходном предложении подлежащее было собственно ремой, а темой был детерминант или другое зависимое, то коммуникативным эффектом той же **дислокации ремы** вокруг темы, т.е. результатом фронтирования глагола будет высвечивание подлежащего. Такая структура характерна для того, что в [28, с. 221] называется **интродуктивным** предложением:

(32) Служил ↗ на Кавказе *один барин* ↘ (пример из [2, с. 200]).

(33) Нападала ↗ на меня в то время *серая тоска* ↘ (пример из [5, с. 175]).

(34) Был ↗ у майора Деева *товарищ* ↘ (пример из [1, с. 119]).

Эти предложения неавтономны, как и предложения, в которых дислокация ремы сопровождается рецессией темы-подлежащего (см. п. 5.2): и те и другие требуют продолжения. Но коммуникативная судьба у подлежащего здесь совершенно иная. В предложении с уже рематическим подлежащим подавление темы придает этому подлежащему дополнительную выделенность: соответствующий объект вводится в рассмотрение.

ПРИМЕЧАНИЕ. Как справедливо указано в [28, с. 221], интродуктивные предложения являются эгоцентрическими – они предполагают наблюдателя или субъекта сознания (см. подробнее [20, с. 425–468]).

Общим свойством обеих дислокаций является использование акцента ИК-6.

### 5.4. Тема-рематическая инверсия; эффект идентификации

В статье [3] рассматриваются различные акцентные варианты (и, соответственно, К-структуры) для английского предложения *John kissed Mary*, из которых для нас представляет интерес вариант с контрастом, который порождает **интерпретацию исчерпывающего перечисления** (exhaustive list interpretation):

(35) John ↘ kissed Mary <exhaustive list interpretation> = ‘John (and only John) kissed Mary’.

В качестве возможного описания коммуникативной семантики этого предложения рассматривалась также перефразировка ‘it was John who kissed Mary’. Перейдем к аналогичному русскому примеру:

(36) а. Джон ↗ | поцеловал Марию ↘ [нейтральная К-структура].

б. Марию ↗ поцеловал | Джон ↘ [К-структура с тема-рематической инверсией].

ЛА-преобразование состоит в том, что тема и рема меняются местами: бывшая тема становится ремой (скорее всего, хотя и необязательно, контрастной, ↘), а вся рематическая составляющая – темой. Дополнительная семантика – исчерпывающее перечисление: ‘только Джон поцеловал Марию’.

В [1, с. 120] в описании семантики этого ЛА-преобразования было использовано понятие презумпции. Предложение (37б) отличается от (37а) тем, что семантика (37б) членится на презумпцию ‘существует <единственный> человек, который ее любит’ и идентифицирующую ассерцию: ‘этот человек – Иван’.

(37)а. Иван ↗ | ее любит ↘ ; б. Любит ↗ ее | Иван ↘ .

Так эксплицируется семантика исчерпывающего перечисления, намеченная в [3].

Для того чтобы данное ЛА-преобразование имело эффект идентификации, должны выполняться следующие условия, касающиеся исходного предложения.

а) Рема исходного предложения должна задавать идентифицирующее свойство объекта, обозначенного темой, или событие, происшедшее с этим объектом,

которое если и не выделяет его, то, в принципе, способно выделить.

б) Исходная тема предложения, которая в результате преобразования становится ремой, должна быть референтной ИГ.

Оба эти условия выполняются в (38), и ЛА-преобразование сопровождается нужным семантическим эффектом:

(38) Вася меня проводит<sub>↘</sub> ⇒ Проводит<sup>↗</sup> меня Вася<sub>↘</sub> = 'существует <единственный> человек, который меня проводит, и этот человек – Вася'.

В примере (39) условие а) не выполняется, и результат соответствующего ЛА-преобразования с трудом поддается осмыслению (поскольку (39а) не выделяет Джона как единственного свободолобца), что отмечено в [3]:

(39) а. Джон любит свободу; б. Свободу<sup>↗</sup> любит | Джон<sub>↘</sub> .

В модальном и кванторном контексте темарематическая инверсия может давать дополнительные – не чисто коммуникативные – семантические приращения:

(40) Разработкой этих правил<sup>↗</sup> должен заниматься математик<sub>↘</sub> ≠ Математик<sup>↗</sup> должен заниматься разработкой этих правил<sub>↘</sub> .

Та же коммуникативная семантика (презюпция единственности) выражается конструкцией примера (41), которая является аналогом клефт-конструкции английского языка (см. о ней [9, с. 446]):

(41) Если кто для меня непонятен, так это моя бабушка = Непонятна<sup>↗</sup> мне моя бабушка<sub>↘</sub> .

Предложение *Моя бабушка мне непонятна*, исходное для (41), никак не выражает идеи о том, что непонятность – это уникальное свойство бабушки.

В [2, с. 211 ff] была рассмотрена К-структура, представленная примером (42а). См. также примеры (42б) и (42в).

(42) а. <Служил на Кавказе один барин.> Звали<sup>↗</sup> его Жилин<sub>↘</sub> .

б. Имя<sup>↗</sup> этой теме – любовь<sub>↘</sub> .

в. <– Давайте я вас провожу. – Нет, спасибо.> Проводит<sup>↗</sup> меня Вася<sub>↘</sub><sup>1</sup> .

Структура предложения (42а) была охарактеризована одновременно с трех точек зрения:

1) с линейно-акцентной точки зрения это предложение с глаголом-темой (*звали*);

2) с логико-семантической точки зрения это предложение со структурой «признак (атрибут) объекта (*звали*) – имя объекта (*его*) – значение признака (*Жилин*)»;

3) с коммуникативной точки зрения: предложение содержит, помимо акцентно выделенной (*звали*), еще одну – **энклитическую** тему (*его*).

Анализ, предложенный в [2], расширяет семантический арсенал теории коммуникативной структуры: наряду с субъектно-предикатным расчленением, двучленным, предлагается использовать параметриче-

ское расчленение – трехчленное, в котором помимо двух обычных коммуникативных компонентов, тема и рема, участвует третий – энклитическая тема. Она указывает на **предмет**; тематический глагол – на **признак** этого предмета, а собственно рема – на **значение признака**. Этот анализ подтверждается тем, что глагольную форму часто можно преобразовать синтаксически в имя признака плюс связка:

(43) Руководил<sup>↗</sup> *этим движением* | Колесов<sub>↘</sub> ⇒ Руководителем<sup>↗</sup> *этого движения* | был Колесов<sub>↘</sub> .

Имя объекта в этих предложениях (в примерах оно выделено курсивом) не является началом, а занимает «вакернагелевскую позицию», т.е. это энклитическая тема: тема-начало тут глагол (NB отличие энклитической темы от рецессивной, т.е. подавленной). Существенный признак энклитической темы – активированность, т.е. предупянутость; отсюда фразовая безударность.

Итак, предложения с глаголом-темой получают ЛА-преобразованием «тема-рематическая инверсия». Трансформационное описание позволяет показать, что именно тематизация глагола порождает а) презюпцию существования ситуации с участником – энклитической темой и б) идентификацию этого участника с участником-ремой.

Тема-рематическая инверсия может не давать эффекта идентификации. Так, в структуре с детерминантно-событийным расчленением (см. тип 2 в разделе 4) тема-рематическая инверсия не дает никаких смысловых приращений – изменение касается только «порядка обхода ситуации». Ср. (44а) и (44б):

(44) а. Но *Аня*<sup>↗</sup> умерла | две недели назад<sub>↘</sub> .

[А. Волос. Недвижимость (2000)].

б. Две недели назад<sup>↗</sup> | умерла *Аня*<sub>↘</sub> .

## 5.5. Тематизация собственно ремы (= «тематическая препозиция»)

Экспрессивная препозиция (п. 5.1) переносит собственно рему в начало, сохраняя неизменным ее фразовый акцент. Между тем тематизация собственно ремы меняет акцент фронтлируемой составляющей на противоположный, и она становится темой. При этом предложение сохраняет свою двучленную структуру, поскольку бывшая тема и остаток ремы становятся единой рематической составляющей. В примерах ниже предложение (б) – результат тематизации собственно ремы в (а):

(45) а. Мой *брат*<sup>↗</sup> | живет в Казани<sub>↘</sub> ; б. В Казани<sup>↗</sup> | живет мой *брат*<sub>↘</sub> .

(46) а. *Дети*<sup>↗</sup> | играют в саду<sub>↘</sub> ; б. В саду<sup>↗</sup> | играют *дети*<sub>↘</sub> .

(47) а. *Коля*<sup>↗</sup> | пришел на рассвете<sub>↘</sub> ; б. На рассвете<sup>↗</sup> | пришел *Коля*<sub>↘</sub> .

(48) а. *Собака*<sup>↗</sup> | укусила мальчика<sub>↘</sub> ; б. Мальчика<sup>↗</sup> | укусила *собака*<sub>↘</sub> .

Следует подчеркнуть различие между тематизацией собственно ремы, о которой идет речь, и темарематической инверсией из п. 5.4. Рассмотрим это

<sup>1</sup> Исходной структурой здесь будет *Вася меня проводит*.

различие, сравнив пример (45) с примером (49) (разбиравшемся в п. 5.4 под номером (36)):

(49) а. Джон ↗ | поцеловал Марию ↘ ; б. Марию ↗ поцеловал | Джон ↘ .

Формальное различие состоит в том, что в примере (49) в начало выносятся вся рема, а в (45) – только именная группа: глагол остается в составе ремы. Различие семантическое состоит в том, что отношение идентификации, которое определяет семантику коммуникативного сдвига в примере (49), возникает на базе событийного глагола – синтаксически близкое преобразование в примере (45) идентификации не порождает: (45б) не значит, что мой брат – единственный, кто живет в Казани.

Рассмотрим теперь известный пример из [31]. Предложение (50а) очевидным образом истинное, в то время как (50б) мы скорее склонны признать ложным:

(50)а. В Австрии ↗ | говорят на немецком языке ↘ ;  
б. На немецком языке ↗ | говорят в Австрии ↘ .

Разница между (50а) и (50б) следующая. Фраза (50а) содержит информацию о том, на каком языке говорят в Австрии – на немецком; а фраза (50б) утверждает статус Австрии как **единственной** страны, где говорят на немецком языке: ее коммуникативная структура порождает семантику исчерпывающего перечисления, поскольку конкретно-референтная ИГ в позиции контрастной, как в примере (49б), ремы выражает идентификацию.

Формально, т.е. на уровне ЛА-преобразования, в (50) тематизации подвергается только собственно рема. Но откуда тогда эффект идентификации, который был описан для другого ЛА-преобразования – для тема-рематической инверсии? Дело в том, что в этом предложении – «сверхтесная» смысловая связь между глаголом и именной группой. Поэтому неважно, что в ЛА-структуре глагол относится к реме; по смыслу он входит в тему.

Такая же «сверхтесная» смысловая связь порождает идентифицирующий эффект в примере (51):

(51) а. *Объект* ↗ | выражается **аккузативом** ↘ .  
б. **Аккузативом** ↗ | выражается *объект* ↘ .

Семантический сдвиг, сопровождающий тематизацию собственно ремы (равно как и тема-рематическую инверсию), может быть весьма существенным. Так, (52а) ≠ (52б); в самом деле, из того, что строители живут в палатках, см. (52а), не следует, что в палатках не живет никто другой, см. (52б):

(52) а. Строители | живут в палатках.  
б. В палатках | живут строители.

Дело в том, что тематическая позиция заставляет понять референциальный статус ИГ в палатках в (52б) как определенный – какового эта ИГ не имеет в (52а). Другой пример ((53б) не означает, что инверсия редко происходит в других местах):

(53) а. Инверсия происходит, как правило, в *независимых высказываниях*; б. В *независимых высказываниях*, как правило, происходит инверсия.

В примере (54) (из [1, с. 110]) тематическая препозиция приводит к изменению сферы действия кванторов (примеры такого рода широко обсуждались в

литературе, начиная с [32]) – в (54а) правила у каждого водителя свои, а в (54б) – единые для всех (т.е. в (54а)  $\forall > \exists$ , а в (54б)  $\exists > \forall$ ):

(54) а. Каждый водитель соблюдает некоторые правила; б. Некоторые правила соблюдает каждый водитель.

Тематическая препозиция собственно ремы предъявляет определенные требования к семантике переставляемых синтаксических групп. Так, в (55) с группой *на столб*, в отличие от (48) с группой *мальчика*, тематическая препозиция практически невозможна:

(55) Машина налетела на столб (пример из [33]).

Субъектно-предикатная инверсия, которая сопровождает тематизацию собственно ремы, имеет когнитивное обоснование, см. [34]. Ее естественно объяснить тем, что обратный порядок слов служит средством выражения **коммуникативной нерасчлененности**: препозитивное подлежащее стремится быть понятым как тема – как в целом предложении, так и в его рематической части; постпозиция подлежащего снимает расчленение (см. п. 5.9).

## 5.6. Контрастная тема

Преобразование «Контрастная тема», т.е. замена обычной темы на контрастную, состоит в том, что начальная именная группа с обычным (и притом факультативным) второстепенным ударением, как в (56а), приобретает контрастное [1, с. 118], как в (56б):

(56)а. Иван ее любит ↘ ; б. Иван ↗ ее любит ↘ .

Семантический сдвиг состоит в том, что вводится в рассмотрение множество объектов, ассоциированное с объектом-началом *a* (в это множество входит и сам объект *a*). *A* в смысл предложения, помимо утверждения о том, что некое свойство *Q* (рема) имеет место для *a*, входит, в статусе **имплицатуры**, компонент ‘для других элементов ассоциированного множества свойство *Q*, возможно, не имеет места’. См. о множестве альтернатив в [35] и о контрасте в [36, 37]. Так что в смысл (56б) входит не только компонент (56а), ‘Иван ее любит’, но и импликация ‘и, возможно, больше никто’, ср. [38, с. 506].

Семантический эффект тематического контрастного акцента обсуждается в [39, с. 775] – на простых примерах. Так, предложение (57а) имеет импликацию (57б), а (57б) ее не имеет:

(57) а. Маша ↗ не пришла ↘ ; б. Маша не пришла ↘ ; в. Пришел кто-то другой.

Еще пример. Предложения (58а) и (58б) близки по смыслу. Однако восходящий акцент в (58б) может быть понят как контрастный, так что (58б) в большей степени несет импликацию, что вера в успех в конце концов имела место (т.е. в последнюю минуту – поверил):

(58) а. До последней минуты ↘ не верил Тарас в успех — .

б. До последней минуты ↗ Тарас не верил в успех ↘ .<sup>2</sup>

<sup>2</sup> Отметим, впрочем, что эти два предложения отличаются друг от друга не только акцентом, но и порядком слов.

Есть серия кванторных наречий и прилагательных, которые недопустимы в позиции контрастной темы [40] (ср. о «неконтрастоспособных» словах в [2, с 19–20 ff]).

(59) а. *Некоторые* ↗ инструменты нам понадобятся ↘ [= ‘а другие нет’; или: ‘а про остальные неизвестно’].

б. \**Все* ↗ инструменты нам понадобятся ↘<sup>3</sup>.

в. Нам понадобятся *все* ↘ инструменты.

(60) а. *Иногда* ↗ он шел домой пешком ↘ <а иногда садился в троллейбус>.

б. \**Всегда* ↗ он шел домой пешком ↘.

в. Он *всегда* ↘ шел домой пешком.

(61) а. *Какой* ↗-*нибудь* молоток найдется ↘ [= есть набор мыслимых качеств молотка; достаточно молотка в том числе с самыми низкими качествами].

б. \**Всякий* ↗ молоток найдется ↘.

в. \**Всего* ↗ мне достаточно ↘ (пример из [42]).

Такая избирательность легко выводится из определения контрастной темы: контраст вводит в ситуацию ассоциированное множество, на фоне которого рассматривается данный объект. Очевидно, слова *все*, *всегда*, *всякий* исключены в позиции контрастной темы, поскольку не порождают ассоциированного множества однородных объектов: *некоторые* ↗ *люди* противопоставлены некоторым другим; а *все* ↗ *люди* уже не противопоставлены никаким другим людям.

### 5.7. Контрастная рема; верификативный контраст

Контраст в реме выражается контрастным понижающимся акцентом ↘ и по смыслу аналогичен (хотя и не тождествен) контрасту в теме. Так, предложение (62б), в отличие от (62а), несет импликацию ‘а не кого-либо другого’:

(62) а. Джон поцеловал Марию ↘ ; б. Джон поцеловал Марию ↘.

Частным случаем коммуникативной структуры с контрастной ремой является структура с **верификативной** ремой. Такая К-структура возникает в том случае, если акцентируется не сама пропозиция, а только ассоциированный с ней предикат утвердительности – ‘имеет место’<sup>4</sup>. К-структура (63), например, возникает в контексте, где о соответствующем событии уже шла речь, но неизвестно, произошло оно или нет:

(63) Джон поцеловал ↘ Марию.

<sup>3</sup> Контрастная тема возможна в контексте отрицания: *Все* ↗ *инструменты нам не понадобятся* ↘. Но это особое, смещенное отрицание, см. [41].

<sup>4</sup> Предикат ‘имеет место’ соответствует тому, что в [43] носит название «сокровенная связка». Если в пропозиции главный логический акцент сосредоточен на предикате ‘имеет место’, то в подчиненной позиции, например, в контексте предиката пропозициональной установки, она может обозначать только факт, ср. *видел, как они купались в Москва-реке и видел, что они купались* ↘ в *Москва-реке* [= что имел место факт].

О структуре с верификативной ремой см. [44]; 2, с. 61–64]. Акцентоносителем в предложении с верификативной ремой является сказуемое. Контраст и в этом случае означает выбор, но тут выбор ограничивается двумя возможностями – *да* и *нет*. Верификативная рема возникает в специальном контексте – например, в полемическом диалоге, см. [18, с. 156].

Верификативная рема означает, что у предложения есть общее отрицание. Так, (64а) имеет соответствующее общеотрицательное – это (64б), а для (65) отрицание под вопросом:

(64) а. Иван подарил ↘ матери свою новую книгу;

б. Иван не подарил ↘ матери свою новую книгу.

(65) Иван подарил матери ↗ свою новую книгу ↘.

Контрастный акцент на предикате ‘имеет место’ может иметь и тематическая составляющая – если она имеет пропозициональное (а не предметное) значение. Пример из [2, с. 63]:

(66) Неудавшаяся ↗ операция только усложнит ↘ положение заложников.

ИГ *неудавшаяся операция* – это результат атрибутивного стяжения [45] пропозиции: (66) означает: ‘если операция не удастся...’

### 5.8. Рематизация приглагольного адвербиала

Сравним два предложения:

(67) а. Родители ↗ *громко* разговаривали ↘ ; б. Родители разговаривали *громко* ↘.

Нейтральную ЛА-структуру имеет предложение (67а), которое членится на группу подлежащего, тему, и группу сказуемого, рему. Между тем более простую семантику имеет предложение (67б), в котором ситуация *родители разговаривали* – пресуппозиция, а ассерция образуется с помощью адвербиала *громко*, который характеризует эту ситуацию.

### 5.9. Рематическая консолидация в субъектно-предикатном комплексе (как противоположность тема-рематическому расчленению)

В п. 5.1 шла речь о бытийных предложениях, для которых нейтральным является инвертированное расположение подлежащего и сказуемого в рематической части. С другой стороны, есть отдельное ЛА-преобразование, которое осуществляет инверсию подлежащего и сказуемого в предложениях, изначально расчлененных на тему и рему, создавая нерасчлененность. Это **рематическая консолидация**:

(68) Мой новый компьютер ↗ сломался ↘ ⇒ Сломался мой новый компьютер ↘.

Для предложения (68), в отличие от предложений типа 3 из раздела 4 (*Наступила весна*) и рематической части бытийных предложений, см. (17б, в), нерасчлененная структура не является исходной.

Рематическая консолидация может быть сопутствующим преобразованием при тематической препозиции собственно ремы (см. п. 5.5). Так, в предложении (69) естественным результатом тематической препозиции собственно ремы является (69б), где ос-

таток ремы и бывшая тема объединяются в один компонент, в единую рему – т.е. подвергаются рематической консолидации. Если рематической консолидации при тематической препозиции собственно ремы не происходит, возникает предложение (69в), с двумя темами, из которых первая оказывается контрастной, – да и рема, скорее всего, контрастная:

(69) а. Хозяин дома↗ | открыл дверь↘ ; б. Дверь↗ | открыл хозяин дома↘ ; в. Дверь↗ | хозяин дома↗ | открыл↘ .

В самом деле, предложение (69в), с контрастной темой, содержит, помимо ассерции ‘хозяин открыл нам дверь’, импликацию ‘хозяин не открыл чего-то еще или не сделал чего-то, что ожидалось’.

### 5.10. Суперпозиция ремы

В [2, с. 161] был описан особый тип коммуникативных структур – структуры с **суперпозицией ремы**. Пример:

(70) Говорящего попугая↘ приобрел на Птичьем рынке↗ студент Порошков↘ .

Будучи сдвинуто со своего естественного места в конце фразы, главное фразовое ударение усиливается, так что его акцент не ИК-1, ↘ , а ИК-2, ↘ .

В предложении (70) есть обычная конечная рема – это подлежащее *студент Порошков*, и есть другая рема в начальной позиции, отмеченная усиленным акцентом, – *говорящего попугая*.

В [14] К-структуру с суперпозицией ремы предлагается трактовать как результат применения преобразования «экспрессивная препозиция» ремы к К-структуре, где уже есть достаточно сложное коммуникативное членение. Так, (70) может быть получено из (71) (NB субъектно-предикатную инверсию):

(71) Студент Порошков↗ приобрел на Птичьем рынке↗ говорящего попугая↘ .

Однако препозиция ремы осуществляется здесь не совсем так, как преобразование «экспрессивная препозиция» ремы описано в п. 5.1. Обычная экспрессивная препозиция ремы упрощает исходное членение, превращая исходную структуру с TR-членением в нерасчлененную. Между тем при суперпозиции ремы число коммуникативных компонентов в предложении остается неизменным.

Суперпонируемой ремой могут стать составляющие самых разных синтаксических типов – дополнение, как в (71), обстоятельство, как в (72), или подлежащее, как в (73) (пример из [14]):

(72) На севере Урала↘ нашли Екатеринбургские археологи↗ кости мамонта↘ .

(73) Сильный взрыв↘ прогремел сегодня↗ на севере Британской столицы↘ .

Предложение (74) трактуется в [2, с. 177] как пример суперпозиции ремы:

(74) <На санитарной линейке умирает Шевелев, полковой командир.> Женщина↘ сидит у его ног — .

Между тем, предложение (74) – это результат обычной экспрессивной препозиции ремы, с обычным для предложений введения в рассмотрение эф-

фектом **топикализации**. Только если глагол не имеет семантики введения в рассмотрение, а предложение синтаксически более сложное, ЛА-структура может трактоваться как содержащая суперпонирующую рему (пример из [2, с. 147]):

(75) <Марья Ивановна пошла ...> Вдруг белая собачка↘ английской породы | залаяла↗ и побежала ей навстречу↘ .

### 5.11. Структура с неингерентной темой

Есть еще одна К-структура с богатой коммуникативной семантикой, которую можно описать как результат ЛА-преобразования (наполненного ясным коммуникативным содержанием) некой нейтральной структуры. Это структура с **неингерентной темой**.

Среди нерасчлененных предложений обратили на себя внимание предложения, у которых главное фразовое ударение сдвинуто к началу фразы (т.е. не падает на конечное слово, как обычно):

(76) а. <Тише.> Бабушка↘ спит; б. <Откройте!> Папа↘ пришел; в. <Пустите!> Я на поезд↘ опаздываю!

В [15, с. 65 ff] непонятный левый сдвиг акцента в примерах типа (76) был интерпретирован как выражение **глобальности** высказывания – т.е. нерасчлененности, отсутствия темы. Далее в [46] было введено понятие **неингерентной темы**: утверждалось, что у этих предложений некоторая тема все-таки есть, но она не входит в состав самого предложения, а находится в предтексте.

Природа предложений с неингерентной темой была раскрыта в [2, с. 182–196], где говорится: «ЛА-контур плюс выбор акцентоносителя <...> выполняют функцию опущенных предикатов пропозициональной установки»; предложение с неингерентной темой – это «разговорный вариант подчиненной пропозиции». Иными словами, в семантической структуре предложения подразумевается пропозициональный предикат – обычно это обоснование или объяснение.

⇒Пример (с. 184):

(77) Худо, брат, жить в Париже: есть↘ нечего (А. С. Пушкин) [= ‘худо, потому что есть нечего’].

Показателен пример (78) (из [2, с. 193]). Здесь фраза *пирожка не подсунет* не может интерпретироваться как содержащая неингерентную тему. Для предложения (78) нормальна ЛА-структура (i), при которой оно утверждает просто **отличие** голода от тетки. А ЛА-структура (ii), с левым смещением фразового акцента, соответствующая неингерентной теме, абсурдна, поскольку требует, чтобы фраза понималась как **обоснование** утверждения *Голод не тетка*:

(78) Голод не тетка: пирожка не подсунет =  
i. Голод не тетка: пирожка не подсунет↘ ;  
ii. Голод не тетка: пирожка↘ не подсунет .

Обратимся к формальной структуре предложения (76а). Акцент на подлежащем *бабушка* не является здесь результатом препозиции слова с конечным рематическим акцентом; нельзя сказать и что в этой фразе «акцент сместился вперед в разговорной речи». Инициальное ударение в нерасчлененном предложении – это специальное оформление **предложений обоснования**.

Предложения с неингерентной темой являются нерасчлененными, т.е. состоят из одной ремы. Однако они отличаются от нейтральных нерасчлененных предложений типа *Пришла весна* или *Умер Гавел*, причем в двух отношениях – и формой, и коммуникативной семантикой. О семантике обоснования только что было сказано. А формальное отличие состоит в том, что в нейтральных предложениях нерасчлененность выражается порядком слов, т.е. линейно, см. преобразование «рематическая консолидация» (п. 5.9), а в неингерентных нерасчлененность выражается прежде всего акцентом: либо словом, которому естественно быть темой, получает рематический акцент (*Бабушка* √ *спит*), либо акцентированное слово передвигается в позицию перед сказуемым (*Есть* √ *нечего*, *Я на поезд* √ *опаздываю*).

Предложения с неингерентной темой не могут быть получены из обычных нерасчлененных изменением порядка слов, т.е. преобразованием «экспрессивная препозиция» темы (п. 5.1). Возьмем предложение *Спит бабушка* √ . Порядок слов в нем понимается как способ выразить контраст. Следовательно, *Бабушка* √ *спит* не есть результат препозиции подлежащего в *Спит бабушка* √ . В [2, с. 187] нерасчлененное *Бабушка* √ *спит* предлагается трактовать как полученное из расчлененного на тему и ремю *Бабушка* √ *спит* √ специальным преобразованием «неингерентная тема», которое обеспечивает предложению особого рода нерасчлененность за счет **специального выбора акцентоносителя**.

Из предложения *Папа пришел* можно построить и нейтральное нерасчлененное предложение (*Пришел папа*), и предложение с неингерентной темой (*Папа* √ *пришел*). А, например, из *Народ безмолвствует* неингерентное нерасчлененное предложение построить можно, см. (79), а нейтральное нерасчлененное – нельзя; предложение *Безмолвствует народ* √ может выражать только такой же контраст, как *Спит бабушка* √, т.е. бессмысленный.

(79) <Дело в том, что> *народ* √ *безмолвствует*.

Согласно [2], акцентоноситель в предложении с неингерентной темой выбирается по формальному правилу, так что «семантической нагрузки выбор того или иного слова не несет». Усматривается иерархия членов предложения с точки зрения их вероятности стать акцентоносителем. Ответ на вопрос «В чем проблема?» может быть: *Едим* √ *плохо*, *Дети* √ *плохо едят* и *Дети кашу* √ *плохо едят*.

Однако специальное оформление обоснования или ответа на вопрос «В чем проблема?» в форме предложения с неингерентной темой не обязательно. А поскольку маркировать семантику обоснования не обязательно, контекстное противопоставление неингерентного нерасчлененного и обычного нерасчлененного предложения может оставаться неясным. См. пример:

- (80) <Концерт отменили.>  
 а. Хазанов (∧) заболел √ .  
 б. Заболел Хазанов √ .  
 в. Хазанов √ заболел.

Разница между (80а) и (80б) очевидна: (80а) означает, что Хазанов играет главную роль в концерте, а в случае (80б) его роль может быть и второстепенной.

А разница между (80б) и (80в) не вполне ясна. Можно думать, что порядок слов, отличающий предложение с неингерентной темой от просто нерасчлененного, например (80в) от (80б), выражает некую **топикальность** именной группы, поставленной в начало предложения, – притом что, согласно ЛА-структуре, темы в обычном смысле в предложении нет. Еще пример:

(81) <Ну, слава богу, приехали. Опять ты дома. (Ласкаясь.)> *Душечка* √ *моя приехала!* [А.П. Чехов. Вишневы сад (1904)].

В этой связи представляют интерес предложения-названия и предложения-заголовки. Эти предложения тетические, т.е. нерасчлененные: подлежащее не данное. Но препозиция подлежащего не несет здесь семантики обоснования. Можно думать, *Грачи* √ *прилетели* (название картины А. Саврасова) отличается от *Прилетели грачи* √ топикальностью ИГ *грачи*, входящей в состав нерасчлененной ремы.

Итак, предложения с неингерентной темой – это специальная **акцентно оформленная** разновидность нерасчлененных предложений.

Согласно [9, с. 448]), препозиция подлежащего в тетическом предложении, свойственная газетным заголовкам (*Гринпис протестует*; *Жириновский разбушевался*), имеет целью топикализацию. Переноса акцента на подлежащее тут не требуется. В предложениях без подлежащего обычно препозитивно дополнение: *Спартак разгромили*; *Файзулину накостыляли*.

## 5.12. Множественная тема; разновидности тем

Как мы видели, в предложении может быть не одна, а две и более тем. Одна из них – как правило, первая – может быть контрастной. Так, (82а) ≠ (82б):

- (82) а. С этой √ проблемой | Иван (∧) | не сталкивался √ ; б. Иван √ | с этой проблемой | не сталкивался √ .

В предложении (83) темы переставить нельзя, поскольку здесь по смыслу контрастной может быть только одна:

- (83) Часть стихов √ | Пушкин (∧) | заменил √ .

В работе [15, с. 72] приводится следующий пример неудачного перевода из Бальзака:

- (84) В 1560 году улица Вьель-Пельтри проходила вдоль левого берега Сены.

Действительно, поскольку в этой фразе *в 1560 году* – контрастное начало, а *улица Вьель-Пельтри* тоже является темой, возникает импликатура ‘Сейчас улица Вьель-Пельтри, быть может, проходит в другом месте’, которая делает предложение аномальным. Если же изменить коммуникативную структуру так, что ИГ *улица Вьель-Пельтри* попадет в состав ремы – *В 1560 году вдоль левого берега Сены проходила улица Вьель-Пельтри*, – то импликатура, порождаемая контрастным началом, станет осмысленной: ‘Сейчас, быть может, вдоль левого берега Сены не проходит улица Вьель-Пельтри’.

В предложении (2), тоже с двумя темами, одна из них является кулисой и имеет акцент ИК-4. Не исключены и две темы с акцентом ИК-3.

Обычной, **ортотонической** теме противопоставлена **атоническая**, которая может не иметь (как *Пушкин* в примере (83)) или не может иметь фразового акцента. Частный случай – **рецессивная** тема [2, с. 21]. Этот термин применяется в п. 5.2 к предложениям с дислокацией ремы и специфическим акцентом ИК-6:

(85) Посади́л → *ded* ре́пку ↘ ⇐ *Дед* → посади́л ре́пку ↘ .

Здесь тема исходного предложения *Дед посадил ре́пку* специально подавлена – путем дислокации ремы и помещения подлежащего в ваккернагелевскую позицию, после первого ударного слова.

Согласно [2], в предложении (43), с тематической инверсией, тоже есть атоническая тема, которая занимает «ваккернагелевскую позицию», т.е. – **энклитическая** (помимо обычной темы-начала, которой является глагол).

Во фразе (86а) подлежащее *строители* занимает ваккернагелевскую позицию. Однако темой, т.е. отдельным компонентом К-структуры, оно здесь не является. Фраза (86а) членится на тему *живут строители* и рему *по всей трассе канала*. Она получается из фразы (86б) преобразованием «левый сдвиг акцентоносителя в теме», которое характеризует предложения, где актуальное членение противоречит синтаксическому, см. [47]:

(86) а. Живу́т → строите́ли | по всей трассе канала ↘ ;  
б. Строи́тели живу́т → | по всей трассе канала ↘ ;  
в. Строи́тели → | живу́т по всей трассе канала ↘ .

Нейтральным коммуникативным вариантом для (86б) является фраза (86в), где соответствие между актуальным и синтаксическим членением исходное – группа подлежащего – тема, группа сказуемого – рема. Отличие (86в) от (86б) касается коммуникативной структуры – в (86в) речь идет о строителях, а в (86б) – о том, где строители живут. Но структура (86б) почти обязательно подвергается левому сдвигу акцентоносителя, что дает (86а).

Другие примеры, где слово попадает в ваккернагелевскую позицию как бы случайно – в результате левого сдвига акцентоносителя в неканонической теме (т.е. такой, где актуальное членение не соответствует синтаксическому):

(87) <В 1948 году Ремарк вернулся в Швейцарию, где и прожил остаток жизни.> Умер → *Ремарк* | 25 сентября 1970 года ↘ [умер *Ремарк* – неканоническая тема, 25 сентября 1970 года – рема] ⇐ *Ремарк* умер → | 25 сентября 1970 года ↘ .

(88) Выгляди́т → *она* | превосходно ↘  
⇐ *Она* выгляди́т → | превосходно ↘ .

(89) Е́сть → *хотелось* | ужасно ↘  
⇐ *Хотелось* е́сть → | ужасно ↘  
⇐ Ужасно *хотелось* е́сть ↘ .

Левый сдвиг может затрагивать не слово, а составляющую. Так, в (90) слово *отец* оказывается в ваккернагелевской позиции, потому что левому сдвигу подвергается акцентонесущая составляющая неканонической темы *отец покупал их*:

(90) <...били в столовой черные стенные башенным боем.> Покупал → *их отец* | давно ↘ (М. Булгаков. Белая Гвардия) ⇐ *Отец* покупал → *их* | давно ↘ .

Предложение с канонической темой – *Отец* → | *покупал их давно* ↘ .

## 6. СЕМАНТИКА КОММУНИКАТИВНЫХ ПРОТИВОПОСТАВЛЕНИЙ

Этот раздел посвящен семантике коммуникативного членения. Обсуждается связь понятий темы/ремы с понятиями данного/нового (п. 6.1) и презумпции/ассерции (п. 6.2). Рассматривается несколько частных видов К-структур и возможности их семантической интерпретации (п. 6.3 и п. 6.4).

### 6.1. Тема и данное, рема и новое; метод вопросов

Есть разные подходы к семантике коммуникативного членения. Изначальным в теории актуального членения было представление о том, что тема выражает данное (т.е. то, что говорящий предполагает известным адресату), а рема – новое (т.е. неизвестное адресату). В [48] приводится, однако, пример, который показывает, что рема не всегда новое, а тема не всегда данное. Рассматривается следующий текст:

(91) В английском языке падежи сохранились только в системе местоимений. *Во французском языке* [тема, новое] наблюдается *такая же картина* [рема, данное].

Тема второго предложения является новым, потому что французский язык впервые вводится в рассмотрение. Рема второго предложения является данным, потому что выражение *такая же картина* представляет собой анафорическую отсылку к старому, т.е. к уже введенному в сознание адресата знанию о сохранении падежей лишь в системе местоимений.

Примеры (92), (93) (из [1, с. 114]) тоже показывают, что рема может быть данным, а тема – новым.

(92) Состоялся матч чехов с канадцами. Выиграли *чехи* ↘ .

(93) Голодная волчиха → встала, чтобы идти на охоту ↘ (А.П. Чехов «Белолобый»).

Другое дело, что в (91) достаточно изменить порядок слов, и мы получим «прямое» соотношение: тема = данное, рема = новое:

(94) *Такая же картина* [тема, данное] наблюдается (и) *во французском языке* [рема, новое].

Имеются частицы *также* и *тоже*, которые можно назвать «коммуникативно обратными» [49] – *также* (и) безударно и маркирует рему как новое, а *тоже* ударно и маркирует рему как старое; отсюда квази-синонимия (95а) и (95б):

(95) а. Гость молчал. Молчал [старое] *также* и *хозяин* [рема – новое].

б. Гость молчал. Хозяин [новое] *тоже* *молчал* [рема – старое].

О специальной коммуникативной функции *тоже* см. [50]. Итак, интерпретацию К-структуры с обращением к понятиям данное/новое приходится отклонить.

Один из подходов к описанию семантики коммуникативного членения – так называемый **метод во-**

**просов**, см. [51]. Предложение *Иван поехал в Японию* отвечает на вопрос «Куда поехал Иван?», а *В Японию поехал Иван* – на вопрос «Кто поехал в Японию?»

Нерасчлененное предложение может быть представлено как ответ на вопрос «Что произошло?». Предложение (96б) – это ответ на вопрос «Маша пришла?», а (96а) – на вопрос «Пришла ли Маша?».

(96) а. Маша<sup>↗</sup> не пришла<sup>↘</sup>; б. Маша не пришла<sup>↘</sup>.

Метод вопросов – это способ выявления ремы, но не способ экспликации коммуникативных значений.

## 6.2. Коммуникативная структура и отрицание

Важную роль в семантической интерпретации коммуникативных различий играет понятие **презумпции**. Коммуникативные различия, выражаемые ЛА-структурой, могут быть представлены с помощью оппозиции «презумпция vs. ассерция». Рематическая часть – это то, что утверждается, ассерция; тематическая – то, что заранее принято как презумпция. Следует иметь в виду, что ассерция и презумпция – это пропозитивные компоненты смысла, пропозиции, а тема и рема – это части предложения.

Будучи связана с оппозицией «презумпция vs. ассерция», коммуникативная структура предложения определяет сферу действия **отрицания**. Так, в (97а), с нейтральной К-структурой, отрицание, выражаемое частицей *не*, которая стоит при глаголе, распространится на все предложение целиком. А в (97б), которое получено в результате тема-рематической инверсии (см. п. 5.4), частица *не* будет стоять при рематическом имени: презумпция ‘есть кто-то, кто любит Марию’ останется вне сферы действия отрицания.

(97) а. Иван<sup>↗</sup> | любит Марию<sup>↘</sup> [нейтральная К-структура] – Иван *не* любит Марию.

б. Марию<sup>↗</sup> любит | Иван<sup>↘</sup> [рема – *Иван*] – Марию любит *не* Иван.

Если некое слово в предложении несет контрастное ударение, то соответствующее ему общеотрицательное предложение имеет присловное отрицание. Так, у предложения (98а), с естественным делением на тему и ремю, смысл описывается единой ассерцией, и оно имеет предикатное отрицание (98б) (о предикатном и присловном отрицании см. [52]). А в предложении (99а) есть контрастное ударение, которое производит расчленение смысла на две пропозиции – ‘Иван куда-то пошел’ [презумпция] и ‘то, куда он пошел, – это концерт’ [ассерция]. Поэтому общеотрицательным для (99а) будет предложение (99б) с присловным отрицанием.

(98) а. Иван<sup>↗</sup> *пошел* на концерт<sup>↘</sup>; б. Иван<sup>↗</sup> *не пошел* на концерт<sup>↘</sup>.

(99) а. Иван пошел *на концерт*<sup>↘</sup>; б. Иван пошел *не на концерт*<sup>↘</sup>.

Сферой действия контраста может быть не только слово, но и составляющая, в том числе – предикация:

(100) Иван не стихи *пишет*, а землю *пашет*.

Для коммуникативной семантики важна не только семантическая, но и **прагматическая презумпция** [53]. Прагматическая презумпция – это суждение, которое, чтобы быть приемлемым в диалоге, должно

быть не только истинно, но и известно адресату. Роль прагматической презумпции видна на следующем примере (ср. [54]):

(101) а. Бернштейн не знал<sup>↗</sup>, что Володя женат<sup>↘</sup>. [С. Спивакова. Не всё (2002)].

б. Бернштейн не знал<sup>↘</sup>, что Володя женат —.

Предложение (101б) обращено к адресату, которому информация ‘Володя женат’ известна. Тогда как фраза (101а), с интонацией незавершенности, **вводит** эту информацию в контекст дискурса. В [55] на примере английского *only* ‘только’ было показано, что семантическая презумпция слова не обязана быть известна адресату, т.е. может не быть прагматической. О том, что семантическая презумпция *только* не является прагматической, см. [53].

## 6.3. Коммуникативная структура предложений с адвербиалами

Оппозиция «презумпция – ассерция» позволяет описать коммуникативные противопоставления в предложениях с адвербиалами. Пример (из [20]):

(102) а. Он затормозил *резко*<sup>↘</sup>;

б. Он *резко* затормозил<sup>↘</sup>.

Ударное и безударное *резко* в (102) играют разную роль в коммуникативной структуре предложения. Оппозиция «презумпция – ассерция» позволяет следующим образом представить это различие. В (102а) у *резко* ударная позиция, рематическая. А в (102б) – безударная: *резко* в (102б) – это **рематический модификатор**. Безударный рематический модификатор выражает дополнительную ассерцию, см. (103ii).

(103) i = Он затормозил [тема, пресуппозиция] & торможение резкое [рема, ассерция];

ii = Он затормозил [ассерция] & торможение резкое [дополнительная ассерция].

Структуры (103i), (103ii) позволяют объяснить, почему (102а) и (102б) по-разному взаимодействуют с отрицанием. У (102а), где *резко* – ассерция, есть простое отрицание:

(104) НЕ (Он затормозил *резко*<sup>↘</sup>) = Он затормозил *не резко*.

А в предложении (102б), где *резко* – рематический модификатор, две ассерции (т.е. конъюнкция ассерций), и потому отрицание затруднено (см. [20, с. 124 ff; 56]):

(105) НЕ (Он *резко* затормозил<sup>↘</sup>) = ‘то ли он затормозил не резко, то ли не затормозил вообще’.

Позиция рематического модификатора возможна не для всех адвербиалов, и изменение порядка слов может давать совершенно иной эффект. Так, предложение (106а), с постпозицией адвербиала *на почту*, имеет неоднозначную К-структуру и двоякое отрицание, см. (106б) и (106в); а препозиция адвербиала, т.е. перенос глагола в финальное положение, дает однозначный акцент на глаголе и отрицание при глаголе, см. (107).

(106) а. Вася пошел *на почту*<sup>↘</sup>.

б. Вася *не* пошел<sup>↘</sup> *на почту*.

в. Вася пошел *не на почту*<sup>↘</sup>.

- (107) а. *Вася на почту пошел* \ .  
 б. *Вася на почту не пошел* \ .

#### 6.4. Частные семантические эффекты сдвига акцента: уменьшительное приращение

Почему *не весит 50 кг* в примере (108) означает ‘весит меньше’? Откуда это «уменьшительное приращение»?

(108) Мешок *не весит* 50 кг [= ‘весит меньше 50 кг’].

Ю.Д. Апресян [57, с. 81–82] признает здесь особое – идиоматическое – взаимодействие глагола с отрицанием. И.М. Богуславский [41, с. 27–29] отвергает идиоматическое отрицание и предлагает следующее объяснение примеру (108). Обычно исходят из того, что (108) является отрицанием (109); между тем, на самом деле, исходным утвердительным для (108) является не (109), а (110):

(109) Мешок *весит* 50 кг \ [= ‘весит ровно 50 кг’].

(110) Мешок *весит* \ 50 кг [= ‘весит 50 кг или больше’].

Естественным – в полном соответствии с логикой – отрицанием для ‘мешок весит 50 кг или больше’ будет ‘мешок весит меньше 50 кг’. Так что с отрицанием в примере (108) все в порядке. Остается объяснить, почему сдвиг фразового ударения, превращающий предложение (109) в (110), приводит к столь существенному изменению смысла.

Просодическая конструкция «**количественная контрастная тема**» – это вариант контрастной темы, осложненный идеей количества в именной группе и модальностью – в предикатной. Увеличительные и уменьшительные приращения, возникающие в результате сдвига ударения на глагол, объяснимы из семантики контрастной темы, которая вводит в рассмотрение ассоциированное множество, см. п. 5.6 и подробнее – в [40].

В теории коммуникативной структуры остаются нерешенные проблемы. Так, перестройки линейно-акцентной структуры, описанные в разделе 5, сопровождаются предсказуемым коммуникативным эффектом лишь при условии, что именные группы, меняющие свою линейную позицию, являются конкретно-референтными. Если речь идет о предложениях с неопределенными, а тем более – с квантифицированными ИГ, то изменение порядка слов меняет сферу действия кванторов в предложении, так что изменяется не только коммуникативная семантика, но и собственно смысл, а часто и истинностное значение предложения. Линейный аспект ЛА-структуры (т.е. порядок слов) требует отдельного рассмотрения.

#### СПИСОК ЛИТЕРАТУРЫ

1. Падучева Е.В. Высказывание и его соотносительность с действительностью. – М.: Наука, 1985.
2. Янко Т.Е. Коммуникативные стратегии русской речи. – М.: ЯСК, 2001.
3. Kuno S. Functional sentence perspective. A case study from Japanese and English // Linguistic inquiry. – 1972. – Vol. 3(3). – P. 260–320.
4. Dahl Ö. Topic and comment. A study in Russian and general transformational grammar // Slavica Gothoburgensia. – 1969. – Vol. 4.
5. Ковтунова И.И. Современный русский язык: Порядок слов и актуальное членение предложения. – М.: Просвещение, 1976.
6. Адамец П. Порядок слов в современном русском языке. – Praha, 1966.
7. Щеглов Ю.К. К понятиям логических субъекта и предиката // Машинный перевод и прикладная лингвистика, 8. – М., 1964. – С. 109–124.
8. Пауль Г. Принципы истории языка / пер. с нем. – М.: Издательство иностранной литературы, 1960.
9. Тестелец Я.Г. Введение в общий синтаксис. – М.: РГГУ, 2001.
10. Selkirk E.O. Phonology and Syntax: the relation between sound and structure. – MIT Press, 1984.
11. Selkirk E.O. Sentence Prosody: Intonation, stress and phrasing // The Handbook of Phonological Theory Blackwell / ed. J. Goldsmith. – London, 1996.
12. Кодзасов С.В., Кривнова О.Ф. Общая фонетика. – М.: РГГУ, 2001.
13. Митренина О.В., Романова Е.Е., Слюсарь Н.А. Введение в генеративную грамматику. – М.: УРСС, 2011.
14. Циммерлинг А.В. Локальные и глобальные правила в синтаксисе // Труды международной конференции Диалог. – 2008. – С. 551–562.
15. Николаева Т.М. Семантика акцентного выделения. – М.: Наука, 1982.
16. Арутюнова Н.Д., Ширяев Е.Н. Русское предложение: Бытийный тип. – М.: Рус. яз., 1983.
17. Стросон П.Ф. О референции // Новое в зарубежной лингвистике, XIII. – М., 1982. С. 55–86.
18. Янко Т.Е. Интонационные стратегии русской речи в сопоставительном аспекте. – М.: ЯСК, 2008.
19. Грамматика 1980 – Русская грамматика. Т. 1–2 / ред. Н.Ю. Шведова. – М.: Наука, 1980.
20. Падучева Е.В. Динамические модели в семантике лексики. – М.: Языки славянской культуры, 2004. – URL: <http://lexicograph.ruslang.ru/TextPdf1/PaduDinamMod2004.pdf>
21. Янко Т.Е. Категории темы, ремы и дискурсивной незавершенности в звучащей речи // Глагольные и именные категории в системе функциональной грамматики. – СПб.: Нестор-История, ИЛИ РАН, 2013. – С. 352–360.
22. Trubetzkoy N.S. Grundzüge der Phonologie (Travaux du Cercle Linguistique de Prague, 7). – Praga, 1939; Трубецкой Н.С. Основы фонологии. – М.: Изд-во иностранной литературы, 1960).
23. Mel'čuk I. Word order in Russian // Слово и язык. Сборник статей к восьмидесятилетию академика Ю.Д. Апресяна. – М.: ЯСК, 2011. – С. 457–482.
24. Paducheva E. Subject-predicate inversion and its cognitive sources // Russian linguistics. – 2010. – Vol. 34. – P. 113–121.

25. Падучева Е.В. К интонационной транскрипции для предложений произвольной синтаксической сложности // Вопросы кибернетики. Семиотические исследования. – М.: АН СССР, 1989. – С. 18–29 (Научный совет по комплексной проблеме «Кибернетика»). (Перепечатано в: Падучева Е.В. Статьи разных лет. – М.: ЯСК, 2009. – С. 259–271).
26. Русский семантический словарь. Т. IV. Глагол / ред. Н.Ю. Шведова. – М.: Азбуковник, 2007.
27. Ицкович В.А. Очерки синтаксической нормы. – М.: Наука, 1982.
28. Арутюнова Н.Д. Предложение и его смысл. – М.: Наука, 1976.
29. Babby L.H. Existential Sentences and Negation in Russian. – Ann Arbor: Caroma Publishers, 1980.
30. Borschev V., Partee B.H. The Russian Genitive of negation in existential sentences // Prague linguistic circle papers. – 2002. – Vol. 4. – P. 185–250.
31. Sgall P., Hajičová E. Focus on focus // Prague bulletin of mathematical linguistics. – 1977. – Vol. 28. – P. 5–34.
32. Chomsky N. Questions of Form and Interpretation // Linguistic Analysis. – 1975. – Vol. 1. – P. 75–109.
33. Restan P. The position of the finite verb in some elementary declarative sentences in modern Russian // The Slavic Verb, Copenhagen. 1981. P. 147–160 (Русский перевод в: «Новое в зарубежной лингвистике», 15. 1985).
34. Падучева Е.В. Коммуникативная расчлененность и пути ее преодоления: инверсия подлежащего // Фонетика и нефонетика. К 70-летию С.В. Коздасова. – М.: ЯСК, 2008. – С. 417–426.
35. Rooth M. A theory of focus interpretation // Natural Language Semantics. – 1992. – Vol. 1. – P. 75–116.
36. Богуславский И.М. Сфера действия лексических единиц. – М.: ЯРК, 1996.
37. Богуславский И.М. Сфера действия начальности и актуальное членение: втягивание ремы // Семиотика и информатика. – 1998. – № 36. С. 8–18.
38. Апресян Ю.Д. Исследования по семантике и лексикографии. Т. I. Парадигматика. – М.: ЯСК, 2009.
39. Lyons J. Semantics. Vol. 1–2. L. etc.: – Cambridge Univ. Press, 1977.
40. Падучева Е.В. Семантика количества и ее отражение в просодии. // Логический анализ языка. Квантификативный аспект языка. – М.: Индрик, 2005. – С. 652–667.
41. Богуславский И.М. Исследования по синтаксической семантике. – М.: Наука, 1985.
42. Филипенко М.В. Об однородных и неоднородных ассоциированных множествах // Научно-техническая информация. Сер. 2. – 1992. – № 5. – С. 29–32.
43. Арутюнова Н.Д. Типы языковых значений: Оценка. Событие. Факт. – М.: Наука, 1988.
44. Адамец П. Порядок слов в современном русском языке. – Прага, 1966.
45. Падучева Е.В. Об атрибутивном стяжении подчиненной предикации в русском языке // Машинный перевод и прикладная лингвистика. – 1980. – № 20. – С. 3–44. (Перепечатано в: Падучева Е.В. Статьи разных лет. – М.: ЯСК, 2009. – С. 145–172).
46. Баранов А.Н., Кобозева И.М. Семантика общих вопросов в русском языке (категория установки) // Изв. АН СССР. Сер. лит. и языка. – 1983. – 42(7).
47. Падучева Е.В. Порядок слов и фразовая просодия // Смыслы, тексты и другие захватывающие сюжеты. Сборник статей в честь И.А. Мельчука. – М. ЯСК, 2012. – С. 473–485.
48. Апресян Ю.Д. Типы коммуникативной информации для толкового словаря // Язык: система и функционирование. – М.: Наука, 1988. – С. 10–22.
49. Падучева Е.В. *Тоже и также*: взаимодействие актуального членения и ассоциативных связей // Ин-т русского языка. Предварительные публикации, 55. – М.: ИРЯ АН СССР, 1974. (Перепечатано в: Падучева Е.В. Статьи разных лет. – М.: ЯСК, 2009. – С. 150–258).
50. Богуславский А. К вопросу о вторичном обозначении определенного содержания в русском связном тексте // Научн. докл. высш. шк. Филол. науки. – 1969. – № 6.
51. Hajičová E. Topic/focus and related research // Prague school of structural and functional linguistics. – Amsterdam/Philadelphia, 1994.
52. Падучева Е.В. Русское отрицательное предложение. – М.: ЯСК, 2013.
53. Падучева Е.В. О производных диатезах отпредикатных имен в русском языке // Проблемы лингвистической типологии и структуры языка / отв. ред. Храковский В.С. – Л.: Наука, 1977. – С. 84–107.
54. Schmerling S. F. Presupposition and the notion of normal stress // Papers from the VII-th meeting of Chicago linguistic society. – Chicago, 1971.
55. Roberts C. Only, presupposition and implicature // Journal of Semantics. 2006. – URL: <http://ling.osu.edu/~croberts/only.pdf>.
56. Horn L.R. A natural history of negation. – Chicago: Univ. of Chicago Press, 1989.
57. Апресян Ю.Д. Лексическая семантика: Синонимические средства языка. – М.: Наука, 1974.

*Материал поступил в редакцию 27.07.15.*

#### **Сведения об авторе**

**ПАДУЧЕВА Елена Викторовна** – доктор филологических наук, профессор, главный научный сотрудник ВИНТИ РАН, Москва  
e-mail: elena.paducheva@yandex.ru

П.Л. Гроховский, В.П. Захаров, М.О. Смирнова, М.В. Хохлова

## Корпус памятников тибетской грамматической традиции\*

*Описывается создание параллельного тибетско-русского корпуса памятников тибетской грамматической традиции, сформировавшейся в VII–VIII вв. н.э. На основе корпуса формируется специальная лексическая база грамматической терминологии, представляющая интерес как для тибетологов, так и для специалистов по общему языкознанию. Корпус может быть использован для лингвистических исследований, преподавания и изучения классического тибетского языка, а также тибетской грамматической традиции.*

**Ключевые слова:** корпусная лингвистика, параллельный корпус, грамматическая разметка, лексико-грамматическая база данных, тибетская грамматическая терминология, тибетская грамматическая традиция.

### ВВЕДЕНИЕ

В настоящей статье описывается создание параллельного тибетско-русского корпуса памятников тибетской грамматической традиции, которая, предположительно, начала свое формирование в VII–VIII вв. н.э., когда были созданы первые дошедшие до нас грамматики «Сумчупа» и «Тагкичжугпа». Эта традиция во многом основывается на буддийских грамматиках, составленных индийскими учеными, и таким образом восходит к индийской грамматической традиции. По методам описания и анализа явлений языка эти грамматики значительно отличаются от западного языкознания. Современные тибетские лингвисты продолжают поддерживать и развивать традицию классического тибетского языкознания.

В состав корпуса входят сочинения, которые наиболее ценятся в тибетской грамматической традиции: 1) два первых трактата «Сумчупа» («Тридцать строф», тиб. *sum su ra*, предположительно VII–VIII вв. н.э.) и «Тагкичжугпа» («Использование знаков», тиб. *rtags kyi 'jug ra*, предположительно VII–VIII вв. н.э.), авторство которых традиционно приписывается создателю тибетской письменности Тхонми Самбхоте (тиб. *thon mi sam b+hoṭa*), 2) грамматики Смтритиджнякирти (санскр. *smṛti-jñānakīrti*) «Врата речи, [подобные] мечу» (тиб. *smra sgo mtshon cha*, XI в. н.э.), 3) комментарий Ситу Махапандиты (*si tu mahA paN+Di ta*) к двум первым трактатам «Прекрасный жемчужный венок – ожере-

лье мудреца» (тиб. *mkhas pa'i mgul rgyan mu tig phreng mdzes*, XVIII в.), 4) комментарий Нгулчу Дхармабахдры (тиб. *dngul chu dharma b+ha dra*) к двум первым трактатам «Устные наставления по сочинению великого ученого Ситу» (тиб. *mkhas mchog si tu'i zhal lung*, XIX в.), 5) комментарий неизвестного автора к двум первым трактатам под названием «Драгоценный венок благих изречений» (тиб. *sum rtags gzhung mchan legs bshad nor bu'i phreng ba*, XVIII/XIX вв.), 6) тибетская грамматика Келзанг Гьюрме (тиб. *skal bzang 'gyur med*) «Ясное зеркало – введение в тибетскую грамматику» (тиб. *bod kyi brda sprod rig pa'i khrid rgyun rab gsal me long*, XX в.).

### КОРПУСЫ ТЕКСТОВ ТИБЕТСКОГО ЯЗЫКА

Несмотря на то, что над автоматической обработкой тибетских текстов трудятся ученые в разных странах (Германия, Великобритания, КНР, США, Япония), до сих пор не выработан единый стандарт корпусного представления тибетского языкового материала.

На последних пяти конференциях Международной тибетологической ассоциации (IATS Seminar) проводилась секция «Тибетские информационные технологии» (Tibetan Information Technology), на которой были представлены проекты применения компьютерных технологий, в том числе и проекты создания и использования корпусов тибетского языка.

Создание подобных корпусов за рубежом только начинается. В Тюбингенском университете им. Эберхарда Карла (Германия) в рамках научно-исследовательской темы 441 под руководством Б. Цайслера в 2002–2008 гг. был реализован проект под номером B11 «Семантические и падежные отношения и межклаузная референция в тибетском языке» [1]. В результате был создан небольшой тибетский корпус, состоящий из четырех текстов – два

\* Корпус разработан в рамках научно-исследовательского проекта РФФИ «Пилотная версия электронного корпуса тибетских грамматических сочинений» (13-06-00621а). Исследование памятников тибетской грамматической традиции выполнено при поддержке гранта СПбГУ 2.38.293.2014 «Тибетская письменная традиция и современность».

текста на доклассическом тибетском и по одному тексту на классическом тибетском языке и ладакхском диалекте, на котором говорят тибетцы, живущие в Индии и Пакистане. В этом корпусе были реализованы частеречная, синтаксическая и семантическая разметки. На сайте проекта 441 как минимум до апреля 2012 г. были доступны фрагменты этого небольшого корпуса, в котором все тексты были также снабжены переводом на английский язык и указанием на источник (оригинальный текст) [2]. Основная цель, для которой создавался данный корпус, это заполнение «пробелов» в тибетской грамматике и, в частности, в семантике глаголов, которая изучалась на основе текстов.

В 2012 г. Н. Хилл и У. Пагель из Школы востоковедения и африканистики при Лондонском университете начали разработку тибетского корпуса объемом 1 млн слогов, тексты которого относятся к трем периодам истории тибетского языка – доклассическому, классическому и современному [3].

Отличие описываемого здесь проекта от вышеперечисленных заключается в том, что в его рамках разработана оригинальная система лингвистической разметки, базирующаяся на теоретическом описании классического тибетского языка [4, 5]. Анализ опубликованных материалов показывает значительные расхождения исследователей в подходах к лингвистическому анализу грамматических категорий классического тибетского языка. Второе отличие связано с характером материалов корпуса – это тексты, посвященные одной из традиционных наук тибетцев – лингвистике. Третья особенность данного проекта – это параллельный тибетско-русский корпус, ориентированный в первую очередь на российских исследователей.

## СТРУКТУРА КОРПУСА

Проект направлен на решение двух задач: построение параллельного корпуса тибетских грамматических сочинений с русским переводом и создание на его основе лексической базы данных грамматиче-

ской терминологии с частотными характеристиками и семантическими отношениями.

Технологический процесс создания корпуса можно представить в виде этапов, главными из которых являются токенизация, т.е. разделение входного текста на составные элементы (словоформы, знаки препинания, цифры и т. п.), разметка текста, загрузка размеченных текстов в структуру корпусного менеджера.

Тибетские тексты и русские переводы в корпусе выровнены по границам предложений тибетской части (см. пример из трактата «Сумчупа» в табл. 1).

В тибетском тексте размечены границы словоформ, что само по себе является нетривиальной задачей, так как традиционная орфография маркирует только границы слогов, лишь в 95 случаях из 100 совпадающие с границами морфем. Существуют различные программы для автоматического выравнивания, например, Hunalign, Vanilla, Abbyu Aligner и др. Однако их настройка для тибетского языка не проста и представляет собой отдельную задачу, решение которой на данный момент отложено. Сравнительно небольшой объем корпуса и необходимость разметки границ словоформ, требующей просмотра всего текста корпуса, делают ручное выравнивание трудоемкой, но выполнимой процедурой.

В настоящий момент тибетская часть корпуса насчитывает около 30 тыс. токенов (слов). Корпус размещен на корпусном портале Санкт-Петербургского университета по адресу <http://corpora.spbu.ru/bonito/index.html>.

## РАЗМЕТКА КОРПУСА

Тибетский текст подвергается морфологической разметке (частеречная разметка, грамматическая аннотация глагольных форм, лемматизация, снятие грамматической омонимии). Была разработана система морфологических тегов для тибетского языка (см. *Приложения 1, 2*). В качестве иллюстрации приведем список тегов для обозначения наиболее употребительных служебных лексем (табл. 2).

Таблица 1

Выравнивание текстов корпуса по предложениям

Текст на тибетском языке	Латинская транслитерация	Русский перевод
རྗེས་འཇུག་ཡི་གེ་བཅུ་པོ་ནི། མིང་གང་གི་ནི་མཐར་ཐུང་བ། དེ་ལ་སྐྱེ་ལོ་བཞི་པ་རྒྱུ་རྒྱུ།	rjes 'jug yi ge bcu po ni// ming gang gi ni mthar sbyar ba // de la A li bzhi pa sbyar//	Что касается десяти конечных графем, то к [ним], добавленным после какого-либо слова, добавляют четвертую гласную.
སྐྱེ་བཅུ་བར་ནི་ཤེས་པར་བྱ།	slar bsdu bar ni shes par bya//	[Это] известно как обобщение.
དང་པོ་གཉིས་ལ་དང་པོ་འཇུག།	dang po gnyis la dang po 'thun//	Первой и второй [конечным графемам] соответствует первая.
གསུམ་པ་བཅུ་ལ་གྲ་དང་རྒྱུ།	gsum lnga bcu la kya dang sbyar//	К третьей, пятой и десятой добавляют куа.
བདུན་པ་ཉིད་ལ་བདུན་པ་སྟེ། ལྷག་མ་རྣམས་ལ་གྲ་རྒྱུ་རྒྱུ། དེ་དག་མི་རྒྱུ་འབྲེལ་བའི་གྲ།	bdun pa nyid la bdun pa ste// lhag ma rnam la gya sbyar ba// de dag i sbyar 'brel pa'i sa//	К седьмой добавляют ту же самую седьмую, К остальным – гуа, И [все] они, с добавлением i – это показатель связи.

Список тегов для обозначения служебных лексем в тибетском языке

Тег	Служебная лексема	Пример
Cj	Союз	dang
Pp	Послелог	drung du
Erg	Показатель эргатива	алломорфы kyis, gyis, gis, s, yis
Com	Показатель комитатива	dang
Dat	Показатель датива	la
Loc	Показатель локатива	na
Dest	Показатель дестинатива	алломорфы tu, du, ra, ru, su
Abl	Показатель аблатива	las
El	Показатель элатива	nas
Comp	Показатель компаратива	алломорфы pas, bas
Gen	Показатель генитива	алломорфы kyī, gyī, gi, 'i, yī
Fin	Конечная частица	алломорфы go, ngo, do, no, bo, mo, 'o, ro, lo, so, to
Top	Выделительная частица	ni
Ind	Неопределенная частица	алломорфы cig, zhig, shig
Emph	Усилительная частица	алломорфы kyang, yang, 'ang
Quant	Слова, выражающие количественные значения ('столько', 'именно' и т.п.)	tsam, kho na, 'ba' zhig, snyed
Pl	Показатель множественного числа	gnams
Quot	Показатель прямой речи	алломорфы ces, zhes

Текстам корпуса также приписываются элементы метаразметки, которые включают информацию о жанре, датировке, авторе текста, его принадлежности к конкретной буддийской школе, например:

● *grammar 1* <file author="Тхонми Самбхота" date="VII-IX века"

*title="sum cu pa" genre="sgra'i rig pa" form="cmixu" >*

● *grammar 2* <file author="Тхонми Самбхота" date="VII-IX век"

*title="rtags kyi 'jug pa" genre="sgra'i rig pa" form="cmixu">*

● *grammar 3* <file author="Смтрпитиджняна-куртму" date="XI век"

*title="smra sgo mtshon cha" genre="sgra'i rig pa" form="cmixu">*

● *commentary 1* <file author="Суму Маханандума" date="XVIII век"

*title="mkhas pa'i mgul rgyan mu tig phreng mdzes" genre="sgra'i rig pa" form="нпроза" >*

● *commentary 2* <file author="Нзулчу Дхармабхадра" date="XIX век"

*title="mkhas mchog si tu'i zhal lung" genre="sgra'i rig pa" form="нпроза" >*

● *commentary 3* <file author="не известен" date="XVIII/XIX век"

*title="sum rtags gzhung mchan legs bshad nor bu'i phreng ba" genre="sgra'i rig pa" form="нпроза">*

● *modern 1* <file author="Келзанг Гьюрме" date="XX век"

*title="bod kyi brda sprod rig pa'i khrid rgyun rab gsal me long" genre="sgra'i rig pa" form="нпроза">*

Морфологическая разметка осуществлялась вручную путем приписывания каждому слову следующей

информации: словоформа в тибетском написании, словоформа в латинской транслитерации, лемма (лексема) в тибетском написании, лемма в латинской транслитерации, тег частеречной разметки, терминологический тег.

В дальнейшем в целях автоматизации морфологической разметки планируется адаптировать для тибетского языка статистический теггер TreeTagger [6]. Корпус, размеченный к настоящему времени вручную, будет использован как тренировочный. Такой прецедент имеется. TreeTagger был использован для разметки корпуса языка дзонгка (тиб. rdzong kha), близкого к тибетскому. Настройка теггера осуществлялась на размеченном вручную корпусе объемом 40 тысяч токенов, оценка результатов работы показала точность выше 90% [7]. Размеченный таким способом корпус нуждается в ручной корректировке, однако на большом корпусе мы получаем существенный выигрыв в трудозатратах. Терминологическая разметка, однако, и в этом случае останется прерогативой специалиста.

Данные разметки представляются в том или ином структурированном виде. Чаще всего используется или линейный формат на языках XML и SGML, или, как в нашем случае, так называемый вертикальный формат, фактически таблица, где каждому токену текста соответствует одна строка, а в каждом столбце (ячейке) записывается определенный морфосинтаксический признак или группа признаков (см. табл. 3). Эта табличная форма может быть представлена в неявном виде, когда ячейки таблицы разделены каким-либо разделителем, обычно знаком табуляции. Далее размеченные тексты загружаются в базу данных корпусного менеджера.

## Фрагмент размеченного текста

Словоформа в тибетской графике	Словоформа в латинской транслитерации	Лемма в тибетской графике	Лемма в латинской транслитерации	Грамматический тег	Терминологический тег
<s>					
<align>					
ད	de	ད	de	P	
ཉི	ni	ཉི	ni	Top	
སུད	sdud	སུད	sdud	VN	Gram L TGrMark
དང	dang	དང	dang	Cj	
འབྱེད་པ་	'byed pa	འབྱེད་པ་	byed	VN	Gram L TGrMark
དང	dang	དང	dang	Cj	
//	//	//	//	Punct	
རྒྱ་མཚན་	rgyu mtshan	རྒྱ་མཚན་	rgyu mtshan	N	Gram GenLex TGrMark
ཚེ་སྐབས་	tshe skabs	ཚེ་སྐབས་	tshe skabs	N	Gram L TGrMark
གདམས་ངག་	gdams ngag	གདམས་ངག་	gdams ngag	N	Gram GenLex TGrMark
ལྷ	lnga	ལྷ	lnga	Num	
འོ	'o	འོ	'o	Fin	
//	//	//	//	Punct	
</align>					
</s>					

## Примечания.

1. <s> и <align> – теги структурной разметки.
2. Теги частеречной разметки см. в *Приложениях 1, 2*.
3. Терминологические теги см. в табл. 5.

## ПОИСК В КОРПУСЕ

Для работы с корпусом используется один из наиболее эффективных корпусных менеджеров Manatee/Vonito. В настоящее время он получил название NoSketch Engine (<http://nlp.fi.muni.cz/trac/noske>). Система состоит из двух частей: загрузчик (Manatee) и графический пользовательский интерфейс, работающий на стороне клиента (Vonito). Программа позволяет загружать предварительно размеченные тексты и строить на их основе корпуса (включая параллельные). Существует расширенная (коммерческая) версия этой системы, которая называется Sketch Engine (<http://www.sketchengine.co.uk>). Она позволяет решать и более сложные задачи, такие как выявление ключевых слов и терминов, построение лексико-семантических групп и др.

Работа с корпусными данными в NoSketch Engine возможна в режимах выдачи конкорданса для заданных элементов корпуса, выдачи коллокаций, составления частотных списков. Конкорданс показывает контексты употреблений для заданного слова или словосочетания. В режиме частотных списков можно получать частотные словники раз-

ных элементов корпуса: лемм, словоформ, грамматических или иных тегов.

В размеченных текстах поиск может осуществляться для разных единиц: словоформ или лемм в тибетской графике, словоформ или лемм в латинской транслитерации, грамматических или терминологических тегов.

Приведен пример выдачи на запрос в режиме конкорданса – список кратких контекстов словоупотребления заданного слова в корпусе (рис. 1). Корпусный менеджер позволяет также переходить и к более широким контекстам.

Для расширенного поиска используется язык регулярных выражений, позволяющий задавать более сложные запросы. Результатом для выражения [wordlat=".\*jug"] будет выдача всех контекстов, в которых встречается словоформа на латинице, заканчивающаяся на слог jug или равная ему (рис. 2).

Система позволяет получать частотные списки на основе загруженных текстов, при этом атрибутами могут выступать различные единицы и даже комбинации единиц. Пример частотного списка приведен в табл. 4.



**Список наиболее частотных тибетских лексических единиц, отмеченных терминологическими тегами**

Лексема в латинской транслитерации	Перевод	Терминологический тег	Частота
sbyor	‘добавлять’	Gram GenLex	123
rjes 'jug	‘финаль’	Gram GenLex TGra	65
yi ge	‘фонема’	Gram L TBas	64
sgra	‘показатель’	Gram GenLex	61
'jug	‘употреблять’	Gram GenLex	58
ming	‘слово’	Gram L TBas	40
tshig	‘высказывание’	Gram L TBas	34
sdud	‘соединять’	Gram L TGrMark	24

### ЛИНГВИСТИЧЕСКИЕ ВОЗМОЖНОСТИ КОРПУСА

Корпус памятников тибетской грамматической традиции может быть использован как в теоретических исследованиях, так и для решения прикладных лингвистических задач. Корпус является источником данных для исследований по лексикографии, семантике и грамматике.

Прикладные лингвисты могут использовать корпус в обучении языку и переводу. Можно отбирать эквиваленты интересующей нас лексической единицы, формировать и сопоставлять принципы перевода имен собственных, устойчивых словосочетаний, грамматических и общенаучных терминов; осуществлять поиск соответствий между грамматическими явлениями и выделять способы их перевода, получив список контекстов для данного явления.

Специальная грамматическая разметка позволяет использовать корпус как ресурс для исследования тибетской грамматической терминологии и общенаучных терминов. Часть тибетских терминов имеет индийское происхождение, часть образовалась в результате терминологизации слов общей лексики. Кроме того, для тибетской грамматической традиции характерно использование буддийской терминологии, которая имеет индийское происхождение, а также общенаучных терминов, распространённых в сочинениях по другим тибетским традиционным наукам (буддийская религиозная доктрина, логика, медицина, ремесло, поэтика, синонимика, стихосложение, астрология и драма). Специальная грамматическая разметка делает возможным поиск эквивалентов для перевода терминов, позволяет сопоставлять употребление грамматических и общенаучных терминов, определять устойчивые термины-словосочетания, анализировать способы формирования терминологических словосочетаний и др.

Кроме того, специалисты по тибетской истории и культуре, а также литературоведы могут использовать корпус для исследования формирования и развития специальной лексики тибетской научной традиции.

### ЛЕКСИЧЕСКАЯ БАЗА КОРПУСА ТИБЕТСКИХ ГРАММАТИЧЕСКИХ СОЧИНЕНИЙ

#### Тибетская грамматическая терминология

Как и в Индии, научное знание в Тибете формировалось в процессе развития, изучения и адаптации религиозной доктрины. Буддизм оказал сильное влияние на развитие тибетской лингвистической традиции и прочих областей протонаучного знания.

В Тибете выделяли десять традиционных наук, которые в свою очередь подразделялись на пять великих (тиб. rig gnas che ba) и пять малых наук (тиб. rig gnas chung ba). Лингвистика входила в число пяти великих наук наряду с буддийской религиозной доктриной (санскр. adhyatmavidyā, тиб. nang gi rig pa), логикой (санскр. hetuvidyā, тиб. gtan tshigs kyi rig pa), медициной (санскр. cikitsavidya, тиб. gso ba'i rig pa) и ремеслом (санскр. silpakarmasthanavidyā, тиб. bzo gnas kyi rig pa). К малым наукам относили поэтику (санскр. kāvyā, тиб. snyan ngag), синонимнику (санскр. abhidhāna, тиб. mngon brjod), стихосложение (санскр. chandam, тиб. sdeb sbyor), астрологию (санскр. jyotiṣa, тиб. skar rtsis) и драму (санскр. naṭa, тиб. zlos gar) [8, с. 34].

Истоки терминологии и моделей описания тибетской лингвистической традиции так же, как и прочих областей протонаучного знания в Тибете, восходят к буддийской религиозной доктрине. В тибетских грамматических сочинениях, как и в трудах по остальным традиционным наукам, широко использовалась терминология, общая для всех индотибетских традиционных наук, преимущественно религиозного и философского характера.

Современное научное языкознание предполагает разграничение обычной лексики и специальной грамматической терминологии. В рамках лингвистических традиций древности в качестве грамматических терминов использовались, как правило, общепотребительные слова в специальном значении.

Часть тибетской грамматической терминологии сформировалась в результате заимствования или калькирования. Начавшаяся в VII в. работа по пере-

воду индийских религиозных текстов привела к заимствованию индийских моделей описания и терминологии. Однако случаи прямого заимствования санскритских терминов встречаются редко, в основном термины переводились на тибетский язык. Активная деятельность по переводу индийских текстов привела к формированию религиозной и философской специальной лексики. В начале IX в. в Тибете были созданы первые списки буддийской терминологии, включавшие тибетские эквиваленты санскритских терминов. Некоторые исследователи отмечают, что точность тибетских переводов и последовательность в использовании религиозных и философских терминов могут позволить реконструировать утраченные оригиналы санскритских текстов. Таким образом, главным источником терминологии тибетской грамматической традиции была санскритская терминология.

Для тибетской грамматической традиции не характерно выделение таких традиционных для западной лингвистики разделов, как фонология, морфология и синтаксис. Базовые термины тибетской грамматической традиции представляют собой обозначения первичных единиц разных языковых уровней [9, с. 24].

Большинство тибетских грамматических сочинений начинают грамматическое повествование с описания тибетского алфавита, различных категорий графем, изложения правил образования графического слога и сочетания графем, а также правил морфологической сочетаемости фонем. Также тибетские грамматики содержат описание служебных морфем и лексем.

Тибетская грамматическая традиция заимствовала индийскую концепцию семи падежей. В индийской традиции категория падежа связана с категорией *карака* (санскр. *kāraḥ*), которая представляет собой промежуточный уровень между семантикой и морфологией. Описание этих категорий относится к семантике (определяются семантически; передают свойство участников групп определенных ситуаций), но выражаются они средствами морфологии – в санскрите это в основном конкретные падежные суф-

фиксы (морфологический уровень). Тибетские грамматисты заимствовали у индийцев данную систему категорий.

### Специальная терминологическая разметка

В рамках проекта было принято решение выработать специальные теги, маркирующие тибетскую грамматическую терминологию. К тегам морфологической разметки корпуса были добавлены элементы традиционного грамматического метаописания – разметка по терминологическим категориям (фонологическая, синтаксическая терминология), собственно в базе данных вводятся научный комментарий, санскритские эквиваленты (для заимствованных терминов), гиперссылки на синонимы, гиперонимы, гипонимы.

Специальная грамматическая разметка в корпусе (табл. 5) позволяет разделить терминологические поля: грамматическую терминологию (тег Gram) и термины традиционных наук (тег GenScien).

Отдельные теги обозначают модели происхождения термина – посредством терминологизации слов общей лексики (тег GenLex) или путем заимствования (тег L).

Базовая грамматическая терминология отмечается тегом TBas. Поскольку тибетским базовым терминам и в целом тибетской терминологии свойственна полисемия, основной задачей было разграничение фонологической терминологии (тег TPhon) и терминов-названий графем (тег TGra).

Также было принято решение разделить термины-названия служебных морфем и лексем (тег TGrMark) и термины падежной грамматики (тег TCGr).

Использование специальных тегов в разметке корпуса позволяет экспортировать из корпуса терминологическую лексику и создать базу данных тибетской лингвистической терминологии, включающую дополнительные комментарии к терминам: указание языка заимствования для заимствованных терминов, приведение эквивалентов на других языках, уточнение способа заимствования (фонетическое заимствование, семантическое заимствование, гибридные термины) и др.

Таблица 5

### Теги для маркировки терминов

Признак классификации	Тег	Значение
Терминологическое поле	Gram	Термин тибетской грамматической традиции
	GenScien	Общенаучный термин, термин традиционных наук
Происхождение термина	GenLex	Термин тибетского происхождения (образован посредством терминологизации слов общей лексики)
	L	Заимствованный термин
Тип терминологии	TBas	Базовый грамматический термин
	TPhon	Фонологический термин
	TGra	Термин-название типа графемы
	TGrMark	Термин-название служебных морфем и лексем
	TCGr	Термин падежной грамматики

## Структура лексической базы

Наполнение лексической базы тибетской грамматической терминологии составляют лексические единицы, отобранные из тибетской части корпуса по заданным тегам, а именно по тегам для маркировки терминов (лингвистических метаданных, см. табл. 4).

В качестве промежуточного формата базы данных предлагается словарный формат Рекомендаций проекта TEI (Text Encoding Initiative) [10, 11]. Разметка TEI основана на синтаксисе языка XML. При разработке формата базы принималось во внимание, что TEI содержит метки для перекрестных ссылок, позволяющие фактически отображать сетевое представление данных в линейном XML-файле.

В разметке TEI описание элементов данных лексической базы представляет собой последовательность тегов верхнего уровня, маркирующих лингвистические термины (ЛТ), содержание которых уточняется дополнительными тегами.

Опишем шаблон представления данных лексической базы в формате XML по спецификации TEI в соответствии с требуемой структурой данных. Этот шаблон имеет несколько блоков и уровней представления.

### Уровень лексической единицы

Лексическая единица базы в XML-представлении начинается с записи вида:

```
<entry n="1" type="lex">
  <term>ལྷ་ལེ</term>
  <pron>A li</pron>
```

где указывается порядковый номер лексической единицы (ЛЕ) (n="1"), ее тип (type="lex" – лексема), а также собственно заголовочное слово ལྷ་ལེ в написании тибетским письмом (тег <term>) и латинской транслитерации A li (тег <pron>).

Далее идёт блок грамматической информации (тег <gramGrp>):

```
<gramGrp>
  <pos> N </pos>
</gramGrp>
```

где указывается часть речи (тег <pos>) и, возможно, дополнительная грамматическая информация (теги <gen>, <flex> и др.).

К уровню лексической единицы относится также блок этимологической информации (тег <etym>), их может быть несколько:

```
<etym n="1">
  <lbl> Сумчупа </lbl>
  <date> VII-IX вв. н.э.</date>
</etym>
```

Атрибут n – порядковый номер блока этимологии, тег <lbl> описывает источник информации, тег <date> – дату или год фиксации термина. Также могут использоваться теги <mentioned> – форма фиксации и <lang> – язык, из которого предположительно было заимствовано слово.

## Уровень связей

Связей может быть несколько в рамках одного термина, каждая из них должна начинаться с тега <rel>, а в качестве атрибута указывается тип связи (напр., type="hyperonym"). Так, для описанного выше ЛТ ལྷ་ལེ (гласная) приводится связанная с ней единица ལེ་ལོ་ ('фонема, графема').

```
<rel type="hyperonym">
  <target>ལེ་ལོ་</target>
</rel>
```

### Уровень примеров

К каждому лингвистическому термину может быть приведено несколько примеров из корпуса:

```
<cit n="1" type="example">
  <quote>ལེ་ལོ་ལྷ་ལེ་ལྷ་ལེ་གཞིས།།</quote>
</cit>
```

Блок с примерами описывается тегом <cit>, где аргумент n – номер примера, а type – его тип. Сам текст заключается в тег <quote>.

## Система ведения лексической базы

Система ведения базы данных предназначена для хранения всех ЛТ в виде, удобном для быстрой и оперативной работы. Используется реляционная СУБД Microsoft SQL. Центральный и наиболее важный компонент системы, ядро – это программный модуль, являющийся связующим звеном между тремя способами представления данных в разрабатываемой системе – XML-представлением, базой данных и пользовательским интерфейсом.

Для упрощения перевода данных XML в базу и обратно структура данных и таблиц внутри базы данных соответствует структуре формата TEI, соответственно, в базе формируется 5 таблиц: Entries, Grammar, Etymology, Relations, Examples. Иерархически, таблицы можно разделить на несколько уровней (как и уровни XML-текста). Самый высший, первый, это уровень входов (ЛТ) – строки в этой таблице ни на что не ссылаются, но на них ссылаются строки из других таблиц.

## ДАЛЬНЕЙШЕЕ РАЗВИТИЕ

В дальнейшем планируются эксперименты по автоматическому извлечению лингвистических терминов из текстов с использованием дистрибутивно-статистического метода [12].

Предполагается увеличить объем корпуса и снабдить его синтаксической разметкой. Перспективным направлением развития корпуса видится создание на его основе более масштабного корпуса специальных текстов на тибетском языке, в том числе текстов, посвященных другим традиционным наукам тибетцев: буддийской религиозной доктрине, логике, медицине, ремеслу, поэтике, синонимике, стихосложению, астрологии и драме.

## ЗАКЛЮЧЕНИЕ

Пилотная версия корпуса тибетских грамматических сочинений будет полезна всем исследователям памятников тибетской грамматической традиции, а также может быть использована для лингвистического исследования тибетского языка, его изучения и преподавания, так как в настоящее время отсутствуют общедоступные корпуса размеченных текстов на тибетском языке, тем более с переводом на русский язык.

Создание частотного словаря лексических единиц тибетской грамматической терминологии и семантический анализ лексической базы позволяют говорить о своего рода лингвистической онтологии для тибетского языка, включающей гипонимы и гиперонимы, многозначные и синонимичные термины. Это дает возможность проанализировать изменение знаний о языке, состав грамматического терминологического поля, а также оценить степень терминологизации слов общей лексики.

## СПИСОК ЛИТЕРАТУРЫ

1. Wagner A., Zeisler B. A syntactically annotated corpus of Tibetan // Proceedings of the 4th International Conference on Language Resources and Evaluation. – Lisboa, 2004. – P. 1141-1144.
2. Semantic Roles, Case Relations, and Cross-Clausal Reference in Tibetan. – URL: <http://www.sfb441.uni-tuebingen.de/b11/b11corpora.html#clarkTrees> (дата обращения 14.04.2012).
3. Tibetan Studies Triumphs Again at the Arts and Humanities Research Council. – URL: <http://www.soas.ac.uk/news/newsitem73472.html> (дата обращения 15.04.2015).
4. Гроховский П.Л. Категории сказуемости и номинализации действия (субстантивно-адъективные формы) в классическом тибетском языке // Очерки по теоретической грамматике восточных языков: Существительное и глагол / под ред. В.Г. Гузева. – СПб.: Издательский дом СПбГУ, 2001. – С. 269–288.
5. Гроховский П.Л. Грамматика имени существительного в классическом тибетском языке // Там же. – С. 76–91.
6. Helmut Schmid. Probabilistic Part-of-Speech Tagging Using Decision Trees // Proceedings of International Conference on New Methods in Language Processing. – Manchester, 1994. – P. 44-49.
7. Chungku Ch., Rabgay J., Faaß G. Building NLP Resources for Dzongkha: A Tagset and A Tagged Corpus. Proceedings of the 8th Workshop on Asian Language Resources (Beijing, China, 21-22 August 2010). – Beijing, 2010. – P. 103-110.
8. Смирнова М.О. История формирования тибетской лингвистической традиции // Вестник Санкт-Петербургского университета. Серия 13. Востоковедение. Африканистика. Вып. 2. – СПб., 2012. – С. 34-41.
9. Смирнова М.О. Базовые термины тибетской грамматической традиции // Там же. Вып. 1. – СПб., 2014. – С. 23-34.
10. TEI P5: Guidelines for Electronic Text Encoding and Interchange / eds. L. Burnard, S. Bauman. – 2010. – 1539 + LXIV p.
11. Захаров В.П. Электронный обменный формат для словарей проекта ТЕИ (Text Encoding Initiative) – СПб.: СПбГУ. РИО. Филологический факультет, 2013. – 80 с.
12. Захаров В.П., Хохлова М.В. Автоматическое извлечение терминов из специальных текстов с использованием дистрибутивно-статистического метода как инструмент создания тезаурусов // Структурная и прикладная лингвистика. Вып. 9. – СПб., 2012. – С. 222-233.

## ПРИЛОЖЕНИЕ 1

### Список тегов для обозначения знаменательных частей речи в тибетском языке

№ п/п	Тэг	Часть речи	Пример	Перевод
1	N	Существительное	bu	‘сын, мальчик’
2	Adj	Прилагательное	dmar po	‘красный’
3	Adv	Наречие	da	‘сейчас’
4	N	Местоимение	kho	‘он’
5	Num	Числительное	gnyis	‘один’
6	V	Глагол	byed	‘делать’
7	Int	Междометие	kye	‘О!’
8	On	изобразительное слово	thrig thrig	‘скрип обуви’

Список тэгов для обозначения форм глагола в тибетском языке

№ п/п	Тэг	Форма глагола	Пример	Перевод
1	VPrs	финитная форма глагола настоящего времени	spong	'отбрасывает'
2	VPst	финитная форма глагола прошедшего времени	spangs	'отбросил'
3	VFut	финитная форма глагола будущего времени	spang	'отбросит'
4	VImp	финитная форма глагола повелительного времени	spongs	'отбрось'
5	VPrsPtcp	причастная форма глагола, образованная от основы настоящего времени с помощью именной частицы	spong ba	'отбрасывающий'
6	VPstPtcp	причастная форма глагола, образованная от основы прошедшего времени с помощью именной частицы	spangs pa	'отбросивший'
7	VPrsPtcpCom	деепричастная форма глагола, образованная с помощью показателя комитатива	spong ba dang	'отбрасывая'
8	VPrsPtcpErg	деепричастная форма глагола, образованная с помощью показателя эргатива	spong bas	'так как отбрасывает'
9	VPrsErg	деепричастная форма глагола, образованная с помощью показателя эргатива с выпадением форманта причастия	spong gis	'так как отбрасывает'
10	VPrsPtcpDat	деепричастная форма глагола, образованная с помощью показателя датива	spong ba la	'когда отбрасывает'
11	VPrsDat	деепричастная форма глагола, образованная с помощью показателя датива эргатива с выпадением форманта причастия	spong la	'когда отбрасывает'
12	VPrsPtcpLoc	деепричастная форма глагола, образованная с помощью показателя локатива	spong ba na	'когда отбрасывает'
13	VPrsLoc	деепричастная форма глагола, образованная с помощью показателя локатива эргатива с выпадением форманта причастия	spong na	'когда отбрасывает'
14	VPrsPtcpEl	деепричастная форма глагола, образованная с помощью показателя элатива	spong ba nas	'когда отбрасывает', 'после того, как отбрасывает'
15	VPrsEl	деепричастная форма глагола, образованная с помощью показателя элатива эргатива с выпадением форманта причастия	spong nas	'когда отбрасывает', 'после того, как отбрасывает'
16	VPrsPtcpAbl	деепричастная форма глагола, образованная с помощью показателя аблатива	spong ba las	'после того, как отбрасывает'
17	VPrsPtcpDest	деепричастная форма глагола, образованная с помощью показателя дестинатива	spong bar	'после того, как отбрасывает'

Материал поступил в редакцию 02.08.15.

**Сведения об авторах**

**ГРОХОВСКИЙ Павел Леонович** – кандидат филологических наук, доцент кафедры монголоведения и тибетологии Санкт-Петербургского государственного университета  
E-mail: plgr@mail.ru

**ЗАХАРОВ Виктор Павлович** – кандидат филологических наук, доцент кафедры математической лингвистики Санкт-Петербургского государственного университета  
E-mail: vz1311@yandex.ru

**СМИРНОВА Мария Олеговна** – аспирант Восточного факультета Санкт-Петербургского государственного университета  
E-mail: 2321781@mail.ru

**ХОХЛОВА Мария Владимировна** – кандидат филологических наук, старший преподаватель кафедры математической лингвистики Санкт-Петербургского государственного университета  
E-mail: khokhlova.marie@gmail.com

**Федеральное государственное бюджетное учреждение науки  
ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ  
ИНФОРМАЦИИ РОССИЙСКОЙ АКАДЕМИИ НАУК**

**предлагает научным работникам, аспирантам и другим специалистам в области естественных, точных и технических наук, желающим быстро и эффективно опубликовать результаты своей научной и научно-производственной деятельности, использовать способ публикации своих работ через систему депонирования.**

«Депонирование (передача на хранение) – особый метод публикации научных работ (отдельных статей, обзоров, монографий, сборников научных трудов, материалов научных конференций, симпозиумов, съездов, семинаров) узкоспециального профиля, разрешенных в установленном порядке к открытому опубликованию, широкое тиражирование которых, как правило, в силу их узкой специализации, не считается целесообразным, а также работ широкого профиля, срочная информация о которых необходима для утверждения их приоритета. Депонирование предусматривает прием, учет, регистрацию, хранение научных работ и обязательное размещение информации о них в специальных информационных изданиях».

Подготовка и передача на депонирование научных работ происходит в соответствии с «Инструкцией о порядке депонирования научных работ по естественным, техническим, социальным и гуманитарным наукам» (М., 2013).

Депонированные научные работы находятся на хранении в депозитарии ВИНТИ РАН, копии работ предоставляются заинтересованным организациям и специалистам на бумажном и электронном носителях и являются официальной публикацией.

Информация о депонированных научных работах включается в информационные издания ВИНТИ РАН, в РЖ ВИНТИ РАН и БД ВИНТИ РАН и аннотированный библиографический указатель «Депонированные научные работы».

Подать научную работу на депонирование можно, обратившись в Отдел депонирования ВИНТИ РАН по адресу:

**125190, Москва, ул. Усиевича, 20.**

**ВИНТИ РАН, Отдел депонирования научных работ.**

**Тел.: 8 (499) 155-43-28, Факс: 8 (499) 943-00-60.**

**e-mail: [dep@viniti.ru](mailto:dep@viniti.ru)**

С инструкцией о порядке депонирования можно ознакомиться на сайте ВИНТИ РАН: <http://www.viniti.ru>