

Зотова А.К.

ОТЧЕТ о работе по теме «ЛИНГВОТИПОЛОГИЧЕСКАЯ БАЗА ДАННЫХ «ЯЗЫКИ МИРА»: ДИАХРОННЫЕ ИЗМЕНЕНИЯ В СЕМЬЯХ И ГРУППАХ ЯЗЫКОВ (Славянские языки)» Объем 15 а.л. Рук. Д.ф.н. Е.И.Ярославцева. в 2011г.

СОДЕРЖАНИЕ ОТЧЕТА

- 1. Направления исследований по теме**
- 2. Обзор исследований 2009-2010 гг.**
- 3. Работа в 2011г.**
 - 3.1. Анализ материала энциклопедии “Языки мира”**
 - 3.1.1. Список статей и типологизация схем описания**
 - 3.1.2. Анализ типовых схем описания**
 - 3.1.3. Анализ текстов статей энциклопедии “Языки мира”**
 - 3.2. Анализ описания языков в БД “Языки мира” в виде рефератов**
 - 3.2.1. Список языков**
 - 3.2.2. Анализ схем описания**
 - 3.2.3. Анализ полноты описания**
 - Неполнота/отсутствие данных**
 - Неточности**
 - Ошибки**
- 4. Сопоставительный анализ описания языков в энциклопедии “Языки мира” и БД “Языки мира”**
 - 4.1. Источники неполноты/неточности/ошибок**
 - 4.2. Локализация данных**
 - 4.3. Анализ терминологии**
- 5. О ручном эксперименте по реконструкции описания Праславянского и Древнерусского языков и диахронных изменений по линии Праславянский → Старославянский → Древнерусский языки на основе данных об их фонетических системах.**
- 6. Перспективы**

1. Отчет содержит материал, иллюстрирующий основные направления исследований по теме «ЛИНГВОТИПОЛОГИЧЕСКАЯ БАЗА ДАННЫХ «ЯЗЫКИ МИРА»: ДИАХРОННЫЕ ИЗМЕНЕНИЯ В СЕМЬЯХ И ГРУППАХ ЯЗЫКОВ (Славянские языки)» Объем 15 а.л. Рук. Д.ф.н. Е.И.Ярославцева. Одновременно он является продолжением публикаций об исследованиях по теме «КОМПЬЮТЕРНАЯ БАЗА ДАННЫХ «ЯЗЫКИ МИРА»: ФОРМАЛЬНО СОВПАДАЮЩИЕ КЛАССЫ ЯВЛЕНИЙ (УНИВЕРСАЛИИ, ФРЕКВЕНТАЛИИ, РАРИТАРИИ И УНИКАЛИИ) В СЕМЬЯХ И ГРУППАХ ЯЗЫКОВ на материале германских (2003-2005 гг.), романских (2005-2007гг.) и славянских языков (2007-2009 гг.).

Работа представляет собой первую попытку исследовать возможности использования лингвотипологической базы данных «Языки мира» для проведения качественного сравнительно-исторического анализа языковых явлений в семьях и группах языков. Исследование принципиально отличается от проводимых количественных компьютерных экспериментов, в которых БД «Языки мира»

рассматривается как корпус описаний, позволяющий производить масштабные расчеты математического или статистического характера.

2. В 2009-2010гг. работа по теме проводилась последовательно и включала в себя несколько этапов:

1. Установление источников исторической и диахронной информации в энциклопедии «Языки мира»;

2. Хронологизация и периодизация основных диахронных типологических процессов и явлений в рамках семьи/группы/подгруппы/отдельного языка (диалекта) (анализ текстов статей о славянских языках энциклопедии «Языки мира»);

3. Построение совокупной иерархии элементов энциклопедического и формального описаний, содержащих информацию о динамике языковых процессов/явлений;

4. Выделение релевантных позиций в модели формального описания языков (реферате) в БД «Языки мира»;

5. Анализ терминологических расхождений между компьютерной и традиционной моделями описания явлений в семьях и группах языков

3. В 2011 г. задачами проекта являлись:

- сопоставительный анализ типовых схем статей энциклопедии «Языки мира»
- сопоставительный анализ моделей описания языков в БД «Языки мира»
- исследование возможности получения исторической и диахронной типологической информации с помощью БД «Языки мира»
- качественная сравнительная оценка исторической и диахронной информации на основе базовых схем описания языков по тексту энциклопедии «Языки мира» и сведений, содержащихся в БД «Языки мира»
- верификация сведений на основе сравнительного анализа описания славянских языков в энциклопедии «Языки мира» и БД «Языки мира»
- качественная сравнительная оценка компьютерной и традиционной моделей описания в семьях и группах языков; выявление их адекватности с помощью ручного эксперимента
- оценка результатов предшествующих компьютерных экспериментов 2007-2009 гг. на основе анализа данных о славянских языках в БД «Языки мира» (ошибки, неточности, лакуны и причины расхождений с данными Энциклопедии «Языки мира»)

3.1. Анализ материала энциклопедии «Языки мира»

Основу статей энциклопедии составляют синхронные описания языков и диалектов на современном этапе. Отдельные статьи посвящены древнему и средневековому состоянию одного и того же языка: Праславянский → Старославянский → Древнерусский → Русский. Следовательно, с помощью сопоставительного анализа можно проследить историю изменений и оценить основные направления динамики диахронных процессов.

Для некоторых языков структурно значимые расхождения между современным и исторически удаленным состоянием отмечаются внутри статей, например наличие/количество монофтонгов, падежей и др. Иногда экскурсы в историю единичны или разрознены и имеют место только в тех случаях, когда речь идет о специфических и/или уникальных явлениях. Следовательно, необходимо проанализировать эти данные и по возможности структурировать их в виде, соотносимом с моделью реферата БД «Языки мира».

Источниками исторической и диахронной типологической информации в Энциклопедии «Языки мира» являются:

- специальные статьи издания энциклопедии, посвященные древним языкам Европы (германские, романские, балтийские языки);
- сопоставительный анализ статей, посвященных различным историческим периодам одного и того же языка/группы языков/диалектов, например: Праславянский → Старославянский → Древнерусский → Русский

3.1.1. Список статей о славянских языках в энциклопедии «Языки мира» и их типовая классификация

Статья «Славянские языки» - схема I для групп языков

ЮЖНОСЛАВЯНСКИЕ ЯЗЫКИ - схема II

- Старославянский и церковнославянский язык (мертвый)
- Болгарский язык
- Македонский язык
- Сербохорватский язык (сербский, хорватский, боснийский языки)
- Словенский язык

ЗАПАДНОСЛАВЯНСКИЕ ЯЗЫКИ - схема II

- Чешский язык
- Словацкий язык
- Серболужицкий язык (с двумя литературными языками верхне- и нижнелужицким)
- Польский язык
- Кашубский язык
- Полабский язык (мертвый)

ВОСТОЧНОСЛАВЯНСКИЕ ЯЗЫКИ - схема II

Древнерусский язык (мертвый)

Древненовгородский диалект (мертвый) - схема III

Русский язык

Украинский язык

Белорусский язык

Малые славянские литературные языки (микроязыки) - схемы I/III

Примечание: данные о Праславянском языковом континууме (раннепраславянском и позднепраславянском языках) содержатся в статье Славянские языки.

3.1.2. Анализ типовых схем описания языков в Энциклопедии «Языки мира»

Для стандартизации описания языков в Энциклопедии «Языки мира» были предложены ТИПОВЫЕ СХЕМЫ СТАТЕЙ:

ТИП I. Статья о семье языков, группе языков, группе диалектов

1. Название
2. Локализация и основные представители
3. Ориентировочное число говорящих
4. Принципы и варианты генетической классификации
5. Хронология разделения — для крупных языковых семей
6. Типичные фонетико-грамматические характеристики

ТИП II. Статья о языке

1.1.0. Общие сведения

1.1.1. варианты названия

1.1.2. генеалогические сведения = ГЕНЕТИЧЕСКАЯ ПРИНАДЛЕЖНОСТЬ

1.1.3. распространение

1.2.0. Лингвогеографические сведения

1.2.1. общий диалектный состав = 1.2.1. ДИАЛЕКТНЫЙ СОСТАВ

1.3.0. Социолингвистические сведения

1.3.1. коммуникативно-функциональный статус и ранг языка = 1.3.1. СТАТУС ЯЗЫКА

1.3.2. степень стандартизации = 1.3.2. ЛИТЕРАТУРНЫЙ ЯЗЫК

1.3.3. учебно-педагогический статус = 1.3.3. ПРЕПОДАВАНИЕ

1.4.0. Тип письменности = 1.4.0. ПИСЬМО

1.5.0. Краткая периодизация истории языка = 1.5.0. ИСТОРИЧЕСКИЕ ПЕРИОДЫ

1.6.0. Внутривидовые явления, обусловленные внешнеязыковыми контактами = 1.6.0.

ВНУТРИСТРУКТУРНЫЕ ИЗМЕНЕНИЯ ИЗ-ЗА КОНТАКТОВ

2.0.0. Лингвистическая характеристика

2.1.0. Фонологические сведения

2.1.1. фонемный состав

2.1.2. просодический состав

2.1.3. позиционная реализация фонем и просодем

2.1.4. слог; наличие и статус долготных противопоставлений

2.2.0. Морфонологические сведения

2.2.1. фонологическая структура морфемы и/или слова; соотношение слога и морфемы

2.2.2. наличие фонологических противопоставлений морфологических единиц и категорий

2.2.3. типы чередований

2.3.0. Семантико-грамматические сведения

- 2.3.1. критерии выделения частей речи; категориальные способы выражения универсальных значений (общая характеристика)
- 2.3.2. характер и способ выражения качественных именных классификаций
- 2.3.3. категория числа и способы ее выражения
- 2.3.4. падежные значения и их выражение; характер категорий принадлежности и их выражение
- 2.3.5. характер и способ выражения качественных глагольных классификаций: за-логовые, видовые или видо-временные значения, наклонения, переходность, по-роды, версии
- 2.3.6. дейктические категории и способы их выражения: категория лица в имени и в глаголе, категория неопределенности/определенности в имени, категория времени в глаголе или в предложении, указание и пространственная ориентация, анафорические средства; выражение отрицания
- 2.3.7. семантико-грамматические разряды слов
- 2.4.0. Образцы парадигм
- 2.5.0. Морфосинтаксические сведения
- 2.5.1. типичная структура словоформы (для языков с развитой морфологией); тенденция к суффиксации и инфиксации; морфологически аномальные разряды слов
- 2.5.2. основные способы и правила словообразования
- 2.5.3. типичная структура простого предложения (или его эквивалента — для языков с нерасчлененной синтаксической структурой) и способы выражения субъектно-объектных отношений; синтаксические позиции; разновидности простого предложения
- 2.5.4. основные правила построения сложных предложений; характерные типы сложных предложений; основные закономерности порядка слов
- 2.6.0. Источник, объем и роль лексических заимствований = 2.6.0. ЛЕКСИЧЕСКИЕ ЗАИМСТВОВАНИЯ
- 2.7.0. Диалектная система = 2.7.0. ДИАЛЕКТНЫЕ ОСОБЕННОСТИ

ТИП III. Статья о диалекте

- 1. Название языка
- 2. Позиция в диалектной группе (отношение к литературному/стандартному языку, особенности лингвистической характеристики)
- 3. Распространение, по возможности — число говорящих
- 4. Функциональная нагрузка

ТИП IV. Статья об отдельном языке объемом менее 0,5 а.л.

- 1.0. Название
- 2.0. Локализация (для мертвых языков — время существования)
- 3.0. Генетическая принадлежность
- 4.0. Характер памятников; тип письменности
- 5.0.0. Лингвистическая характеристика
 - 5.1.0. Фонологические сведения (включая сведения об ударении, тоне и т.д.)
 - 5.1.1. слог
 - 5.1.2. типы чередований
 - 5.2.0. Морфология
 - 5.2.1. морфологический тип языка (агглютинативный, флективный и т.д.)
 - 5.2.2. состав и характер морфологических категорий; способы их выражения
 - 5.2.3. основные способы словообразования
 - 5.3.0. Синтаксис
 - 5.3.1. структура простого предложения (его эквивалента)
 - 5.3.2. характерные типы сложного предложения
 - 5.4.0. Генетическая и ареальная характеристика лексики
- 6.0. Сведения о диалектах

Примечание 1: шрифтом выделены позиции, название которых было изменено на первом этапе формализации текста статей энциклопедии (компьютерный варианты модели реферата БД «Языки мира» М1)

Примечание 2: цветом выделены позиции, отсутствующие в окончательной версии реферата БД «Языки мира»

3.2. Анализ описания языков в БД “Языки мира” в виде рефератов

3.2.1. Список языков представленных (список А) и **отсутствующих в БД «Языки мира»** в виде рефератов (**список Б**):

список А

1. Белорусский язык
2. *Булгарский язык* → Болгарский язык
3. Кашубский язык
4. Македонский язык
5. Полабский язык (Мертвый)
6. Польский язык
7. Русский язык
8. Серболужицкий язык
9. *Сербскохорватский язык* → Сербохорватский язык
10. Словацкий язык
11. Словенский Язык язык
12. Старославянский язык
13. Украинский язык
14. Чешский язык

Примечание 1: термин *булгарский* отсутствует в энциклопедических описаниях этого языка. Единственным вариантом иного обозначения в Энциклопедии “Языки мира” является самоназвание, которое приводится в соответствующей форме – Български език. Стандартным термином, употребляемым для данного языка принято считать “Болгарский язык”.

Примечание 2: термин *сербскохорватский* представляет собой один из вариантов русского обозначения наряду с такими как хорватско-сербский, сербский, которые используются лишь окказионально. Стандартным термином, употребляемым для всего данного ареала принято считать “Сербохорватский язык”.

список Б:

Праславянский язык в составе статьи Славянские языки (схема I).

Древнерусский язык (мертвый) (схема II)

Древненовгородский диалект (мертвый) (схема III)

3.2.2. Анализ схем описания

В процессе подготовки компьютерной версии БД исходный вид содержательной информации статей энциклопедии подвергался поэтапной обработке с целью ее

максимальной адаптации к формату Интернета. Таким образом, были получены два новых варианта модели реферата, отличающиеся как от первоначальной модели, так и друг от друга (варианты М1 и М2).

Вариант М1 - переходный от полного представления материала статей Энциклопедии “Языки мира” и **Вариант М2** - окончательный вид модели реферата в Интернет версии БД (выделен цветом).

1.1.0. СОКРАЩЕННОЕ НАЗВАНИЕ ЯЗЫКА

1.1.1. ВАРИАНТЫ НАЗВАНИЯ

1.1.2. ГЕНЕТИЧЕСКАЯ ПРИНАДЛЕЖНОСТЬ

1.1.3. РАСПРОСТРАНЕНИЕ

1.2.1. ДИАЛЕКТНЫЙ СОСТАВ

1.3.1. СТАТУС ЯЗЫКА

1.3.2. ЛИТЕРАТУРНЫЙ ЯЗЫК

1.3.3. ПРЕПОДАВАНИЕ

1.4.0. ПИСЬМО

1.5.0. ИСТОРИЧЕСКИЕ ПЕРИОДЫ

1.6.0. ВНУТРИСТРУКТУРНЫЕ ИЗМЕНЕНИЯ ИЗ-ЗА КОНТАКТОВ

2.1.1. ФОНЕМНЫЙ СОСТАВ

2.1.2. ПРОСОДИЧЕСКИЕ ЯВЛЕНИЯ

2.1.3. ФОНЕТИЧЕСКИ ОБУСЛОВЛЕННЫЕ ПРОЦЕССЫ

2.1.4. СЛОГ

2.2.1. ФОНОЛОГИЧЕСКАЯ СТРУКТУРА

2.2.2. ФОНОЛ. ПРОТИВОПОСТАВЛЕНИЯ МОРФОЛ. КАТЕГОРИЙ

2.2.3. МОРФОЛОГИЧЕСКИ ОБУСЛОВЛЕННЫЕ ЧЕРЕДОВАНИЯ

2.3.0. МОРФОЛОГИЧЕСКИЙ ТИП ЯЗЫКА

2.3.1. КРИТЕРИИ ВЫДЕЛЕНИЯ ЧАСТЕЙ РЕЧИ

2.3.2. ИМЕННЫЕ КЛАССИФИКАЦИИ

2.3.3. ЧИСЛО

2.3.4. ПАДЕЖНЫЕ ЗНАЧЕНИЯ

2.3.5. ГЛАГОЛЬНЫЕ КАТЕГОРИИ

2.3.6. ДЕЙКТИЧЕСКИЕ КАТЕГОРИИ

2.3.7. ЧАСТИ РЕЧИ

2.4.0. ПАРАДИГМЫ

2.5.1. СТРУКТУРА СЛОВОФОРМЫ

2.5.2.СЛОВООБРАЗОВАНИЕ

2.5.3.ПРОСТОЕ ПРЕДЛОЖЕНИЕ

2.5.4.СЛОЖНОЕ ПРЕДЛОЖЕНИЕ

2.6.0. ЛЕКСИЧЕСКИЕ ЗАИМСТВОВАНИЯ

2.7.0. ДИАЛЕКТНЫЕ ОСОБЕННОСТИ

Примечание: цветом выделены те позиции, которые не вошли в окончательный формат модели реферата в Интернет версии БД

Отличие обеих компьютерных версий от первоначальной модели, максимально соответствующей схемам статей энциклопедии, состоит в том, что в них содержится расширенный набор характеристик (подчиненных строк в позициях). Степень детализации по отдельным позициям достигает значительных различий. При этом в силу универсальной ориентации схем формального описания некоторые специфические/редкие (но интересные и значимые для языка и его истории явления) оказываются не представленными или представленными в неявном виде.

Пример поэтапного свертывания информации при переходе от текстов статей Энциклопедии “Языки мира” к формализованному описанию в БД “Языки мира”.

Этап первый:

Тексты статей энциклопедии включают в себя полную и разнообразную информацию. В частности, при наличии противопоставления по *долготе-краткости* гласных, отмечается, является ли оно самостоятельным (чешский язык) или *сочетается с* противопоставлением по *закрытости-открытости, напряженности* и т. д. То есть, при переходе от исходного текста статей к формализованному представлению это уточнение сведений о фонемном составе могло бы иметь вид:

2.1.0. Фонологические сведения

2.1.1. фонемный состав

Гласные

подъем

верхн./верх.пониж./верх.-ср./нижн.ср/ср./нижний

ряд

отсутствует

долгота

долгие/краткие

долгие/краткие/сверхкраткие

лабиализация

лабиализованные/нелабиализованные

назальность
назализованные/неназализованные

открытость
открытые/закрытые

редукция
редуцированные/нередуцированные

Однако, окончательный вариант этого раздела модели в Интернет версии БД имеет следующий вид:

- 1 2.1.1. ФОНЕМНЫЙ СОСТАВ
- 2 .О
- 3 .гласные
- 4 .О
- 5 ..подъем
- 6 ...верхн./верх.пониж./верх.-ср./нижн.ср/ср./нижний

- 15 ..ряд
- 16 ...отсутствует

- 22 ..долгие/краткие
- 23 ..долгие/краткие/сверхдолгие
- 24
- 26 ..лабиализованные/нелабиализованные
- 27 ..назализованные/неназализованные
- 28 ..напряженные/ненапряженные

- 30 ..открытые/закрытые

- 33 ..редуцированные/нередуцированные

- 38 ..фарингализованные/нефарингализованные
- 39 ..долгие/краткие/сверхкраткие/беглые
- 40 ..долгие/среднедолгие/краткие/сверхкраткие

Примечание 1: информация о долготе/краткости представлена непоследовательно с разрывом или с нарушением подчиненности, если строки 39 и 40 являются подвариантами строки 38.

- 22 ..долгие/краткие
- 23 ..долгие/краткие/сверхдолгие
- 24
- [характеризация по другим признакам]

- 39 ..долгие/краткие/сверхкраткие/беглые
- 40 ..долгие/среднедолгие/краткие/сверхкраткие

Примечание 2: нарушена иерархия соподчиненности строк и, следовательно, фактическая информация, содержащаяся в строках с 22 по 40, переходят в категорию классифицирующей. В остальных разделах классификация признаков и характеристик проводится единообразно, в частности:

Примечание 3: в других разделах реферата эти же свойства имеют равное положение в иерархии, например:

2.1.2. ПРОСОДИЧЕСКИЕ ЯВЛЕНИЯ

признак гармонического уподобления

долгота
ряд
подъем
назальность
огубленность
палатальность

Примечание 4: такого рода погрешности определили неудачу эксперимента по проверке гипотезы о возможности определения универсалий, раритарий, фреквенталий и уникалий в Славянских языках на основании формальных свойств рефератов БД (2007-2009гг).

Этап второй: переход от текста статьи энциклопедии к формализованному описанию языка.

Во избежание повторов в текстах статей энциклопедии «Языки мира» используется последовательная система отсылок от одних пунктов к другим. Подобная система совершенно отсутствует в БД «Языки мира». Во многих случаях это затрудняет поиск фактических данных, поскольку сложная иерархическая структура некоторых позиций, разнородность формулировок в строках с классифицирующей информацией (комбинация терминологических и нетерминологических словосочетаний) и большой объем детализирующей информации (число подчиненных строк в позициях) не позволяют устанавливать связи, если отсутствуют однозначные терминологические соответствия.

Например:

Исходный вид и рекомендуемая для статьи Энциклопедии «Языки мира» версия позиции

2.2.3. ЧЕРЕДОВАНИЯ ФОНЕМ

отсутствие чередований
локализация
внутри морфемы
на стыках морфем и слов

функциональный тип

морфологические (сопутствующие) = нефонологические

грамматические (смыслоразличительные) = фонологические

вид чередований

вокалические (**апофония**)

количественные (долготные)

с нулем

качественные

по тембру

аблаут

умлаут

по подъему
по дополнительному признаку (звонкие/глухие, с толчком без толчка, преглоттализированные/простые и т.п.)

консонантные

смешанные (консонантно-вокалические)

чередования комплексов

тоновые

регистровые (высокий/низкий и т.п.)

контурные (восходящий/нисходящий и т.п.)

акцентные

Вариант этой же позиции в окончательной Интернет версии БД «Языки мира»

2.2.3.МОРФОЛОГИЧЕСКИ ОБУСЛОВЛЕННЫЕ ЧЕРЕДОВАНИЯ

.О

.отсутствии чередований

.функциональный тип

..сопутствующие

..смыслоразличительные

...словообразовательные

...в имени

...в глаголе

...словоизменительные

...в имени

...в глаголе

.вид чередований

..вокалические

..консонантные

..смешанные

..тоновые

..чередования комплексов

Данные в реферате “Старославянский язык” (БД «Языки мира»)

2.2.3.МОРФОЛОГИЧЕСКИ ОБУСЛОВЛЕННЫЕ ЧЕРЕДОВАНИЯ

.функциональный тип

..смыслоразличительные

...словообразовательные

...в имени

...в глаголе



.вид чередований



..вокалические

...качественные



2.5.2.СЛОВООБРАЗОВАНИЕ

.деривация	<input checked="" type="checkbox"/>
..аффиксация	
постфиксы	<input checked="" type="checkbox"/>
...префиксы	
...суффиксы	
..конверсии	

Термин *аблаут* локализован в позиции

2.1.3. ФОНЕТИЧЕСКИ ОБУСЛОВЛЕННЫЕ ПРОЦЕССЫ

.О

.гласные

..структурные позиции

..комбинаторные позиции

...отсутствие чередований

...аблаут

Как видно данные об одном и том же языковом явлении, выраженные разными терминами и находящиеся в различных позициях, не соотносятся друг с другом. Это приводит к неполноте информации (незаполненные строки в реферате), хотя в неявном виде информация может присутствовать в других позициях.

Примечание: подробная интерпретация расхождений (например в случае терминов и выражаемых ими понятий, таких как *аблаут*, *апофония*, *внутренняя флексия*) не приводится в связи с ограниченным объемом отчета.

3.2.3. Анализ полноты описания

Сравнительный анализ текстов статей энциклопедии “Языки мира” и информации, содержащейся в Интернет версии БД “Языки мира”, показал, что кроме отмеченных выше отсутствующих описаний языков и позиций, в БД встречаются случаи отсутствия данных, неточности и ошибки.

Например, при сравнении рефератов Старославянский язык и Русский язык, в частности отметим:

Отсутствие данных и неточности:

2.5.3. ПРОСТОЕ ПРЕДЛОЖЕНИЕ

.строй

..номинативный

..линейный порядок членов предложения

..свободный

..основной

- ...SVO
- .согласование
- ..актант определяет форму предиката
- ...субъектное согласование
- ...по лицу
- ...по падежу
- ...по роду
- ...по числу
- ..атрибутивное согласование
- ...по лицу
- ...по падежу
- ...по роду
- ...по числу
- 2.5.4.СЛОЖНОЕ ПРЕДЛОЖЕНИЕ
- .О
- ..линейный порядок компонентов
- ...не фиксирован
- ..особенности подчиненного компонента
- ...оформление сказуемого
- ...зависимые личные формы
- ...нефинитные формы
- ...финитные формы
- ...абсолютные обороты
- ..структура относительного предложения
- ...главное предшествует придаточному
- ..тип построения
- ...сочинение/подчинение

.тип связи	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
..союзная/бессоюзная		<input type="checkbox"/>	<input checked="" type="checkbox"/>	
..союзы	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
...знаменательные слова		<input type="checkbox"/>	<input checked="" type="checkbox"/>	
...самостоятельные служебные элементы		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
...союзные формы глагола		<input type="checkbox"/>	<input type="checkbox"/>	

Примечание: в связи с ограниченным объемом отчета подробный анализ указанных неточностей не приводится.

4. Сопоставительный анализ описания языков в энциклопедии “Языки мира” и БД “Языки мира”

В тексте энциклопедии “Языки мира” содержатся номера и названия позиций типовой схемы статьи. В некоторых позициях имеются только номера. Часть информации, тем не менее, требует квалифицированного анализа для отнесения в ту или иную позицию реферата БД “Языки мира”.

Так, в частности при проведении эксперимента по реконструкции диахронных изменений в системе вокализма в Славянских языках по линии Праславянский → Старославянский → Древнерусский → Русский, было обнаружено, что содержащаяся в тексте на естественном языке с помощью традиционной терминологии обширная информация оказывается размещенной в различных разделах реферата. Сведения могут присутствовать в неявном виде (в виде набора признаков и характеристик) и могут быть обнаружена в результате тщательного поиска.

Кроме того, практически отсутствуют формальные средства выражения вариативности. Например, различия в ударном и безударном вокализме в языке или отдельных диалектах языка могли бы быть переданы путем выбора более чем одной строки в одном и том же подразделе реферата. Но поскольку в целом описание вариантов не систематизировано, выбор более чем одной характеристики приводит к искажению данных по сравнению с реальной картиной в языке.

4.1. Источники неполноты/неточности/ошибок

Информация в статьях подразделяется на два типа. Часть сведений/явлений обязательна для всех статей о языках. Отсутствие информации данного типа значимо, т.к. оно свидетельствует об отсутствии явления. В случае отсутствия информации не должно

оставаться неясности: а) данных нет, потому что вопрос не изучен; б) данными не располагает автор; в) соответствующие сведения приводятся в другой позиции статьи. Таким образом, информация может отсутствовать в статье только в случае абсолютной неизученности данного явления.

В структуре реферата БД “Языки мира” явные указания об отсутствии данных не систематизированы. Часть позиций содержит в себе строку “отсутствует/отсутствуют”, но это не относится к случаям отсутствия информации, которые оговариваются в статьях. Например, отсутствие данных о числительном в реферате “Старославянский язык”, может свидетельствовать как об отсутствии данного класса, так и о спорности, неустановленности и отсутствии данных о числительном как о явлении.

СТАРОСЛАВЯНСКИЙ

2.3.7. ЧАСТИ РЕЧИ

.знаменательные

..глагол

...финитные формы

...нефинитные формы

..местоимение

..наречие

..существительное

..числительное - не отмечено в реферате

При этом в других разделах реферата, в том числе в описании парадигм содержатся данные:

2.4.0. ПАРАДИГМЫ

.глагол

..финитные формы

.числительное

..парадигма не совп.с именной по составу категорий

Другой тип информации относится к сведениям о явлениях, которые необязательны для характеристики, но составляют специфику данного языка. Наличие/отсутствие таких явлений в языке не отражено в структуре реферата, например отсутствие данных о дифтонгах/трифтонгах в реферате не может свидетельствовать об их отсутствии в данном языке. Например:

СТАРОСЛАВЯНСКИЙ

2.1.1. ФОНЕМНЫЙ СОСТАВ

.гласные

..общая схема монофтонгов

...различия по лабиализации *отсутствуют данные в реферате*

....отсутствуют
в переднем ряду
в среднем ряду
в заднем ряду
в переднем и среднем ряду
в переднем и заднем ряду
в среднем и заднем ряду
во всех рядах

..дифтонги **отсутствуют данные в реферате**
 ..только восходящие
 ..только нисходящие
 ..восходящие/нисходящие
 ..долгие/краткие
 ..нисходящие/равновесные
 ..восходящие/нисходящие/равновесные

4.2. Локализация данных

В силу универсальной ориентации типовых схем некоторые специфические/редкие (но интересные и значимые для языка явления) могут не найти места в рамках схемы. Если в тексте статьи такие специфические грамматические явления могут быть указаны или подробно охарактеризованы в другом месте статьи или в примечаниях за пределами схемы, то в реферате такая возможность исключена. Например, такое существенное определение как *вокалический строй*, отражающее специфику славянских языков на ранних этапах их формирования и отличающее их от *консонантного строя* развившихся на его основе более поздних славянских языков, не находит своего выражения в модели реферата, т.к. соответствующие термины отсутствуют. В отличие от приведенных выше объективных причин отсутствия данных в рефератах (см. выше) в этом случае отсутствие специальной терминологии объясняется причинами иного порядка.

Модель реферата в целом основана на принципах морфологической классификации языковых фактов и явлений. Этим объясняется то, что данные о языке структурируются по формальным критериям: характер морфем, способ их соединения (пре- или постпозиция). Способ соединения грамматических морфем имеет непосредственное отношение к синтаксису: агглютинация – фузия (морфонология), соотношение морфемы и слова (изоляция, аналитизм/синтетизм словообразования и словоизменения) и т.д.

Морфологическая типология и соответствующая классификация языков характеризует не конкретные языки, в данном случае славянские, а основные структурные явления и тенденции, существующие в разных языках, в которых всегда представлено несколько морфологических типов.

Очевидно, что оптимальная универсальная модель описания языков (идеальная унифицированная модель реферата) должна сочетать в себе разнотипные классификационные критерии, максимально соответствовать конкретным языковым структурам и базироваться также на фонологических и историко-генетических основаниях. В существующей модели возможности классифицировать языки по типу фонетического строя практически отсутствуют. Косвенно сведения об особенностях вокалического и консонантного строя славянских языков можно было бы получить по позициям о типах письменности, но данные о письменности в окончательном формате реферата также отсутствуют. Едва ли не единственным остается вариант последовательного сопоставительного анализе позиций

2.1.1. ФОНЕМНЫЙ СОСТАВ

..общая схема монофтонгов

..число монофтонгов

...

по линии от позднеиндоевропейской системы вокализма до системы вокализма старославянского, поскольку по фонетическому и морфологическому строю он был близок праславянскому языку позднего периода.

Тесно связанные явления, представленные в позиции 2.1.2. ПРОСОДИЧЕСКИЕ ЯВЛЕНИЯ, не могут быть источником информации, т.к. данные в них в реферате о старославянском языке отсутствуют. О проблемах, связанных с данными об исторической эволюции гласных, даже если она хорошо установлена и засвидетельствована в языке, будет также сказано в 4.3. Анализ терминологии.

Таким образом, составление рефератов и интерпретация содержащейся в них информации требует достаточно высокой лингвистической квалификации и должно опираться на тщательный анализ материала, часто выходящего за рамки отдельной статьи о конкретном языке. В противном случае не исключены значительные потери информации, неточности и ошибочность данных.

4.3. Анализ терминологии

В энциклопедии “Языки мира” используется общепринятая современная терминология. Предположительно, терминология в статьях энциклопедии должна быть унифицирована и совпадать с терминологией в модели реферата БД “Языки мира”. Однако, поэтапная формализация содержания статей и унификация схем описания привели к потерям, неполноте и неточностям в описании целого ряда языковых явлений.

Анализ текстов статей показал, что терминология, характерная для описания славянских языков, не всегда уточняется в пределах одной статьи, например в общей статье тома “Славянские языки” или других статьях о группах языков. Имеются случаи как полной синонимии, например:

первое смягчение = первая палатализация
вторичное смягчение = вторая палатализация

так и частичной, например: аблаут = апофония

Имеются расхождения в терминологии между статьями энциклопедии “Языки мира” и моделью реферата БД “Языки мира”, например:

музыкальное ударение = количественное ударение

предпоследний слог = пентультима

Иногда синонимия является далеко не полной, как это происходит с термином *слоговая интонация*, который используется в энциклопедии “Языки мира” и отсутствует в реферате БД. Этот термин нельзя однозначно трактовать как синоним термина *музыкальное ударение*, который имеется в реферате. Таким образом, описываемое с его помощью явление может быть охарактеризовано не вполне адекватно с помощью целого набора признаков в составе других позиций. Процедура поиска информации может быть проиллюстрирована с помощью эксперимента по реконструкции данных о *слоговой интонации*, фрагмент которой приводится ниже.

5. О ручном эксперименте по реконструкции описания Праславянского и Древнерусского языков и диахронных изменений по линии Праславянский → Старославянский → Древнерусский языки на основе данных об их фонетических системах.

В истории славянских языков древнейшее музыкальное дарение находилось в сложном взаимодействии со слоговой интонацией (акут — циркумфлекс) и долготой гласных и в дальнейшем преобразовалось в большинстве славянских языков в силовое ударение. К ранней праславянской эпохе относится явление изменения ударения в пределах словоформы, которое привело к появлению подвижного ударения в морфологических парадигмах.

Для иллюстрации процесса реконструкции описания Праславянского и Древнерусского языков и диахронных изменений по линии Праславянский → Старославянский → Древнерусский языки на основе данных об их фонетических системах приведем один из фрагментов динамической модели реконструкции данных о *слоговой интонации* с помощью ручного эксперимента.

Термины *акут* и *циркумфлекс* употребляется в описании *слоговой интонации* Праославянского языка в статье “Славянские языки”, “Старославянский язык”, “Древнерусский язык” в качестве стандартного (по анализу других источников) и существенного для отражения специфики Славянских языков. В модели реферата эти термины отсутствуют в явном виде. Частично их описание восстанавливается по косвенным сведениям в других позициях.

В модели реферата термины имеют характеристики *акутный* и *циркумфлексный* в составе позиции

2.1.2.ПРОСОДИЧЕСКИЕ ЯВЛЕНИЯ

.ударение

..тип

..вид

..носитель ударения

..фиксированность

.тон

..регистровые признаки

..контурные признаки

...акутный/грависный

...ровный/циркумфлексный/нисходящий

...восходящий/циркумфлексный/прерывистый

Но по схеме они являются характеристиками тона, а не ударения, хотя являются сложным видом и того и другого.

Таким образом, информация по таким существенным языковым явлениям как *акут* и *циркумфлекс* (крайне важным для характеристики не только славянских языков) и динамике их эволюции, не находит явного выражения. Поиск по сложной траектории через другие признаки, требует обращения к целому ряду других позиций и является длительным и неэффективным.

Из определения *циркумфлекс* явствует:

циркумфлекс =

циркумфлексное ударение

реализуется в виде двояконаправленного движения голосового тона

нисходяще-восходящее

восходяще-нисходящее

является разновидностью музыкального ударения

противопоставляется

острому ударению

тупому ударению

СИНОНИМЫ

циркумфлексная слоговая интонация

облечённый тон

облечённое ударение

В истории славянских языков различаются
 старое циркумфлексное ударение
 на слогах с кратким дифтонгом
 новое циркумфлексное ударение
 развилось на месте старого акутового (острого) ударения

В славянских языка циркумфлексное ударение представлено в праславянском, старославянском, сербскохорватском языках, где оно реализуется на слогах с долгой гласной. В современном русском отсутствует.

Адекватная модель формализованного описания могла бы иметь вид:

2.1.2. ПРОСОДИЧЕСКИЕ ЯВЛЕНИЯ

ударение

вид

музыкальное

акут { в модели отсутствует }

старый { в модели отсутствует }

новый { в модели отсутствует }

носитель ударения

фонема

слог

ударный

или

.тон

..число тонов

..регистровые признаки

...

..контурные признаки

...

акут { в модели отсутствует }

старый акут { в модели отсутствует }

новый акут { в модели отсутствует }

циркумфлекс { в модели отсутствует }

старый циркумфлекс { в модели отсутствует }

новый циркумфлекс { в модели отсутствует }

Ударение исторически изменчиво и в процессе развития языка один тип может сменяться другим. Поэтому включение терминов *акут* и *циркумфлекс* отчасти восполнило бы лакуны, образовавшиеся в результате исключения из модели реферата позиций 1.1.1. – 1.1.6., т.к. наличие этого явления в языке является свидетельством уникальности вокализма, явной или опосредованной генетической или структурной связи и синхронных или исторических контактах данного языка с другим(и) языками, например, музыкальное ударение имеется в литовском (контакты), сербском (генетическое родство) и т.д.

6. Перспективы

На данном этапе ясно, что получение исторической и диахронной типологической информации и построение общей модели анализа такой информации на основе рефератов БД “Языки мира” в принципе возможно.

Практические результаты экспериментов позволяют пополнять БД “Языки мира”. С одной стороны это приведет к более полному описанию отдельных языков (редактирование рефератов) с помощью данных из совокупности статей энциклопедии “Языки мира” о семье/группе, полученных в результате их детального анализа. С другой стороны, восполнит неточности и неполноту, вызванную имеющейся структурой модели реферата.