

А.К. Зотова (Институт языкознания РАН)

A.K. Zotova (Institute of Linguistics, Russian Academy of Sciences)

**Специализированные электронные ресурсы: терминосистема базы данных
«Языки мира» и проблемы разработки русско-английского синонимического
терминологического указателя**

**Specialised e-resources: the terminological system of the “Languages of the world”
database and the problems of the Russian-English synonymic terminological index**

Аннотация

В статье описываются исследования лингвотипологической базы данных «Языки мира», созданной в ИЯз РАН. Основное внимание уделяется проекту русско-английского синонимического терминологического указателя, работа над которым включает сбор обширного фактического материала и тщательный анализ терминосистемы базы данных. Проект может представлять интерес для специалистов в области сравнительно-исторического языкознания, прикладной лексикографии и служить источником лингвистической терминологии, в том числе для дидактических целей.

The paper outlines a series of studies on the linguistic typological database “Languages of the world” (DB) developed in the Institute of Linguistics, the Russian Academy of Sciences. Special attention is paid to the project of the DB’s Russian-English synonymic terminological index. The project involves a large number of empirical data collection as well as the DB’s terminological system thorough analysis. The research could be of interest for comparative linguists, applied lexicographers and serve as a source of linguistic terminology for didactics purposes.

Ключевые слова

База данных, лингвистическая терминология, лингводидактика, терминосистема, языки мира

Database, linguistic terminology, language pedagogy, terminological system, languages of the world

Цель статьи — представить основные направления исследований на материале лингвотипологической базы данных (БД) «Языки мира», проводимых в секторе прикладной лингвистики Института языкознания РАН. База данных «Языки мира» разрабатывается в Институте языкознания на протяжении ряда лет [1; 2; 7; 8]. База включает в себя формализованные и структурированные описания языков мира в виде специально разработанных универсальных моделей (т.н. *рефератов*) и основана на издаваемой в Институте языкознания РАН энциклопедии «Языки мира».

На этапе практической реализации Интернет-версии БД актуальным становится вопрос о справочном аппарате, а именно об указателях к ней. Как известно, БД «Языки мира» разрабатывалась в Институте языкознания РАН на основе оригинальных статей, написанных авторами для Энциклопедии «Языки мира». Таким образом, оба источника представляют собой результаты исследований нескольких поколений лингвистов. Помимо ценности содержащейся в них фактической информации, эти исследования иллюстрируют широкую палитру школ, направлений и традиций описания языков — тюркологии, германистики, славистики и др. Этот тип информации представляет значительный самостоятельный, в том числе и исторический интерес.

За годы своего существования БД претерпела несколько этапов усовершенствования, связанных главным образом с использованием ее в качестве специализированного электронного ресурса и адаптацией к возможностям Интернета. Объем и исходный вид содержательной информации статей энциклопедии «Языки мира» подвергался поэтапной обработке. В результате были получены два варианта модели реферата, отличающиеся как от первоначальной схемы, так и друг от друга. Вариант 1 — переходный от текста статей Энциклопедии с сохранением полного формата типовой схемы описания языков (но с измененными и/или сокращенными названиями позиций), и вариант 2 — оконча-

тельный вид модели реферата в Интернет-версии БД, сокращенный за счет таких позиций как:

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.

Одновременно с работой по совершенствованию БД, она используется в качестве материала для сравнительно-исторических, типологических и прикладных исследований. В частности, сотрудниками сектора прикладной лингвистики исследовались формально совпадающие классы явлений (универсалии, фреквенталии, раритари и уникалии) в семьях и группах языков (германских, романских, славянских). Затем был проанализирован потенциал БД на предмет получения сведений об исторических и диахронических изменениях на материале славянских языков [3; 4]. Важно отметить, что эти исследования принципиально отличаются от количественных компьютерных экспериментов, в которых БД рассматривается как корпус описаний, позволяющий производить масштабные расчеты математического или статистического характера [6].

Поиск форм представления исторической и диахронической информации показал, что необходима тщательная экспертиза как формата представления, так и фрагментов содержания БД. Были проанализированы источники неточностей и ошибок, связанные с адаптацией описания явлений и признаков, имеющихся в Энциклопедии, к требованиям реферата и затем модели реферата в его окончательном виде. Требования унифицирован-

ного описания приводят к заданной в модели реферата формальной классификации и искусственному отнесению признаков к той или иной позиции, что в ряде случаев препятствует адекватному представлению информации. Универсальная ориентация схемы модели определяет ее основное направление — от общего к частному. Однако эксперименты показали, что при этом некоторые специфические и/или редкие, но интересные и значимые с точки зрения как синхронии, так и диахронии явления оказываются не представленными или представленными в неявном виде, а их поиск затруднен или невозможен. Так, например, такое существенное определение, как *вокалический строй*, отражающее специфику славянских языков на ранних этапах их формирования в отличие от *консонантного строя*, развившегося на его основе в более поздних славянских языках, не находит своего выражения в модели реферата, т.к. соответствующие термины отсутствуют. Косвенно сведения об особенностях вокалического и консонантного строя славянских языков можно было бы получить по позициям о типах письменности, языковых контактах и др., но эти позиции в окончательную версию реферата не включены.

Другой пример, связанный с терминологией, локализацией данных и принципами классификации, заложенными в основу модели реферата, — термины *акут* и *циркумфлекс*. Эти термины крайне важны для описания эволюции славянских и ряда других языков. Специфика обозначаемых ими явлений (сложный вид тона и ударения) указывает на связь с определенным периодом общей истории и сохранении их в части современных славянских языков. Эти термины могли бы восполнить лакуны, возникшие в результате отсутствия в модели реферата позиций 1.1.1. — 1.1.6. Описываемые ими явления свидетельствуют об уникальности вокализма, явной или опосредованной генетической или структурной связи, синхронных или исторических контактах между языками, например, музыкальное ударение имеется в литовском (контакты) и сербском (генетическое родство) и т.д. Во многих случаях уточнение терминологии с помощью возможных синонимов и английских эквивалентов может стать одним из инструментов решения проблемы коррекции данных. Естественным следствием стала разработка синонимического терми-

нологического русско-английского указателя на сопоставительном материале описаний языков в БД и Энциклопедии «Языки мира».

В настоящее время ведется работа по изучению терминосистемы БД в проекции ее на терминологию Энциклопедии «Языки мира». Анализ осуществляется путем последовательного сопоставления описаний языков в БД и Энциклопедии «Языки мира» и включает в себя анализ терминологии оригинальных статей, написанных авторами для Энциклопедии, справочного аппарата к отдельным томам и терминологии соответствующих разделов рефератов БД. Анализируются все типы описания языков — полного или частично структурированного (согласно одной из четырех типовых схем) текста статей энциклопедии, а также варианты модели описания языков в БД.

Одновременно разрабатывается лексикографическая модель указателя и формируется словник. В первую очередь выделяется *фактографическая* информация, собственно термины/термины-синонимы, и эквиваленты *классифицирующей* информации — элементы формулировок, которые используются в рефератах в виде отдельных строк. Соотношение общего числа строк модели реферата (3821) и строк, относящихся к каждому из этих типов информации, составляет приблизительно 1:10, т.е. 400 и 3400 соответственно. Например:

2.1.1. ФОНЕМНЫЙ СОСТАВ

...

3 .гласные

...

5 ..подъем

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

...

11...верхний/верхний пониженный/средний/нижний

12...верхний/средний откр./средний закр./нижний

13...верхний/средний/нижний откр./нижний закр.

14...средний закрытый/нижний

15...ряд

16...отсутствует

...

19...передний/средний/задний

20...передний/средне-передний/средне-задний/задний

...

2 ...различия по ряду

43...в нижнем подъеме

44.....отсутствуют

...

65...различия по напряженности

67...различия по назализации

69...в переднем и заднем ряду

70...в среднем и заднем ряду

71...различия по фарингализации

72....во всех рядах

73...различие схем

74...у беглых и небеглых

75...у кратких и долгих

...

82...у фарингализованных и нефарингализованных

83...число монофтонгов

84...один

85...два

86...три

109...дифтонги

110...только восходящие

...

- 115...восходящие/нисходящие/равновесные
- 116...трифтонги
- 117...общее число дифтонгов и трифтонгов**
- 118...один
- 119...два
- 120...три
- ...
- 140...тридцать два ¹**

Как видно, строки, содержащие собственно термины, наряду с элементами формулировок, уточняющих их качественные и количественные характеристики, образуют единый квази-текст, который является фрагментированным каркасом терминосистемы БД.

При разработке макро- и микроструктуры указателя рассматриваются различные варианты словника, их преимущества и недостатки. Обобщается и классифицируется фактический материал, анализируются возможные формы его представления. Возможны следующие варианты типов указателя:

1. Русско-английский терминологический указатель с учетом формата модели реферата (указатель тезаурусного типа) [5], например:

Абруптив[ные] < ларингальные признаки < шумные < согласные < фонемный состав — **Abruptives**.

Достоинства данного варианта заключаются в том, что он облегчает поиск термина, так как позволяет сохранить его исходную (измененную/падежную и др.) форму на русском языке. Недостатки: избыточность и повторяемость формулировок, например, неоднократное употребление словосочетаний типа *морфологическое/синтаксическое выражение* в нескольких позициях реферата (2.3.2., 2.3.3., 2.3.4. и т.д.).

¹ Как видно, строки, содержащие собственно термины, наряду с элементами формулировок, уточняющих их качественные и количественные характеристики (выделены в тексте), образуют единый квази-текст, который является фрагментированным каркасом терминосистемы БД.

2. Алфавитный русско-английский синонимический указатель лингвистических терминов с формированием словника путем фильтрации терминов из всего корпуса описания в формате модели реферата, например:

Абессив — абессив, лишительный падеж, изъявительный падеж. *Abessive*.

Аблатив — творительный падеж, относительный падеж, отделительный падеж, удалительный падеж, внешне-местный падеж. *Ablative*.

3. Алфавитный русско-английский синонимический указатель лингвистических терминов с элементами толкового и переводного глоссария.

Параллельно разрабатывается оптимальная структура базовой лексикографической единицы указателя. Работа осуществляется одновременно в практическом и теоретическом направлениях. Продолжается сбор материала для формирования общего корпуса словаря в виде электронной картотеки терминов и терминологических словосочетаний на русском и английском языках. Помимо двух основных источников — рефератов БД и текстов статей о языках в соответствующих томах Энциклопедии «Языки мира», материалом служат многочисленные англоязычные публикации и специализированные электронные ресурсы. Во избежание ошибок и неточностей при подборе эквивалентов, а также для снятия многозначности в случаях несоответствия объема терминов (родовидовые отношения), признано целесообразным, наряду с минимальной русско-английской терминологической парой и ее синонимами в линейном ряду (кальки, транслитерация, традиционные варианты лингвистических школ и др.), включать в единицу словника возможные дополнительные элементы. Одним из таких элементов может служить вертикальный микроконтекст в виде гиперонимов и/или гипонимов. В ряде случаев такой контекст позволит заполнить терминологические лакуны (не обнаружение устойчивых переводных эквивалентов) или объяснить выбор потенциальной контекстной замены на основе принципа «часть — целое», что согласуется с общим форматом базы данных, где терминология используется наряду с обычной лексикой на разных иерархических уровнях — в составе соподчиненных строк модели реферата.

Сбор материала для формирования общего корпуса словаря осуществляется в виде электронной картотеки терминов и терминологических словосочетаний на русском и английском языках. Работа включает в себя идентификацию терминов с учетом парадигматических отношений в иерархической структуре реферата БД, отбор терминов для формирования синонимических рядов на русском и английском языках и классификацию терминологических единиц с точки зрения их частотности и/или актуальности для современного научного узуса. Другими источниками материала являются учебные и терминологические словари, справочники и энциклопедии, а также оригинальные специальные тексты на русском и английском языках (научные статьи, аннотации, рецензии, рефераты и т.д.).

К числу основных трудностей при разработке макро- и микроструктуры указателя можно отнести разнородность и многоаспектность лексики позиций реферата, а также повторяемость терминологических словоупотреблений, которые не являются синонимами и имеют различные контекстные эквиваленты. Например, термин *postposition* в значениях ‘постпозиция’ и ‘энклитика’ в разделе реферата 2.3.6. Дейктические категории и ‘послелог’ в разделе 2.3.7. Части речи.

При подборе переводных эквивалентов ряда терминов, найденных в оригинальных источниках на английском языке, обнаруживаются лакунарные терминологические образования, не имеющие регулярных однозначных русских соответствий в виде отдельных терминов и/или устойчивых словосочетаний. Общее количество таких случаев окончательно не установлено. Одним из возможных решений может стать применение типичных для практики перевода специализированных и научно-технических текстов техник — описания, интерпретации, транслитерации, например, передача оппозиции *motile/inert* эквивалентной парой ‘мотильный/инертный’ для характеристики субъектно-объектных отношений в структуре простого предложения с линейным порядком членов типа VSO в позиции 2.5.3. Простое предложение.

В теоретическом плане качественная оценка традиционных описаний языковых явлений (с учетом несовпадения принципов классификаций и различия школ, терминологии

гических традиций и времени появления терминов) и терминосистемы БД позволяет проследить динамику и основные тенденции развития терминообразующих моделей. В большинстве случаев наблюдается унификация форм с помощью использования латинизированных терминов, калькирования иноязычных, преимущественно английских и немецких, конструкций, в том числе синтаксическое калькирование. Структура атрибутивной пары *прилагательное + существительное* преобладает над парой *существительное + существительное*, например, *verbal paradigm vs verb paradigm*, или *morphemic alterations*, которым в модели реферата соответствуют словосочетания ‘парадигма глагола’ и ‘чередование морфем’ соответственно.

Каждый из элементов сопоставляемой пары *статья Энциклопедии — реферат БД* может рассматриваться под разными углами зрения. По времени создания они могут отстоять друг от друга весьма значительно. Имеет место несовпадение принципов классификаций, традиций описания (различные школы) и используемой терминологии. Интерпретация терминологии, характерной для отдельных лингвистических традиций, затрудняется тем, что ее расшифровка может содержаться в разных статьях или в одной из общих статей тома о семье/группе языков. В общем случае терминологический указатель имеет вид краткого списка сокращений, как например, в томе энциклопедии «Тюркские языки»:

Сокращения грамматических терминов

дат.-вин. – дательно-винительный афф[икс]

вр[емя]

буд[ущее]

давнопрош[едшее]

наст[оящее]

прош[едшее]

деепр[ичастие]

л[ицо]

накл[онение]

усл[овное]

отриц [ательная форма]

п[адеж]

вин[ительный]

дат[ельный]

дат.-местн. – дательно-местный

и т.д. Всего 45 единиц с элементами иерархии.

Результаты теоретического анализа терминологических расхождений между моделями описания явлений в семьях и группах языков могут быть использованы для решения практических задач:

— верификации сведений о языках и выявление адекватности их описания в БД «Языки мира»;

— повышения качества дальнейшей работы по пополнению и редактированию БД «Языки мира» на этапе перехода от текстового описания к формальному представлению данных о языках;

— оптимизации применения БД «Языки мира» в дидактических целях при интерпретации сведений из формальных описаний, в том числе с использованием иных, чем это предусмотрено типовой моделью реферата, терминов и классификаций.

Именно такой потенциал предполагается создать и сделать его отличительной особенностью указателя. Для этого планируется включать в него не только унифицированную терминологию, принятую в БД, но и термины, традиционно используемые авторами, представляющими различные лингвистические школы. Крайне важно при этом использовать определения, характеризующие уникальные явления, свойственные отдельным языкам, в том числе редким, исчезающим и мертвым.

Таким образом, разработка дополнительных указателей может расширить и оптимизировать возможности использования БД «Языки мира» Института языкознания РАН в исследовательских, справочных и дидактических целях.

Литература

1. Виноградов В.А., Новиков А.И., Ярославцева Е.И. База данных «Языки мира» как инструмент лингвистического исследования // Вопросы языкознания, № 3, 2003.
2. Журинская М.А., Новиков А.И., Ярославцева Е. И. Энциклопедическое описание языков мира (теоретические и прикладные аспекты). М., 1986.
3. Зотова А.К. Лингвотипологическая база данных «Языки мира»: Диахронные изменения в семьях и группах языков (Славянские языки). 2012 г. // http://www.ilingran.ru/zotova/2011_report.
4. Зотова А.К. Синхронный и диахронный аспекты базы лингвотипологических данных «Языки мира» // Материалы Международного фонетического конгресса по фонетике. Таллин, 1987.
5. Никитина С.Е. Тезаурус как способ описания и представления языка науки (на материале лингвистической терминологии). Дисс. канд. филол. наук. М., 1983.
6. Поляков В.Н., Соловьев В.Д. Компьютерные модели и методы в типологии и компаративистике. Казань, 2006.
7. Ярославцева Е. И. Грамматикон и база данных «Языки мира» // Проблемы прикладной лингвистики 2001. М., 2002.
8. Ярославцева Е. И. Географический и генетический указатели к базе данных «Языки мира» // Проблемы прикладной лингвистики. Вып. 2. М., 2003.