

Из каких элементов состоит метаязык лингвистики?¹

What kind of elements does the metalanguage of linguistics consist of?

Крылов С. А. (krylov-58@mail.ru)

Институт востоковедения РАН, Институт системного анализа РАН

Обсуждаются типы металингвистических баз данных, рассматриваемых как инструмент систематизации метаязыка лингвистики. Сделана попытка дать классификацию систематизируемых металингвистикой предметов. Перечисляются способы упорядочения металингвистических объектов. Особое внимание уделено подробному исследованию типов семантических элементов, фигурирующих в составе металингвистических моделей.

§ 1. Металингвистические базы данных как компьютерный инструмент инвентаризации элементов метаязыка лингвистики

Вопрос о том, из каких элементов состоит специальный метаязык лингвистики, может решаться разными путями, среди которых — и автоматизированное извлечение устойчивых словосочетаний из корпуса лингвистических текстов², и концептуальная систематизация уже извлечённых из текста терминов³, в том числе в виде формальной онтологии⁴, и автоматический анализ соотношений между экспликансами (терминами, входящими в тексты дефиниций) и экспликандумами (понятиями, разъясняемыми с помощью этих дефиниций)⁵.

Один из возможных подходов к вышеуказанной проблеме заключается в практическом эксперименте по составлению предметных указателей к книгам и статьям по лингвистике в формате компьютерных баз данных (БД). В данном докладе делается попытка очертить круг содержательных проблем, встающих при попытке описать то, какие содержательные объекты стоят за теми единицами, которые состав-

ляют «входы» в предметные указатели⁶ к книгам по лингвистике. Проблема словников указателей обсуждается в §§ 2–3, § 5 и в Приложении; проблема упорядочения предметов в указателе — в § 4.

Информационно-технологический продукт, служащий инструментом систематизации элементов метаязыка лингвистики, называется металингвистической базой данных (МБД). Типам МБД посвящён § 2 данного доклада.

§ 2. Содержательные разновидности металингвистических баз данных

МБД, разрабатываемые в интегрированной информационной среде StarLing⁷, служат инструментом систематизации знаний о лингвистике (а не напрямую о языке), однако косвенно способствуют также систематизации сведений о языке. Можно выделять две разновидности МБД: (1) метанаучные (МН-) МБД (входы в которые являются металингвистическими проекциями научных текстов по лингвистике) и (2) метаобъектные (МО-) МБД (входы

¹ Данная работа получила поддержку РГНФ (грант № 08-04-00190а).

² См., напр., [Митрофанова и Захаров 2009].

³ См., напр., Никитина 2010.

⁴ См., напр., [Соколова и др. 2008].

⁵ См., напр., [Шелов 2009].

⁶ «Предметные указатели» понимаются в широком смысле этого слова (то есть как указатели любых элементов специального метаязыка науки (в данном случае лингвистики), представляющих информационный интерес для пользователя, желающего отыскать в тексте информацию об интересующем его предмете. «Предметные указатели» в узком смысле слова — это то же, что терминологические указатели.

⁷ Об этой системе см.: [Крылов, Старостин 2006]; [Крылов, Старостин 2007].

в которые являются металингвистическими проекциями языковых сущностей).

Входами в МО-МБД служат, например, характеристики языковых общностей (имена языков и народов, языковых ареалов, эпох бытования языка); нарицательные лингвистические термины (ЛТ); имена языковых единиц (в том числе имена таксономических классов внеязыковых сущностей).

§ 3. Объекты систематизации в металингвистике

Следует прежде всего проводить различие между онтологическим (материальным) уровнем, на котором можно выделить объектное множество (оригинал, универсум) с существующими в нём отношениями и гносеологический (эпистемологический, идеальный) уровень, на котором выделяется модельное множество (модель, теория) с заданными на нём отношениями. Эту модель и строит металингвист, воплощающий её в виде грамматики, словаря, указателя, графа, дерева, атласа и т. п.

Исходные универсумы в лингвистике таковы (см. подробнее Приложение ниже).

- I. Универсум языковых явлений
 - IA. Общелингвистический универсум
 - IB. Частнолингвистический универсум
 - IC. Универсум речевых событий
- II. Универсум собственно лингвистики
- III. Мир лингвистических моделей

В целом внутри класса III (модельных объектов) выделяются подклассы объектов, прямо корреспондирующие с объектами класса I. Отличие состоит в том, что объекты класса I первичны, т. е. существуют в долингвистическом мире (до и независимо от лингвистов и их деятельности), тогда как объекты класса III (модели) вторичны, т. е. они суть плоды интеллектуальной деятельности лингвистов. Что касается объектов класса II, то они не относятся ни к онтологии, ни к модельным объектам, так как принадлежат миру субъектов познавательной деятельности.

МН-МБД описывают объекты класса II, а МО-МБД — объекты класса III. Однако объекты класса III (гносеологические объекты) являются моделями объектов класса I (онтологических объектов); благодаря этому МО-МБД косвенно способствуют систематизации знаний о языке.

§ 4. Параметры сортировки в металингвистических базах данных

Основные параметры сортировки объектов лингвистических универсумов и их представления в виде организованных систем таковы.

1. Алфавитное упорядочение. Применяется почти ко всем категориям элементарных объектов (кроме дат).

Разновидности: прямое и обратное (обратное применяется обычно к множествам слов или морфологических единиц).

2. Тематическое упорядочение. Основывается на некоторой заранее известной смысловой классификации и иерархизации инвентаризуемых единиц. Ср. тезаурусные словари, генеалогические перечни языков.

3. Хронологическое упорядочение. Возможно для дат появления текстов, дат публикации словарей и грамматик и т. п.; но также и для имён лингвистов при упорядочении их биографий по датам жизни.

4. Квантитативное (статистическое) упорядочение (ср. словарные статьи в частотных словарях). Объектом статистической иерархизации чаще становятся слова, но бывает и собственно металингвистическая иерархизация, напр., имён лингвистов и их работ по степени цитируемости.

Основные способы упорядочения могут осложняться «гнездованием».

5. Речевое (синтагматическое) упорядочение. Линейный порядок фиксации единиц в БД иконически отображает хронологический (линейный) порядок их появления в тексте. Ср. пословные морфологические разборы и т. п.

6. Дерево (в частности, линеаризованное). Ср. деревья синтаксического разбора.

7. Граф (в частности, дерево). Ср. классификационные деревья, семантические графы; таблицы (напр., «парадигм») и т. п.

Классификации могут быть основаны либо на древовидном, либо на универсальном принципе; реально распространена также комбинированная классификация объектов. Оба типа допускают табличное представление. Тематический порядок всегда отражает ту или иную классификацию объектов. Блок-схемы как особый вид классификации представляют собой связанные ориентированные графы.

§ 5. Применение металингвистических баз данных для составления словаря семантического метаязыка русских грамматических теорий

Одной из любопытных частных задач, решаемых в русле изложенного подхода к металингвистике, является инвентаризация и систематизация

семантических элементов (далее СЭ), т. е. элементов семантического метаязыка (далее СМ) русских грамматических теорий на материале корпуса специальных текстов, воплощающих эти теории (далее КГТ). Такая работа базируется на интегральном лингвистическом подходе к СМ, разработанном в ряде работ автора (1985–2010 гг.).

В состав СМ включается несколько типов единиц.

1) дефиниционно-семантические элементы (ДСЭ), то есть слова и конструкции того естественного (или «полу-естественного») языка, на котором строятся толкования единиц, встречающиеся в КГТ (нередко толкования в текстах из КГТ заключены в марровские кавычки); разновидностью ДСЭ являются глоссы, используемые при переводе речевых отрезков с незнакомых или малознакомых языков.

2) категориально-семантические единицы (КСЕ), то есть слова и словосочетания, называющие те или иные типы явлений действительности, обозначаемых единицами изучаемого естественного языка-объекта. КСЕ составляют лексикон того СМ, на котором строятся суждения о семантике в КГТ. КСЕ могут быть простыми и составными.

2.1) Примеры простых КСЕ («категориально-семантических элементов», далее КСЭ) — *ситуация, положение вещей, факт, действие, деятельность, состояние, процесс, качество, количество, признак, свойство, предмет, субстанция, существо, вещество, человек, лицо, деятель, объект, агенс, паиенс, субъект, орудие, способ, место, время, причина, цель, условие, уступка, следствие, мера, параметр, атрибут, характеристика, характер, оценка, множество, элемент, предел, достижение, момент, период, длительность, кратность, интенсивность, мера, сравнение, сопоставление, тождество, бытие, обладание, локализация, пространство, отношение, часть, целое, деталь, пол, возраст, скорость, форма, размер, протяжённость, содержание, отличие, сходство, мужчина, женщина, ребёнок, детёныш, город, река, житель, средство, транспортное средство, страна, небесное тело, явление природы, следование, предшествование, одновременность, фаза, конец, начало, продолжение, отсутствие, наличие, лишение, наделение, владелец, имущество* и т. п.

2.2) Составные КСЕ («категориально-семантическими комплексами», далее КСК) имеют неэлементарную структуру. В их составе можно выделить стержневой (ключевой, опорный) элемент (см. ниже СЭ, набранные неподчёркнутым **полу-жирным**) и его модификатор (распространитель, уточнитель) (см. ниже СЭ, набранные подчёркнутыми шрифтами).

Модификаторы бывают двух типов: характеризующий и параметрический.

2.2.1) Характеризующий модификатор (см. ниже ФСЭ, подчёркнутые воднистой линией) выде-

ляет некоторую разновидность или подкласс того, что обозначено стержневым компонентом КСК. Ср. *действие физическое, действие речевое, действие ментальное, состояние физическое, состояние психическое, состояние устойчивое, состояние временное, процесс предельный, процесс непредельный, предмет считаемый, предмет не считаемый, предмет одушевлённый, предмет неодушевлённый* и т. п.

2.2.2) Параметрический модификатор (см. ниже СЭ, подчёркнутые пунктиром) позволяет выделить некоторый параметр, атрибут, признак или составную часть того, что обозначено стержневым компонентом КСК. Типичные примеры таких КСК — это: *действие # орудие; действие # результат; действие # интенсивность; действие # способ; или предмет # размер; предмет # форма; предмет # функция; предмет # устройство; или лицо # пол; лицо # возраст; лицо # национальность; лицо # профессия* и т. п.

В реальных текстах возможно одновременное появление характеризующего модификатора и параметрического: *предмет считающий # способ счёта (штуками, парами, дюжинами)*.

3) функционально-семантические элементы (ФСЭ) представляют собой модифицирующие компоненты составных лингвистических терминов, обозначающих языковые или речевые единицы. ФСЭ обозначают семантические характеристики языковых и речевых единиц. Они призваны дать семантическую спецификацию тех единиц, которые обозначаются стержневыми компонентами соответствующих терминов. ФСЭ употребляются в составе трёх типов лингвистических терминов — (3.1.) собственно семантических терминов (ССТ); (3.2.) семантически ориентированных терминов (СОТ), не являющихся собственно семантическими; (3.3.) ономаσιологических СОТ.

(3.1) В составе ССТ ФСЭ обозначают те или иные семантические характеристики внутренней стороны знаков. Стержневыми компонентами ССТ обычно бывают термины *значение, функция, смысл, семантика, употребление, использование*.

(3.2) В составе СОТ, не входящих в состав ССТ, ФСЭ обозначают подклассы двусторонних (значимых) единиц языка в целом, выделяемые по признаку наличия тех или иных (соответствующих) семантических характеристик у внутренней стороны этих значимых единиц. Стержневыми компонентами таких СОТ могут быть разные лингвистические термины, но чаще всего ими бывают термины *слова, глаголы, предикаты, имена, существительные, прилагательные, наречия, предлоги, союзы, частицы, предложения, конструкции, словосочетания, обороты, высказывания*.

(3.3) В составе ономаσιологических СОТ (СООТ) ФСЭ используются в качестве заполнителей второй валентности собственно ономаσιологиче-

ских терминов. В качестве носителей этой валентности обычно выступают термины *имена, названия, обозначения, наименования*, а также глаголы *означать, именовать, называть, обозначать, выражать, передавать* и конверсивные по отношению к ним глаголы *именоваться, обозначаться, выражаться*.

ФСЭ бывают адъективными и субстантивными. ФСЭ в составе большинства СООТ являются субстантивными (исключение составляют такие СООТ, как *реляционные имена, одушевлённые имена, оценочные обозначения, соматические выражения* и т. п.). ФСЭ в составе большинства СОТ, не входящих в состав ССТ, являются адъективными (исключение составляют термины типа *глаголы движения, глаголы чувства, глаголы речи, прилагательные цвета, предикаты оценки* и т. п.).

Субстантивные ФСЭ тождественны соответствующим КСЕ (и потому не нуждаются в отдельной инвентаризации), а адъективные ФСЭ представляют собой адъективные (обычно регулярные, но изредка — супплетивные или квазисупплетивные) дериваты от соответствующих КСЕ. Типичные примеры адъективных супплетивных (или квазисупплетивных) дериватов: *действие* → *акциональный, состоящее* → *статальный, отношение* → *реляционный, восприятие* → *перцептивный, бытие* → *экзистенциальный, вещество* → *субстанциальный, повтор* → *мультипликативный (итеративный), событие* → *эвентуальный* и т. п.

Приложение. Какие «предметы» могут обозначаться словами, составляющими входы в предметные указатели?

I. Языковые явления

IA. Общелингвистический
универсум

(IA.1.) Мир языковой системы: языковая система и ее подсистемы; языковые единицы (ЯЕ); отношения между ЯЕ; члены отношений между ЯЕ (именуемые по их ролям в системе); функции ЯЕ; способы выражения значений; классы ЯЕ; члены классов ЯЕ; языковые структуры; части языковых структур (именуемые по их ролям в структуре); логические связи языковых явлений.

(IA.2.) Речевая динамика: речевые процессы и речевая деятельность (типы, аспекты и компоненты); речевые события, речевые действия (типы, аспекты и компоненты); речевая способность (типы, аспекты и компоненты); речевое варьирование (типы и проявления).

(IA.3.) Языковое функционирование.

(IA.4.) Языковые изменения.

(IA.5.) Языковые сходства и различия.

(IA.6.) Исторические отношения между языковыми общностями.

IB. Частнолингвистический
универсум

(IB.1.) Исторические языковые общности.

(IB.1.1.) Языки.

(IB.1.2.) Надъязыки: генеалогические единства (группы, семьи, макросемьи); контактные единства (языковые союзы); ареальные единства (территориальные множества языков).

(IB.1.3.) Подъязыки (варианты одного языка): диалекты (территориальные, профессиональные, сословные); исторические этапы языка; жанровые подъязыки.

IB.2. Ареалы распространения языков: континенты, регионы, страны, провинции; населённые пункты.

(IB.3.) Частнолингвистические единицы.

(IB.3.1.) Знаковые единицы.

(IB.3.1.1.) Лексические единицы: слова и их варианты (семантические и формальные); фразеологизмы разной протяжённости и их варианты (семантические и формальные), паремии (пословицы, поговорки, загадки, крылатые слова и выражения, ходячие афоризмы).

(IB.3.1.2.) Морфологические единицы.

(IB.3.1.2.1.) Знаменательные морфологические единицы (основы; корни).

(IB.3.1.2.2.) Грамматические (служебные) морфологические единицы (форманты, аффиксы, флексии и их эквиваленты).

(IB.3.1.3.) Конструкции (синтаксические и лексико-синтаксические).

(IB.3.1.4.) Иероглифы (цифры и другие идеографические символы).

(IB.3.2.) Фигуры выражения (фонемы, варианты фонем, буквы, слоги).

(IB.3.3.) Фигуры содержания (семы, понятийные категории, индивидуальные сущности⁸ и т. п.).

IV. Мир речевых событий

(IV.1.) Словесность (множество зафиксированных текстов на некотором языке).

(IV.1.1.) Специальная устная словесность (фольклор): памятники фольклора; сказители; регионы распространения традиции; даты фиксации произведения; персонажи.

(IV.1.2.) Специальные памятники письменности (филологический мир): памятники письменности; внешние субъекты

⁸ Индивидуальные сущности соотносимы с объектами действительности (инвентарь которых фиксируется, в частности, в энциклопедиях и других источниках знаний о мире).

коммуникации (авторы; адресаты; переписчики; переводчики; комментаторы; рецензенты и критики); даты (создания; публикации; переиздания; перевода).

(IB1.3.) Мир неспециальных устных и письменных связанных текстов (дискурсов) и их связанных частей (высказываний): тексты, высказывания, участники речевых актов (говорящий, слушающий, аудитория), а также разнообразные характеристики связанных текстов и высказываний.

IB2. Вхождения речевых знаков-экземпляров (tokens) (ВЭ).

(IB2.1.) ВЭ двусторонних знаков;

(IB2.2.) ВЭ мыслей-представлений или референтов.

(IB2.3.) ВЭ звучаний-представлений (произнесений и написаний), т. е. сигналов

II. Универсум собственно лингвистики

(II.1) лингвисты (в том числе лингвисты-непрофессионалы);

(II.2) лингвистические школы и направления;

(II.3.) лингвистические кружки, общества, ассоциации и т. п.;

(II.4) места, где протекает деятельность лингвистов (континенты, страны, провинции, города)⁹

⁹ «Места» можно подразделить на кабинетные (учреждения) vs. полевые (соотносимые с объектами из универсума IB.2, см. выше).

(II.5) учреждения, где протекает деятельность лингвистов (научные: академии, институты, отделы, секторы и т. п.; учебные: вузы, школы; кафедры; ведомства; редакции; фирмы и т. п.).

(II.6.) Лингвистические события.

(II.6.1.) Памятники лингвистической и пре-лингвистической (филологической, фольклорной, диалектографической) деятельности (ПЛД): грамматики, словари, издания памятников, словарные картотеки и т. п.; даты (создания; публикации; переизданий; переводов) самих ПЛД.

(II.6.2.). Цитирование этих ПЛД (ссылки на ПЛД; отклики, содержащие ссылку на данный ПЛД; даты публикации откликов; авторы откликов).

III. Мир лингвистических моделей

III.1. Описания языков (словари, грамматики и т. п.)

III.2. Описания текстов и речевых отрезков: хрестоматии текстов, издания памятников, продукты транскрипции и транслитерации, переводы, фонетические сонограммы, комментарии, глоссы, формальные представления в виде морфологических и синтаксических «разборов», синтаксических графов (в частности, деревьев зависимостей и составляющих), цепочки трансформационного вывода, толкования примеров и т. п.

III.3. Описания ЯЕ: словарные статьи, правила (законы), исключения к ним и т. п.

Литература

1. *Митрофанова О. А., Захаров В. П.* Автоматизированный анализ терминологии в русскоязычном корпусе текстов по корпусной лингвистике // Кибрик А. Е. (ред.), Компьютерная лингвистика и интеллектуальные технологии. Вып. 8 (15). По материалам международной конференции «Диалог'2009» (Бекасово, 27–31 мая 2009 г.). М.: РГГУ, 2009, с. 321–328 (<http://www.dialog-21.ru/dialog2009/materials//49.htm>).
2. *Шелов С. Д.* Мир лингвистической теории сквозь призму ее терминологии (к итогам лингво-компьютерного анализа одного лингвистического текста) // Исследование познавательных процессов в языке (Серия: Когнитивные исследования языка. Выпуск V). М., 2009.
3. *Никитина С. Е.* Семантический анализ языка науки. На материале лингвистики. М.: Либроком, 2010. — 146 с.
4. *Соколова Е. Г., Кононенко И. С., Загорюлько Ю. А.* Опыт систематизации знаний и интернет-ресурсов для Портала знаний по компьютерной лингвистике // Кибрик А. Е. (ред.), Компьютерная лингвистика и интеллектуальные технологии. Вып. 8 (15). По материалам международной конференции «Диалог'2009» (Бекасово, 27–31 мая 2009 г.). М.: РГГУ, 2009, с. 465–470 (<http://www.dialog-21.ru/dialog2009/materials/html/72.htm>).
5. *Крылов С. А., Старостин С. А.* Интегрированная информационная среда StarLing и её использование в сфере корпусной лингвистики // Лауфер Н. И., Нариньяни А. С., Селегей В. П. (ред.), Компьютерная лингвистика и интеллектуальные технологии. Труды международной конференции «Диалог'2006» (Бекасово, 31 мая — 4 июня 2006 г.), М.: Издательство РГГУ, 2006. С. 303–307 (<http://www.dialog-21.ru/dialog2006/materials/html/Krylov.htm>).
6. *Крылов С. А., Старостин С. А.* Создание и переработка лексических баз данных в интегрированной информационной среде StarLing // Кобозева И. М., Нариньяни А. С., Селегей В. П. (ред.), Компьютерная лингвистика и интеллектуальные технологии. Труды международной конференции «Диалог'2007» (Моск. обл., 1–6 июня 2007 г.), М., Наука, 2007 (<http://www.dialog-21.ru/dialog2007/materials/html/49.htm>).