

Информация энциклопедического характера в прикладном семантическом словаре

On encyclopaedic data in an applied semantic dictionary

Семенова С. Ю. (Sonya_sem@mail.ru)

ИНИОН РАН, Москва

Включение в лингвистический словарь сведений об обозначаемых реалиях, практикующееся в современной лексикографии, отвечает духу когнитивного направления, для которого характерен интерес к целостному восприятию информации. В статье вопрос об энциклопедической информации обсуждается применительно к АОТ-ориентированному семантическому словарю, в котором лексика представлена структурированно, и круг сведений, как лингвистических, так и расширенных, жестко связан с форматами словаря. Рассматривается система энциклопедических функций словаря РУСЛАН; обсуждаются вопросы их заполнения и расширения их круга. Также рассматривается ряд лексических классов и единиц, для которых в словаре целесообразно указание конкретных видов энциклопедической информации.

1. Об энциклопедической информации в традиционных словарях

Разграничение лингвистической информации о слове и внелингвистической информации об обозначаемой реалии составляет одну из традиционных проблем в лексикографии. Оно связано с более общей и «вечной» проблемой вычленения собственно предмета лингвистики и такими ее границами, как сосюрсовское противопоставление внутренней и внешней лингвистики, или, например, более локальное, возникшее в прикладных задачах разделение участников ситуации на семантические и прагматические. В лексикографической деятельности наряду с разграничением стоит и вопрос об объединении данных обоого рода. Например, в одной из последних своих работ В. Г. Гак отмечал тенденцию в общей лексикографии, обратную тенденции к рассеянию информации по разным словарям — включение разнообразных сведений о реалиях в толковые или переводные словари классического типа, стирание грани между языковым и энциклопедическим словарем, шаги к созданию своего рода «энциклопедии языка» [3]. Как отмечено в более давней работе этого ученого, соединение обоих видов информации может осуществляться двумя способами: 1) разработка двух словарей (примерно для одного и того же словника) — языкового и энциклопедического; 2) соединение сведений

обоого рода в одной словарной статье. Последнее в традиционном словаре может быть осуществлено либо путем структурированного выделения энциклопедической информации, в специальной зоне, в виде стандартизованных пояснений, определений, либо более свободно, в том числе — в виде примеров, цитат, дающих представление о конкретных воплощениях описываемой сущности [2].

В отечественной лексикографии включение энциклопедических данных практикуется не только в словарях, предназначенных широкой читательской аудитории и выполняющих просветительские функции, но и в словарях, адресованных в первую очередь лингвистам и базирующихся на четком очерчивании сферы данной науки. Например, такая информация содержится в ряде статей, подготовленных для Интегрального словаря русского языка [13]; в рамках его концепции этнолингвистические (культурологические) аспекты рассматриваются как лингвистически существенные для интегрального описания лексемы [там же, с. 4]. Энциклопедические сведения приведены для таких лексем, как *письмо* (в значении «текст») — данные о способах почтовой коммуникации; *журнал* (как «периодическое издание») — данные о видах периодических изданий, типовом оформлении. Для лексемы *месяц* (как «часть года») указываются сведения астрономического характера; для лексем *месяц* (как «видимый образ луны») и *небо* (в значении «высшие силы») приводятся культурологические материалы;

для лексемы *облако* («скопление в небе водяных капель») даны определенные естественнонаучные сведения; для имени *пробка* («крышка») указывается типовой материал изготовления. В некоторых статьях информация, расширенная по сравнению с минимальным толкованием, введена в поле значения. Например, для слова *забор* названы его части, способ строительства, обычные размеры и материалы.

2. Энциклопедическая информация в автоматических словарях

Соотношение двух видов данных актуально и в компьютерной лексикографии, в том числе в разработках, использующих формализмы для представления сведений о слове. Энциклопедическая информация может «официально» включаться в лексическую базу данных, образуя соответствующую зону, может принципиально оставаться «за кадром», может принимать в себя части толкования, которые не отражаются в лингвистической зоне в силу ограниченности ее форматов, а может, наоборот, отразиться в собственно лингвистической зоне.

В структурированной базе данных предметных имен системы «Лексикограф» [5], не нацеленной на явное представление энциклопедической информации и оперирующей, в основном, с членами толкования, можно, тем не менее, увидеть некоторые следы взаимодействия двух видов данных. Так, судя по примерам в [5], выбор типового предиката, указываемого при описании артефакта, может опираться на некое более расширенное, «энциклопедичное» видение смысла, чем то, которого требует, скажем, «узкое» традиционное толкование предметного имени через родовое; сама задача указания такого предиката предполагает обращение к онтологическому знанию, пусть и обыденному. Так, идеи укрытия, защиты для имени *палатка* (например, туристическая) наследуются от свойств родового имени *помещение* и опускаются при «именном» толковании с интегрирующим и дифференцирующим компонентами («*палатка1* — временное помещение из натянутой на остов ткани, шкуры»; см. словарь

С.И. Ожегова), а задача указания предиката требует эксплицировать их. С другой стороны, некоторые части толкования, не отражаемые в формализме базы (например, «топологический» аспект — наличие двух или трех колес у *велосипеда*; в МАСе и словаре С.И. Ожегова этот факт включен в толкование), либо должны быть вообще опущены, либо уйти в зону комментария, что фактически означает отнесение этой информации к энциклопедии.

Тенденция к сращиванию лингвистической и внелингвистической информации находится в русле когнитивного направления, для которого важны механизмы понимания и их моделирова-

ние, и менее существенно разделение понимаемого на языковые и бытийные составляющие.

Идея сочетания в словарных продуктах лингвистических сведений о предметных именах и энциклопедических сведений о пространственных свойствах обозначаемых предметов, высказывалась, например, в статье [4], относящейся к серии работ, посвященных грамматике вербализации зрительных впечатлений.

Концепция сочетания в прикладной системе сведений двух видов — терминологического словаря и декларативных знаний о предметной области, слияния тезауруса (именно как лингвистического продукта) и онтологии, обсуждается в [8].

Определенный тип объединения лексических и понятийных связей в рамках одного формализма реализуется в тезаурусных системах класса WordNet, в частности, в компьютерном тезаурусе русского языка, в существенной мере опирающегося на идеологию WordNet [1 и др.]. В данном случае традиционные тезаурусные связи («выше-ниже», «часть-целое») несут больше онтологическую нагрузку, а лексические свойства описываются дополнительными по сравнению с классическим тезаурусом связями — валентностными, деривационными и некоторыми другими.

Основная направленность прикладных компьютерных словарей — автоматическая обработка текста, в конечном счете, понимание текста и использование содержащейся в нем информации. Очевидно, что «интеллектуальность» обработки, возможность интерпретирования текста, раскрытия аллюзий, логического вывода должна увеличиваться при расширении словарной информации, при опоре процессоров на энциклопедические данные наряду с собственно словарными.

3. Энциклопедическая информация в словаре РУСЛАН

3.1. Состав энциклопедических функций

Оба вида сведений, языковых и экстралингвистических, представленных в одном словарном продукте, предполагается использовать в рамках Информационно-лингвистической модели обработки текста, предложенной Н. Н. Леонтьевой. Основным семантическим словарным ресурсом системы, разрабатываемой в соответствии с этой моделью, является словарь РУСЛАН [6, 7 и др.; 12 и др.]. Работы в рамках данной модели неоднократно поддерживались грантами РГНФ и РФФИ; в настоящий момент в РГНФ подан проект 09-04-00302а «Взаимодействие словаря и текста в процессе построения ситуативных структур».

В словаре РУСЛАН, параллельно с зонами семантических характеристик, валентностей, тезаурусных связей, лексической сочетаемости, формализованного представления ситуаций и некоторых других, предусмотрена зона энциклопедической информации. Как и лингвистические данные, энциклопедическая информация отражается в формализованном виде, путем использования системы жестких шаблонов.

Это осуществляется с помощью ряда энциклопедических функций, аналогичных лексическим. В настоящее время в метаязыке словаря, разработанном Н.Н. Леонтьевой, предусмотрены функции: *Anti* (антоним — функция, которая для одних лексем может интерпретироваться как энциклопедическая, для других как обычная лексическая); *Cap* (главный, выделенный элемент обозначаемой сущности); *Param* (типовой параметр, по которому характеризуется сущность); *Mesur* (типовая единица измерения — например, если сущность представляет собой измеряемую величину); *Sing* (единичный элемент, если сущность концептуализируется как множество или масса). Имеются также функции БОЛЬШОЙ_РАЗМЕР; ИСХОДНАЯ_ТОЧКА; КОНЕЧНАЯ_ТОЧКА; ПРИНАДЛЕЖНОСТЬ; СОСТАВНАЯ_ЧАСТЬ. Примеры использования функций при описании слов или лексем: *Anti* (*гуманитарный*) = *естественный* (о науках); *Anti* (*холодный*) = *горячий*; *Cap* (*страна*) = *столица*; *Param* (*потолок*) = *высота*; *Mesur* (*температура*) = *градус*; *Sing* (*снег*) = *снежинка*; БОЛЬШОЙ_РАЗМЕР (*камень*) = *глыба*; ИСХОДНАЯ_ТОЧКА (*жизнь*) = *рождение*; КОНЕЧНАЯ_ТОЧКА (*жизнь*) = *смерть*; ПРИНАДЛЕЖНОСТЬ (*воин*) = *оружие*; СОСТАВНАЯ_ЧАСТЬ (*диссертация*) = *глава*.

Кроме того, в словаре есть «географические» энциклопедические функции для нарицательных имен соответствующего семантического поля; это функции ВЫШЕ_ГЕО, НИЖЕ_ГЕО, ЖИТЕЛЬ:ВЫШЕ_ГЕО (*море*) = *океан*; НИЖЕ_ГЕО (*море*) = *озеро*; ЖИТЕЛЬ (*город*) = *горожанин* и т.п. (Топонимическая лексика, для которой эти функции также актуальны (ЖИТЕЛЬ (*Тула*) = *туляк*; ЖИТЕЛЬ (*Германия*) = *немец* и т.п.), составляет отдельную, отраслевую базу данных). Функции ВЫШЕ_ГЕО и НИЖЕ_ГЕО позволяют не только обозначать иерархии, но и отражать буквальную пространственную вложенность географических объектов: ВЫШЕ_ГЕО (*Сибирь*) = *Российская Федерация*).

При значительном количестве функций, заложенных в метаязык РУСЛАНа, в настоящее время энциклопедическая информация регулярным образом указывается лишь для небольшой части лексики, и объем энциклопедических сведений предполагается расширить. Это требует осмысления сфер применения имеющихся функций, разграничения между ними, введения новых функций, доработки форматов, а также выявления лексических классов, для которых целесообразно указание онтологической информации. Далее в статье мы коснемся этих вопросов; рассмотрение будет иметь пока предварительный характер.

3.2. Энциклопедическая информация и часть речи

Энциклопедическая информация в разной мере актуальна для разных частеречных классов. Традиционно ее приводят, в основном, для номинативной лексики, в том числе для предметных имен. Круг словарных статей в [13], для которых указаны энциклопедические данные, подтверждает это. В самом деле, номинативная лексика, с одной стороны, с ее относительно устоявшимся дроблением на бытийные классы, а с другой стороны, с ее соответствием декларативному, статичному знанию, на котором обычно основываются онтологии, более всего подходит для отработки методик энциклопедических описаний.

Кстати, что также соответствует выбору лексем в [13], в энциклопедической информации часто «нуждаются» существительные, семантические полнозначные (а не тяготеющие к местоименности или метаязыковым единицам, типа *вопрос*, *предмет*, *цель* и т.п.), несколько выходящие за рамки самой общей лексики и отчасти, «наполовину», принадлежащие сферам техники, культуры, экономики, других отраслей. (Специальных, ввелингвистических сведений также может требовать отраслевая номенклатура, но узкоспециальные термины предметных областей мы здесь не рассматриваем, и в словарь РУСЛАН такая лексика попадает в очень незначительном количестве).

Для неноминативной лексики вопрос требует отдельного исследования. Что касается глагольной лексики, то круг глаголов, «требующих» энциклопедическую информацию, по-видимому, невелик. Для определенных пластов производной глагольной лексики, сложившейся по механизмам словообразовательных лексических функций (*финансы* -> *финансировать*), энциклопедические данные подчас можно приводить при описании мотивирующего слова. Названия базовых, «исконных» действий и состояний (*бежать*, *говорить*, *понимать* и др.) и их приставочные дериваты, в основном, не требуют энциклопедических пояснений; базовые глаголы интересны именно лексическим значением. В конкретном словаре РУСЛАН для многих глаголов желательные информативные элементы описания могут быть размещены в перспективных и пока малоосвоенных зонах ситуаций, прагматических импликатур; тем самым зона энциклопедии разгружается. Но энциклопедическая информация все же полезна при описании неких специальных действий, типа названий ремесел, технологий — *сваривать*, *верстать* и др.; она, например, может отражать аспекты, параметры технологических процессов.

Для прилагательных и наречий в РУСЛАНе энциклопедическая информация, в основном, ограничивается пока указанием антонимии (см. п. 3.4), хотя для ряда лексических единиц, кроме того, могут быть указаны внеязыковые сведения — культурологические, например, описывающие социальные сте-

реотипы (для слов *культурный, интеллигентный, взаимы, ва-банк* и других), либо элементы научных сведений (например, для составных слов научной сферы, типа *электромагнитный, злокачественный, полнозначный* и т.п.).

3.3. Энциклопедическая функция «Параметр»

Одной из важных в принципе и относительно частотных в РУСЛАНе разновидностей энциклопедической информации является информация о типовых параметрах обозначаемых объектов. В основном эта функция применяется при описании номинативной лексики. Изначально функция *Param* предназначалась для указания только количественных параметров: *Param (холм) = высота*; *Param (штанга) = вес* и т.п. В дальнейшем целесообразным представляется расширить понимание этой функции, описывать с ее помощью также количественные, не градуируемые в числовой форме признаки сущностей, но способные, как и количественные, принимать варьирующиеся значения при варьировании в объекте: *Param (плод) = спелость*; *Param (краска) = цвет*; *Param (автомобиль) = марка* и т.п. Расширение понятия «параметр» в неколичественную сферу, с сохранением идеи варьируемости значений внутри данного класса объектов, является естественной ступенью абстрагирования от исходной, количественной интерпретации данного понятия [11]. Функция *Param* удобна для «анкетирования» описываемой лексики, помогает лексикографу структурировать онтологические знания об обозначаемом предмете, уяснить, какие свойства для данного объекта существенны (так, абзац как элемент письменного текста характеризуется *содержанием, координатами на странице, длиной, шрифтовым оформлением*).

Указание в зоне энциклопедической информации типовых количественных параметров для данной сущности позволяет, например, при текстовом анализе уточнять конкретный вид параметра в конструкциях с составными количественными прилагательными: *пятилетний ребенок = > возраст*; *пятилетняя дискуссия = > продолжительность*; *пятилетний план = > срок*. Указание типовых неколичественных признаков также способствует распознаванию смысловых связей между текстовыми единицами.

В дальнейшем представляется полезным разделение в словаре энциклопедических свойств на постоянные и переменные для обозначаемой сущности: свойства ее как таковой и свойства, отличающие данный экземпляр в ряду ему подобных. Так, для сущности *август* постоянными свойствами будут продолжительность в 31 день, принадлежность к летнему сезону, наличие погодных предвестников осени (при этом восьмой порядковый номер в ка-

лендарном году считается элементом толкования; см. МАС или словарь С.И. Ожегова), а в качестве переменного свойства, например, может быть указан признак «оценка по погоде», принимающий разные значения в разные годы («теплый», «дождливый» и т.п.).

Переменная информация естественным образом выражается при помощи функции *Param*. Если эта функция обозначает признак с небольшим и «перечислимым» кругом значений, то эти значения полезно добавить в ее формат. Так, для признака «род» (грамматический) целесообразно перечислить значения «мужской», «женский», «средний»; для бинарного признака «истинность» — указать варианты «истина» и «ложь».

Совокупность постоянных свойств будет распределена между разными энциклопедическими функциями, в том числе может быть задействована та же функция *Param*, но с указанием фиксированного значения. Различие между постоянными и переменными свойствами аналогично различиям между прямой и параметрической диатезами некоторых глаголов с валентностью на содержание, в том числе глаголов принятия решений типа *выбирать, рисковать* [10 и др.]. Нотацию для отражения постоянных и переменных значений в РУСЛАНе еще предстоит разработать.

К функции *Param* тематически близка функция *Mesur*. Последняя применяется для представления имен количественных параметров *температура, вес* и др., но уместна также при описании сущностей, не параметризовавшихся до статуса величин, но связанных с величинами, так называемых «квазипараметров» [11]. Например, единица измерения *градус* характеризует не только *температуру*, но и погодные явления, такие как *мороз, жара, заморозки*: *Mesur (мороз) = градус* (ср. атрибутивные конструкции типа *мороз 20 градусов*), а мерами веса характеризуется не только сам параметр *вес*, но и предметные сущности *груз, багаж*, имеющие валентность на значение данного параметра: *Mesur (груз) = килограмм; тонна*. Иногда описание требует указания прототипического масштаба единиц измерения: вряд ли *груз* будет характеризоваться *граммами*, а *гиря* — *тоннами*; *размах* скорее всего будет измеряться в *метрах*, а, например, не в *сантиметрах* и не в *световых годах*. Определенное множество имен параметров (*расстояние, время, ток, частота* и др.) нейтрально к масштабу величин — либо в силу своей общезначимости (*время*), либо в силу научной «беспристрастности» к масштабу величины (*ток*), и для этого множества в словаре указываются либо основные «антропоцентрические» единицы — *метр, килограмм, минута, час*, либо основные (т.е. не кратные 10 в n-ой степени) физические единицы — *ампер, ватт, герц* и т.п.

Детализированное описание средств выражения параметрической информации полезно для

сплошного анализа текста, а также для задач типа Information extraction, для извлечения параметрической информации из текста и представления в виде структур, подобных традиционным информационным триадам или объектно-характеристическим таблицам, для совершения математических и логических операций над извлеченными из текста данными [11].

Вообще, параметрическими структурами (принимаемыми широко, а не только количественно) охватываются значительные слои знаний, выражаемых естественным языком; в определенном смысле параметрическая структура, мыслимая либо как двухчастная (признак — значение), либо как трехчастная /«триада»/ (объект-признак — значение), либо четырехчастная (объект-признак-условие — значение), но с неизменным разделением на главные, разнородные по своей природе части — «левую» и «правую» — на признак и его значение, гомоморфна структуре суждения, предложения, также в главном своем членении двухчастной — субъектно-предикатной; темо-рематической.

3.4. Функция «Антоним» — лексикологические и энциклопедические аспекты

Широко практикующимся при ведении словаря РУСЛАН является указание антонимов. В данном словаре не предусмотрено деление антонимов на лексические и энциклопедические; для простоты оба типа лексем, находящихся в противопоставлении к данной, вносятся в энциклопедическую зону. Лексические антонимы понимаются традиционно — это лексемы, толкования которых на уровне примитивов отличаются отрицанием одного из семантических компонентов или расположением значения соответствующего компонента симметрично на противоположных полюсах шкал «много-мало», «хорошо-плохо» (*горячий — холодный; добрый — злой*). Неточные лексические антонимы, кроме того, отличаются дополнительными оттенками значений, например, незначительным нарушением симметрии на полюсах (*горячий — прохладный; добрый — недоброжелательный* и т.п. [9, с. XV]).

Антонимы энциклопедические составляют некий дополнительный, объемлющий пласт лексических пар, для которых нарушаются те или иные элементы приведенного определения. Для них, например, вместо прототипически статичного положения на шкалах может отмечаться противоположная направленность процессов: *анализ — синтез, индукция — дедукция* и др. Противопоставляться могут символические представители подлинных оппозиций: *зеленый — красный* (о сигналах светофора); *белый — красный* (о политических силах в Гражданскую войну); за первой парой стоит бук-

вальное противопоставление разрешения и запрета; за второй — основное политическое противостояние (за частную собственность или против нее). Противопоставляемые единицы могут быть разными по семантической природе; по своей прототипической концептуализации; так, в паре *ядерный — обычный* (о вооружениях) прилагательное *ядерный* означает «относящийся к процессам, происходящим в атомном ядре, к использованию энергии атомного ядра» (МАС), а *обычный* — «такой, как всегда, постоянный, привычный» (МАС).

Для энциклопедических антонимов важно противопоставление референтов, практикующееся в реальной действительности.

В отдельных случаях противопоставления, как энциклопедические, так и лексические, могут быть исчерпывающими, когда в данном классе существей возможны только две данные альтернативы: *рай — ад; дескриптивный — нормативный* (о словаре); *цветной — черно-белый* (об изображениях); *аристократия — плебс* (т.е. не-аристократия).

В подавляющем большинстве случаев полной логической взаимодополняемости не отмечается (например, кроме *ядерного* и *обычного* оружия, еще имеется *химическое, биологическое* и др.; кроме *начала* и *конца* имеется *середина* и т.д.). В качестве антонимичных энциклопедических пар выбираются те, для которых противопоставления прагматически актуальны; подспудно учитывается тот факт, что действительность, исходно многообразная, способна выстраиваться под некое главное бинарное противостояние. Так, противопоставление *либерал — коммунист*, по-видимому, является главным для современной российской политической сцены, хотя имеются также сторонники других мировоззрений. В сфере наук наиболее представительной, видимо, является оппозиция *гуманитарные — естественные*, связанная с противопоставлением природы и общества, хотя существуют, кроме того, науки *точные, технические, прикладные*. Кроме *морского* и *сухопутного* транспорта имеется также *речной, воздушный*, но оппозиция *морской — сухопутный*, содержащая идею отрицания («флот» либо «нефлот»), является при классификации транспорта наиболее принципиальной.

Приведение антонимов в словаре полезно для структуризации возможных контекстов описываемой единицы, для раскрытия ее парадигматики.

Проблему при описании антонимии составляет обусловленная ею потребность мелко дробить слово на значения; аристотелевский принцип, согласно которому разные антонимы указывают на разные значения исходной лексемы, едва ли полностью реализуем в прикладном словаре, в котором значения должны распознаваться процессором, и потому их не должно быть много. В РУСЛАНе деление слова на значения, конечно, практикуется, но оно является укрупненным.

Заметим, что при описании антонимии выход за рамки чистой лексикологии практикуется и в других прикладных системах. Так, в русском тезаурусе типа WordNet антонимия рассматривается на уровне синсетов, т.е. по-видимому возможны ситуации, когда антонимия не будет точной для всех абсолютно лексических пар внутри противопоставляемых синсетов [1].

3.5. О других энциклопедических функциях

Функции *Сар*, *СОСТАВНАЯ_ЧАСТЬ* при описании предметной лексики могут, в частности, использоваться для отражения топологических знаний о предмете: *Сар* (*кепка*) = *kozyрек*; *СОСТАВНАЯ_ЧАСТЬ* (*цирк /как помещение/*) = *арена*; *купол*. Пример с именем *кепка* показывает, что энциклопедическая зона может вмещать элементы толкования. Подобная информация в РУСЛАНе не отражается в семантической зоне, где описание смысла представляет собой пересечение таксономических классов.

При работе с функциями *Сар*, *СОСТАВНАЯ_ЧАСТЬ*, *ПРИНАДЛЕЖНОСТЬ* встают проблемы их различения: какой элемент обозначаемой сущности считать выделенным (допускается только один выделенный элемент), а какие — «рядовыми»? Не является ли, например, *арена* не просто составной частью, а главным элементом *цирка*? Функция *СОСТАВНАЯ_ЧАСТЬ* уместна, если объект концептуализируется как составной; например, если это техническое устройство: *СОСТАВНАЯ_ЧАСТЬ* (*компьютер*) = *монитор*. Функция *ПРИНАДЛЕЖНОСТЬ*, например, используется, если описываемым объектом (ее аргументом) является некий действующий субъект, «собственник»: *ПРИНАДЛЕЖНОСТЬ* (*художник*) = *палитра*.

Разграничение сфер употребления этих близких энциклопедических функций входит в более общую задачу разграничения единиц метаязыка. Ряд разграничений среди семантических характеристик и смысловых отношений, составляющих основной инструментарий для описания смысла, приведен в [12]; различительная работа с энциклопедическими функциями еще предстоит.

Анализ лексики показывает, что круг энциклопедических функций, структурированно отражающих знания о мире, может быть несколько расширен в РУСЛАНе. Например, может указываться материал, из которого из которого изготавливается предмет; материал может быть постоянным и переменным: *Subst* (*сабля*) = *сталь* (постоянное свойство); *Subst* (*посуда*) = *фарфор*; *металл*; *пластик* и т.п. Могут указываться химические составляющие: *Content* (*абрикос*) = *каротин*. Может отражаться сфера применения сущности:

Usage (*барий*) = *рентгенодиагностика*; *Usage* (*французский /язык/*) = *дипломатия*. Может указываться типовая манифестация: *Manif* (*недовольство*) = *социальный протест*; *Manif* (*воспаление*) = *отек*.

Наряду с функциями *Сар* и *СОСТАВНАЯ_ЧАСТЬ* может быть введена функция «разновидность»: *РЗВДН* (*ткань*) = *ситец*; *шелк*; *драп* и т.п.; *РЗВДН* (*оружие*) = *холодный*; *огнестрельный*; *ядерный* и т.п. Отчасти эта функция будет дублировать тезаурусные связи, но в то же время, она даст возможность «собирать» терминологические словосочетания, в частности, адъективные, с помощью которых обозначается большое число понятий, иерархически подчиненных данному. (Для РУСЛАНе включение словосочетаний актуально, поскольку в настоящее время тезаурусная зона словаря в основном отражает связи между изолированными словами).

Быть может, целесообразно также введение иерархических связей по предметным областям, подобных связям среди географических объектов: *ВЫШЕ_ЛИНГВ* (*французский*) = *романский*; *ВЫШЕ_БИОЛ* (*мандарин*) = *цитрусовый* и т.п. Этот вопрос зависит от дальнейших путей развития системы, от того, с какие отраслевые задачи могут оказаться возложенными на нее. Отраслевые иерархии могут отличаться от обыденных, определяемых подчас не научными классификациями, а прагматикой. Например, для обыденной картины мира тезаурусная связь *ВЫШЕ* (*собака*) = *домашнее животное* представляется более значимой, чем связь *ВЫШЕ_БИОЛ* (*собака*) = *млекопитающее*.

Энциклопедическая зона, помимо фиксации определенного круга бытийных знаний, снабжает словарь рядом дополнительных лексических связей, и эти связи обогащают его как поисковую систему, гипертекстовую систему, например, увеличивают количество потенциальных входов. В массе своей лексические связи между лингвистической и энциклопедической частями словаря могут поставлять материал для различных информационных экспериментов, выявления новых соотношений. Так, интересный пример несимметричности лексических единиц есть в работе [5], где сравниваются имена *медведь* и *берлога*: для имени *медведь берлога* входит в энциклопедическую информацию (место зимней спячки), а для имени *берлога медведь* — явный семантический компонент.

Конечно, АОТ-ориентированный словарь, создаваемый небольшим коллективом, в настоящее время может вобрать лишь ограниченные сведения онтологического характера. В перспективе, для более существенного расширения информации о мире, в словаре должен быть задействован развернутый аппарат представления знаний.

И очевидно, что энциклопедическая зона связана с прагматикой; отбирая энциклопедические сведения для словаря, лексикограф, с одной стороны, должен исходить из того, что возможно выразить имеющимися средствами, а с другой стороны, должен предсказывать информационную значимость помещаемого материала для потенциального пользователя или потенциальных приложений.

Литература

1. Азарова И. В., Митрофанова О. А., Синопальникова А. А. Компьютерный тезаурус русского языка типа WordNet // Компьютерная лингвистика и интеллектуальные технологии. Труды Межд. конференции Диалог'2003 (Протвино, 11–16 июля 2003 г.). — М.: Наука, 2003. — С. 43–50.
2. Гак В. Г. Проблема создания универсального словаря (энциклопедический, культурно-исторический и этнолингвистический аспекты) // Национальная специфика языка и ее отражение в нормативном словаре: Сб. статей. — М.: Наука, 1988. — С. 119–125.
3. Гак В. Г. От лингвистического словаря к энциклопедии языка // Проблемы русской лексикографии. Тезисы докладов международной конференции Шестые Шмелевские чтения 24–26 февраля 2004 г. — М., 2004. — С. 17.
4. Кобозева И. М. Как мы описываем пространство, которое видим: типы и ранги объектов // Труды Межд. Семинара Диалог'96 по компьютерной лингвистике и ее приложениям. Пушкино, 4–9 мая 1996 г. — М., 1996. — С. 109–111.
5. Красильщик И. С., Рахилина Е. В. Предметные имена в системе «Лексикограф» // НТИ. Сер. 2. — М., 1992. — № 9. — С. 24–31.
6. Леонтьева Н. Н. К теории автоматического понимания естественных текстов. Часть 2. Семантические словари: состав, структура, методика создания — М.: Изд-во МГУ, 2001.
7. Леонтьева Н. Н. Автоматическое понимание текстов: системы, модели, ресурсы: Учеб. пособие для студ. лингв. фак. вузов — М.: Изд. Центр «Академия», 2006. — 304 с.
8. Нариньяни А. С. Кентавр по имени Теон: тезаурус + онтология // Труды Межд. семинара Диалог'2001 по компьютерной лингвистике и ее приложениям. — Аксаково, 2001. — Т. 1. — С. 184–188.
9. Новый объяснительный словарь синонимов русского языка. Первый выпуск. 2-е изд., испр. — М.: Школа «Языки русской культуры», 1999.
10. Падучева Е. В. «Риск — благородное дело»: о системе значений глагола рисковать // Труды Межд. семинара Диалог'2001 по компьютерной лингвистике и ее приложениям. — Аксаково, 2001. — Т. 1. — С. 189–194.
11. Семенова С. Ю. Параметризация как метод познания и как языковой механизм // Логический анализ языка. Квантификативный аспект языка — М.: Индрик, 2005. — С. 466–476.
12. Семенова С. Ю. Если семантический класс широк для слова (К представлению лексики в машинном словаре) // Компьютерная лингвистика и интеллектуальные технологии: Труды Межд. конф. «Диалог 2007» (Бекасово, 30 мая — 3 июня 2007 г.) — М.: Изд. центр РГГУ, 2007. — С. 492–501.
13. Семиотика и информатика. Выпуск 32. Материалы к Интегральному словарю современного русского литературного языка (образцы словарных статей). — М., 1991.