

Семантические поля словаря РОСС: опыт заполнения, анализ дескриптивных возможностей

(Материалы к унификации словарных описаний)

С.Ю.Семенова^[1]

ИНИОН РАН

Решение задач информационного анализа текста в существенной степени зависит от организации словарных данных. К разряду словарных систем, нацеленных на использование в таких задачах, относится Русский ОбщеСемантический словарь (РОСС), разрабатываемый в течение нескольких лет коллективом под руководством Н.Н.Леонтьевой в рамках системы "ПОЛИТекст" (ранее эти работы проводились в Институте США и Канады РАН, в настоящее время они базируются в НИВЦ МГУ) [1-6]. В работах [1,2] подробно рассмотрена структура словаря, его форматы, язык описаний. В [3-6] приведены методики описания отдельных классов лексики – грамматических (прилагательные, наречия) и семантических (параметрические, информационные слова).

Вкратце, словарная статья РОСС содержит сведения о семантических (таксономических) классах, в которые попадает слово (или лексема), о грамматических свойствах, валентностях, таксономии актантов, синтаксической модели управления, о дериватах, тезаурусных связях, лексических функциях, устойчивых словосочетаниях, о некоторых энциклопедических (онтологических) свойствах обозначаемой словом сущности, а также английские эквиваленты слова, отраслевые пометы, эмпирический показатель информационной значимости (вес), примеры типового употребления. Значения многозначных слов, четко различающиеся по отражаемым в словаре признакам, описываются разными статьями. Словарь имеет развитую систему шаблонов, позволяющих составителю выбирать дескрипторы из предлагаемого списка; лишь незначительная часть полей имеет свободный формат. Пример словарной статьи - описание слова "вектор" - приводится в Приложении.

Дополнением к словарю РОСС, охватывающему, в основном, полнзначные слова, служат словари терминологических словосочетаний, служебных слов и устойчивых оборотов (в частности, составных предлогов), географических названий, названий организаций, собственных имен и персоналий; в совокупности все это составляет словарный ресурс системы "ПОЛИТекст". Форматы описания этих единиц имеют много общих черт с форматом РОСС.

Словарь, насчитывающий к настоящему моменту около 7000 входов, является коллективным продуктом; его статьи составляются разными авторами. Общая концепция словаря и стоящей за ним смысловой грамматики, структура, язык описаний, значительная часть словарного наполнения созданы Н.Н.Леонтьевой; последние версии программной оболочки разработал А.В.Сокирко; в разные периоды над статьями, наряду с автором, работали А.С.Панина, М.Г.Шаталова, О.А.Штернова и др. лингвисты; в составлении статей участвовали также студенты РГГУ, интересующиеся компьютерной лексикографией. Работа над словарем авторским коллективом, а также (и это главное!) нацеленность данного

продукта на использование при автоматическом анализе текста, требуют унификации словарных описаний, выработки общих правил написания статей, единообразного употребления предлагаемых дескрипторов разными авторами. (Заметим, что проблема унификации описаний в той или иной мере стоит перед всеми разработчиками коллективных лексикографических систем, см., напр., [7, Предисловие]).

В 1999 г. в РОСС был введен массив объемом более 1500 слов разных частей речи, относящихся к сфере общей и политической лексики, а также к предметной области "Информатика и вычислительная техника". При этом практиковалось преимущественно "сквозное" описание отобранной для ввода лексики - по алфавиту, без предварительного разграничения по частям речи или тематическим классам. Параллельно с описанием новых слов была проделана определенная унификационная работа, в том числе по частичному редактированию ранее введенных статей. Однако словарь продолжает пополняться, и становится необходимой постоянная работа по систематизации охватываемого им материала.

В данной статье хотелось бы обсудить методические моменты, связанные с заполнением семантических полей словаря.

К собственно семантическим полям в ныне действующей версии РОСС относятся поле КАТ (крупная таксономическая категория слова), поле СХ (семантических характеристик - основное и обязательное для заполнения семантическое поле словаря), поле ВАЛ, где указываются валентности - семантические роли участников ситуации (для предикатных слов), и совокупность полей СХ_i (i=1,2,...), описывающих семантику актантов. Другие поля словаря можно классифицировать как грамматические, тезаурусные, энциклопедические и т.д. В Приложении семантические поля выделены графически.

Рассмотрим сложившиеся в процессе работы над словарем принципы заполнения полей КАТ, СХ и СХ_i. Унификация представления информации в поле ВАЛ, связанная с осмыслением номинализации валентностей, прямого и метафорического понимания классических семантических ролей (филлморовских падежей и др.), заполнением лакун во множестве ролей, проведением грани между актантом и сирконстантом, является темой отдельной работы.

Поле КАТ - крупная таксономическая категория слова

Крупная категория слова/лексемы - характеристика, которая в определенной мере является частичечной. Существительным в поле КАТ чаще всего приписывается константа ЭТК_ОБ (этикетка объекта), глаголам и предикативам - ЭТК_СИТ (этикетка ситуации), основной массе прилагательных и наречий - ЭТК_ПРИЗН (этикетка признака), местоимениям - дескриптор МЕСТ (местоимение, местоименность). Однако задание категорий не является грамматическим, в большей мере смысловыми являются категории ЭТК (этикетка), АСПЕКТ, ОТНОШЕНИЕ, ОПЕРАТОР.

Неопределенная категория ЭТК используется при описании существительных, которые обозначают сущности и понятия, отличающиеся высокой степенью абстракции: "мысль", "образ" и т.п. (ср. абстрактные этикетки с "хорошими" ЭТК_ОБ - предметными именами и именами организаций или, например, с периферийными ЭТК_ОБ - названиями символов, шрифтов и др. полуабстрактных информационных или геометрических сущностей: "фигура", "буква", "кириллица", "тире" и др.). Если же существительное являет собой лексическую функцию S₀ от некоторого глагола ("изучение" от "изучать"), то ему присваивается глагольная категория ЭТК_СИТ.

К АСПЕКТным могут относиться параметрические слова разных частей речи ("вес", "весить", "тяжелый 1", "налегке" и т.п.), а также темпоральные и фазовые слова ("вовремя", "начинать", "одновременный", "новый" и др.).

К ОПЕРАТОРАМ отнесены кванторные слова ("произвольный", "повсеместный", "навечно" и др.) или наречия-интенсификаторы ("очень", "гораздо", "значительно" и др.). Идея, лежащая в основе такой номинализации - действие, производимое над некоторой изначальной сущностью (или множеством однородных сущностей).

К категории ОТНОШЕНИЕ, помимо предлогов и союзов, переходящих непосредственно в имя смыслового отношения, отнесены некоторые глаголы - это в большинстве своем стативные глаголы типа "принадлежать", "обозначать", "весить" и др., в том числе - категоризируемые обычно с помощью лексической функции Copul: "являться", "быть". Здесь довлеет идея связанности сущностей; в имя отношения эти слова переходят не сразу, а на этапе коррекции.

Категория МЕСТ понимается несколько шире, нежели обычная частиречная характеристика; наряду с грамматическими местоимениями, местоименными также считаются формально полнозначные слова, для которых в тексте требуется восстановить связь с антецедентом: "разновидность", "вопрос" и др., дейктические слова "нижеследующий", "настоящий 2" и др.

Дескрипторы поля КАТ могут образовывать логические формулы (в общем виде, дизъюнкции конъюнкций), что позволяет комбинировать характеристики, классифицируя слово с разных сторон, например, с грамматической и смысловой. Так, фазовый глагол "кончатся" характеризуется логическим произведением "глагольной" и "аспектной" констант: $KAT(\text{"кончатся"}) = ЭТК_СИТ \& АСПЕКТ$; кванторное слово "всегда" трактуется одновременно как квантор (оператор), и как аспектное слово: $KAT(\text{"всегда"}) = ОПЕРАТОР \& АСПЕКТ$.

Если в одной статье из прагматических соображений оказываются совмещенными разные значения/употребления слова (из-за невозможности разделить употребления по формальным признакам, ввиду лексикализованности отдельных употреблений, малой актуальности каких-то употреблений для конкретного типа анализируемых документов, вследствие оставления "на потом" более детального описания слова и т.п.), то в данном поле может использоваться дизъюнкция констант; ср., напр., конкретное пространственное и переносное употребления слова "вектор" (в Приложении). Вообще в в РОССе как прикладном словаре нежелательно разmultiplication значений: наличие нескольких статей одного слова затрудняет анализ текста. Поэтому часто применяется, так сказать, "мягкое" дробление, когда с помощью дизъюнкций дескрипторов в семантических полях описываются разные коннотации и разные типы употреблений слова.

Наряду с перечисленными категориями в описаниях могут использоваться и дополнительные дескрипторы - для выделения прагматически значимых объектов. Например, в настоящей версии словаря отдельной константой помечаются еще и географические объекты: $KAT(\text{"регион"}) = ЭТК_ОБ \& ГЕО$.

При обработке текста поле КАТ полезно для настройки системы на конкретный вид обработки (анализ /глагольных/ ситуаций, отслеживание изменения ситуаций, поиск параметрической информации, установление межфразовых связей и т.д.). Из совокупности подобных задач, по сути дела, и складывается семантический анализ текста.

Поле СХ - семантические характеристики слова

Поле семантических характеристик так же, как и поле КАТ, заполняется по шаблонам с помощью специально разработанной (как часть смысловой грамматики) системы констант, которые обозначают важные с точки зрения прагматики элементарные /или более сложные/ смыслы. Примеры констант (стандартных СХ словаря РОСС): ПРЕДМ [предмет, предметность], ВЕЩ(ЕСТ)ВО, ВЛАСТЬ, ОРГ [организация, учреждение], ПРОЦС [процесс, процессность], СОБИР [собираемость] и др.

В описаниях СХ применяются и такие характеристики, которые могут выступать в качестве наименований участников ситуации у предикатных слов: ПРИЧИНА, ПАЦИЕНС, ЧАСТЬ, РЕЗУЛЬТАТ и др. Такие константы интерпретируются в словаре, главным образом, как бинарные смысловые отношения (СО) между словом-заголовком (С) и его i-м актантом (Ai) [8]. Как наименования семантических ролей они используются в поле ВАЛ /ВАЛ ("говорить") = АГЕНТ (A1,С), СОДЕРЖ (A2,С), АДРЕСАТ (A3,С)/, но могут выражать смысл и самого слова-заголовка: СХ("получаться") = РЕЗУЛЬТАТ. Особенно естественно с помощью СО описываются так называемые метаслова, к которым относятся, в частности, сами названия ролей, а также их синонимы: СХ("часть", "фрагмент", "отрывок" и т.п.) = ЧАСТЬ.

В качестве дескрипторов в поле СХ могут применяться и лексические функции (ЛФ) - если они выражают смысл слова-заголовка: Ver для "истинный", Won для "интересный" и т.п. Параметрические слова большого и малого полюсов количественной шкалы (прилагательные и наречия) характеризуются, например, при помощи ЛФ Magn и AntiMagn: СХ ("горячий 1") = ...& Magn; СХ ("холодный 1") = ... & AntiMagn. Если слово образовано по стандартному механизму от других частей речи - "противостояние" от "противостоять", "новизна" и "новость" от "новый" и т.п., то в числе прочих дескрипторов в поле СХ указывается соответствующая ЛФ (Si, Ai, i=0,1,... и т.п.) от исходного слова: СХ ("важно 2") = ...& Pred0/предикатив/("важный"). Тем самым, в словаре в определенной мере отражается этимология слова-заголовка (дериваты от него тоже приводятся в словаре как аргументы ЛФ, но уже в другом поле; см. в Приложении дериват "векторный" от "вектор").

Всего сейчас в словаре около 160 констант: примерно 50 СХ, 70 СО и 40 ЛФ. Очевидно, что конечный (хотя и не малочисленный) набор констант не может охватить все онтологические явления, и при описании многих слов приходится использовать логические формы (конъюнкции, а иногда и дизъюнкции), составленные из стандартных констант. Эта процедура требует выработки определенных соглашений, в частности, между разными авторами словарных статей.

Укажем некоторые примеры цепочек СХ, используемых при описании составных (при данном наборе стандартных СХ) таксономических классов существительных.

Названия отраслей науки ("биология", "математика" и т.д.) кодируются конъюнкцией СХ: ОДЕЯТ & ИНТЕЛ - 'область деятельности, интеллектуальная'. Мероприятия (научные, политические и пр. - "конференция", "съезд" и т.д.) характеризуются как ДЕЙСТВ[ие] & ОРГ - 'действие, обладающее статусом организации'.

Текстам, документам присваивается СХ НОСИНФ - "носитель-информации", письменные знаки тоже интерпретируются как информационные объекты: СХ ("запятая") = ЧАСТЬ & НОСИНФ; Программные продукты описываются как 'артефакты, информационные, интеллектуальные': СХ ("сортировщик 2") = АРТ& ИНФ & ИНТЕЛ. Объекты, предназначенные для хранения информации ("фрейм", "файл", "дискета" и т.п.), кодируются как 'вместилища, абстрактные или информационные': СХ ("дискета") = ПРЕДМ & ВМЕСТЛ & НОСИНФ; СХ ("файл") = АБСТР & ВМЕСТЛ & ИНФ.

Части тела человека могут представляться с помощью цепочек ПРЕДМ & ЧАСТЬ & СУЩЕСТ; последняя СХ обозначает то, что связано с бытийной сферой человека: СХ (голова 1) = ПРЕДМ & ЧАСТЬ & СУЩЕСТ; СХ (рука) = ПРЕДМ & ИНСТР & ЧАСТЬ & СУЩЕСТ.

Продукты питания характеризуются как вещества, связанные со сферой человека: СХ ("хлеб") = ВЕЩВО & НАЗН [назначение] & СУЩЕСТ.

Описания СХ глаголов, предикативов, прилагательных и наречий (конъюнктивные цепочки) начинаются с общих, категориальных констант типа ДЕЙСТВ [ие], ПРИЗНАК и др. (см. ниже), отражающих таксономию этих слов как представителей частей речи.

Цепочки СХ глаголов начинаются одной из констант ДЕЙСТВ, СИТУАТ[ивность], ПРОЦЕСС, ПРЕДИК[ативность]. Эти таксономические категории пересекаются с известными глагольными классификациями, например, с вендлеровской. СХ ДЕЙСТВ приписывается глаголам, обозначающим в исходной диатезе целеполагающее действие; СХ СИТУАТ присваивается стативным состояниям типа "спать", глаголам восприятия и нек. др. предикатам, "констатирующим" ситуации; СХ ПРОЦЕСС приписывается, например, предикатам монотонного изменения типа "возрастать", "таять"; СХ ПРЕДИК - связкам и параметрическим глаголам типа "весить". В описаниях также участвуют другие константы, обозначающие более конкретную семантику глагольной лексемы: СХ ("идти 1") = ДЕЙСТВ & ДВИЖ; СХ ("видеть 1") = СИТУАТ & ВОСПР; СХ ("видеть 2" /понимать/) = СИТУАТ & ИНТЕЛ и т.д.

При моделировании семантики глагольных классов, так же как и классов существительных, необходимы соглашения между авторами статей об использовании

одинаковых дескрипторов. Например, речевые акты описываются как 'действия, коммуникационные' /плюс, быть может, какие-то конкретизации/: СХ ("заявлять") = ДЕЙСТВ & КОММУНИК.

Прилагательные и наречия также классифицируются прежде всего категориально: как ПРИЗНАКИ или ОЦЕНКИ. ПРИЗНАКИ - это относительные и нечетко выраженные качественные прилагательные и наречия ("надомный", "домашний", "по-домашнему"), а ОЦЕНКИ - отчетливо выраженные качественные ("длинный", "интересно"). Некоторые многоаспектные относительные прилагательные, отличающиеся широким набором реляций, связывающих их с определяемым словом (ср. "морской" - в море /"рыба"/, возле моря /"курорт"/, для функционирования в море /"корабль"/ и т.п.), характеризуются как ОПР[еделения]; ср. многообразие и определенность семантики в паре слов "морской"(СХ = ОПР) и "приморский" (СХ = ПРИЗНАК) [4].

Категориальные СХ дополняются содержательными: СХ("автоматический") = ПРИЗНАК & ПРИНЦ_Д [принцип действия]; СХ("фактически") = ПРИЗНАК & МОДЛ [модальность] & Ver. Пример соглашений, действующих на классах прилагательных - описание прилагательных цвета с помощью СХ ВОСПР[иятие]: СХ ("фиолетовый") = ПРИЗН & ВОСПР.

Для предикативов используется категориальная константа СИТУАТ плюс конкретика: СХ ("можно") = СИТУАТ & МОДЛ; СХ ("жарко" /на улице/) = СИТУАТ & Magn & Pred0 (жаркий). Таким образом, в поле СХ реализуется компонентное описание значения слова и фактически моделируется его компонентная семантическая структура. В целом, по мере наращивания объема словаря наблюдается тенденция к увеличению длины дескриптивной формулы, к дроблению классов, к вовлечению в описание все большего числа компонентов.

Однокомпонентные описания - редкость в РОССе. Примеры – описания существительных, у которых таксономический класс "попадает в яблочко", имеется точная родовая константа и смысл слова далее не поддается конкретизации при данном множестве дескрипторов: СХ ("фабрика") = ОРГ.

В описаниях СХ слова-заголовка цепочками констант используются элементы синтаксиса [1]. Константы КАУЗ (каузатор, каузация), ОТСУТ (отсутствие), СВЯЗАН (связанность), ЧАСТЬ могут выступать как операторы, "навешиваемые" на стоящие справа от них

стандартные СХ. Например, описание СХ ("запуск")= ДЕЙСТВ & КАУЗ & ФУНКЦ интерпретируется как 'действие, каузатор - функционирования', а запись СХ ("пересылка") = ДЕЙСТВ & КАУЗ & ДВИЖ & ЛОК может прочитываться как 'действие, каузатор-движения и [некоторой] локализации'. Отношение СВЯЗАН, используемое в качестве СХ, позволяет отразить в описании тот факт, что часть семантических характеристик лексемы находится на некотором периферийном, не основном уровне ее семантической структуры; таким периферийным компонентом описания может быть, например, константа ДВИЖ для слова "делегация" /СХ ("делегация") = ОДУШ & СОБИР & СВЯЗАН & ДВИЖ/ или константа ИНСТР [инструмент] для глагола "резать" /СХ ("резать") = ДЕЙСТВ & СВЯЗАН & ИНСТР/. У СХ ОТСУТ "сферу действия" составляет только одна, соседняя справа СХ: СХ ("останавливать 1" (автомобили) = ДЕЙСТВ & КАУЗ & ОТСУТ & ДВИЖ - 'действие, каузирующее отсутствие движения').

Если в одной статье фактически соединяются разные значения/употребления слова, то в поле СХ, так же, как в поле КАТ, используются дизъюнкции логических произведений, соответствующих семантике отдельных лексем (см. в Приложении "вектор").

Очевидно, что информация о СХ слова актуальна на всех этапах информационного анализа текста - от восстановления актантов слова синтаксическим анализатором и построения изолированных семантических узлов до организации пользовательского взаимодействия с базами текстовых фактов.

Поля СХi - семантические характеристики актантов

Форматы этих полей такие же, как у поля СХ. Формальное отличие в практике заполнения сводится к тому, что в описаниях актантов преимущественно используются не конъюнкции, а дизъюнкции дескрипторов - перечисляются альтернативные варианты категорий, по которым может "пробегать" актант слова. Заполнение этой группы полей подчас требует скрупулезного изучения сочетаемости слова-заголовка. Исчерпывающее описание таксономии участников в общем случае вряд ли возможно; часто в СХi предсказываются лишь наиболее частотные либо наиболее массовые классы.

Примеры длинных (и, скорее всего, неполных) цепочек категорий дают прилагательные, которые характеризуются широкой сочетаемостью с существительными (первым и часто единственным актантом прилагательного считается определяемое слово): СХ1("домашний") = ДЕЙСТВ V ОДЕЯТ V ПРЕДМ V СИТУАТ V ОДУШ V ВЕЩВО /"домашние": "работа", "хозяйство", "туфли", "обстановка", "учитель", "пыль" и т.п./. Длинную цепочку порождает, к примеру, и второй актант глагола "продавать": слишком много классов объектов могут служить товаром. Если цепочка а priori кажется почти что бесконечной, бывает целесообразно классифицировать актант с помощью дескриптора ЛЮБОЙ (таковой возможен только в полях СХi), - тогда актант будет искаться в тексте только по грамматическим признакам. Широкий круг классов для актанта можно обозначить также, исключив какие-то невозможные или маловероятные категории с помощью оператора ОТСУТ; таким образом можно, например, задать второй актант у исходного значения глагола "разрабатывать": СХ2("разрабатывать 1"/создавать/) = ОТСУТ & ОДУШ.

Разумеется, формульное описание семантики актанта может быть и однокомпонентным; примером тому служит единственная константа ОДУШ в качестве СХ1 для большинства речевых актов ("приказывать", "спрашивать" и т.п.). Другие, достаточно регулярные, но непрототипические категории агенса для такого представительного глагольного тематического класса, как речевые акты (организация - коллективный агенс, "говорящий" орган, интеллектуальный артефакт, носитель информации, ср.: "Ставка Верховного приказывает"; "сердце приказывает", "программа спрашивает пользователя", "газета

спрашивает своих читателей" и т.п.), по возможности отражаются в словаре, но в дальнейшем они войдут в грамматику регулярных преобразований.

У слов-заголовков, характеризующихся узкой лексической сочетаемостью, CXi подбираются иногда для небольшой группы сочетающихся слов или даже для одного слова: CX1 ("зеркально") = ДЕЙСТВ & КАУЗ & ВОСПР, а конкретный контекст ("отражать", "отображать") указывается в поле словосочетаний или в поле лексических функций. (Актантом типового наречия считается распространяемый им глагол/глагольная группа.) При этом само "узкое", лексикализованное, сочетание может быть весьма частотным, как частотно, например, в экономической литературе сочетание "покупательная способность" / CX1("покупательный") = МОДЛ/.

За многообразием характеристик в дизъюнктивных цепочках полей CXi могут стоять разные языковые явления, в том числе:

- семантическая компрессия: "скорость поезда" /CX1j("скорость") = УСТР[ойство]/ -> "скорость движения поезда" /CX1j = ПРОЦЕСС & ДВИЖ/ (переменная j пробегает по членам дизъюнктивной цепочки);
- изменения в реляциях, связывающих многоаспектное прилагательное с определяемым именем: "московская делегация" /из Москвы, CX1j("московский") = ОДУШ & СОБИР & СВЯЗАН & ДВИЖ/ vs "московская погода" /имеет место в Москве, CX1j = ЯВЛЕН/;
- метонимические (диатетические) сдвиги употреблений слова: "заправлять бензин (в бак)" / CX2j("заправлять 1") = ВЕЩВО/ vs "заправлять бак (бензином)" /CX2j = ВМЕСТИЛИЩЕ/; детальное исследование этого лингвистического явления принадлежит Е.В.Падучевой [9];
- метафорические переходы: "вектор скорости" /CX1j("вектор") = ПАРАМЕТР/ -> "вектор политических сил" /CX2j = СОЦИАЛЬНОЕ/.

Наблюдающиеся в подобных примерах вариации типов употреблений, как уже отмечалось, далеко не всегда возможно и целесообразно выделять в отдельные статьи прикладного словаря.

Заметим, что в настоящей версии словаря пока не используются форматы, позволяющие отражать в рамках одной статьи сведения о зависимости таксономии одних актантов слова от таксономии других; классы разных актантов описываются независимо, в виде параллельных цепочек альтернатив. Формализованное описание взаимной координации этих классов, а также координации грамматических форм актантов заданы в идеологии РОССа, но еще не реализованы программно.

Информация о семантике актантов играет важную роль при уточнении синтаксических связей, при сборке узлов, обозначающих различные факты действительности. Семантические узлы, в которых актанты вершин восстановлены по совокупности смысловых и грамматических признаков, естественно, характеризуются более высоким показателем достоверности, чем узлы, собранные только на основании моделей управления. Для процедур сборки значимы как совпадение, так и определенным образом исчисленная близость таксономических категорий.

Семантическая информация об актантах служит одной из опорных точек для разрешения лексической многозначности в том случае, если значения "разнесены" по разным статьям. Так, по таксономии 2-го и 3-го актантов можно различить пространственное и информационное значения глагола "фиксировать", у которых поверхностные, предложно-падежные формы выражения участников ситуации могут совпадать, ср.: "фиксировать 1" - "повязку" (CX2 = ПРЕДМ) "на ране" (CX3 = ПРЕДМ) и "фиксировать 2" - "внимание общественности" (CX2 = АБСТР & ВОСПР) "на фактах нарушения законодательства" (CX3 = СИТУАТ & ВРЕД); "фиксировать 1" - "конечность" (CX2 = ПРЕДМ & ЧАСТЬ) "в горизонтальном положении" (CX3 = ЛОК) и "фиксировать 2" - "факты нарушений" (CX2 = СИТУАТ & ВРЕД) "в протоколе" (CX3 = НОСИНФ).

Описание многозначности в прикладном словаре осложняется необходимостью распределять дескрипторы между статьями разных лексем слова таким образом, чтобы по возможности избегать коллизий при автоматическом распознавании значений. Иногда для этого приходится дробить категории, иногда - намеренно огрублять описания, пренебрегая какими-то пересекающимися классами.

В примере с глаголом "фиксировать" для обеспечения непересекаемости можно, скажем, оговорить недопустимость СХ НОСИНФ в качестве СХЗ первого значения глагола: СХЗ ("фиксировать 1") = ПРЕДМ & ОТСУТ & НОСИНФ (такая оговорка существенна, поскольку многие носители информации /"книга", "дискета" и т.п./ трактуются одновременно и как предметы; тем самым мы как бы выбрасываем из рассмотрения экзотические употребления "фиксировать 1" типа "зафиксировать этикетку на дискете").

Статьи разных лексем слова могут, кроме того, различаться и по значениям других полей - по грамматике, лексической сочетаемости, тезаурусным связям, пресказывающим разное контекстное окружение у лексем. В целом вопросы представления лексической многозначности в прикладном словаре, равно как и ее автоматического снятия в процессе текстового анализа, составляют тему для отдельного обсуждения.

Благодарность: Автор пользуется возможностью выразить глубокую признательность Н.Н.Леонтьевой за плодотворное и деликатное руководство лексикографическими работами, а также за полезные замечания, высказанные ею по прочтении данной статьи.

Приложение. Пример словарной статьи

ЗГЛ = ВЕКТОР

ГХ [грамматич. хар-ка] = СУЩ: ИГ

ВЕС = 3

ПО [предметная область] = наука; полит

АНГ = vector

ЛФ [лексические функции] = S0 (векторный)

ЭНЦ [энциклопед. функции] = Anti (скаляр)

АСЦ [ассоциации] = отрезок; луч 2; тензор; равнодействующая

КАТ = ЭТК_ОБ V ЭТК

СХ = ПРОСТР[анственность] & НАПРАВЛ[енность] V

АБСТР[актное] & НАПРАВЛ & РЕЗУЛЬТАТ

ВАЛ = 1. СОДЕРЖ[ание], A1, C;

2. ИДЕНТ[ификатор], A2, C

СХ1 = ПАРАМЕТР V ЯВЛЕН V ВЕЛИЧ V СИТУАТ V СОЦ[иальное] V АБСТР

СХ2 = ИМЯ V ЦБК [цифро-буквенный комплекс]

ГХ1 = косв. доп.: Родит.

ГХ2 = прим. опр.: Именит.

ИЛЛ [иллюстрации] = 1. вектор скорости;

вектор напряженности магнитного поля E;

вектор x->;

вектор прерывания (информ.);

2. вектор общественных настроений;

вектор политических сил.

Литература

Леонтьева Н.Н. Русский общесемантический словарь (РОСС): структура, наполнение. // НТИ. Сер. 2. - 1997. - N 12. - С.5-20.

Леонтьева Н.Н. Категоризация единиц в русском общесемантическом словаре (РОСС). // Труды Международного семинара Диалог'98 по компьютерной лингвистике и ее приложениям. - С.519-532.

Семенова С.Ю. Поиск параметрической информации в тексте: алгоритмический и лексикографический аспекты // Труды Международного семинара Диалог'96 по компьютерной лингвистике и приложениям. - М., 1996. - С.227-230.

Семенова С.Ю. Прилагательные в семантическом словаре одной прикладной системы. // Труды Международного семинара Диалог'98 по компьютерной лингвистике и ее приложениям. - С.553-564.

Сокирко А.В. Исследование слов с характеристиками 'информация' и 'носитель информации' в русском общесемантическом словаре. // Там же. - С.565-573.

Семенова С.Ю. Наречия и предикативы в прикладном семантическом словаре // Труды Международного семинара Диалог'99 по компьютерной лингвистике и ее приложениям. - Таруса, 1999. - С.256 - 264.

Апресян Ю.Д. и др. Новый объяснительный словарь синонимов русского языка. Первый выпуск. - М., 1999. - 512 с.

Леонтьева Н.Н. О статусе валентностей в информационном анализе текста // Семиотика и информатика. Вып.36. - М., 1998. - С.41-50.

Падучева Е.В. О роли метонимии в концептуальных структурах // Труды Международного семинара Диалог'99 по компьютерной лингвистике и ее приложениям. - Таруса, 1999. - С.215 - 225.

[1] Работа выполнялась при поддержке Российского Фонда Фундаментальных Исследований и Российского Гуманитарного Научного Фонда; Проекты РФФИ: 99-06-80296 и РГНФ: 96-03-12103в.