

ЕСЛИ СЕМАНТИЧЕСКИЙ КЛАСС ШИРОК ДЛЯ СЛОВА (К ПРЕДСТАВЛЕНИЮ ЛЕКСИКИ В МАШИННОМ СЛОВАРЕ)¹

IF THE SEMANTIC CLASS IS TOO BROAD FOR A LEXEME (TOWARDS MEANING REPRESENTATION IN A COMPUTER DICTIONARY)

Семенова С.Ю. (nsomin@ipiran.ru), ИНИОН РАН

В словаре, ориентированном на автоматическое понимание текста, основным путем описания лексического значения является приписывание слову семантического класса (или нескольких классов). Качество описания значений во многом определяется словарным инструментарием, который включает совокупность предлагаемых лексикографу классов. Отбор классов для словаря представляет собой сложную задачу на стыке лингвистики и информационной науки. Очевидно, что любой ограниченный набор классов будет недостаточным для описания лексического "континуума", включающего лексику периферии классов, лексику индивидуальной семантики и малых семантических групп. В данной работе рассматриваются некоторые пути представления смысла такой лексики с помощью конкретной системы классов – семантических характеристик и смысловых отношений метаязыка машинного словаря РУСЛАН. Затрагиваются вопросы пополнения метаязыка, интерпретации смысла самого метаобозначения класса, сопоставления близких классов, соотнесения ядра и периферии класса.

1. К постановке проблемы

Задачи автоматического понимания текста, такие как построение семантического представления, диалог на естественном языке, перевод с опорой на семантическое представление, интеллектуальный информационный поиск, пополнение баз данных из текста и др., немислимы без словарной семантической информации. В настоящее время в словаре, ориентированном на автоматическое понимание, основным путем описания лексического значения является причисление слова к семантическому классу (или одновременно к нескольким классам). Мера точности и подробности описания значений во многом определяется словарным инструментарием, который включает совокупность предлагаемых лексикографу классов. Отбор классов для словаря – семантических, тематических, онтологических, (либо семантических характеристик, семантических полей, дескрипторов, помет – терминология варьируется в разных системах) представляет собой сложную задачу на стыке лингвистической и информационной наук.

Очевидно, что любой ограниченный набор классов будет недостаточным для описания лексического "континуума", в который кроме слов из относительно четко определяющихся представительных классов: УСТРОЙСТВ, ДОЛЖНОСТЕЙ, ВЕЩЕСТВ, ДВИЖЕНИЙ, ДЕЙСТВИЙ, ПРОЦЕССОВ, ЭМОЦИЙ и др. (*принтер* – УСТРОЙСТВО, *директор* – ДОЛЖНОСТЬ, *металл* – ВЕЩЕСТВО и т.п.), входят:

- слова, относящиеся к периферии классов; например, имена мифических существ (*ангел*, *амур* и др.) периферийны по отношению к классу имен деятелей – "должностей";

- слова индивидуальной семантики (которые могут быть прагматически весьма важны, например глагол *рисковать*: для более-менее подробного таксономического представления его смысла желательна расшифровка присущей ему модальности; желательна таксономическая категория, отличная от прототипической глагольной категории ДЕЙСТВИЕ (например, это может быть категория ИНТЕРПРЕТАЦИЯ); для описания второго участника уместна "штучная" семантическая роль типа 'угроза'; см. анализ этого слова в [6]);

- слова из небольших семантических групп, не входящих в перечень классов словаря; так, семы 'накопление' или 'опора', нужные для описания слов типа *запасать*, *облакачиваться* и др., в силу своей малой представительности могут отсутствовать в метаязыке словаря.

Слова (лексемы) трех последних категорий условно назовем промежуточными (конечно, они могут выступать таковыми лишь по отношению к фиксированной системе семантических классов).

¹ Работа выполняется в продолжение проекта РГНФ 04-04-00185а "Разработка грамматики концептуальных отношений для автоматического извлечения знаний из текстов"

Возникает проблема, как представлять промежуточные слова в словаре. Расширять метаязык, вводя дополнительные классы? Но пополнение метаязыка составляет определенную проблему, поскольку, вообще говоря, оно требует пересмотра уже имеющихся описаний. Желательно также придерживаться принципа экономии метаязыка – для минимизации пересечения классов и, соответственно, субъективизма в использовании единиц. Естественно и требовать, чтобы классы, предъявляемые лексикографу, были массовыми, пригодными для описания сразу множеств слов, а значит, их самих не должно быть слишком много.

Оставлять промежуточные слова недоопределенными? Но эти слова могут нести большую информационную нагрузку, которая неизбежно останется “за кадром” при текстовом анализе. “Недописание” (когда не хватает имеющихся форматов) практикуется в действующих информационных системах, например библиотечных; оно неизбежно при любой дискретизации информации, но для научной разработки ценно стремление к уменьшению неопределенности.

Пытаться построить описание, используя имеющиеся классы некими дополнительными, расширительными путями, например вводя дополнительные элементы синтаксиса, дополнительные договоренности? Рассмотрим данную возможность на примере конкретного словаря.

Отметим, что тот же подход к представлению смысла – отнесение слова (лексемы) к семантическому классу – применяется и при семантической разметке текстовых корпусов (особенно при разметке автоматизированной) [9] и соответственно, возникают те же проблемы, связанные с нехваткой классов. Кроме того, в корпусах более остро, чем в АОТ-ориентированном словаре, стоит вопрос об идентификации (путем приписывания класса) именно отдельных лексем, соответствующих реальным текстовым словоупотреблениям. Ведь АОТ-ориентированный словарь на настоящий момент представляет собой лишь дискретный отпечаток гораздо более сложной лексической системы языка, в нем неизбежны объединения значений, лакуны в совокупностях значений слова, а корпуса содержат натуральные фрагменты речевых потоков, которые требуется классифицировать сквозным образом. В корпусе операция семантической разметки, вообще говоря, раскладывается на два этапа – сначала требуется отнести словоупотребление к лексеме, а затем уже лексему к классу. При создании систем АОТ первая из задач в идеале предоставляется алгоритму семантического анализа, который в том числе должен снижать или уменьшать многозначность словоупотреблений в тексте, а вторая полностью возлагается на словарь.

2. О семантической инструментарию словаря РУСЛАН

Итак, хотя проблема представления смысла ограниченным набором классов представляется общей, рассмотрим ее на примере конкретного продукта – АОТ-ориентированного семантического словаря РУСЛАН [2, 3, 5, 10-12]. Словарь РУСЛАН, нацеленный на использование при разных этапах обработки текста, кроме собственно семантической, содержит информацию морфологическую, синтаксическую (набор валентностей и модели управления), сочетаемостную, тезаурусную, энциклопедическую, словообразовательную, информацию о переводных эквивалентах слова, об отраслевой принадлежности и некоторую другую.

В словаре РУСЛАН основным (и наиболее проработанным) способом представления лексического значения является применение конъюнкций стандартных, заложенных в систему семантических характеристик (СХ) и смысловых отношений (СО).

Семантические характеристики, безаргументные и тем самым “предметные” единицы метаязыка, служат обозначениями различных семантических классов и полей (АРТЕФАКТЫ, ЯВЛЕНИЯ, СИТУАЦИИ, ИНФОРМАЦИЯ и др.).

Смысловые отношения, у которых выделяются два (как правило несимметричных) аргумента (РЕЗЛТ (А, В) = А является результатом В; ФУНКЦИЯ (А, В) = А есть функция В; ЧАСТЬ (А, В) = А есть часть В и др.), предназначены в первую очередь для именованности валентностей предикатных слов, но ряд из них может служить при описании классов – например, для денотации разнообразных “функций” и “частей”. Одни СО обозначают почти исключительно имена ролей, другие пригодны для обозначения представительных лексических классов. Например, СО АГЕНТ (.), СУБЪЕКТ (.), ОБЪЕКТ (.), ПАЦИЕНС (.), АДРЕСАТ (.), СОДЕРЖАНИЕ (.) служат главным образом именами ролей; из лексики, кроме самих этих метаслов, с их помощью можно описать, видимо, лишь слова из их парадигматического круга, весьма небольшого: *пациент, радиослушатель,;* *содержимое, суть* и т.п. В тексте падежные СО выражаются не столько лексикой, сколько конструкцией, синтаксической позицией при соответствующих предикатах (ПАЦИЕНС аккузативом при переходном глаголе, АДРЕСАТ дативом при глаголе речи, СОДЕРЖАНИЕ – конструкцией ЧТО+Придаточное и т.п.). Другие СО, те же ФУНКЦИЯ и ЧАСТЬ, а также ВРЕМЯ (.), ПАРАМЕТР (.), ФОРМА (.) более подходят для обозначения лексических классов; им естественно сопоставить семантические поля – лексику поля времени, имена величин, имена фигур.

Наряду с СХ и СО в конъюнкциях могут использоваться лексические функции (ЛФ), в том числе словообразовательные. Например, смысл слова *изобретатель* – общую семантическую характеристику слова можно

описать с помощью конъюнкции СХ ОДУШЕВЛЕННОЕ, СО ФУНКЦИЯ и ЛФ S1: СХ (*изобретатель*) = ОДУШ & ФУНКЦИЯ & S1(*изобретать*). Поле с соответствующим названием (СХ) – обязательное поле семантического словаря РУСЛАН; собственно об его заполнении и идет, главным образом, речь в данной работе.

Представление смысла в виде конъюнкций дескрипторов соответствует идеологии компонентного анализа; конъюнкты СХ и СО можно считать машинными аналогами семантических компонентов слова.

Метаязык словаря в целом, и набор СХ и СО в частности, предложены Н.Н.Леонтьевой [3 и др.]. Список СХ и СО приведен в [3]; в основном он соответствует традиционным способам крупного членения действительности и лексики. При этом изначально совокупность СХ и СО была ориентирована на обработку общественно-политических текстов; в силу этого, например, существенно детализирована группа СО, названных по именам сложных предлогов, которыми богаты официальные документы (В-СООТВЕТСТВИИ, В-КАЧЕСТВЕ, В-СОСТАВЕ и др.), детализированы СХ социальной сферы (ВЛАСТЬ, ГОСУДАРСТВЕННОЕ, АДМИНИСТРАТИВНОЕ). С другой стороны, исходно не были детализированы классы, нужные, например, для описания природных существей – растений, микроорганизмов, заболеваний, разновидностей физических действий и т.п. Натуральные классы, как общелексические (а в последние годы словарь дрейфовал в сторону описания общей лексики), так и более специальные, служащие для описания номенклатур, например, соответствующие классификациям в биологии, по мере необходимости - при настройке словаря на предметную область - могут быть добавлены в метаязык, конечно, при учете вышеназванных соображений об его экономии.

В некоторой мере имеющаяся система СХ и СО отражает идеологию ‘признак-значение (или ‘параметр-значение’), характерную для научного мировоззрения в целом. Ряд дескрипторов РУСЛАН (СФЕРА, ТЕМА и др.) указывают на признак, а ряд, например, ЛФ Magn, Bon (а особенно ЛФ с отрицанием: AntiMagn, AntiBon, AntiVer) - на значение. СХ ФИНАНСЫ или ИНТЕЛЛЕКТУАЛЬНОЕ можно рассматривать как значения социального признака СФЕРА. При лексикографической работе полезно оперировать обеими типами единиц; в компонентном представлении желательно, чтобы для лексемы были указаны не только признаки, но и значения, тогда конкретность описания повысится. (Конечно, кроме случаев противопоставления признаков и значений в метаязыке имеются единицы другой природы, идущие от иных способов концептуализации действительности; можно выразить отношения ‘часть-целое’, ‘причина-следствие’ и некоторые другие).

Слово разбивается на лексемы, описываемые в отдельных словарных статьях; некоторые приемы представления лексической многозначности в словаре РУСЛАН приводятся в [5]; общие принципы – ограничение количества значений и снабжение описания каждой выделенной лексемы алгоритмически различимыми контекстами (в полях описания валентностей, тезаурусных связей, лексических функций, текстовых иллюстраций и др.). Дробление на лексемы не может быть мелким и подробным, поскольку это сделало бы автоматическое распознавание значений крайне затруднительным.

Развитие словаря, кроме собственно лексикографических описаний, предполагает развитие метаязыка, которое должно опираться на анализ возможностей сложившегося метаязыка, ведь добавление новых единиц желательно предварить осмыслением сфер употребления старых. Некоторые частные исследования на этом поприще уже проводились, например, в [13] сравниваются близкие по содержанию СХ ИНФОРМАЦИЯ и НОСИТЕЛЬ-ИНФОРМАЦИИ. Вопросы изучения грамматики концептуальных отношений, в том числе взаимосвязей классов СХ и СО, составили тему отдельного исследовательского проекта [4]. Добавление дескрипторов неизбежно, но надо понять, что из существенного, первоочередного для покомпонентных описаний отсутствует в словаре.

Актуальна также выработка методик представления лексики, не входящей точным, очевидным образом в имеющиеся классы – ведь такая лексика неизбежно встретится при ограниченном наборе крупных классов. Если отсутствует прямой класс, можно пытаться обойтись другими метаязыковыми единицами; при этом интересно проследить, какие ассоциативные механизмы срабатывают при описании лексики обходными путями. Для слов, укладываемых в классы, семантическое описание тоже можно и должно развивать. Например, помимо моделирования компонентной структуры, в которой компоненты-конъюнкты формально равноценны, желательно соблюдение одного из основных традиционных принципов толкования – выделение общего родового, интегрирующего признака и указание признаков дифференцирующих, отражающих индивидуальность слова в данном классе. В РУСЛАНе в настоящее время сложились некоторые конкретные приемы, позволяющие приближаться к этой цели. Вопросам такого рода – манипулированию имеющимися классами - и будет в основном посвящено дальнейшее изложение.

Ещё одним направлением развития данного словаря стала проработка альтернативных механизмов представления смысла. В настоящее время Н.Н. Леонтьевой и ее учениками проводится апробация другого способа описания значений – с помощью разложения связанной со словом ситуации, например, ситуации, обозначаемой предикатным словом, на ряд более простых ситуаций: двухаргументных элементарных ситуаций (ЭСИТ), упорядоченных, например, по времени; уточняется соответствующий формализм [3 и др.].

3. Отражение смысла с помощью широкого класса

Поскольку метаязык РУСЛАН включает крупные таксономические категории, близкие к частеречным (ПРЕДМЕТ, АБСТРАКТНОЕ; СИТУАЦИЯ, ПРОЦЕСС, ДЕЙСТВИЕ; ОПРЕДЕЛЕНИЕ, ПРИЗНАК, ОЦЕНКА; ОТНОШЕНИЕ и нек. др.), то класс будет приписан каждому слову; каждое слово получит хотя бы общее описание (существительные – по крайней мере предметы, явления, абстракции, или ситуации, глаголы – действия, процессы, либо ситуации, прилагательные и наречия – признаки, оценки или определения, служебные слова – отношения; конечно, и между этими парами-тройками “родственных” категорий необходима выработка разграничительных правил, см. ниже). Но основное информационное (и эстетическое) достоинство словаря всё же должно состоять в уместной детализации описаний, в отражении ограниченными, алгоритмически считываемыми средствами индивидуальных семантических свойств. В РУСЛАНе для этого используются прежде всего более конкретные, тематические СХ и СО.

Отметим, что проблема соотношения крупных таксономических классов, мотивированных грамматически (в частности, аспектологически), и классов тематических, онтологических, отражающих более “внешнее” с точки зрения лингвистики, неграмматизированное разнообразие мира, обсуждается в работах по систематизации лексического материала, например, в работах Е.В.Падучевой, посвященных исследованию глагольной лексики в рамках проекта “Лексикограф” [7,8 и др.].

Также отметим, что принцип замены недостающего класса верхним, объемлющим классом соблюдается не только в лексикографии, но и в определенных видах информационной деятельности, в частности, в каталогизации (предметной обработке) литературы. Например, если в предметный каталог по каким-то причинам не введена рубрика “Контент-анализ” (например, до сих пор не было соответствующей литературы), то для описания произведения на эту тему можно использовать рубрики, которые выступают как обобщающие. При этом обобщение может осуществляться по разным аспектам понятия, которому посвящено произведение: если имеется в виду автоматический контент-анализ, то обобщающей рубрикой будет, например, комплекс “Текст – Автоматический анализ”, если преобладают политологические вопросы, то “Информационно-аналитическая деятельность”, если акцентирована ручная вычислительная сторона метода, то “Лингвистика прикладная” и т.п. (Приведенные рубрики взяты из предметного каталога библиотеки ИНИОН РАН). Причем предметизатор, в отличие от лингвиста-лексикографа, имеет дело с более неопределенным материалом, нежели лексический состав языка; вводя рубрику и связывая ее с имеющимися, он должен предсказывать потенциальную наполняемость рубрики, а тем самым меру перспективности обозначаемого ею научного направления.

В РУСЛАНе использование широкого класса практикуется не только при прямом описании семантики (поле СХ), но и при описании семантических ролей слова. Как было отмечено вначале, одно из назначений двухаргументных смысловых отношений в словаре – именование семантических ролей предикатных слов. В номинации ролей также, как и при приписывании классов лексеме, встречаются вопросы – далеко не всегда удастся обойтись стандартным набором валентностных СО: АГЕНТ(,), ОБЪЕКТ(,), ПАЦИЕНС(,), АДРЕСАТ(,), СОДЕРЖАНИЕ(,) и т.п., не всегда удастся и метко “окрестить” участника, если его смысловая связь со словом отличается от перечисленных.

Для наиболее затруднительных или неопределенных случаев в метаязыке имеются неуточненные СО ПАКТ(,), В-АКТ(,), ОГРН(,) – соответственно, первый актант, второй актант, ограничение. В частности, таким “безмянным” образом – первый и второй актанты – могут быть описаны участники предикатов *равный, больше* (в употреблении типа *Величины x и y равны между собой; Величина x больше y*); неуточненные СО используются при описании служебных слов и оборотов, например союзов [2]. СО ОГРАНИЧЕНИЕ может быть употреблено как неопределяемый гипероним к более конкретным СО, например к СО СОДЕРЖАНИЕ, ТЕМА или СПЕЦИАЛИЗАЦИЯ, когда употребление последних спорно, периферийно: *четверка* – (ОГРН) -> *по геометрии*.

Итак, для каждого слова в РУСЛАНе (и для каждого его актанта) найдется класс, хотя бы широкий. Но как, если класс в самом деле широк, обогатить описание, сделать его релевантным для нетривиального информационного текстового анализа?

Рассмотрим несколько конкретных приемов.

4. Применение цепочек дескрипторов

Определенной систематической детализации описаний можно достичь путем построения в поле СХ цепочек дескрипторов – обозначений классов. Правила использования и сочетания СХ и СО удобно вырабатывать при прицельных описаниях конкретных семантических полей. Например, проводилось практически сквозное описание одушевленной нарицательной лексики, и при этом сложились некоторые правила использования единиц метаязыка [12]. Так, родовой дескриптор ОДУШЕВЛЕННОСТЬ для имен ряда должностей и широко понимае-

мых профессий заменяется на СО ДОЛЖНОСТЬ (для имен типа *сотрудник* или *бухгалтер*); т.е. данное смысловое отношение выступает как имя класса – подкласса СХ ОДУШ. СО ДОЛЖНОСТЬ нередко сочетается с СХ ВЛАСТЬ (*президент, директор*). Большинство других обозначений человека описываются как конъюнкции СХ ОДУШ с дескрипторами, лишь частично обслуживающими сферу одушевленной лексики: ОДУШ & ОТНОШЕНИЕ (для имен родства и подобных: *мать, друг*); ОДУШ & ОЦЕНКА (для аксиологической лексики: *талант, гений, негодяй*); ОДУШ & СПЕЦИАЛИЗАЦИЯ (для названий узких специальностей, для профессиональных номенклатур: *биохимик, плиточник*); ОДУШ & ФУНКЦИЯ (главным образом, для обозначения всевозможных S1, в том числе не профессиональных – статичных, а эпизодических: *завещатель, обидчик*); ОДУШ & СОЦИО (для имен, выражающих социальный статус человека: *пенсионер, блокадник*); ОДУШ & СУЩЕСТВОВАНИЕ (для отражения биологических аспектов номинации: *блондин, слепой*; обобщенно обозначается физическое, биологическое существование). Применяются и более длинные цепочки метаязыковых единиц, например, модель ОДУШ & ФУНКЦИЯ & ВРЕД – для обозначения деятелей криминальной сферы: *угонщик* и подобное; модель ОДУШ & СОЦИО & РЗНВД & ИНТЕЛ (“разновидность интеллектуальная”) – для обозначения различных приверженностей человека: *демократ, патриот, меломан* и др. Фактически на классе обозначений человека и его представительных подклассах подобные цепочки исчислены; если в метаязыке имеются единицы, отражающие более индивидуальные компоненты смысла слов из приведенных подклассов, они добавляются в цепочки: СХ (*гений*) = ОДУШ & ОЦЕНКА & Von & Magn; СХ (*бухгалтер*) = ДОЛЖНОСТЬ & СВЯЗАН & ФИНАНСЫ. По стандартным началам цепочек можно идентифицировать класс и подкласс, а по индивидуальным окончаниям – дифференцирующие признаки лексем.

Отметим, что среди одушевленной лексики немало и “хороших”, “непромежуточных” имен (например, имен должностей), поскольку социальная сфера, к которой они относятся, структурирована в данном словаре. Цепочки применимы и к “хорошим” именам (цепочки конкретизируют значение), и к именам, для которых не хватает прямых классов (“приверженностям” и др.), когда они служат обходным способом выразить семантику.

Существенный материал для анализа дескриптивных возможностей словаря составляет адъективная лексика [10]. Для описания прилагательных предусмотрены стандартные начала цепочек: ОЦЕНКА &...- для качественных прилагательных, связанных со шкалами: *высокий - низкий, добрый - злой* и т.п.: ПРИЗНАК &... - для основной массы адъективных лексем, в том числе относительных, характеризующихся более-менее фиксированным типом реляции: *желеобразный* (признак по консистенции), *приморский* (по локализации), *сухопутный* (по типу среды) и т.д.; ОПРЕДЕЛЕНИЕ &... - для относительных прилагательных с широким набором реляций – по разным аспектам смысла мотивирующего слова: *морской* (ср.: *транспорт, рыба, курорт, Министерство, воздух* и др.). После первого знака “&” в цепочках следуют уточняющие классы, в той мере, в какой они наличествуют в метаязыке. Для прилагательного *морской* уточнителем будет СХ ГЕОГРАФИЧЕСКОЕ, поскольку море - в наиболее широком смысле - географический объект: СХ (*морской*) = ОПР & ГЕОГР & A0 (*море*).

5. Применение дескрипторов в качестве операторов

Для ряда дескрипторов: ОТСУТСТВИЕ, СВЯЗАН, КАУЗАЦИЯ, ЧАСТЬ, РАЗНОВИДНОСТЬ, ОПРЕДЕЛЕНИЕ практикуется использование в качестве операторов, “навешиваемых”, подобно кванторам, на стоящие справа дескрипторы [11]. Дескрипторы ОТСУТСТВИЕ и СВЯЗАН позволяют в определенной мере отражать модальность и ранг компонентов. Например, СХ ОТСУТСТВИЕ моделирует компонент с отрицанием: СХ (*молчать*) = СИТ & ОТСУТ & КОММУНИКАЦИЯ.

СО СВЯЗАН, служебное, не отягощенное конкретной смысловой нагрузкой, кроме выражения ассоциированности, обычно служит для понижения ранга компонента; показывает, что компонент не выражает непосредственную принадлежность слова к стоящему справа классу, а лишь связь слова с данным классом: слово *бухгалтер* не входит в класс финансов, но связано с этим классом (“по должности”); *кружка* не является непосредственно элементом класса форм, но сема формы таким образом отмечается в ее представлении: СХ (*кружка*) = АРТЕФАКТ & ВМЕСТИЛИЩЕ & СВЯЗАН & ФОРМА; глагол *нет* не есть глагол восприятия (кстати, для его точной идентификации в словаре пока нет подходящего класса, тем самым он – “плохой”), но с помощью СО СВЯЗАН можно “притянуть” его, например, к классу восприятий, имеющемуся в метаязыке: СХ (*нет*) = ДЕЙСТВИЕ & СВЯЗАН & ВОСПРИЯТИЕ.

Применение СО СВЯЗАН позволяет в случае отсутствия прямого класса сослаться на класс, ассоциированный с недостающим.

СХ КАУЗАЦИЯ, применяемая как оператор, помогает моделировать отчасти в поле СХ ситуативное описание смысла, тем самым также выступает как конкретизатор: СХ (*рождение*) = СИТ & СТАДИЯ & КАУЗАЦИЯ & СУЩЕСТВОВАНИЕ.

СО ЧАСТЬ и РАЗНОВИДНОСТЬ выступают как две стороны более общего партитивного концептуаль-

ного отношения; соответственно, они обозначают маркированные компоненты пар “часть-целое” и “элемент-множество”. СО ЧАСТЬ при употреблении как оператор указывает на класс целого: СХ (*отдел*) = ЧАСТЬ & ОРГАНИЗАЦИЯ; СО РАЗНОВИДНОСТЬ, наоборот, указывает на класс элемента (см. выше пример описания идеологических приверженностей, когда класс элемента характеризуется как ИНТЕЛЛЕКТУАЛЬНОЕ).

СХ ОПРЕДЕЛЕНИЕ применяется как обобщенная характеристика относительных прилагательных, см. ниже; подчиненный конъюнкт указывает на класс мотивирующего слова: СХ (*зимний*) = ОПРЕДЕЛЕНИЕ & ВРЕМЯ.

При описаниях часто, особенно если нет прямого класса, возникает вопрос, какой аспект значения выделить и отразить, а какие оставить в тени. Например, с чем больше СВЯЗАНО действие, обозначаемое глаголом *упаковывать* – со ВМЕСТИЛИЩЕМ (например, *упаковывать в коробку*) или со СРЕДСТВОМ (*в бумагу*)? Обе эти категории, связанные метонимически [7 и др.] можно отразить, например, описывая альтернативные СХ третьего участника. Но что кроме ДЕЙСТВИЯ, имея лишь набор массовых классов, можно внести в компонентную структуру самого глагола (в его собственное поле СХ?). Быть может, отразить идею КАУЗАЦИИ, но что каузируется? По-видимому, каузируется ЛОКАЛИЗАЦИЯ, понимаемая весьма обобщенно: был предмет *вне* некоторой среды - упаковки, а стал *внутри* нее. Таким образом, СХ (*упаковывать*) можно представить как ДЕЙСТВИЕ & КАУЗАТОР & ЛОКАЛИЗАЦИЯ; представление ДЕЙСТВИЕ & СВЯЗАНО & ВМЕСТИЛИЩЕ/СРЕДСТВО менее удачно, поскольку эти категории будут продублированы при характеристике участника. (Оговоримся, что приведенные варианты заполнения поля СХ не являются истинами в последней инстанции, они безусловно допускают коррекцию; их следует рассматривать прежде всего как иллюстрации сложившихся частных методик.)

Вопрос о том, чтобы выделить наиболее характерное в лексическом значении (и именно это характерное затем непосредственно внести в описание или “объять” классом) – ставился и в других прикладных лексикографических системах. Например, при описании предметных имен в системе “Лексикограф” [1] в формализованное представление значения вносится типовой предикат, связанный с обозначаемым предметом, и этот предикат в ряде случаев также приходится выбирать из альтернативных.

6. Использование возможностей других полей

Для отражения индивидуальных свойств лексем в РУСЛАНе можно, кроме поля семантических характеристик, использовать шаблоны других полей. Например, для имени *кружка* в поле энциклопедических функций можно указать, что выделенным, “главным” элементом (энциклопедическая функция *Сар*) будет ручка: ЭНЦ (*кружка*) = *Сар* (*ручка*), для имени *гора* прагматически важным параметром будет *высота*: ЭНЦ (*гора*) = *Param* (*высота*) и т.п.

Можно задействовать и зону тезаурусной информации; в ней можно указать гипероним, близкий по иерархической лестнице: ВЫШЕ(*кружка*) = *посуда*; в тезаурусном поле ассоциаций можно назвать единицы парадигматического ряда: АСС (*кружка*) = *стакан; чашка* и т.п.

Для “отраслевых” слов можно использовать поле “Проблемная область” (ПО), где предусмотрен набор отраслевых помет - например, чтобы показать, что определенные специальности человека относятся к разряду научных: ПО(*грамматист*) = “наука”.

Вынесение части информации из чисто семантического поля СХ в другие, “внешние” поля соответствует известным представлениям о компонентном анализе, как о методе, связывающем лингвистическое и металингвистическое знание.

7. Расширенное и суженное понимание классов

Этот прием рассмотрим на примере применения смысловых отношений. Иногда приходится прибегать к расширенному пониманию имеющихся СО. Например, чтобы обозначить 3-го участника глагола *ввергать* (Агент X ввергает Пациенса Y-а -> во что?), приходится абстрагированно истолковывать идею конечной точки (имеется прототипически пространственное СО КОН-Т – “конечная точка”), которое может быть истолковано, как некое конечное абстрактное состояние; от исходного понимания конечной точки как сущности пространственной и фазовой остается только фазисность, которая представляется доминирующей для ситуации, обозначаемой данным глаголом).

Практикуются случаи обратного перехода - от абстрактного понимания идеи, заложенной в СО, к конкретному. Например, СО ИСТОЧНИК было введено изначально как абстрактное, как источник информации, но почему бы (за неимением точного физического класса) не использовать этот дескриптор в его прямом значении при описании таких предметных имен, синонимичных, как *источник* (водный), *родник*, *ключ*? Это, кстати, один из случаев упомянутого вначале применения падежного СО для классификации лексики.

8. Конкретные случаи описания промежуточной лексики

Рассмотрим некоторые конкретные случаи, затруднительные для описания имеющимися классами. Промежуточную лексику в избытке поставляют прилагательные. Значительная их часть характеризуется более индивидуальными семантическими признаками, чем приведенные крупные классы. Например, как отразить крупными классами идею самомнения (для прилагательных личностной оценки *тщеславный, надменный, заносчивый*), или идею замкнутого пространства, отсутствия “дыр”; само понятие “дыры”, важное с точки зрения топологии (*герметичный*), как отразить идею деятельности на расстоянии, при отсутствии непосредственного контакта (*заочный*)?

В рамках поля СХ подобные прилагательные часто остаются недоопределенными; для продолжения цепочек выхватывается лишь наиболее общее в значении, например, лишь аксиологическая оценка: СХ (*заносчивый*) = ПРИЗНАК & AntiBon; лишь пространственность и неуточненное ОТСУТСТВИЕ: СХ (*герметичный*) = ПРИЗНАК & ПРОСТРАНСТВЕННОЕ & СВЯЗАН & ОТСУТСТВИЕ; лишь социальное и связанное с локализацией: СХ (*заочный*) = ПРИЗНАК & СОЦИАЛЬНОЕ & СВЯЗАН & ЛОКАЛИЗАЦИЯ.

Часть информации отразится в других полях, например, для *заочный* в поле ЭНЦ можно указать, что задействован параметр *расстояние*: ЭНЦ (*заочный*) = Param (*расстояние*). (Кстати, упоминание расстояния в статье слова *заочный* иллюстрирует тот случай, весьма нередкий, когда для отражения второстепенного смыслового признака средства есть, а для более важного (здесь - отсутствия контакта) их пока нет; аналогичные случаи постоянно возникают в предметизации, и получается, что поисковый образ документа односторонен).

Для ряда других промежуточных слов (например, прилагательных личностной оценки: *легкомысленный, лживый* и др.) перспективно уточнение смысла описанием ситуаций – ведь такие оценки в принципе толкуются через поведение.

Случаи нехватки дескрипторов часто встречаются при описании глагольной лексики, изобилующей конкретными смысловыми нюансами. Это настолько важный грамматический класс, что его компонентное представление требует отдельных рассмотрений (не говоря о других типах словарного представления, например, ситуационного). Здесь ограничимся лишь несколькими примерами глаголов, нестандартных для описания имеющимися средствами; как правило, они описываются некими обходными путями, при расширительном понимании смысла дескрипторов: *впадать* (о реке) – СВОЙСТВО & ПРОСТРАНСТВЕННОСТЬ & СВЯЗАН & КОН-Т; *вдыхать* – ДЕЙСТВИЕ & СТАДИЯ & СВЯЗАН & СУЩЕСТВОВАНИЕ & СВЯЗАН & ВЕЩЕСТВО & СВЯЗАН & ВМЕСТИЛИЩЕ (стадия противопоставляется выдоху, под веществом имеется в виду – воздух, под вместилищем - легкие); *нумеровать* – ДЕЙСТВИЕ & ИНФОРМАЦИЯ & КАУЗАЦИЯ & ИДЕНТИФИКАТОР. Нестандартен глагол *всходить* (о семенах) – его можно описать, например, с помощью ЛФ CausIncep, но понимание этой ЛФ не будет буквальным – всходя, семена “возникают и начинают быть” лишь в определенном отношении. Трудны для описания, например, глаголы физической сферы *облакачиваться, дотрагиваться* (пока не хватает единиц для отражения идей опоры и контакта). Определенную трудность представляет описание глаголов интерпретации: *вдаваться* (в детали), *упорядочивать* – не вполне понятно, как лучше отразить идею структурирования, в том числе мелкого; идею порядка. Упомянутый выше глагол *рисковать* в РУСЛАНе пока описывается с помощью СО МОДАЛЬНОСТЬ: СХ (*рисковать*) = СИТУАЦИЯ & МОДАЛЬНОСТЬ. Косвенно описываются глаголы, связанные с идеей накопления: СХ (*запасать*) = ДЕЙСТВИЕ & СВЯЗАН & БОЛЬШЕ & СВЯЗАН & ПОСЛЕ.

К числу промежуточных слов относятся обозначения особых, маркированных видов ментальности, связанных с нарушением постулата физического присутствия. Например, слова с семантикой кажимости: *видение, сниться* (постулат физического существования нарушен, но в то же время нельзя применить дескриптор AntiVer в его прямом понимании). Сейчас для их описания применяется СО МОДАЛЬНОСТЬ – подразумевается модальность обобщенная, неуточненная. При описании одушевленной лексики встает вопрос, как отразить божественное, сакральное. Высшая сила может описываться, например, как ОДУШ & АБСТРАКТНОЕ.

9. О содержании классов

Практика словарных описаний вообще, а промежуточной лексики в особенности, подводит к осознанию необходимости анализировать концептуализацию самих классов, номенклатуру сущностей, охватываемых ими. Сравнивать изначально задуманное содержание классов и реальное их наполнение, сложившееся за годы работы над словарем. Исследовать пересечения классов, выявлять иерархические отношения между ними. Полезно попарное сопоставление близких классов, уточнение сфер их употребления. В целом анализ концептов, стоящих за обозначениями классов, составляет постоянное поле исследований (это касается и конкретных классов РУСЛАНа, и подобных им классов вообще). Исследования можно проводить как на лексикографическом мате-

риале, анализируя и словник, и накопленные словарные статьи, так и на материале текстовом, интерпретируя реальные высказывания с помощью имеющихся СХ и СО.

Можно привести некоторые частные соображения, касающиеся сравнения дескрипторов. Например, в паре СО СУБЪЕКТ – АГЕНТ за СО АГЕНТ закрепляется именование первого участника целеполагающего действия – активного участника, а за СУБЪЕКТОМ – именование участника (X) тех ситуаций, где целеполагание нарушается. Ситуации нарушения целеполагания в принципе можно задать экстенционально, перечислением: первый участник малоконтролируемых действий: X <-дышит, или малоконтролируемых состояний сознания: X <- верит; второй участник состояний, выражаемых предикативами: трудно работать -> X-у; участник прототипических происшествий [7]: Ветер (X) <- порвал провода и др.

Аналогично могут сопоставляться члены пары ОБЪЕКТ – ПАЦИЕНС, где основным дифференцирующим признаком будет наличие/отсутствие изменений во втором участнике переходного глагола: если изменения есть (мыть -> посуду), то участник - ПАЦИЕНС, если нет (созерцать -> жизнь), то ОБЪЕКТ. Интересно при этом исчислять периферийные случаи, отличные от общего правила, подчас требующие соглашений между лексикографами. Вообще, договоренности между разработчиками об использовании метаязыка – одна из обязательных сторон лексикографической практики.

Примерами пар, для которых целесообразен сравнительный анализ интенционалов, являются также ДОЛЖНОСТЬ - СПЕЦИАЛИЗАЦИЯ, ИНСТРУМЕНТ – С-ПОМОЩЬЮ, ПАРАМЕТР - ПРИЗНАК, ВРЕМЯ - СТАДИЯ, ИНФОРМАЦИЯ – НОСИТЕЛЬ (пара рассматривалась в [13]; при этом обсуждалась возможность применения информационных технологий для анализа сфер употребления), ПРИНАДЛЕЖНОСТЬ – ЧАСТЬ, РАЗНОВИДНОСТЬ – В-СОСТАВЕ, AntiBon - ВРЕД и некоторые другие. Желательно выработать (или уточнить) собственно толкования обозначений классов. Очевидно, что сферы употребления некоторых близких классов могут быть жестко разделены лишь волевым решением. Не исключено, что в результате сравнительного анализа некоторые единицы будут сочтены избыточными.

Обращение к текстам, в частности в Интернете, показывает, например, что СО ФУНКЦИЯ, выражающее общую идею действия, выполняемого с участием данной сущности, стандартным образом переходит в нижестоящее СО ДОЛЖНОСТЬ - при описании ситуаций социальной жизни, общественного труда (работать бухгалтером), в СО НАЗНАЧЕНИЕ - в мире артефактов (кондиционер служит для охлаждения воздуха) и в СО СВОЙСТВО - в сфере природных объектов (солнце дарит жизнь; лимон тонизирует). За СО ФУНКЦИЯ остается описание должностных и прочих обязанностей (Есть такая профессия – Родину защищать), а также, как за гиперонимом, интерпретация фактов и лексем, не вписывающихся в стандартные модели: всевозможные S1, отличные от должностей (см. примеры выше), факты с нарушением постулата истинности (Он только числится бухгалтером – формальная функция), случаи использования предметов не по назначению (Топить можно и ассигнациями – ассигнации здесь служат топливом) и другие. (Конечно, подобные примеры не всегда связаны напрямую со способами описания конкретных слов (которые рассматриваются в словаре обычно лишь в прототипическом употреблении), но эти примеры помогают прояснить онтологические отношения между концептами, находящиеся “над лексикой”).

10. О пополнении метаязыка

Несколько слов о пополнении метаязыка. Конечно, словарь всегда открыт для этого, но с тем условием, чтобы ввод нового дескриптора был мотивирован.

Для описания предметных областей могут добавляться отраслевые классы. Наиболее легкие случаи для словарной системы в целом – когда потенциальный новый класс есть строго вкладывающийся подкласс класса существующего (ЖИВОТНОЕ как подкласс ПРЕДМЕТОВ, УСЛУГА как подкласс ДЕЙСТВИЙ, БОЛЕЗНЬ как подкласс СИТУАЦИЙ и т.п.). Несколько сложнее, если новый класс частично пересекается со старыми, является результатом другой концептуализации (например, ТОВАР пересекается и с ПРЕДМЕТАМИ, и с ВЕЩЕСТВАМИ, и с ИНФОРМАЦИЕЙ и прочее; ИНТЕРПРЕТАЦИЯ может быть связана с МОДАЛЬНОСТЬЮ, ИНТЕЛЛЕКТУАЛЬНЫМ, ВОСПРИЯТИЕМ, ОЦЕНКОЙ); при вводе подобной мировоззренчески новой сущности может потребоваться значительный пересмотр былых описаний.

Опыт составления статей показывает, что из общих концептов в метаязык, по-видимому, целесообразно добавить обозначение опоры (основы) – без этого затруднительно описывать соответствующее семантическое поле (опираться, облакачиваться, фундамент, базис и др.), а также обозначение контакта (для dotрагиваться, очный/заочный и др.) - гиперонима по отношению к опоре (в физической сфере) и к имеющейся СХ КОММУНИКАЦИЯ (в социуме).

Для описания общей лексики могут быть добавлены широкие натуральные классы - ЗВУК, агрегатные состояния ТВЕРДОЕ и ГАЗ (ЖИДКОЕ есть) и т.д.

Для детализированного описания предметной лексики полезно отражение разновидностей СО ЛОКАЛИЗАЦИЯ – введение примерно таких СО, как ВНУТРИ (,), ВВЕРХУ(,), СПРАВА (,), ГРАНИЦА (,), ОРИЕНТИР (,) или подобных [14]; возможно и введение топологического класса ОТВЕРСТИЕ (или подобного). Примеры слов, для описания валентности которых целесообразно СО ОРИЕНТИР: *вращаться* (Земля *вращается* –(ОРИЕНТИР)-> *вокруг Солнца*); *колебаться* (о температуре) *около нуля*; в настоящее время в РУСЛАНе для описания этого участника может использоваться либо более широкое СО ЛОКАЛИЗАЦИЯ, либо малосемантизированное СО ОГРАНИЧЕНИЕ.

11. Место заключения

Подведем итог. В словаре РУСЛАН основными приемами семантического описания лексики, недостаточно согласующейся со стандартными классами, являются:

- а) приемы заполнения поля СХ лексемы:
 - указание широкого класса, например частеречного или другого категориального;
 - применение цепочек дескрипторов; состав цепочек, предназначенных для описания конкретных семантических групп, обговаривается разработчиками;
 - применение операторов типа СВЯЗАН, понижающих ранг подчиненного оператору класса и вовлекающих в описание имеющиеся смежные классы;
 - для слов-дериватов, например относительных прилагательных, указание класса мотивирующего слова;
 - расширенное, метафорическое понимание (и применение) классов;
 - наоборот, суженное использование обозначений абстрактных классов для описания ряда конкретных лексем, например, графически совпадающих с метасловом – обозначением класса;
- б) использование инструментария других полей – энциклопедических функций, тезаурусных связей, ситуаций.

Данная техника должна быть дополнена развитием и коррекцией метаязыка семантических описаний. Некоторые рекомендуемые дополнительные классы названы; в целом же задача оптимизации метаязыка требует детального изучения имеющихся классов, исчисления их лексического состава, исследования их иерархий и пересечений, сопоставления их конкретного и абстрактного понимания, выработку договоренностей о применении. Учитывая, что классы данного словаря в основном совпадают с общепринятыми семантическими таксонами, исследования и методическая работа такого рода могут быть полезны и для разработки других словарных продуктов.

Список литературы

1. Красильщик И.С., Рахилина Е.В. Предметные имена в системе “Лексикограф” // НТИ, сер. 2. – 1992. – N 9. – С. 24 - 31.
2. Леонтьева Н.Н. О статусе валентностей в информационном анализе текста // СИИ. – Вып. 36. – М., 1998. – С. 41 – 50.
3. Леонтьева Н.Н. К теории автоматического понимания естественных текстов. Часть 2. Семантические словари: состав, структура, методика создания – М.: Изд-во МГУ, 2001.
4. Леонтьева Н.Н. О грамматике концептуальных отношений // Компьютерная лингвистика и интеллектуальные технологии. Труды Международной конференции “Диалог-2006”. – М.: Изд-во РГГУ, 2006. – С. 342.
5. Леонтьева Н.Н., Семенова С.Ю. Об отражении полисемии в прикладном семантическом словаре // “Компьютерная лингвистика и интеллектуальные технологии”. Труды Международного семинара Диалог’2002. (Протвино, 6-11 июня 2002 г.) – М.: Наука, 2002. – Т.2. - С. 489 - 496.
6. Падучева Е.В. “Риск – благородное дело”: о системе значений глагола *рисковать* // Труды Международного семинара Диалог’2001 по компьютерной лингвистике и ее приложениям. - Аксаково, 2001. - Т.1. -С. 189 - 194.
7. Падучева Е.В. О параметрах лексического значения глагола: таксономический класс участника // Русский язык в научном освещении. N 1 (3). 2002. – С. 87 - 111.
8. Падучева Е.В. О параметрах лексического значения слова: онтологическая категория и семантический класс // Проблемы русской лексикографии: Тезисы докладов Международной конференции Шестые Шмелевские чтения 24-26 февраля 2004 г. – М., 2004. – С. 75 - 83.
9. Рахилина Е.В., Кобрицов Б.П. и др. Многозначность как прикладная проблема: лексико-семантическая разметка в национальном корпусе русского языка // Компьютерная лингвистика и интеллектуальные технологии. Труды Международной конференции “Диалог-2006”. – М.: Изд-во РГГУ, 2006. – С. 445 - 450.

10. Семенова С.Ю. Прилагательные в семантическом словаре одной прикладной системы // Труды Международного семинара Диалог'98 по компьютерной лингвистике и ее приложениям. – Казань, 1998. - С. 553 – 564.
11. Семенова С.Ю. Семантические поля словаря РОСС: опыт заполнения, анализ дескриптивных возможностей (Материалы к унификации словарных описаний) // Труды Международного семинара Диалог'2000 по компьютерной лингвистике и ее приложениям. – Протвино, 2000. – Т.2. – С. 308 – 316.
12. Семенова С.Ю. О фреймовом подходе к описанию одушевленной лексики в прикладном словаре // Scripta linguisticae applicatae. Проблемы прикладной лингвистики. Выпуск 2.- М.: “Азбуковник”, 2004. - С.56 – 70.
13. Сокирко А.В. Исследование слов с характеристиками “информация” и “носитель информации” в русском общесемантическом словаре // Труды Международного семинара Диалог'98 по компьютерной лингвистике и ее приложениям. – Казань, 1998. - С. 565 – 573.
14. Соколова Е.Г., Болдасов М.В. Принципы построения семантических аннотаций содержания изображений // Компьютерная лингвистика и интеллектуальные технологии. Труды Международной конференции “Диалог-2006”. – М.: Изд-во РГГУ, 2006. – С. 474 - 480.