

ЭНЦИКЛОПЕДИЧЕСКАЯ ИНФОРМАЦИЯ О СЛОВЕ И ЛЕКСИЧЕСКИЕ КЛАССЫ

ENCYCLOPAEDIC DATA ON A WORD AND LEXICAL CLASSES

Семенова С.Ю. (Sonya_sem@mail.ru)

*Институт научной информации по общественным наукам РАН,
Российский государственный гуманитарный университет*

Аннотация

Работа посвящена вопросу о пополнении прикладного формализованного семантического словаря энциклопедической информацией. Исследование, с одной стороны, направлено на развитие конкретного лексикографического продукта (машинного семантического словаря РУСЛАН), с другой стороны, ставит своей теоретической целью выяснение, описания каких лексических единиц целесообразно снабжать онтологической информацией и какого рода информация релевантна для этого. Путем сплошного просмотра фиксированного массива общей лексики объемом около 10.000 слов выявлен ряд лексических классов, соотносимых с конкретными типами энциклопедической информации. В качестве массовых способов структуризации энциклопедических знаний, ассоциирующихся с общей лексикой, проявили себя указание параметров обозначаемой реалии и перечисление разновидностей реалии. Эти и некоторые другие типы информации могут формализовываться при помощи аппарата энциклопедических функций, предложенного Н.Н.Леонтьевой. В работе исследуются возможности применения и развития этого аппарата при описании общей лексики.

Исследование выполняется при поддержке Российского гуманитарного научного фонда; проект РГНФ-10-04-12108 в.

Данное исследование выполняется в развитие работы [1], в которой рассматриваются способы дополнения русского прикладного семантического словаря внелингвистической (энциклопедической; онтологической) информацией об обозначаемых словами реалиях. Внесение внелингвистических сведений в лингвистический словарь, в частности прикладной, поднимает ряд вопросов; в их числе такие вопросы, как:

- 1) Какого рода сведения об обозначаемых реалиях актуальны (в целом, априори) для традиционного круга задач автоматического текстового анализа - для автоматического перевода, избирательного извлечения информации, сжатия текста,

построения семантического представления и других задач, на реализацию которых нацелен словарь?

2) Каковы могут быть формализмы для представления энциклопедических сведений?

3) Для описания каких лексических единиц и классов возможны и целесообразны энциклопедические расширения?

В данной работе мы преимущественно будем рассматривать 3-й вопрос – о целесообразных типах энциклопедической информации при описании тех или иных лексических единиц. При этом первые два вопроса также частично будут затронуты.

С практической стороны работа ориентирована прежде всего на развитие конкретного словарного продукта – словаря РУСЛАН, создаваемого под руководством Н.Н. Леонтьевой с 1990-х годов [2], на пополнение этого словаря новой информацией и на расширение его формализмов, хотя, как кажется, вопрос о систематизации отражения онтологических знаний в лингвистическом словаре, применительно к конкретным лексическим категориям, представляет интерес в более широком, теоретическом плане.

В словаре РУСЛАН, метаязык и шаблоны которого разработала Н.Н. Леонтьева, для размещения внелингвистической информации предусмотрена специальная зона (ЭНЦ). Данные представляются в ней как аргументы энциклопедических функций, таких как Cap (главный, маркированный элемент обозначаемой сущности: Cap (*оркестр*) = *дирижер*); Sing (единичный элемент, «квант»: Sing (*жидкость*) = *капля*), Param (типовой параметр, по которому характеризуется обозначаемая сущность: Param (*потолок*) = *высота*) и другие. При том, что устроенные таким образом функции несут, безусловно, некоторую информацию о действительности, они прежде всего вводят в словарь лексические связи, дополнительные к тем, которые представлены в нем в других зонах – к тезаурусным и словообразовательным связям, лексическим функциям-коллокамтам, связям между членами устойчивых оборотов. В [1] предложен ряд новых (для словаря РУСЛАН) энциклопедических функций того же типа: Subst – материал изготовления: Subst (*ситец*) = *хлопок*; Usage – сфера типового использования сущности: Usage (*наркоз*) = *хирургия* и нек. др. В [1] отмечается также, что из имеющихся функций наиболее продуктивной для описания лексики показала себя функция Param; в [3] ставится вопрос о дополнительной структуризации этой функции для представления с ее помощью не только лексической, но и фактографической информации, в том числе количественной.

Вопрос о методиках отбора энциклопедических сведений в лингвистический словарь, насколько нам известно, в отечественной

литературе не получил до настоящего времени подробного систематизированного освещения, хотя некоторые работы, как в области традиционной, «бумажной», так и в области компьютерной лексикографии, тем не менее, можно назвать.

В.Г. Гак, поднимавший эту проблему в рамках **традиционной научной лексикографии**, высказывал мысль о перспективности объединения лингвистического словаря и общей энциклопедии, а также о включении в переводный словарь лингвострановедческих сведений [4 и др.].

Привлечение энциклопедических сведений практикуется и в конкретных словарных описаниях, отражающих результаты лексико-семантических исследований. Так, в материалах к Интегральному словарю современного русского литературного языка [5] ряд статей существительных снабжен информацией культурологического характера, касающейся обыденной деятельности человека, обыденных астрономических представлений и т.п.; такая информация приводится в энциклопедической зоне.

Некоторые знания о фрагментах действительности имеются и в Новом объяснительном словаре синонимов русского языка [6 и др.]: указание бытийных нюансов (в терминах НОСС – семантических различительных признаков) необходимо для объяснения различий в сферах употребления синонимичных единиц. Сведения о существенных для различения синонимов свойствах реалий размещаются в зоне значения синонимического ряда: «**Зданиями** могут быть названы только достаточно большие сооружения, построенные по определенному архитектурному замыслу» ([6, С. 82], статья синонимического ряда «ДОМ 1, ЗДАНИЕ и др.»; «... **снедь** может и подаваться на стол, и лежать в сумке или в холодильнике» [там же, С. 96], статья син. ряда «ЕДА 1, уходящ. СНЕДЬ и др.»; «**Орнамент** тесно связан с национальной культурой, с той или иной эпохой» [там же, С. 446], статья ряда «УЗОР 1.1, (спец. ОРНАМЕНТ)»). По сути дела, связи, проводимые от <снеди> к <холодильнику> или от <орнамента> к <культуре> - это связи энциклопедического плана, не отражаемые в толкованиях и не имеющие также устойчивых сочетаемостных манифестаций (в этом плане связь от <здания> к <архитектуре> представляется более «лингвистичной»: например, возможно употребление атрибутивной конструкции *здание старинной/современной архитектуры*). (Отметим, что характер связи между сущностями /лингвистическая vs энциклопедическая/ может зависеть от направления «стрелки». Любопытный пример с единицами <медведь> и <берлога> приведен в [7]: для имени *медведь* <берлога> – сущность энциклопедического

порядка, а для имени *берлога* ‘медведь’ – семантический компонент; близкая ситуация наблюдается и при соотнесении сущностей <снедь> и <холодильник>).

В прикладной области можно назвать работу [8], отчасти предваряющую структуру РУСЛАН; в ней аппарат энциклопедических функций предлагалось применить при описании терминов узкой предметной области («Электроизмерительные приборы»); в работе показано, как с помощью энциклопедических функций можно отразить различные свойства реалий: принцип действия прибора, способ включения и др.

В [9] описывается проект объединенного ресурса, включающего словарь тезаурусного типа и онтологию (формальную модель предметной области), которые были бы способны функционировать как единое целое во многих приложениях, связанных с обработкой ЕЯ.

В [10] рассматривается задача автоматического пополнения онтологии сведениями из традиционного энциклопедического издания - «Российского энциклопедического словаря» (М., 2001), представленного в электронной среде. Результатом должна стать онтология, опирающаяся на многопрофильное профессиональное энциклопедическое описание; разработанная авторами технология применима также к конвертированию в форму онтологий традиционных семантических лексикографических произведений.

Соотношение общелексических описаний и описаний предметных областей постоянно находится в зоне внимания специалистов по прикладной лингвистике. Например, сопоставлению идеографической структуры лингвистического тезаурусного ресурса RussNet и логико-понятийной схемы предметной области («Кадровые назначения») посвящена работа [11], где с опорой на этот общелексический тезаурус и на <предикатную> модель предметной области решается одна из задач направления Information extraction.

В настоящем исследовании мы попробуем выяснить, какие виды энциклопедической информации «навевает» сплошной просмотр общей лексики; для просмотра взят словник «Краткого русско-французского учебного словаря» А.А.Зализняка (около 10000 слов; М., 1961); словник этого словаря, небольшой по объему, неоднократно применялся и зарекомендовал себя в качестве источника обыденной лексики в семантических исследованиях (например, при разработке БД «Лексикограф» [7] он изначально использовался как опорный массив).

Просмотр лексики в существенной мере опирался на ассоциативный, интроспективный отбор онтологических сведений, хотя при рассмотрении единиц словника привлекался также энциклопедический справочный материал – главным образом материал «Советского энциклопедического словаря» (в дальнейшем СЭС; М., 1987). Данный метод исследования является лишь предварительным; он нацелен на выявление как релевантной лексики, так и релевантного онтологического материала. В дальнейшем метод может быть подкреплён более пристальным изучением структуры материала справочных изданий, использованием корпусных данных при изучении контекстов описываемой лексики, а также более основательно апробирован путем систематизированного внесения энциклопедических описаний в словарь.

Прежде чем перейти к рассмотрению лексических классов, попробуем уточнить, какого рода информацию мы будем считать энциклопедической (внелингвистической, онтологической). К таковой будем относить информацию о реалии, а также о комплексе «слово-реалия»,

во-первых, выходящую за пределы прототипического толкования, т.е. некой минимальной информации, объясняющей значение слова или лексемы (примерные, «усредненные» варианты толкований можно почерпнуть из толковых словарей, таких, например, как словарь С.И.Ожегова и МАС);

во-вторых, выходящую за рамки описания валентностной структуры слова, поверхностно выражающейся предложно-падежными моделями управления;

в-третьих, выходящую за пределы классических тезаурусных связей ВЫШЕ и СИНОНИМ (при этом родовые имена, реализующие тезаурусную связь ВЫШЕ, могут дублироваться в толкованиях в форме интегрирующего компонента, если толкование имеет родо-видовую структуру);

в-четвертых, выходящую за пределы устойчивых лексических связей, выражаемых стандартными лексическими функциями.

Так, не к энциклопедической (а к лингвистической, семантической) информации будут отнесены сведения о том, что *велосипед* приводится в движение педалями, а *август* – восьмой по порядку месяц года; эти данные непосредственно входят в толкования имен. Не будет отнесена к энциклопедической информация о том, что некто может оказаться *арестованным* - > за [прототипически плохой поступок], поскольку для глагола *арестовывать* причина является участником, выраженным предложно-падежной конструкцией. Не будут считаться

энциклопедическими факты, что *вода* – это *жидкость*, а *лингвистика* также называется *языкознанием* (классические связи по тезаурусу). Не будет энциклопедическим знанием то, что сильный *дождь* часто называют *проливным*; такое знание описывается лексической функцией. Напротив, к энциклопедическим сведениям о названных реалиях можно отнести, например, то, что у велосипеда привод от двух педалей осуществляется через цепную передачу (СЭС); что август имеет продолжительность 31 день, что арестованные, как правило, обращаются за помощью к адвокатам, что вода – компонент большинства технологических процессов (СЭС), что диаметр дождевых капель – от 6-7 до 0,5 мм (СЭС), что в древности серьезный вклад в языкознание внес Аристотель (данный факт также отмечен в СЭС) и т.п. Вопрос состоит в отборе подобных сведений для конкретных лексических единиц и в способах размещения их в формализованном словаре.

Также следует отметить, что нами не будет рассматриваться узкоспециальная научно-техническая энциклопедическая информация, т.е. «сложная» информация, которая может быть связана с обозначаемыми реалиями в предметных областях – например, информация о структуре аминокислот при описании единицы *белок* или сведения по теории относительности при описании единицы *время*. Для нас объектом рассмотрения будут сведения о реалиях, характерные для более обыденной, так называемой «школьной» картины мира. При этом наивную картину мира, воплощающую непосредственное восприятие действительности, мы считаем недостаточной – словарь нацелен на анализ деловой прозы, которая может содержать определенные специальные знания – так что будем ориентироваться на некий «средний» по глубине уровень.

Отметим, кроме того, что задача отбора в словарь сведений о реалии связана с общеинформационной задачей структуризации знаний, развертывания «карт» предметных областей, соотношения тезаурусных и иных структур для представления информации. Можно надеяться, что размышления об энциклопедическом «окружении» общей лексики помогут и структуризации предметных областей; нас, в частности, интересует структура такой предметной области, как компьютерная лингвистика.

Конечно, не все пласты общей лексики подлежат энциклопедическим описаниям. Перечислим вначале некоторые отрицательные случаи.

Например, среди субстантивной лексики такие описания едва ли применимы для общих, родовых имен («межотраслевых омонимов» или «метаслов») типа *база*, *аппарат*, *функция*, *анализ* и

т.п., если они рассматриваются как изолированные единицы. Конкретные значения они, как правило, приобретают в составе терминологических сочетаний в предметных областях (*база данных, летательный аппарат, равномерно непрерывная функция, математический анализ* и т.п.), и уже в областях, а не в общем словаре могут сопровождаться подробными описаниями, в которых энциклопедическая составляющая, возможно, будет превалировать над лингвистической.

Вряд ли целесообразно энциклопедическое описание наименований сущностей из сферы восприятия: видимых, слышимых, осязаемых явлений: *темнота, треск, привкус* и др.; наименований субъективно осязаемых эмоций: *радость, ярость* и др.; такого рода сущности изначально связаны не с объективным научным или практическим знанием, которое оформляется и классифицируется как энциклопедическая информация, а с труднообъективизируемым миром субъекта восприятия.

Энциклопедические описания, по-видимому, маловероятны для ряда представителей одушевленной нарицательной лексики, в том числе для имен отношений: *земляк, автор, враг* и др. (хотя полностью исключить бытийных знаний нельзя; например, можно привести некоторые сведения по авторскому праву, некоторые сведения из этической сферы и т.п.). Для названий профессий (часто являющихся словарными дериватами от глаголов профессиональной деятельности) энциклопедическая составляющая (указание особенностей профессий) может приписываться родственному неодушевленному названию деятельности - например, не имени *сварщик*, а имени *сварка* или глаголу *сваривать* (если же имя деятеля не является производным от имени деятельности, например, *букинист*, то присвоение энциклопедической информации ему возможно).

Вряд ли подлежат энциклопедическим расширениям названия национальностей и жителей стран, которые встречаются в рассматриваемом словнике (*француз, болгарин* и др.); поскольку, по сути, такие имена относятся к отраслевой, географической лексике, и лингвострановедческие сведения, связанные с образовавшими эти имена топонимами, уместны в географическом, а не в общелексическом словаре.

Определенное количество лексики освобождается от энциклопедических описаний за счет того, что часть информации о реалиях объемлется валентностями слова. Например, для ряда имен деятелей можно выделить энциклопедический признак «возможность более узкой специализации», и этот признак в тексте может реализовываться в виде участника, выражаемого моделью

управления: *следователь* -> по особо важным делам; *артист* -> оригинального жанра и др. (Приведем для сравнения имена, формально не имеющие данной валентности, но по онтологическим свойствам допускающие специализацию: *рабочий*, *архитектор* и др.; специализация у них может, например, выражаться композитами: *рабочий-слесарь*, *архитектор-реставратор* и т.п.; для таких имен возможность специализации может быть отражена в энциклопедической зоне).

Теперь перейдем к рассмотрению лексических классов, для которых определенные типы энциклопедической информации, наоборот, выявляются. Будем говорить о лексике субстантивной, которая в наибольшей степени подвержена энциклопедическим описаниям в словарях.

Некоторый уже накопленный опыт ведения энциклопедической зоны показывает, что наиболее регулярным типом энциклопедической информации для имен служит информация о параметрах (характеристиках, признаках, свойствах), по которым прототипически характеризуется обозначаемая сущность; соответственно, наиболее массовой формой выражения онтологической информации в РУСЛАНе служит энциклопедическая функция Param, вводящая в описания параметры; среди них количественные («высота», «емкость» и т.п.) и неколичественные («форма», «цвет», «материал» и т.п.).

Функция Param удобна для своего рода анкетирования лексики на предмет выявления актуальных бытийных свойств референтов [1, 3]. Она как бы «призывает» лексикографа исчислить и собрать воедино разнообразные мыслимые свойства обозначаемой сущности.

В самом деле, перечень характерных свойств (X обладает свойствами Y_1, Y_2, \dots, Y_n) составляет один из фундаментальных типов декларативной информации о реалии, наряду с отнесением к родовому понятию (X есть Z) и исчислением разновидностей (X подразделяется на W_1, W_2, \dots, W_m). При этом вторая и третья формы, вообще говоря, могут дублироваться в традиционных лингвистических описаниях: вторая форма, по сути, дублируется в толковании, которое в большинстве случаев имеет родо-видовую структуру, а третья может либо непосредственно дублироваться в тезаурусе (если там указываются нижестоящие понятия), либо дублироваться в инверсных толкованиях разновидностей (если X подразделяется на W_1, W_2, \dots, W_m , то W_1 есть X, W_2 есть X, ..., W_m есть X). Для первой формы (перечня свойств) прототипически только некое базовое, выделенное свойство (условно Y_1) войдет в толкование (как дифференцирующий компонент при родо-видовом

представлении), а остальные свойства (Y_2, \dots, Y_n) как раз и составят содержание энциклопедической информации.

При просмотре словаря можно назвать некоторые лексические классы, членам которых присущ примерно одинаковый набор параметров. В словаре параметры могут приписываться каждому слову отдельно (как это реализуется в РУСЛАНе в настоящее время), а могут также приписываться классу; тем самым может быть задана своего рода «энциклопедическая» грамматика.

Так, среди предметных сущностей, резервуары характеризуются емкостью, формой, материалом: ЭНЦ (*канистра*) = {Param (*емкость*); Param (*форма*); Param (*материал*)}. (Очевидно, что такое описание при семантическом анализе текста поможет интерпретировать фрагменты типа *плоская / пластмассовая / пятилитровая канистра*).

Для внутренних помещений (комната, коридор и др.) актуальны параметры площадь, форма, высота потолка, расположение окон, освещенность и, быть может, еще некоторые другие. В тексте эти параметры могут выражаться, например, фрагментами: *низкий / длинный / темный коридор; просторная / солнечная комната* и т.п.

Для предметов декоративного назначения (абжур, браслет и др.) существенны размер, цвет, форма, материал изготовления (*плоский белый абжур, золотой браслет* и т.п.).

Для предметов одежды и обуви важны параметры размер, фасон, материал (в том числе вид ткани, кожи, волокна): *туфли на высоком каблуке, замшевые сапоги, ботинки 45-го размера, шерстяная блуза* и т.п.

Для строений актуальны высота, этажность, форма, цвет, материал (*трехэтажный особняк; высокий кирпичный дом* и т.п.).

Конечно, объектам, в принципе относящимся к классам, также присущи индивидуальные сочетания свойств; так референт лексемы *алмаз 2* (в значении «украшение»), в отличие от других украшений, не нуждается в характеристике по признаку «материал» (материал, очевидно, входит в толкование), но характеризуется по признаку «ценность».

Среди других классов, «подверженных» характеристике с помощью параметров, можно назвать, например, плодово-ягодные растения (актуальны ареал культивирования, сроки созревания плодов, содержание полезных веществ и др. параметры); строительные материалы (актуальны состав, прочность, сфера применения и др.); информационные продукты (существенны объем, авторство и др.). Список легко может быть продолжен;

параметрический путь описания объектов действительно показывает себя продуктивным.

Желательно, чтобы исчисление свойств опиралось не только на интроспекцию, но и на статистические, корпусные данные. Например, материал НКРЯ показывает, что для сущности <комната> кроме названных выше параметров внутренних помещений статистически значимо также указание обитателей: *моя/твоя* и т.п. *комната* (значимо по корпусу также указание назначения комнаты, но оно выражается лингвистически, с помощью модели управления: *комната* → *для отдыха*).

Отметим, что на актуальность того или иного параметра способна проливать свет определительная конструкция с количественным прилагательным. Как, например, показать, что высота для потолка более «энциклопедична», нежели громкость для треска? Это можно сделать, обратившись к корпусу, в котором имеется множество вхождений конструкции *N-метровый потолок* (*четырёхметровый, пятиметровый, ...*) и несопоставимо меньше случаев вхождения сочетания *N-децибелльный треск* (так, в Яндексе в мае 2010г. на десятки тысяч страниц, где встретилось сочетание *N-метровый потолок* пришлось не более 500 страниц, где есть сочетание *N-децибелльный треск*).

Формально указание наименований типовых параметров с помощью функции Param часто уместно в рамках простановки полезных для интерпретации текста лексических связей, типа связи *ботинки* → *размер*: ЭНЦ (*ботинки*) = Param (*размер*). Но за эти рамки может быть совершен выход, если, например, указывать не только наименования параметров, но и их значения, актуальные для представляемых в словаре реалий. Так, при описании имен месяцев можно сообщать их продолжительность в днях: ЭНЦ = {Param (*сентябрь*) = *продолжительность*; *продолжительность* (*сентябрь*) = («точно», 30, «день»)}; для имени физического явления *свет* можно указать значение скорости: ЭНЦ (*свет*) = {Param (*скорость*); *скорость* (*свет*) = («приблизительно», $3 \cdot 10^8$, «метр/секунда»)} и т.п. Значения параметров можно разделять на постоянные (для описываемой сущности) и на переменные; для переменных можно указывать область допустимых значений: ЭНЦ (*дождь*) = {Sing (*капля*); Param (*капля*) = *диаметр*; *диаметр* (*капля*) = («от-до», 0.5, 7 «миллиметр»)}.

В РУСЛАНе путем несложной доработки формата функции Param можно отразить ряд способов задания количественных значений параметров: точные, интервальные, приблизительные, отчасти описательные, например: ЭНЦ (*комнатный 2*) = {Param (*температура*); *температура* (*комнатный 2*) = («приблизительно»,

20, «градус»}); ЭНЦ (*античный*) = {Param (дата); дата (*античный*) = («меньше или равно», 5, «век»)} и т.п.

Получающиеся структуры гомеоморфны информационным триадам «объект-параметр-значение»; основные варианты заполнения триад в зависимости от способа задания количественного значения исчислены в [12].

Дополнение наименования параметра его значениями – это уже переход от указания лексических связей к полноценной фактографии в словаре; таким образом можно запечатлеть существенную часть тех знаний о действительности, которые концептуализируются как параметрические.

Другие энциклопедические функции РУСЛАН воплощают альтернативные способы структуризации информации. Например, предложенная в [1] функция РАЗНОВИДНОСТЬ отражает раскрытие онтологических знаний по типу ветвления. Давно «укоренившаяся» в словаре функция СОСТАВНАЯ_ЧАСТЬ в определенной мере аналогична ей; она применяется, например, для декомпозиции сущностей, концептуализируемых как собирательные; составляющие элементы могут описываться как однородные: СОСТ_ЧАСТЬ (*алфавит*) = *буква* и как разнородные: СОСТ_ЧАСТЬ (*бакалея*) = {крупы; мука; соль; чай; кофе; пряности (СЭС)}. Функция РАЗНОВИДНОСТЬ и функция СОСТАВНАЯ_ЧАСТЬ (в ее разнородном варианте) соответствуют перечислительному способу раскрытия информации. Функция Content, применяемая пока только для указания химического состава описываемых реалий (Content (*абрикос*) = *каротин*), но пригодная также для отражения информационного содержимого (которое, как известно, по метафорическому механизму соответствует содержанию вещественному), отвечает стратегии развертывания знаний вглубь.

Перечисление разновидностей и раскрытие содержания составляют, вслед за указанием параметров, еще два типа энциклопедической информации, относительно регулярно подкрепляющихся лексическим материалом.

Например, перечисление разновидностей уместно при описании слов-денотатов более-менее отчетливо делимых областей деятельности: РЗВДН (*авиация*) = {гражданская ~; военная ~}. Применимо оно также при характеристике многих предметных артефактов, для которых деление на виды определилось различным функциональным назначением, традиционными особенностями производства и другими факторами: РЗВДН (*вагон*) = {пассажирский ~; грузовой ~ и др.}; РЗВДН (*сахар*) = {сахарный песок; сахар-рафинад}.

Некоторую проблему составляет конкуренция энциклопедической функции РЗВДН с зоной тезаурусных связей, если в последней, наряду со связями ВЫШЕ и СИНОНИМ, представлять связь НИЖЕ. Если же связь НИЖЕ, проставляемая вручную, отсутствуют в тезаурусной зоне (а в РУСЛАНе, например, она отсутствует), и эта связь может только автоматически генерироваться по словарю, то очевидно, что значимого пересечения с функцией РЗВДН не будет; последняя, в основном, будет фиксировать словосочетания, а по словарю будут генерироваться изолированные лексемы.

Раскрытие содержания как путь изложения энциклопедической информации может быть актуальным для ряда слов, обозначающих некий внешний, интерпретирующий или обобщающий заголовок по отношению к поясняющему этот заголовок содержанию. Это, например, слова с семантикой оценки, интерпретации: *авантюра, катастрофа, агрессия* и др., или слова с валентностью на содержание, не выражающейся предположно-падежными моделями: *аргумент, благосостояние, рекомендация 2* (совет специалиста) и т.п. Диагностическим контекстом для такого круга слов может служить возможность их подстановки на место субъекта в вопросах: *В чем состоит?; В чем заключается? В чем проявляется (авантюра, благосостояние, ...)?*. Ряд слов этого круга входит в параметрическую неколичественную лексику, подчиняясь также ее диагностическим контекстам: *узнать, сообщить, выяснить (аргументы, рекомендации и т.п.)*. В словаре признаки интерпретационности и наличия внутреннего содержания целесообразно отражать в лингвистической зоне (в РУСЛАНе, например, это удобно сделать в зоне семантических характеристик: СХ (*авантюра*) = ДЕЙСТВИЕ & AntiBon & СВЯЗАН & СОДЕРЖАНИЕ; семантическая характеристика СОДЕРЖАНИЕ как раз и будет показателем интерпретационности).

Раскрытие содержания пока (в очень ограниченной мере, путем выхватывания из целостного сценария неких изолированных существенных элементов, по сути дела, ключевых слов) позволяет осуществить функция Content: Content (*благосостояние*) = {уровень жизни; обеспечение; предметы; услуги; условия} (см. соотв. статью СЭС). В принципе, при развитии формализмов словаря раскрытие понятия сможет представлять собой, например, описание типовой структуры или типового сценария обозначенной реалии, а также, возможно, сведения о единичных (ярких) реализациях. (Отметим при этом, что не все слова данного круга, отвечающие названным контекстам, «нуждаются» в энциклопедической информации. Например, слова *мероприятие, способ, подход*, тяготеющие к

метасловам и лишены конкретики, слабо соотносятся с какими-либо четкими сценариями).

Еще один тип информации проступает благодаря коннотациям и переносным значениям, имеющимся у некоторых лексических единиц. Переносные значения по метафорическим путям указывают на существенное, маркированное в сознании носителя языка свойство исходной реалии, которое может быть выделено среди прочих свойств и зафиксировано в энциклопедической зоне; в частности, путем учреждения соответствующей энциклопедической функции. Любопытный пример дает бестиарная лексика, у которой метафорические трансформации регулярно отражаются во фразеологии: *Metaphorical (акула) = хищный*; ср.: *акулы мирового бизнеса*; *Metaphorical (слон) = {громоздкий; неуклюжий}*; ср. *слон в посудной лавке* и т.п.

Информация, поставляемая переносными значениями, занимает промежуточное положение между лингвистической и энциклопедической (она бытийна, но выявляется при анализе лингвистического материала, то есть «отлингвистична»), и она может размещаться в энциклопедической зоне, например, для единообразия. Вообще, в реальном жестко структурированном словаре энциклопедическая зона может использоваться и для фиксации некоторых собственно лингвистических сведений [1]. Словарь предоставляет свои форматы для описания комплекса <слово-реалия> в целом, и внутри его зон возможны некоторые чисто прагматические перемещения.

Опора на переносные значения в определенном смысле может быть противопоставлена роли различительных признаков в Новом объяснительном словаре синонимов: в НОСС лингвистические различия внутри синонимических рядов выявляются при исходном сопоставлении реалий, а в ситуации с метафорическими значениями лингвистический механизм метафоризации, наоборот, помогает экплицировать отдельные бытийные свойства.

В заключение отметим, что отраженный в работе анализ лексики и ассоциирующихся с ней типов онтологической информации направлен в первую очередь на обогащение лингвистического словаря. Работа не претендует ни на разработку подробной онтологии, ни на создание объединенной словарно-онтологической энциклопедии; ее цель – развить некий дополнительный аппарат в рамках собственно лингвистического продукта и нащупать методику его использования при представлении конкретных лексических классов.

Ряд вопросов остаются пока неотвеченными.

Например, еще не исчислены случаи, когда целесообразно приписывание энциклопедической информации несубстантивным единицам, т.е. прилагательным, глаголам, наречиям. Этот вопрос интересен тем, что традиционно энциклопедические сведения бывают соотнесены с субстантивной частью лексикона, а релевантность/нерелевантность прочих частей речи не обсуждается. В словаре РУСЛАН все же совершается некоторый выход за рамки субстантивов. Так, в [1] подробно рассмотрена функция *Anti* (т.е. антоним), понимаемая в метаязыке данного словаря и как традиционная лексическая (*Anti (мягкий) = жесткий*), и как энциклопедическая - антоним по прагматике, не являющийся точным лексическим антонимом, но выражающий некое широко практикуемое противоположение: *Anti (математический) = гуманитарный*. Эта функция массово применяется в том числе при описании прилагательных. Раскрытие содержания понятия также, в принципе, может применяться при описании адъективной лексики, например, аксиологической. Так, можно пытаться частично формализовать и представить с помощью сценариев, что такое *честный, воспитанный* и т.п. человек. Указание параметров реалии, в принципе, может использоваться в представлении глагольной лексики, к примеру, при описании некоторых технологических операций: ЭНЦ (*сваривать*) = {Param (*источник_энергии*); Param (*степень_механизации*) и др.; см. статью «Сварка» в СЭС} (Здесь возникает необходимость принять методическое решение: чему в словаре общей лексики приписывать энциклопедические данные, глаголу (*сваривать*), являющемуся исходной частью речи при обозначении операции, или его S0-деривату (*сварка*) - субстантивному термину, этимологически вторичному, но относящемуся к грамматическому классу, традиционно выступающему носителем энциклопедических свойств? В энциклопедическом словаре, где представляются только субстантивные именные группы, такой вопрос не мог бы возникнуть, а в общелексическом он правомерен). В целом соотношение несубстантивной лексики и энциклопедической информации требует дальнейшего рассмотрения.

Еще один вопрос касается принципов отбора информации. Конечно, в реальном структурированном словаре представляется только то, что может быть представлено с помощью имеющихся форматов. Но как далее ориентироваться среди представимого? Черпать материал из авторитетных энциклопедических изданий? Этот подход методичен и перспективен; он, например, реализуется при построении онтологии универсального характера на основе «Российского энциклопедического словаря» [13, 10 и др.]. Но для

общелексического словаря есть свои особенности. Во-первых, его словник даже на субстантивных классах заведомо не совпадает со словником энциклопедии; в последнем представлена только та часть лексики, которая традиционно ассоциируется с конкретным научным и научно-практическим знанием. Во-вторых, энциклопедия в целом не ориентирована на отражение обыденных культурологических сведений, весьма полезных в общем словаре. Если в целом и ориентироваться на профессиональную энциклопедию, то как действовать при описании расходящихся с ней лексических классов? Игнорировать для них энциклопедическую зону? А если не игнорировать, то как отбирать информацию? Опирается на корпусы, на ассоциации информантов и самого лексикографа? По-видимому, разумным является совмещать все эти тактики, выработав определенную меру их сочетания при реализации описаний в конкретном словаре.

Литература

1. Семенова С.Ю. Информация энциклопедического характера в прикладном семантическом словаре // Компьютерная лингвистика и интеллектуальные технологии : По материалам ежегодной Международной конференции «Диалог 2009» (Бекасово, 27-31 мая 2009 г.). Вып. 8 (15). – М.: РГГУ, 2009. С. 444-450.
2. Леонтьева Н.Н. К теории автоматического понимания естественных текстов. Часть 2. Семантические словари: состав, структура, методика создания. М.: Изд-во МГУ, 2001.
3. Семенова С.Ю. О представлении энциклопедической информации в прикладном семантическом словаре в форме параметров обозначаемой реалии // Горизонти прикладної лінгвістики та лінгвістичних технологій : Матеріали міжнародної наукової конференції MegaLing'2009. – К.: Довіра, 2009. С. 99-100.
4. 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.
5. Семиотика и информатика. Выпуск 32. Материалы к Интегральному словарю современного русского литературного языка (образцы словарных статей). М., 1991.
6. Новый объяснительный словарь синонимов русского языка. Первый выпуск. 2-е изд., испр. М.: Школа «Языки русской культуры», 1999.

7. Красильщик И.С., Рахилина Е.В. Предметные имена в системе «Лексикограф» // НТИ. Сер.2. М., 1992. С. 24-31.
8. Леонтьева Н.Н., Волковыская Е.В., Копылова О.В., Молчанова Т.В., Штернова О.А. Словарь энциклопедических функций и его роль в автоматическом индексировании // НТИ. Сер. 2., № 7, 1978.
9. Нариньяни А.С. Кентавр по имени Теон: тезаурус + онтология // Труды Международного семинара Диалог '2001 по компьютерной лингвистике и ее приложениям. Аксаково, 2001. Т.1. С. 184-188.
10. Бочаров В.В., Пивоварова Л.М., Рубашкин В.Ш. Логико-лингвистический анализ текстов определений в энциклопедических и толковых словарях // Горизонти прикладної лінгвістики та лінгвістичних технологій : Матеріали міжнародної наукової конференції MegaLing'2009. К.: Довіра, 2009. С. 76-77.
11. Азарова И.В., Гребеньков А.С., Ландо Т.М. Использование маркеров актантных позиций при анализе деловых текстов для расширения логической схемы предметной области // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 4-8 июня 2008 г.). Вып. 7 (14). М.: РГГУ, 2008. С. 11-16.
12. Семенова С.Ю. Алгоритм извлечения информации о параметрах из текстов рефератов и первичных документов // НТИ, Сер. 2., № 6, 1991. С. 22-32.
13. Рубашкин В.Ш. Онтологии – проблемы и решения. Точка зрения разработчика // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2007» (Бекасово, 30 мая – 3 июня 2007 г.). М.: Издательский центр РГГУ, 2007. С. 481-885.