

# НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ  
ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 7

Москва 2000

## ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

УДК 81'374[002:004]; 025.43:[002:004]

В. Ш. Рубашкин, Д. Г. Лахути

### Семантический (концептуальный) словарь для информационных технологий

Часть 3. Методы формирования и ведения словаря

*Рассматриваются методы формирования и ведения семантического (концептуального словаря), описанного в первых двух частях статьи, прежде всего методы формализации описаний объектов на примере фрагмента системы понятий "средства транспорта". Характеризуются общие принципы организации системы базовых признаков, лежащей в основе такого словаря. Оцениваются возможности реализации большого концептуального словаря; описываются разработанные средства его ведения.*

Настоящая статья продолжает серию публикаций ([1; 2]; см. также [3]), посвященных проблемам формализации реальных понятийных систем и создания концептуального словаря.

До недавнего времени работы по исследованию и формализации понятийных систем были представлены двумя слабо связанными друг с другом направлениями. Первое — это работы, посвященные теории и методологии формализации лексических подсистем. Второе — движимая практически-

ми потребностями, так сказать "полевая", работа по описанию и систематизации достаточно больших пластов реального языкового материала.

Конструктивная часть первого из названных направлений концентрируется в рамках двух научных дисциплин. Это лингвистическая семантика и философская логика.

В лингвистической семантике (см., например, [4-6]) выбирались первоначально лексические поля, обладающие логической прозрачностью и

обозримостью отношений между понятиями: термины родства, названия цветов и оттенков цвета, предметов обихода (мебель, одежда, кухонная утварь и т. п.), глаголы движения, термины психоэмоциональных состояний и мыслительных актов и др. Анализ здесь, как правило, весьма тщателен и подробен, охватывает как парадигматический, так и сочетаемостный, как собственно понятийный, так и узуально-лексический аспекты употребления лексических единиц. В лучших своих образцах он охватывает обширный и весьма нетривиальный языковой материал<sup>1</sup>. Вместе с тем, этот анализ обычно ограничивается содержательной систематизацией лексических подсистем и не доводится до того уровня формализации смысла, который необходим для использования лингвистических описаний в современных вычислительных моделях. В каком-то смысле сюда примыкают и работы, посвященные критическому анализу понятийного аппарата конкретных научных дисциплин. Необходимость анализа и модельной реконструкции понятийного аппарата возникает по внутренней логике развития науки. Общеизвестно, что именно интерес к понятийным основаниям математики стимулировал возникновение и развитие символической логики. Через некоторое время интерес к языковым проблемам переместился в теоретическую физику и далее в социологию, геологию, археологию, лингвистику. Однако и здесь — пожалуй, за исключением оснований математики — обсуждение почти всегда ведется на уровне “методологии науки”, а сколь-нибудь полная формальная реконструкция соответствующих понятийных систем мыслится как дело “прекрасного будущего”.

**Философская логика** в некотором смысле начинается там, где лингвистическая семантика заканчивает: она берет уже предварительно описанные (или интуитивно тривиальные) лексические подсистемы. Исходный материал здесь весьма ограничен. Обычно это всего несколько, максимум десятков-два, терминов. Это, например, термины, фиксирующие последовательность событий во времени, термины пространственных отношений, термины, определяющие логические модальности событий (*возможно* и *необходимо*), термины, в которых формулируются оценки или нормы и др. Целью анализа в этом случае является строгая логическая реконструкция отдельных лексических подсистем. Соответственно, результаты анализа, если они доведены до конца, оформляются как прикладное исчисление предикатов: “логика времени”, “логика пространства” (см. напр. [8]), “логика оценок” и т. п. Существенно, что в каждом таком исчислении всё языковое окружение формализуемого лексического фрагмента выводилось за рамки анализа и если в нем и было представлено, то лишь как определяемая на уровне языковой интуиции “область значений” переменной некоторого типа (ср., например, событийную переменную в модальной логике, используемую при рассмотрении свойств выражений вида  $Mx$  —  $x$  возможно и  $Nx$  —  $x$  необходимо). Выбор материала здесь мотивирован отчасти

его доступностью для точного анализа — логической обработанностью на уровне языковой интуиции, а отчасти — чисто философским интересом. Скажем, построения логики оценок можно рассматривать как продолжение традиционных философских исследований по этике, — но принципиально другими средствами.

Второе — “полевое” — направление характеризуется меньшими притязаниями на построение теоретических моделей и принципиально эмпирическим и утилитарным подходом к языковому материалу. Понимаемое в широком смысле, оно может быть отождествлено со всей практической лексикографией. В более узком смысле это, во-первых, практика создания ориентированных на человека **идеографических словарей** и, во-вторых, имеющая непосредственное отношение к нашей теме практика разработки **информационно-поисковых тезаурусов**, связанная с развитием технологии документальных ИПС (см., например, [9, 10]).

## ФОРМАЛИЗАЦИЯ ОПИСАНИЙ ОБЪЕКТОВ

Объектная лексика — центральная, ядерная часть понятийной системы [3, с. 51 и др.]. Практическую методологию формализации объектной лексики удобнее и нагляднее рассмотреть на конкретном примере. В качестве такового возьмем понятийную подсистему **названий транспортных средств** с сопутствующей терминологией. Удобство ее состоит в том, что, с одной стороны, эта подсистема несет на себе все признаки профессиональной терминологии, с другой — понятна не включенному в профессию рядовому носителю языка. Исходную (разумеется, далеко не полную) совокупность терминов можно представить следующим рядом:

*поезд, вагон, тепловоз, электровоз, судно, баржа, катер, лайнер, автобус, троллейбус, такси, автомобиль, (“мерседес”, “вольво”, “жигули”, “форд”, ...), шлюпка, дрезина, трамвай, грузовик, самосвал, кузов, кабина, колесо, гусеница-2<sup>2</sup>, самолет, велосипед, мотоцикл, буксир-1, телега, повозка, тягач, гужевой транспорт, автотранспорт, дирижабль, воздушный шар, вертолет, паром, сани, буер, байдарка, мопед, моторная лодка, подводная лодка, бензовоз, пикап, планер, танкер, ...*

С каким исходным материалом сталкивается инженер по знаниям, имея перед собой задачу логической реконструкции подобной лексической подсистемы? В его распоряжении имеются определения соответствующих терминов, составленные при участии специалистов данной предметной области. Такие определения могут быть найдены в общих или специализированных толковых и энциклопедических словарях<sup>3</sup>, в какой-то (меньшей) своей части — в терминологических стандартах. Нужно заметить, что профессионально ориентированные

<sup>1</sup>Ср., например, [7] и многие другие работы того же направления.

<sup>2</sup>Здесь и далее мы указываем номер лексического значения в соответствии с нумерацией значений, используемой в Словаре [13].

<sup>3</sup>Мы будем использовать, главным образом, определения, предлагаемые таким достаточно авторитетным изданием, каковым считается *Политехнический словарь* [11]. Используем также *Советский энциклопедический словарь* [12] и общезысловый толковый словарь С. И. Ожегова [13].

толковые и энциклопедические словари имеют целью дать не только номинальное определение термина, но и — коротко — самые существенные знания об определяемом предмете. Для целей построения концептуального словаря, в котором должны формализоваться в первую очередь объемные и — в строго оговоренных случаях — ассоциативные отношения между понятиями [2, 3], энциклопедические толкования, как правило, избыточны<sup>1</sup>.

Содержание работы, которая должна быть при этом выполнена, сводится к четырем следующим этапам:

1. Содержательный анализ словарных определений и вычленение из них базовых признаков и других информационно значимых элементов.

2. Построение иерархической системы базовых признаков ("деревя признаков").

3. Логическая реконструкция словарных определений в терминах построенной на втором этапе системы базовых признаков. (При этом построение формальных определений<sup>2</sup> должно осуществляться только в терминах допустимых и заранее специфицированных логических схем).

4. Перевод полученных толкований на язык описания лексики, принятый в системе ведения концептуального словаря<sup>3</sup>.

Относительно первого этапа следует заметить, что в этой технологии инженер по знаниям, как правило, не берет на себя ответственность за содержательную (профессиональную) корректность определений и толкований. Эта ответственность всецело возлагается на составителей словарей и стандартов. Критическая позиция ограничивается тем, что в сомнительных и неясных случаях инженер по знаниям может обращаться к разным словарным источникам, сопоставляя и оценивая их в основном с точки зрения логической прозрачности и последовательности. И лишь в случаях очевидного несоответствия предлагаемого словарем определения языковой интуиции или его явной логической несостоятельности инженер по знаниям может корректировать словарные определения, обращаясь по мере необходимости за консультацией к специалисту. Однако в любом случае на него возлагается вычленение из текста словарного определения информационно значимых для концептуального словаря компонентов и отделение собственно определяющей части словарной статьи, задающей набор необходимых и достаточных признаков определяемого объекта, от дополняющих ее элементов — информации справочного и пояснительного характера<sup>4</sup>. На него возлагается также восстано-

вление тех характеристик определяемого предмета, которые считаются очевидными для читателя-человека и поэтому не формулируются явно, хотя они и существенны для установления места данного понятия в классификационном ряду. Проиллюстрируем сказанное вполне типичным примером.

*трамвай* = вид городского рельсового транспорта; вагон или несколько вагонов, один из которых (моторный) приводится в движение тяговым электродвигателем мощностью 30–70 кВт. Т. получает энергию из сети постоянного тока напряжением 500–600 В от контактного провода, подвешенного на высоте 5,5–6 м. Рельсы служат обратным проводом. Т. передвигается со скоростью до 40 км/час по рельсовому пути. Для уменьшения шума в буксах, рессорах ставят спец. прокладки, на колесах — резиновые и др. покрытия [11].

Необходимыми с точки зрения установления объемных отношений между понятиями и описания лексической сочетаемости здесь можно считать, видимо, признаки 'рельсовый', 'городской', 'электрический', 'состоит из вагонов'. Указание мощности двигателя, питание его из сети постоянного тока, напряжение питания, высота контактного провода, и т. п. с точки зрения целей формализации понятийной системы являются избыточными. Эти характеристики определяют не столько данный класс транспортных средств, сколько используемые для его реализации технические решения, а они со временем могут существенно меняться. (Вряд ли мы будем использовать другой термин только потому, что в трамваях начнут применять, например, асинхронные тяговые двигатели или изменят напряжение питания.) С другой стороны, в приведенном описании нет явного указания на то, что по умолчанию это 'пассажирский транспорт'; из этого определения также нельзя извлечь информацию о том, что (пассажирский) трамвай движется по определенному (имеющему номер) маршруту. (Фраза *Прошел десятый трамвай* скорее будет означать 'прошел трамвай десятого маршрута', чем количество наблюдавшихся говорящим поездов.)

Приведем информационно значимые части некоторых определений<sup>5</sup>.

*автомобиль* = транспортное средство на колесном ходу с собственным двигателем для перевозки по безрельсовым путям...;

<sup>1</sup> Так, в *Политехническом словаре* отбор материала для словарных статей мотивируется следующим образом: "Основная цель Словаря — дать в сжатой форме объяснение (выделено мной — В. Р.) понятий и терминов... Статья содержит, как правило, определение понятия или термина (дефиницию), во многих случаях этимологическую справку, краткое описание процесса, механизма, прибора, аппарата, материала, закона, правила и т. д., их назначение или область применения." [11, с. 5]

<sup>2</sup> Далее будем именовать их "формальными толкованиями".

<sup>3</sup> Все касающиеся последних двух пунктов смотри в [2].

<sup>4</sup> Эта работа может осуществляться в режиме диалога с программой, обеспечивающей хотя бы частичный и приближенный семантико-синтаксический анализ текстов определений. Такая программа должна пытаться автоматически распознавать и выделить элементы толкования и предложить эксперту — специалисту предметной области — подтвердить либо отвергнуть каждый из результатов ее работы, либо дать возможность эксперту указать в тексте определения те дополнительные элементы, которые он посчитает нужным включить в систему базовых признаков данной предметной области и, соответственно, в формальное толкование определяемого термина. (Сказанное, очевидно, означает, что технологически вычленение компонентов толкования целесообразно совместить с построением предварительной "заготовки" формального определения данного термина). О значимости таких средств поддержки для создания большого концептуального словаря см. в последнем разделе данной статьи.

<sup>5</sup> В соответствии со сказанным выше тексты словарных определений существенно сокращены и в некоторых случаях отредактированы. (В качестве источников использованы названные выше словари.)

**автобус**=автомобиль с кузовом вагонного типа... [11];

**автобус**=многоместный автомобиль для перевозки пассажиров [12];

**автобус**=многоместный (9-170 пассажиров) автомобиль с кузовом гл. обр. вагонного типа... [13];

**троллейбус**=безрельсовый электрический транспорт; питание получает через контактную сеть [по умолчанию — пассажирский — В. Р. и Д. Л.];

**мотоцикл**=двух- или трехколесное транспортное средство, снабженное двигателем внутреннего сгорания (рабочим объемом 49,8 см<sup>3</sup> и более.);

**поезд**=сформированный и сцепленный состав из вагонов с одним или несколькими действующими локомотивами или моторными (самоходными) вагонами...;

**дирижабль**=управляемый аэростат, летательный аппарат легче воздуха с [винтовым] двигателем...;

**судно**=плавучее сооружение для перевозки грузов и пассажиров... и пр., а также для военных целей;

**баржа**=несамоходное грузовое судно, перемещаемое буксиром или толкачом...;

**танкер**=наливное самоходное судно...;

**теплоход**=судно, приводимое в движение двигателем внутреннего сгорания;

**конвейер**=транспортная грузовая машина непрерывного действия;

**шлюпка**=мелкое беспалубное гребное или моторное судно.

Содержательный анализ такого рода определенных и терминологии, которая в них используется, позволяет вычленил ряд существенных дифференциальных признаков (противопоставлений): наземный — водный — воздушный; рельсовый — безрельсовый; самоходный — несамоходный; электрический — парусный — дизельный — атомный — ...; пассажирский — грузовой; транспорт непрерывного действия — ...; и др.

На втором этапе инженер по знаниям должен систематизировать и логически организовать совокупность базовых свойств, выделенных на первом этапе, имея в виду достижение двух целей.

Во-первых, это — логическая полнота набора значений для каждого из вводимых признаков. В определениях (как и в деловых текстах вообще) широко используется практика **умолчаний**. При этом “обычное”, “само собой разумеющееся” значение признака зачастую не вербализуется, а вербализуются, так сказать, лишь отклонения от нормы. Скажем, такая характеристика судна, как наличие подводных крыльев всегда будет зафиксирована явно, тогда как противопоставляемая ей характеристика “обычных” судов (*водоизмещающее*) почти никогда не будет указана. То же самое в противопоставлении транспортных средств для специальных грузов: сравним наливное судно и судно для перевозки навалочных грузов, с одной стороны, и судно для генеральных грузов — с другой. В других случаях может просто не существовать термина, оформляющего нужное противопоставление, — скажем, есть термины *электрический двигатель*, *реактивный двигатель*, *двигатель внутреннего сгорания*, но нет термина *мускульный*

*двигатель*, который необходим для того, чтобы описать отличие, скажем, *велосипеда* от *мотоцикла*. Весьма характерный пример такого рода — отсутствие термина, противопоставляемого определению *транспорт непрерывного действия*; его необходимо сконструировать в процессе построения концептуального словаря (*свободно перемещающийся транспорт* или *транспорт дискретного перемещения* или ...?). Другой аспект логической полноты — это полнота самого набора признаков, достаточность его для разграничения всех классов объектов, рассматриваемых в данной предметной области.

Во-вторых, это — системность (взаимозависимость) признаков. Здесь имеется в виду прежде всего установление связей признаков по условиям применимости, т. е., построение дерева признаков [3]. И, далее, установление между значениями признаков имплицативных и ассоциативных связей.

Результат, который может быть получен для нашего примера (фрагмент терминосистемы), показан в Приложении 1.

На третьем этапе должны быть построены определения исходных терминов. Языковые средства, используемые для построения определений (допустимые логические схемы определений), должны быть предварительно специфицированы — например, так, как это сделано в [2]. Проиллюстрируем это на нашем материале.

*дирижабль*=АЭРОСТАТ[14.1]&

С АВТОНОМНЫМ\_ИСТОЧНИКОМ\_ЭНЕРГИИ [Т3.1]<sup>1</sup>

*судно*=[ТРАНСПОРТНОЕ\_СРЕДСТВО]\_ВОДНОЕ[Т2.2]

*баржа*=

[ТРАНСПОРТНОЕ\_СРЕДСТВО]\_ВОДНОЕ[Т2.2]&  
НЕ\_САМОДВИЖУЩЕЕСЯ[Т3.2]& ГРУЗОВОЕ[Т7.1]

*танкер*=

[ТРАНСПОРТНОЕ\_СРЕДСТВО]\_ВОДНОЕ [Т2.2] &  
САМОДВИЖУЩЕЕСЯ[Т3.1] &  
С\_АВТОНОМНЫМ\_ИСТОЧНИКОМ\_ЭНЕРГИИ[Т4.1] &  
НАЛИВНОЙ\_[ТРАНСПОРТ][Т9.2]

*автомобиль*=

НЕРЕЛЬСОВОЕ\_ТРАНСПОРТНОЕ\_СРЕДСТВО [Т10.1] &  
С\_АВТОНОМНЫМ\_ИСТОЧНИКОМ\_ЭНЕРГИИ[Т4.1]

*автобус* (x)=

АВТОМОБИЛЬ (x) &  
ПАССАЖИРСКОЕ\_[ТРАНСПОРТНОЕ\_СРЕДСТВО]  
[Т7.2] (x) &  
Эу(Э v (ИМЕТЬ\_ЧАСТЬЮ (x, y, v) & v=ДА) &  
КУЗОВ\_ВАГОННОГО\_ТИПА (y))

*троллейбус*=

НЕРЕЛЬСОВОЕ\_ТРАНСПОРТНОЕ\_СРЕДСТВО [Т10.1] &  
БЕЗ\_АВТОНОМНОГО\_ИСТОЧНИКА\_ЭНЕРГИИ[Т4.1] &  
ЭЛЕКТРОТРАНСПОРТ[Т5.4]

*шлюпка*=

[ТРАНСПОРТНОЕ\_СРЕДСТВО]\_ВОДНОЕ[Т2.2] (x) &  
МАЛОЕ[Т17.2] (x)

&

Эу(Э v (ИМЕТЬ\_ЧАСТЬЮ (x, y, v) & v=НЕТ) & ПАЛУБА (y))

Помимо сказанного, нужно иметь в виду, что в концептуальном словаре могут (и должны) фиксироваться имплицативные и ассоциативные связи не

<sup>1</sup>В квадратных скобках после терминов указаны индексы, соответствующие системе индексации понятий, принятой в Приложении 1. Референциальные индексы в формулах определений в случаях, где они очевидны, опускаются.

только между семантическими примитивами (базовыми признаками), но и между производными — вводимыми с помощью формальных толкований — понятиями. Ограничимся по этому поводу только одним примером. Вполне очевидное для современного носителя языка<sup>1</sup> суждение *Всякое судно, движимое мускульной энергией, есть малое судно* в словаре может быть формализовано, например, по следующей логической схеме:

[ТРАНСПОРТНОЕ\_СРЕДСТВО]\_ВОДНОЕ[Т2.2] &  
[ИСТОЧНИК\_ЭНЕРГИИ]\_МУСКУЛЬНЫЙ[Т5.1]—  
[ТРАНСПОРТНОЕ\_СРЕДСТВО]\_ВОДНОЕ[Т2.2] (x)  
& МАЛОЕ[Т17.2]

Чтобы включить в концептуальный словарь эту связь, нужно выполнить следующие действия:

1) добавить в словарь концепты, имеющие семантический тип И-толкование [2, п. 2.2], соответствующие левой и правой части данной эквивалентности;

2) в словарную статью первого из названных концептов добавить имплицативную связь со ссылкой на второй концепт [2, п. 2.8].

Аналогичным образом, если в концептуальный словарь включено определение вида *Весло есть движитель судна с мускульным источником энергии*, то можно рассчитывать, что система анализа текста, использующая такой словарь, будет в состоянии понять, что, скажем, в тексте *Они совершили погоню на веслах* не идет речь о путешествии на теплоходе, на лыжах, верхом и т. д.

## ОБЩАЯ ОРГАНИЗАЦИЯ СИСТЕМЫ БАЗОВЫХ ПРИЗНАКОВ

Задача инженера по знаниям не просто в том, чтобы логически реконструировать относительно автономные терминосистемы разных предметных областей, но еще и в том, чтобы интегрировать их в единую формализованную модель понятийной системы, в которой (модели) определено место каждой из терминологических подсистем и, соответственно, эксплицитно определены логические отношения между любыми двумя — сколь угодно далекими в смысловом и профессиональном отношении — терминами. Для такого объединения придется вводить концепты, которые обычно не вербализуются в рамках профессиональных терминосистем, но необходимы для точного описания смысловых связей конкретных терминов. Речь идет о смысловых единицах, функционально аналогичных так называемым семантическим маркерам, используемым в некоторых лингвистических теориях

для описания семантической сочетаемости лексики (*'одушевленность'* — *'неодушевленность'*, *'дискретность'* — *'континуальность'*, *'обладание сознанием'* и т. п.). Однако здесь, во-первых, имеется в виду существенно более детальная классификация действительности<sup>2</sup>, а во-вторых, используются эти смысловые единицы в рамках принципиально иной модели — модели "дерева признаков" [3]. Будем называть их признаками — категоризаторами. Поясним это на конкретных примерах<sup>3</sup>.

В интегрированной понятийной системе должна быть, в частности, определена сочетаемость любого имени объекта с характеристиками из таких часто используемых групп, как *'агрегатное состояние'*, *'физические свойства'*, *'вещественный состав'*, *'геометрическая форма'*, *'социальная/биологическая функциональность'* и т. п. Так, к концептам типа *'театр'*, *'правительство'*, *'город'*, *'аэропорт'*, *'кинофильм'*, и т. п. никакие характеристики, относящиеся к первым четырем группам (*'вес'*, *'плотность'*, *'прозрачность'*, *'цвет'*; *'медный'*, *'деревянный'*, *'кварцевый'*, *'твердый'*, *'жидкий'*, *'площадь'*, *'конический'*, *'спиральный'*, ...), вообще не применимы. Для другой очень большой группы концептов — *часы*, *голодильник*, *тигр*, *автомобиль*, *директор* — не применимы только характеристики *'вещественного состава'*. Соответственно, нужно включить в дерево признаков признак-категоризатор *Характеризуется материалом?\** (со значениями ДА и НЕТ).

В качестве более детального примера рассмотрим общую категоризацию характеристик по *'геометрической форме'*. Здесь оказываются необходимыми следующие признаки-категоризаторы. (1) Характеризуется ли вообще объект геометрической формой? (2) Если ответ на этот вопрос ДА, то дальнейшее деление определяется вопросом характеризуется ли объект *регулярной геометрической формой*. Для очень большого числа объектов предопределен отрицательный ответ на этот вопрос: ср., например, *'решетчатый микроскоп'*, *'круглый крокодил'*, *'кубический пиджак'* и т. д. Чтобы зафиксировать в словаре эту негативную сочетаемость, можно задать имплицативные связи, например, от таких концептов, как *'техническое устройство'*, *'живое существо'*, *'предмет одежды'* к концепту *'иметь нерегулярную геометрическую форму'*. Что касается *'регулярных геометрических форм'*, то они далее подразделяются на *'многогранники'* (с дальнейшей конкретизацией понятиями *'призма'*, *'пирамида'*, *'параллелепипед'*, *'куб'*, ...) <sup>4</sup>, *'тела вращения'* (с дальнейшей конкретизацией понятиями *'цилиндр'*, *'конус'*, *'шар'*, *'тор'*, *'форма ста-*

<sup>1</sup> Стоит обратить также внимание на зависимость истинности этого суждения от исторической эпохи!

<sup>2</sup> Здесь следует заметить, что, сталкиваясь с задачей детальной и логической последовательной категоризации действительности, сразу обнаруживаешь, насколько традиционная философия, в которой хотелось бы искать и находить ответы на этот счет, не готова такие ответы давать (и пока просто не осознает существования такой проблемы). Положение дел здесь напоминает то, в котором находилась лингвистика, когда на нее пришлось взглянуть глазами разработчика компьютерных лингвистических моделей: достаточно детальные описания языка были налицо, но полнота и логическая последовательность этих описаний оказались для инженерной лингвистики совершенно неудовлетворительными. Поэтому от разработчика концептуального словаря в этом плане требуются еще и содержательные решения — и, следовательно, достаточная философская культура. Речь идет не больше и не меньше, как о построении того, что можно назвать *инженерной онтологией!*

<sup>3</sup> Организация некоторой части данной подсистемы словаря показана в *Приложении 2*. Не будет лишним отметить, что близкий к описываемому здесь принцип организации словарного материала принят в полностью не зависимом проекте исследовательского коллектива Computer Research Laboratory (CRL), New-Mexico State University (все доступные материалы на сайте <http://crl.nmsu.edu>).

<sup>4</sup> Эти конкретизирующие понятия, конечно, в нужных случаях тоже упорядочиваются иерархически или вводятся посредством формальных толкований.

кана', ...), 'прочие регулярные формы' (с дальнейшей конкретизацией понятиями типа 'спиральный', 'решетчатый', 'змеевик', 'серпообразный', 'трубообразный', 'Т-образный', 'П-образный', и т. п.). (3) Существенным, конечно, является деление характеристик формы на 'планарные (плоские) формы' и 'объемные формы', причем внутри последних различаются 'полые' и 'не имеющие полостей' (сравним, например, *положить карандаш в коробку* и *'положить карандаш в доску'*). (4) Если необходимо детализировать сочетаемость имени объекта с такими, скажем, характеристиками протяженности, как 'длина', 'ширина', 'высота', 'толщина', 'глубина', придется вводить узкоспециальные признаки-категоризаторы, такие, как 'наличие выделенного продольного направления (характеризуется длиной?)' (ср. *длина трубы* и *'длина часов'*).

Понятно, что для описания сочетаемости конкретных терминов, даже при использовании механизма наследования свойств, требуется достаточно кропотливая работа; однако сама ее возможность обусловлена присутствием в словаре признаков-категоризаторов этого типа. И, в частности, чтобы интегрировать в единую систему понятий такую специализированную подсистему, как *транспортные средства*, нужно включить в словарь целую цепочку понятий, последовательно сужающих, начиная с общего понятия 'материальный объект', множество именуемых объектов и — в этой же иерархии — определяющих объемные отношения вводимых понятий. Здесь нужны такие противопоставления, как 'естественный объект' — 'объект культуры'; 'объект материальной культуры' и 'социальный субъект'; деление объектов материальной культуры по масштабно-уровневому основанию: 'сооружения и технические системы', 'технические устройства', 'узлы', 'детали', 'квизидетали' (см. Приложение 2).

## ВОЗМОЖНОСТИ РЕАЛИЗАЦИИ

Сказанное выше отвечает на вопрос, что нужно делать для логически последовательной формализации профессиональной терминологии. Остается вопрос, **выполнима ли** реально такая работа, учитывая, с одной стороны, необходимость детального и квалифицированного экспертного анализа большого числа достаточно узких терминологических подсистем и, с другой стороны, принимая во внимание, что минимальный общий объем концептуального словаря, имеющего практическую ценность с точки зрения эффективного использования в информационных технологиях, должен быть оценен порядком  $10^5$  лексических единиц. Достаточную же универсальность — с точки зрения охвата профессиональной терминологии — он приобретет при объеме порядка  $10^6$  единиц и более. Еще больше усложняет проблему то очевидное обстоятельство, что терминология развивается и изменяется вместе с развитием и изменениями, происходящими в науке, технике, народном хозяйстве. Иногда — как видно на примере информационных технологий — эти изменения могут происходить достаточно быстро и радикально. В определенной мере это

относится и к языку повседневного общения, где на наших глазах радикальные сдвиги в организации социальной жизни приводят к появлению и, соответственно, исчезновению целых пластов лексики. Так что здесь необходима еще и служба перманентного слежения за развитием языка. Все это приводит к выводу, что подобная работа не может быть выполнена в организационной форме одного локализованного во времени и пространстве проекта.

Организационное решение, на наш взгляд, должно состоять в другом: не в создании новых организационных структур, а в расширении и частичной переориентации функций уже существующих и успешно работающих словарно-терминологических служб. В таком качестве сейчас выступают, во-первых, редакции издательств и научные коллективы, осуществляющие подготовку и актуализацию общеязыковых и профессионально ориентированных толковых и энциклопедических словарей, и, во-вторых, службы, занимающиеся стандартизацией профессиональной терминологии. И те, и другие (словарные редакции в особенности) имеют хорошо отлаженную систему работы с корпусом экспертов, в их распоряжении огромный объем соответствующих материалов, они же осуществляют слежение за развитием языка. Таким образом, проблема состоит в том, чтобы, прежде всего, убедить общество и сами названные службы в необходимости и целесообразности такого расширения их функций<sup>1</sup> (для этого нужно, очевидно, найти такие формы использования конечного продукта, при которых эти службы будут экономически заинтересованы в его создании и широком распространении, так же, как они сейчас заинтересованы в распространении традиционных словарей). Другая — и весьма нетривиальная — сторона проблемы состоит в том, чтобы обеспечить эту работу адекватным набором средств методической и технологической поддержки. В рамках данной статьи остановимся именно на этой стороне дела. Ввиду отсутствия практического опыта и слабой научной разработанности вопроса здесь, вообще говоря, необходимо отдельное и достаточно подробное обсуждение. Предваряя таковое, попробуем дать краткое перечисление возможных средств поддержки.

1. Эффективное использование современных информационных технологий для работы с накопленным (традиционным) языковым материалом — прежде всего технологий быстрого поиска в полнотекстовых базах данных и технологий СУБД, — а также средств организации и поддержки совместной работы над проектом (поддержка "рабочих групп" средствами сетевых технологий).

2. Дистрибутивно-статистический анализ текста. Решаемые задачи: выделение ключевых понятий по данному корпусу текстов, относящихся к определенной предметной области; частичная автоматическая атрибуция (приписывание словарных характеристик) выделенных понятий по линейному и синтаксическому контексту.

3. Автоматический семантико-синтаксический анализ словарных определений с целью выделения из них основных семантических компонентов, включаемых в формальное толкование термина (см. также примечание 4 на с. 3).

<sup>1</sup> Принципиально важным и весьма обнадеживающим в этом отношении является то обстоятельство, что сами терминоведы хорошо осознают необходимость "логико-понятийного" анализа предметной области в процессе создания терминологических словарей, равно как и расширение отображаемого в таких словарях набора ассоциативных связей (ср., например [14, 15]).

4. Разработка системы регламентированного диалога для словарного описания понятий на языке специалиста предметной области. Такого же типа диалог нужен для выделения и описания частных систем семантических примитивов (базовых признаков) в ограниченных предметных областях. Здесь должны в максимальной степени использоваться те "заготовки" и частичные описания, которые могут быть получены средствами, указанными в п.п. 2 и 3.

## СРЕДСТВА ВЕДЕНИЯ СЛОВАРЯ

Для пополнения и редактирования демонстрационной версии словаря разработана программная система Gentest, ориентированная на администратора словаря и поддерживающая следующие функции:

- ввод новых словарных статей;
- редактирование введенных словарных статей;
- просмотр словарных статей;
- специальные способы просмотра.

Формирование словарной статьи вновь вводимого понятия производится в режиме регламентированного диалога. Проблема здесь состоит в том, что лишь два словарных признака — семантическая категория и тип — являются обязательными. Применимость всех прочих словарных признаков — будем называть их *дополнительными* — зависит от семантической категории, типа и, возможно, также от значения других словарных признаков. Таким образом, схема словарного описания определяется непосредственно в процессе формирования словарной статьи. Соответственно, используется процедура диалога, управляемого вводимыми данными. Техника реализации такого диалога в общем виде описана в [16]. В данной системе для его реализации использован следующий прием. Все дополнительные словарные признаки помещаются в общий список — таблицу. Порядок следования признаков в списке определяется зависимостями по условиям применимости: если значение признака  $P_i$  используется для определения условия применимости признака  $P_j$ , то  $P_i$  должен предшествовать  $P_j$  в списке. В этой же таблице указаны условия применимости каждого словарного признака в виде конъюнкции значений некоторых из предшествующих признаков. В процессе ввода, после того, как администратор словаря определил значение очередного словарного признака, перевычисляются условия применимости следующих признаков и предъявляется первый, оказавшийся релевантным. Указанная процедура повторяется до исчерпания списка дополнительных признаков. Поскольку в рабочем формате словаря одни и те же поля могут использоваться для представления разных словарных признаков (семантика поля может быть переменной), в таблице дополнительных признаков указан также способ реализации признака. Эта информация используется также при просмотре словарной статьи: здесь возникает задача интерпретации содержимого полей. Поэтому процедура вывода словарной статьи для просмотра строится аналогично процедуре формирования словарной статьи: перебор всех дополнительных словарных признаков с последовательным перевычислением условий их применимости, выбор очередного релевантного признака и определение

способа его реализации в структуре данных. Во всех случаях, когда набор значений признака перечислим, поддерживается предъявление соответствующего списка значений. Если список возможных значений формируется как подмножество концептов словаря (например, условие применимости признака выбирается из списка объектных концептов — СК=3; базовый признак для концепта, представляющего единицу измерения, — из подкласса "наименования числовых признаков" — СКТ=1.4 и т. д.), предъявляемый список значений формируется с помощью соответствующих фильтров. Указание на способ формирования списка возможных значений также помещается в таблицу дополнительных признаков.

В режимах просмотра можно выбрать один из следующих способов просмотра словаря:

- просмотр собственно словарной статьи (предъявляется набор словарных характеристик концепта);
- просмотр дерева признаков;
- просмотр развертки [1, с. 81] — для имен объектов.
- тестирование словаря.

В режиме просмотра дерева признаков показывается условие применимости, относящийся к нему куст признаков и значения одного из признаков этого куста. Администратор имеет возможность перемещать это "окно" по дереву признаков в любом желаемом направлении.

Понятно, что в такой системе описания достаточно непростой задачей оказывается логический контроль данных при вводе и, соответственно, контроль корректности процедур редактирования словаря. Особого внимания в этом плане требуют следующие пункты.

1. При удалении концепта и его словарной статьи предварительно нужно проверить наличие ссылок на удаляемый концепт в других словарных статьях. Очевидно, удаление такого концепта — пока не удалены все ссылающиеся на него — будет некорректным. Попытка удаления в таких случаях блокируется и выдается перечень всех концептов, зависимых от удаляемого.

2. Удаление/изменение дополнительных словарных признаков производится в порядке, обратном порядку их ввода: при необходимости удалить/изменить некоторый признак должны быть удалены и заново определены все следующие за ним.

3. Тщательного контроля требует также процедура добавления *имплицативных связей* [3]. Каждый концепт-консеквент должен удовлетворять условию совместимости с концептом-антецедентом. Соответственно, после добавления в словарную статью очередной имплицативной связи список выбираемых концептов-антецедентов должен перевычисляться. Следовательно, соответствующим образом должен быть определен фильтр, формирующий список выбираемых значений данного признака.

Опция тестирования словаря позволяет вычислять и просматривать определяемые словарем объемные и ассоциативные отношения между двумя объектными понятиями. (Для наглядного представления объемных отношений используются диаграммы Эйлера). Тестируемые понятия могут быть заданы администратором, либо выбираться

случайным образом. В последнем случае администратор задает количество случайно выбираемых пар.

## СПИСОК ЛИТЕРАТУРЫ

1. Рубашкин В. Ш., Лахути Д. Г. Семантический (концептуальный) словарь для информационных технологий. Ч. 1. // НТИ. Сер. 2.— 1998.— № 1.— С. 19–24.
2. Рубашкин В. Ш., Лахути Д. Г. Семантический (концептуальный) словарь для информационных технологий. Часть 2. Логическая интерпретация понятий и их описание в концептуальном словаре // НТИ. Сер. 2.— 1999.— № 5.— С. 1–12.
3. Рубашкин В. Ш. Представление и анализ смысла в интеллектуальных информационных системах.— М.: Наука, 1989.
4. Никитин М. В. Курс лингвистической семантики.— СПб: Научный центр проблем диалога, 1996.
5. Васильев Л. М. Семантика русского глагола.— М.: Высшая школа, 1981.
6. Wierzbicka A. Lexicography and conceptual analysis.— Ann Arbor, 1985.
7. Апресян Ю. Д. Лексическая семантика.— М.: Наука, 1974.
8. Зиновьев А. А. Логическая физика.— М.: Наука, 1972.
9. Белоногов Г. Г. и др. Компьютерная лингвистика в системе НТИ // Вопр. информ. теории и практики. Вып. 58.— М.: ВИНТИ, 1989.
10. Лахути Д. Г. Автоматизированные документально-фактографические информационно-поисковые системы // Итоги науки и техники. Информатика. Т. 12.— М.: ВИНТИ, 1988.
11. Политехнический словарь.— М.: Сов. энциклопедия, 1980.
12. Советский энциклопедический словарь.— М.: Сов. энциклопедия, 1981.
13. Ожегов С. И. Словарь русского языка.— М.: Русск. яз., 1991.
14. Герд А. С. Научно-техническая лексикография. // Прикладное языкознание.— СПб: Изд-во С.-Петербурга. ун-та, 1996.
15. Лейчик В. М. Прикладное терминоведение и его направления // Прикладное языкознание.— СПб: Изд-во С.-Петербурга. ун-та, 1996.
16. Рубашкин В. Ш. Формирование массивов данных путем диалога с системой классификации // Вопр. информ. теории и практики. Вып. 47.— М.: ВИНТИ, 1982.— С. 82–99.

Продолжение прилож. 1

№	Признаки и их значения	УП	Прочие связи, синонимы и примечания
T4	С автономным источником энергии?*	T3.1	
T4.1.	С автономным источником энергии*		
T4.2.	Без автономного источника энергии*		
T5	По типу источника энергии*	T3.1	
T5.1.	Мускульный*		→ T 4.2
T5.2.	Ветровой*		→ T 4.2
T5.3.	Топливосжигающий*		→ T 4.1
T5.4.	Электротранспорт		
T5.5.	Атомный		→ T 4.1
T6	По назначению*	T1.2	
T6.1.	Собственно транспортные*		
T6.2.	Вспомогательные и специальные*		
T7	По типу груза*	T6.1	
T7.1.	Грузовой		
T7.2.	Пассажирский		
T7.3.	Грузопассажирский		
T8.	По типу движителя и способу опоры на поверхность*	T2.1	
T8.1.	Колесный		
T8.2.	Гусеничный		
T8.3.	Шагающий		
T8.4.	Саный		
T9.	По виду груза*	T7.1.	
T9.1.	Сухогрузные		
T9.2.	Наливные		
T10.	По типу дорожного полотна*	T8.1.	
T10.1.	Рельсовый		
T10.2.	Нерельсовый		
T11.	По положению в водной среде*	T2.2.	
T11.1.	Надводные суда		
T11.2.	Подводные суда		
T12.	По способу опоры на водную поверхность*	T11.1.	
T12.1.	Водоизмещающие		
T12.2.	На подводных крыльях		
T12.3.	Глиссирующие		
T13.	По способу создания подъемной силы*	T2.3.	
T13.1.	Аэростат		= Самолет
T13.2.	Аэроплан		
T13.3.	Вертолет		
T14.	Грузоподъемность	T7.1. V T7.3	СКТ-4.1
T15.	Пассажировместимость	T7.2. V T7.3	СКТ-4.1
T16.	Водоизмещение	T12.1	СКТ-4.1
T17.	По размерам судна*	T2.2	
T17.1.	Малое		
T17.2.	Большое*		
...	...		

### Приложение 1

№	Признаки и их значения	УП	Прочие связи, синонимы и примечания
T0.	Транспортное средство	...	
T1.	По способу транспортировки*	T0	
T1.1.	Транспорт непрерывного действия		
T1.2.	Транспорт дискретного перемещения*		
T1.3.	Подъемные и погрузочно-разгрузочные устройства		
T2.	По среде перемещения*	T1.2	
T2.1.	Наземный		= судно
T2.2.	Водный		= летательный аппарат
T2.3.	Воздушный		
T2.4.	Космический		
T2.5.	Комбинированный		
T3.	Самодвижущееся?*	T1.2	
T3.1.	Самодвижущееся		
T3.2.	Не самодвижущееся*		



№	Признаки и их значения	УП	Прочие связи, синонимы и примечания
0.	Универсум (Объект)	—	
1.	Материальный: Идеальный?*	0	
1.1.	Материальный объект		
1.2.	Идеальный объект		
2.	Дискретность	0	= Счетность — Счетный — Несчетный
2.1.	Дискретный объект		
2.2.	Континуум		
3.	Характеризуется материалом*?	1.1.	
3.1.	Характеризуется материалом*		
3.2.	Не характеризуется материалом*		
4.	Характеризуется геометрической формой?*	1.1.	
4.1.	Характеризуется геометрической формой*		
4.2.	Не характеризуется геометрической формой*		
5.	Характеризуется агрегатным состоянием?*	3.1.	
5.1.	Характеризуется агрегатным состоянием*		
5.2.	Не характеризуется агрегатным состоянием*		
6.	Объект-Квазиобъект*	1.1.	
6.1.	Объект		
6.2.	Квазиобъект*		→ 3.2.
7.	Тип квазиобъекта*	6.2.	
7.1.	Статический квазиобъект*		
7.2.	Динамический квазиобъект*		
8.	Локализованность*	2.1.	
8.1.	Локализованный*		
8.2.	Не локализованный*		
9.	Масштаб объекта*	1.1.	
9.1.	Космический объект*		
9.2.	Геообъект*		
9.3.	Макрообъект*		
9.4.	Микрообъект*		→ 3.2.
10.	Гомогенность*	3.1.	
10.1.	Гомогенная система		
10.2.	Гетерогенная система		
11.	Стехиометричность*	10.1.	= Дальтоиды = Бертоллиды
11.1.	Стехиометрического состава		
11.2.	Нестехиометрического состава		
12.	Химическая элементарность*	11.1.	
12.1.	Химические элементы		
12.2.	Химические соединения		
13.	Дисперсность*	10.2.	
13.1.	Дисперсная система		
13.2.	Не дисперсная система*		
14.	Регулярность формы*	4.1.	
14.1.	Регулярной формы		
14.2.	Нерегулярной формы*		
15.	Тип регулярной формы*	14.1	
15.1.	Тело вращения		Тело, ограниченное плоскостями
15.2.	Многогранник		
15.3.	Прочие регулярные формы*		
16.	Правильность формы*	13.2.	
16.1.	Правильный		
16.2.	Неправильной формы*		

№	Признаки и их значения	УП	Прочие связи, синонимы и примечания
17.	Планарность*	4.1	
17.1.	Планарная форма*		= Плоская геометрическая фигура
17.2.	Объемная форма*		
18.	Выделенность продольного направления*	4.1	
18.1.	С выделенным продольным направлением*		
18.2.	Без выделенного продольного направления*		
19.	Длина	18.1	СКТ-4.1
20.	Характеризуется высотой*?	4.1	СКТ-4.1
20.1.	Характеризуется высотой*		
20.2.	Не характеризуется высотой*		
21.	Высота <sup>1</sup>	20.1	СКТ-4.1
22.	Площадь поверхности*	17.1	
23.	Полый?*	17.2	
23.1.	Полый—ДА*		
23.2.	Полый—НЕТ*		
...	...		
101.	Естественный: Искусственный*	1.1	
101.1.	Естественный объект*		
101.2.	Объект культуры*		
102.	Обладает сознанием?*	101.2	
102.1.	Социальный субъект*		
102.2.	Объект материальной культуры*		
103.	Вид социального субъекта*	102.1	
103.1.	Социальный индивид*		→ 3.2.
103.2.	Социальная система*		
104.	Тип социальной системы*	103.2	
104.1.	Социальная целостность*		
104.2.	Функционально специализированная система*		
105.	Масштаб объекта материальной культуры*	102.2	
105.1.	Техническая система*		→ 3.2.
105.2.	Техническое устройство*		→ 3.2.
105.3.	Узел*		→ 3.2.
105.4.	Деталь*		→ 3.1.
105.5.	Квазидеталь-1		→ 3.1.
105.6.	Квазидеталь-2		→ 3.2.
106.	Технические устройства по назначению*	105.2	
106.1.	Преобразователи энергии*		
106.2.	Устройства для добычи, получения и переработки материалов*		
106.3.	Устройства для формообразования*		
106.4.	Устройства для хранения, переработки и передачи информации*		
106.5.	Транспортные устройства		
106.6.	Устройства не-посредственного жизнеобеспечения*		
...	...		

<sup>1</sup>Высота в значении протяженность тела по вертикали (перечень всех значений слова см. в [13])

Материал поступил в редакцию 18.04.2000.