

# ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

УДК 81'374.2:004

В. Ш. Рубашкин, Д. Г. Лахути

## Семантический (концептуальный) словарь для информационных технологий. ч. 1

*Описывается семантический (концептуальный) словарь, предназначенный для использования в информационных технологиях, связанных с пониманием текста (обнаружение и исправление ошибок в текстах, распознавание письменного и устного текста, аннотирование и реферирование, текстовый информационный поиск, извлечение фактов из текста на естественном языке и др.). В данной части статьи описывается общий подход к построению такого словаря, основанный на использовании логического языка, соотношения логического и естественного языка, принципы организации концептуального словаря, функции, реализуемые с его помощью, принципы семантической классификации лексики.*

### 1. ОБЩАЯ КОНЦЕПЦИЯ СЕМАНТИЧЕСКОГО СЛОВАРЯ

#### 1.1. Назначение

Словарь, описываемый в настоящей статье, можно с равным правом называть концептуальным или семантическим, исходя из того, что характеристика “семантический” может относиться как к понятиям (концептам), рассматриваемым как инвариантные относительно естественных языков, так и к значениям слов и выражений определенного естественного языка, в какой-то мере отражающим его специфику (подробнее об этом см. в разделе 1.3).

В самом общем виде можно сказать, что назначение семантического словаря — обслуживание компьютерных информационных технологий. Технически семантический словарь должен быть реализован как специфический информационно-вычислительный ресурс, стандартным образом подключаемый к любым информационным технологиям — так, как подключаются к ним, например, системы проверки правописания.

Если говорить конкретнее, многие из современных информационных технологий — грамматические корректоры, системы автоматического перевода, технологии распознавания письменного текста (OCR) и устной речи (Speech Recognition), технологии СУБД и вообще технологии, в тех или иных пределах решающие задачу “понимания” текста, — все они и ряд других технологий достаточно определенно ощущают исчерпанность традиционно используемых лингвистических средств и методов (как правило, это морфологический и синтаксический анализ). Дальнейшее совершенствование этих технологий требует принципиального расширения используемого инструментария. Речь идет, например, о расширении пока весьма ограниченного набора опций грамматического контроля в грамматических корректорах; о мощных средствах смыслового развертывания (“избыточного индекс-

сирования”) запросов в информационно-поисковых системах; об использовании — наряду с морфологическими и синтаксическими — семантических фильтров в системах OCR и Speech Recognition и т. д. Семантические методы, конечно, не следует считать единственным направлением совершенствования информационных технологий, но верно и то, что без семантических методов, используемых самостоятельно или в сочетании с другими методами — например, методом аналогий, — существенное продвижение в указанных направлениях скорее всего окажется невозможным.

Нужно также иметь в виду, что привлечение новых методов — как это обычно бывает — открывает возможность постановки принципиально новых задач. Отнюдь не единственным, но впечатляющим примером может служить задача разработки средств, обеспечивающих прозрачный для смысла доступ к базам данных.

#### 1.2. Отношение к естественному языку

Отправным пунктом любых рассуждений и построений, касающихся концептуального словаря, должно быть признание того, что единицами описания в нем являются НЕ слова, а понятия.

Основным источником слов принято считать корпус текстов, написанных на данном языке. При этом технически слово вычленяется из текста достаточно просто: это часть текста, ограниченная с двух сторон пробелами (или пробелом и знаком препинания). После создания действующих моделей словоизменения формирование словника по данному корпусу текстов превратилось в почти рутинную процедуру (“почти” — так как остается еще проблема омонимии словоформ: например, словоформы *для* или *при* могут быть распознаны не только как предлоги, но и как глаголы *длить* или *переть* и т. п.).

Источники, которые могут быть использованы для формирования первичного инвентаря понятий

некоторой предметной области (не говоря уж о языке в целом), далеко не столь очевидны. Эффективной процедуры выделения понятий из текста не существует. Основным источником для инвентаризации понятий скорее принято считать профессиональную компетенцию специалиста. В этом смысле очевидным предшественником формализованного семантического словаря являются традиционные толковые и энциклопедические словари. Накопленный в их совокупности огромный содержательный материал, несомненно, должен широко использоваться при создании формализованного семантического словаря. Однако этот материал должен быть критически переосмыслен и принципиально реконструирован. Главное различие состоит в том, что традиционные словари ориентированы на интеллект человека, апеллируют к его знаниям и его языковой компетенции, тогда как формализованный семантический словарь сам является инструментом моделирования языковой компетенции и энциклопедических знаний. В традиционных словарях смысловые различия и смысл как таковой не столько формализуются, сколько **разъясняются** — путем перефразирования, использования более или менее развернутых описаний, предъявления типовых контекстов употребления. Из всего этого достаточно сложно вычленивать четко определенную, пригодную для формализованного описания систему отношений между понятиями. Именно эту работу и надлежит проделать создателю семантического словаря.

Но проблема не только в этом. Профессиональные толковые и энциклопедические словари совершенно игнорируют те смысловые единицы, которые считаются “очевидными”, “само собой разумеющимися”, т. е. заведомо известными читателю. Скажем, в (русском) энциклопедическом словаре можно найти статьи ‘внутренние болезни’, ‘внутренние войска’, ‘внутренний рынок’, ‘внутренняя речь’ и т. д., но нет словарной статьи для слова *внутренний* или отношений *находиться внутри*, *находиться над*, *находиться под* и т. п. Точно так же никакому составителю энциклопедического словаря не придет в голову включить в него словарные статьи, скажем, для числительных *один*, *два*, *три*, ..., *первый*, *второй*, ... Обнаружение и описание этих “очевидных” смыслов для составителя формализованного семантического словаря превращается в отнюдь не простую проблему. В общезыковом толковом словаре соответствующие слова, конечно, представлены, но там основное внимание обращается на то, чтобы зафиксировать многозначность слова (скажем, для русского слова *два* кроме основного значения будет отмечено еще значение ‘учебная оценка “неудовлетворительно”’), а не на определение/описание смысла.

Кроме того, описание понятийной системы именно как **системы** требует введения в словарь единиц, никогда не употребляемых в речи, — единиц, так сказать, внутрисистемных. Поясним это на двух небольших примерах. В общественных науках широко используются понятия ‘индивид’ (личность) и ‘социальная группа’. Однако в языке отсутствует специальное название для объединения этих двух классов. Между тем, в семантическом словаре такая единица — можно определить ее выражением ‘социальный субъект’ — необходима, поскольку ей должны быть приписаны некото-

рые характеристики, общие для обоих классов — например ‘быть (или не быть) юридическим лицом’, ‘иметь адрес’, ‘осуществлять целенаправленные’, и т. п. Другой пример — необходимость ввести в понятийную систему общее название для таких реальных, как *отверстие*, *щель*, *трещина*, *канавка*, ... — всего того, что характеризуется формой и размером, но не характеризуется материалом (ср. *алюминиевая втулка* и ?*алюминиевое отверстие*). Так появляется внутрисистемный термин ‘квазидеталь 2’. Потребность в таких “системоорганизующих” единицах конструктор семантического словаря обнаруживает постоянно.

### 1.3. Семантика вывода и семантика перевода

Семантический словарь должен поддерживать два принципиально разных типа процедур:

- 1) процедуры перевода текста с естественно-го языка на язык концептуального представления (“язык представления знаний”) и обратно;
- 2) процедуры манипулирования в рамках уже построенного концептуального представления.

Процедуры первого типа, очевидно, должны работать в составе лингвопроцессора информационного типа совместно с процедурами морфологического и синтаксического анализа; при анализе они имеют дело с семантическим истолкованием единиц естественного языка и — что весьма существенно — должны обеспечивать разрешение большей части лексических и синтаксических неоднозначностей.

Процедуры второго типа должны работать в составе машины (ограниченного) логического вывода; их основная задача — установление отношений между понятиями и суждениями.

С учетом этого приходим к тому, что правильнее говорить не об одном семантическом словаре, а, по крайней мере, о трех взаимосвязанных словарях.

- **Собственно концептуальный (понятийный) словарь.** Единицами описания в нем являются понятия. Каждое понятие может быть представлено стандартным термином и внутренним техническим идентификатором — например, числовым ключом. Описания в этом словаре ориентированы на формализацию правил построения и правил вывода для выражений семантического языка. (В том числе и правил вывода, опирающихся на энциклопедические знания, фиксируемые в самом словаре.) С точки зрения многообразия национальных языков единицы этого словаря (и весь он в целом) суть языковые универсалии.

- **Толковый словарь.** В нем для каждого слова должны быть указаны его понятийные эквиваленты. В отличие от традиционного (“книжного”) толкового словаря, многозначность здесь лишь фиксируется ссылкой на единицы другого уровня. Сами же смысловые различия и связи отдельных значений могут быть установлены лишь путем обращения к понятийному словарю. Существенно, что объектом понятийного толкования могут быть не только слова, но и словосочетания. Существенно также и то, что отнюдь не все слова естественного языка получают здесь собственно понятийное (“терминологическое”) толкование. Такие служебные элементы естественного языка, как предлоги, местоимения и другие анафорические элементы, наречия типа *очень*, *весьма* чаще всего не имеют в

логическом представлении собственно лексического эквивалента, но необходимы для правильного установления семантической структуры. Они могут получить в толковом словаре не понятийную, а функциональную ('указатель роли в предикатной структуре', 'модификатор значения' и т. п.) интерпретацию.

• **Обратный толковый словарь.** Он должен обеспечивать переход от понятия ко всем его способам выражения в данном национальном языке (или во всех национальных языках, работа с которыми поддерживается данной системой словарей). Обратный толковый словарь будет, очевидно, основным инструментом при семантическом синтезе. В этот же словарь, по идее, должны быть интегрированы средства описания узуальной сочетаемости слов, представляющих заданный смысл, (наподобие лексической функции Map в модели "Смысл ↔ Текст").

#### 1.4. Модельная база

Устройство собственно понятийного словаря определяется выбором языка семантического представления. Единственный класс языков, в которых обеспечивается полная и последовательная формализация смысловых связей, — это языки логического типа. Поэтому оптимальной представляется такая последовательность шагов по разработке семантического словаря:

1) систематизация языковой интуиции относительно семантики текста и семантики термина;

2) построение логического языка, формализующего сформулированные содержательные представления;

3) принятие прагматически оправданных ограничений, относящихся к правилам построения и правилам вывода в логическом языке;

4) установление способов логической интерпретации для разных категорий терминов и, соответственно, состава словарных описаний.

Таким образом, все элементы описания, используемые в собственно понятийном словаре, должны иметь логическую интерпретацию.

#### 1.5. Организация семантического словаря

Единицы семантического словаря распределены по семантическим категориям: свойства и отношения, статические характеристики и процессы, дескриптивные и функциональные ("строевые") единицы и др.; внутри категорий — по семантическим типам. Состав словарной статьи переменной и зависит от семантической категории и семантического типа. Кроме того, различаются семантические примитивы и производные термины. Для производных терминов приводятся формальные толкования.

В собственно концептуальном (понятийном) словаре определяются следующие виды отношений между понятиями:

(1) объемные отношения для понятий в пределах одной категории (включение, пересечение, несовместимость);

(2) ассоциативные отношения (часть — целое, объект — функция, пространственная смежность и др.);

(3) функциональные отношения (значение — наименование признака, признак — условие применимости, процесс — изменяющееся свойство, понятие — способы его вербализации и т. п.).

Ядро словаря образует иерархическая система понятий — семантических примитивов ("дерво признаков"), в которой характеристики отдельных понятий и отношения между ними устанавливаются с использованием механизма наследования свойств.\*

С точки зрения информационных технологий, обращающихся к словарю, он представляет собой библиотеку функций, возвращающих указанную характеристику заданного понятия, проверяющих наличие указанного отношения между понятиями или обеспечивающих переход от слов/словосочетаний к понятиям и обратно. Словарь может быть реализован стандартным способом — как DLL-модуль, включающий одновременно программный код, обеспечивающий вычисление словарных функций, и все компоненты самого словаря, оформленные как совокупность ресурсов. Основными критериями оптимизации словарных ресурсов являются, очевидно, минимизация времени доступа и, затем, минимизация объема используемой рабочей памяти.

#### 1.6. Формирование и пополнение

Средства ведения словаря могут быть ориентированы на администратора словаря — квалифицированного специалиста в области инженерии знаний — либо на пользователя информационных технологий — специалиста в своей предметной области. В первом случае семантический словарь представлен как реляционная база данных, структурные элементы которой определяются в терминах словарных характеристик. Во втором случае поверх такой базы данных должен быть сформирован интерфейс пользователя, позволяющий управлять словарем как базой знаний — в терминах, принятых в данной предметной области. Соответственно этому вычлняются два последовательных этапа в разработке средств ведения словаря. Ясно, что на первом этапе может быть использована стандартная технология баз данных, тогда как второй этап требует разработки специальной методологии, прецеденты которой в настоящее время отсутствуют.

#### 1.7. Теоретические и методические проблемы

Основные теоретические и методические проблемы, возникающие при создании словаря, суть следующие:

1) установление функционально и модельно мотивированной классификации лексики и состава словарных описаний;

2) установление границы между семантически примитивами и семантически производными понятиями;

3) определение способов описания семантических моделей управления;

4) развитие объективной методологии формирования концептуального словаря большой размерности (порядка 100 тыс. единиц).

\* См. Рубашкин В. Ш. Представление и анализ смысла в интеллектуальных системах. — М.: Наука, 1989; с. 60-68.

## 2. ТЕКСТОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ФУНКЦИИ СЕМАНТИЧЕСКОГО СЛОВАРЯ

Основные направления использования концептуального словаря для совершенствования и развития современных информационных технологий охарактеризованы в таблице. Здесь можно различить два случая:

Информационная технология	Требования к семантическому компоненту лингвопроцессора и концептуальному словарю
Семантический анализ делового текста в целом	<ol style="list-style-type: none"> <li>1. Разрешение лексической и синтаксической омонимии (выбор семантически наиболее удовлетворительного варианта синтаксического разбора предложения).</li> <li>2. Установление отношений ко-референции.</li> <li>3. Восстановление смысловой неполноты (лексически не выраженные отношения, эллипсис).</li> <li>4. Поддержка анализа "по прецедентам".</li> </ol>
Грамматический корректор	<ol style="list-style-type: none"> <li>1. Поддержка синтаксического анализа.</li> <li>2. Проверка семантических согласований в синтаксических связях.</li> <li>3. Проверка наличия семантически обязательных элементов в предложении.</li> <li>4. Поддержка анализа "по прецедентам".</li> </ol>
Информационно-поисковые системы	Поддержка процедур смыслового варьирования запросов.
Автоматический перевод	<ol style="list-style-type: none"> <li>1. Разрешение лексической омонимии (в особенности омонимии предлога)</li> <li>2. Поддержка синтаксического анализа.</li> <li>3. Поддержка перевода "по прецедентам".</li> </ol>
Автоматическое реферирование и аннотирование	<ol style="list-style-type: none"> <li>1. Выбор существенных тем через построение семантического представления исходного текста.</li> <li>2. Возможность генерализации тем, явно сформулированных в тексте.</li> </ol>
OCR и Speech Recognition	<ol style="list-style-type: none"> <li>1. Разрешение лексической омонимии (через выбор семантически наиболее удовлетворительного варианта синтаксического разбора предложения).</li> <li>2. Поиск подходящего прецедента.</li> </ol>
Технологии баз данных	<ol style="list-style-type: none"> <li>1. Поиск в текстовых столбцах по смысловому образцу.</li> <li>2. Пополнение баз данных посредством извлечения знаний из текстовых сообщений.</li> <li>3. Организация прозрачного для смысла доступа к большой совокупности БД (в частности, в Интернет).</li> </ol>

первый — непосредственное использование концептуального словаря в информационных технологиях — в информационно-поисковых системах, в автоматическом реферировании и аннотировании;

второй — использование концептуального словаря для поддержки процедур информационно-лингвистического анализа делового текста, результаты которых непосредственно и используются в информационных технологиях, например, улучшение результатов синтаксического анализа за счет дополнительных семантических средств разрешения лексической и синтаксической неоднозначности.

## 3. СЕМАНТИЧЕСКАЯ КЛАССИФИКАЦИЯ ЛЕКСИКИ (семантические категории и типы)

Исходные противопоставления для построения семантической классификации лексики можно попытаться извлечь из трех основных источников. Это:

- лингвистическая точка зрения (части речи);
- точка зрения логики;
- точка зрения современных информационных технологий.

Взгляд "от лингвистики" дает два наиболее существенных противопоставления:

- Имени и глагола. В понятийной плоскости оно может быть интерпретировано как противопоставление статических и динамических (процессы, действия) описаний.
- Знаменательной и служебной лексики. Имеется в виду, что первая может быть тем или иным способом соотнесена с реалиями внешнего мира, тогда как вторая скорее обслуживает внутренние потребности ("строй") языка. Этому соответствуют, например, утверждения, что предлоги или союзы не имеют собственного денотата.

Взгляд "от логики" предлагает:

- Аналогичное принятому в лингвистике противопоставление знаменательной и служебной лексики. К последней в логике чаще всего относят логические связки и кванторы.
- Деление на "индивиды" и "предикаты".
- Отсутствующее в лингвистике противопоставление свойства и отношения.
- Противопоставление элементарное ("атомарное") — производное (конструируемое средствами языка из элементарных единиц; вводимое определениями). Хорошо известно, что это противопоставление длительное время было в центре внимания теории познания, а затем стало предметом дискуссий при построении логических моделей естественного языка ("логический атомизм", "логический эмпиризм").

Взгляд "от информационных технологий" (если говорить конкретно — главным образом, от технологии баз данных) обращает внимание на характеристики, которые и в лингвистике и в логике оставались в тени:

- на противопоставление признак ("атрибут", технически — имя столбца таблицы) — значение признака;
- на необходимость классифицировать атрибуты по области допустимых значений (числовые, строковые и задаваемые списком — последние можно определить как "классификационные").

Объединяя и критически систематизируя все перечисленное, получаем следующий набор противопоставлений, существенных для выбора способа категоризации лексических единиц. Противопоставления, так сказать, "верхнего" уровня, в наибольшей степени влияющие на контекстную сочетаемость ("дистрибуцию") лексических единиц:

- знаменательная—служебная лексика;
- статика — динамика;
- свойство — отношение;
- наименование — значение признака;
- большая — малая чувствительность к контекстному окружению.

Заметим, что мы не включили в этот перечень противопоставление **индивиды — предикаты** и противопоставление **семантические примитивы — семантически производные термины**. Первое противопоставление для естественного языка мы считаем вообще непродуктивным. Наша позиция относительно этого пункта может быть выражена одним тезисом: **чистых имен в естественном языке нет**. Соответственно этому, мы строим логический язык, ориентированный на формализацию содержания делового текста, как язык без (индивидуальных объектных) имен; допустимые в нем индивидные константы — это числа и строки символов.

Это приводит к следующему перечню основных семантических категорий:

1. **Наименования признаков;** (1)
2. **Имена объектов;** (3)
3. **Служебные термины;** (4)
4. **Термы свободной сочетаемости;** (5)
5. **Процессы и действия;** (6)
6. **Статические отношения.** (7)

(После названия указан внутренний идентификатор семантической категории, используемый в спецификациях словарных описаний).

#### 4. ГЛУБИННАЯ И ПОВЕРХНОСТНАЯ СЕМАНТИКА

Это противопоставление, как уже было сказано в п. 1.3, вытекает из принципиально разной природы двух видов семантических процедур — логической обработки на уровне уже сформированного семантического представления, с одной стороны, и семантического анализа естественного языка, — с другой. В ЕЯ-тексте присутствуют элементы, которых в принципе не может быть в логическом представлении или которые не имеют в логическом представлении собственно лексического эквивалента, но существенные и необходимые для правильного установления семантической структуры. В качестве характерных примеров можно указать:

- 1) предлоги;
- 2) личные местоимения и другие слова, используемые в функции анафорических элементов;
- 3) указательные местоимения (*этот, тот, другой, такой (же),...*);
- 4) наречия типа *очень, весьма*;
- 5) слова и конструкции, представляющие в естественном языке единицы измерения.

Рассмотрим эти примеры подробнее.

**Предлоги.** В логическом представлении предлогов, разумеется, нет. Однако без учета семантики предлогов семантическая структура предложения не может быть правильно установлена. Предлог выполняет две основные семантические функции:

функцию указателя семантической роли управляемого предлогом существительного как актанта при предикате — глаголе или существительном, управляющем предложной группой:

*прибыл ИЗ Москвы, прибыл В Москву.*

Здесь предлоги (вместе с падежным оформлением предложной группы) маркируют ролевые позиции существительного (*начальный/конечный пункт движения*) относительно глагола *прибывать*;

функцию "свернутого в предлог" подразумеваемого отношения между теми же элементами. В семантическом представлении содержание этого отношения желательнее восстановить и выразить лексически:

- корпус с отверстиями* (ИМЕТЬ\_ЧАСТЬЮ);
- отверстие в корпусе* (БЫТЬ\_ЧАСТЬЮ);
- книга на столе* (НАХОДИТЬСЯ\_НА);

и т. п.

Кроме того, в некоторых случаях предлогу с семантической точки зрения следует приписать нулевую семантическую значимость и рассматривать его просто как средство синтаксического присоединения "хозяина" к "слуге". Разумеется, соответствующая синтаксическая связь получает ту или иную семантическую интерпретацию. Но эта интерпретация здесь осуществляется по общим правилам интерпретации синтаксических связей и игнорирует способ ее предложного оформления. Для этих случаев характерна возможность синонимического перифразирования конструкции с использованием другого предлога или вообще без предлога:

*весом в 3 тонны=весом 3 тонны.*

В большинстве случаев выбор правильной семантической интерпретации предложной связи определяется ее синтаксическим и семантическим контекстом, что предполагает, в свою очередь, необходимость детального и достаточно изощренного словарного описания.

**Анафорические элементы.** Ясно, что в логическом представлении не существует ничего, лексически эквивалентного словам *он, который* и т. п. При анализе текста эти слова порождают процедуру отыскания антецедента и — в собственно логическом представлении — процедуру отождествления референциальных индексов. В семантическом же описании этим элементам может быть сопоставлен общий термин типа *нечто* ("любой предмет речи") — без какого бы то ни было категориального уточнения ("объект — процесс", "свойство — отношение", "атрибут — значение"). При этом имеется в виду, что при интерпретации непосредственных контекстных связей, а затем при установлении референциального тождества с антецедентом категориальная и содержательная неопределенность будет снята.

**Указательные местоимения.** Как показано в цитированной работе В. Ш. Рубашкина (§ 3.13),

указательные местоимения в основном своем значении выполняют семантическую функцию **индикатора референции**. Это означает, что они существенны только на этапе перехода от естественного языка к семантическому представлению текста. В самом семантическом представлении — после того как установлено референциальное тождество/различие соответствующих знаменательных имен (имени, сопровождаемого индикатором референции и установленного процедурами анализа его antecedента), — указательные местоимения элиминируются, а знаменательным именам приписываются референциальные индексы, фиксирующие результаты анализа.

**Наречия типа очень.** Сочетаются с **нормативно оценочными значениями** *малая величина, большая величина*, либо с терминами, содержащими эти значения и толкуемыми по схеме:

$$P(x, \nu) \& M(\nu) \text{ и } P(x, \nu) \& B(\nu),$$

т. е., со словами типа *легкий, тяжелый, холодный, горячий, ...* (см. далее, п. 4.3). Синтаксическое присоединение наречия *очень* к прилагательным указанного типа семантически должно интерпретироваться как **усиление нормативной оценки**:

$$\text{тяжелый} \Rightarrow \text{BEC}(o, \nu) \& B(\nu);$$

$$\text{очень тяжелый} \Rightarrow \text{BEC}(o, \nu) \& \text{OB}(\nu).$$

При этом сам усилительный элемент в семантическом представлении не присутствует; он сам и его семантика присутствуют только в процедуре семантической интерпретации соответствующих естественных языковых конструкций.

**Единицы измерения.** Эти, казалось бы, узкоспециализированные элементы весьма существенны именно для деловых текстов. Логическая модель измеримой величины (числового признака) не предполагает явной спецификации единицы измерения. В схеме вида

$$\text{BEC}(x, \nu)$$

предполагается, что числовое значение признака уже приведено к стандартной (или общепотребительной) единице измерения. Но это означает, что такая стандартизация должна быть предусмотрена в процедуре перевода. Соответственно, единицы измерения должны иметь семантическое описание,

включающее, по крайней мере, помету, указывающую, является ли данная единица измерения стандартной или требует приведения; в последнем случае должен быть указан переводной коэффициент или — в общем случае — формула пересчета, скажем, формула

$$\nu^{\circ} K = 273,3 + \nu^{\circ} C$$

для пересчета температуры из шкалы Цельсия в шкалу Кельвина.

С чисто теоретической точки зрения, видимо, было бы последовательнее интерпретировать предлоги и другие подобные единицы так же, как и другие словарные единицы национального языка, интерпретируемые тем или иным набором альтернативных значений. Отличие от толкований знаменательной лексики здесь, однако, в том, что значением, скажем, предлога может быть не лексическая, а грамматическая единица в логическом языке — указатель роли (имя валентности), а значением единицы измерения — коэффициент или алгоритм ее пересчета. Впрочем, технически все это можно представить как обычную понятийную интерпретацию, сопоставляющую слову один или несколько понятийных эквивалентов, — если ввести в понятийную систему некоторое число специализированных **квазипонятий**, толкования которых включают названные выше вещи.

Так или иначе, необходимо признать, что, в отличие от единиц глубинной семантики — так сказать, дескриптивных понятий (*вес, лед, остров, пи-мезон, ...*), которые с точки зрения разнообразия национальных языков суть языковые универсалии, элементы поверхностной семантики суть языково зависимые элементы. Самый очевидный пример здесь — **предлоги**. Сам набор предлогов в разных языках различен. В языках с развитой падежной системой (таких, как русский) интерпретация предлога может зависеть и от падежа управляемого слова. Поэтому построить языково независимое описание здесь невозможно. **Единицы измерения** более универсальны, но и здесь есть очевидная (не столько языковая, сколько культурная) национальная специфика.

\* \* \*

Статья подготовлена при поддержке Российского фонда фундаментальных исследований (грант 97-06-80203).

*Материал поступил в редакцию 26.12.97.*