

ISSN 0548-0027

НАУЧНО-ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ



СЕРИЯ 2

Информационные процессы и системы

СОДЕРЖАНИЕ

ОБЩИЙ РАЗДЕЛ

Нестеров П. В., Белоногов Г. Г., Кузнецов Б. А. Компьютерная лингвистика в ВИНТИ 1

ИНФОРМАЦИОННЫЕ СИСТЕМЫ

Терещенко С. С. Центры информационного анализа, экспертизы, прогнозирования и коммерческой деятельности: функции, технология, организационная структура, подготовка кадров 4
Ефремов И. А. Адаптивная персональная информационная система и ее применение 17

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

Падучева Е. В. Отпредикатные имена в лексикографическом аспекте . . . 21
Рахилина Е. В. Предметные имена и имена предикатов 29

СПРАВОЧНО-ИНФОРМАЦИОННЫЙ РАЗДЕЛ

Марчук Ю. Н. Концепции в современной технологии 32

№5
1991

Е. В. Рахилина

ПРЕДМЕТНЫЕ ИМЕНА И ИМЕНА ПРЕДИКАТОВ

Обсуждаются некоторые особенности семантико-синтаксического поведения предметной лексики в предложно-падежных конструкциях и, в связи с этим, — способы представления семантической информации о предметном имени в словаре.

В статье затрагивается проблема лексикографического описания значительного пласта лексики, который, тем не менее, редко становится предметом лингвистических исследований. Речь пойдет о предметных лексемах, для которых на сегодняшний день остается открытым даже вопрос о способе их представления в словаре.

Действительно, основа словарного представления глагола — это его валентная структура. С одной стороны, она является своеобразным «каркасом» толкования, т. е. задает схему толкования предиката, а с другой — естественным образом вводит всю необходимую синтаксическую информацию о глаголе, так как валентная структура в конечном счете описывает управление глагольной лексемы и, таким образом, связывает семантику и синтаксис в лексикографическом описании глагола (ср. понятие модели управления, соотношения семантических и синтаксических валентностей предиката и др. в теории «Смысл ↔ Текст» [1, 2]).

Что же касается семантического описания предметных имен, то его «каркас» никак не определен — о предметных именах, мы, в сущности, не знаем ничего, кроме того, что они непредикатные и, следовательно, валентностей не имеют. Поэтому описание семантики и синтаксиса сочетаний типа *угол альфа*, *стаканчик из-под мороженого*, *болт с гайкой* и др. может вызывать трудности, если считать невозможным приписывать предметным именам какие-либо валентности.

Между тем, достаточно широко (в особенности, в работах, близких к формально-логическому направлению) распространена точка зрения, согласно которой все предметные имена являются предикатными (см., например, [3] и мн. др.). Следуя этому подходу, мы должны были бы рассматривать не синтаксическое отношение между непредикатными лексемами *угол* и *альфа*, а отношение между предикатом *быть углом* и его аргументом *альфа*. Лингвистически этот путь малопродуктивен прежде всего вот в каком отношении: если считать, что на семантическом уровне представления никак не моделируется разница между предметными и предикатными словами естественного языка, то каким образом на других уровнях языка возникает противопоставление таких классов лексем, как, например, имена и глаголы?

Кроме того, а priori не ясно, что такое, вообще говоря, *быть углом*. Валентность здесь вводится как попытка отразить синтаксические связи лексемы и в сущности не имеет отношения к ее семантической структуре — в принципе, формальной логике синтаксическая структура лексемы более интересна, чем семантическая. Таким образом, получается, что это, так сказать, валентности совсем другой природы, чем те, о которых мы привыкли говорить, описывая предикат *рубить* или *пилить*. Безусловно, существенным обстоятельством здесь является то, что в формальной логике (и в опирающейся на ее традиции аппарат «формальной» лингвистики) речь идет не о валентности лексемы, которая заполняется лексемой же, а о предикате, места

которого могут быть заполнены терминами, т. е. сущностями совершенно другого, не языкового уровня — референтами. Таким образом, эти отношения моделируют только некоторые синтаксические и референциальные свойства предметного имени, но не его лексическую семантику: последняя в этом случае становится заложницей чисто логического подхода, не способного продвигать ее вперед.

Другой путь решения этой проблемы предопределен некоторыми теоретическими принципами, принимаемыми авторами модели «Смысл ↔ Текст». Согласно этой модели некоторые группы имен признаются лексикографически связанными с предикатами. Они реализуют стандартный способ заполнения валентностей этих предикатов. Например, *строитель* = «тот, кто строит», *строение* = «то, что построили», так что для предиката *строить* первую его валентность описывает имя *строитель* (S_1 от *строить*), а вторую — *строение* (S_2 от *строить*).

Такой взгляд на отношение между именем и предикатом позволяет моделировать одновременно и семантику, и синтаксис имен типа *строитель* или *строение*. Действительно, как моделировать отношение *строитель* → *города*? Если считать, что *строитель* — «тот, кто строит X », то очевидно, что оставшаяся свободной объектная валентность на X предиката *строить* (*строить что*) «передается» имени S_1 . Отглагольное имя приобретает возможность наследовать какие-то валентности исходного глагола. Следовательно, согласно модели «Смысл ↔ Текст» отглагольное имя *строитель* содержит переменные в толковании и имеет валентности благодаря «семантической производности» от предиката (*строить*). Обратим здесь внимание на то, что при этом семантическая структура отглагольного имени отличается от семантической структуры предиката: предикату (например, предикату *строить*) соответствует структура, содержащая в вершине предикат (*строить* = «[некоторым способом] каузировать и начать существовать»), а имени — структура, содержащая в вершине терм (*строитель* = «тот, кто строит»).

Если пойти по этому пути, то легко обнаружится, что подобная семантическая производность по отношению к какому-либо предикату свойственна очень большому числу предметных имен. Прежде всего это относится к именам артефактов. В самом деле, всякий артефакт — это некоторое приспособление, устройство, сооружение, предназначенное для использования в каком-то виде человеческой деятельности. Поэтому предикат, описывающий эту деятельность, естественно укладывается в структуру толкования, в целом общую для всех артефактов: *ложка* = «приспособление для еды, такое что ...»; *кресло* = «приспособление для сидения, такое что ...»; *зонтик* = «устройство, предназначенное для предохранения человека от дождя/солнца, такое что ...» и т. д. (Здесь мы, конечно, отвлекаемся от разнообразной информации другого рода, входящей в лексикографическое представление предметного имени, — цвет, размер, материал и мн. др.).

Таким образом, в толкованиях подавляющего большинства предметных имен будут содержаться семантические переменные, унаследованные от тех предикатов, с которыми эти имена оказываются семантически связанными. В этом смысле предметные имена (по крайней мере, очень значительная их часть) являются предикатными. Заметим, что последнее утверждение касается не одних только артефактов. Например, такие имена натуральных классов, как названия культурных растений, очень естественно связываются с предикатами сажать или сеять, а имена плодов этих растений — с предикатом *есть* (ср., например, [4]).

Преимущества подобного описания предметных имен — прежде всего в описании семантики их синтаксиса. Рассмотрим, например, отношение посессивности: в нормальном прагматическом контексте *моя комната* = 'комната, в которой я (обычно) живу', *мой борщ* = 'борщ, который я сварил' или 'борщ, который я ем'. Конечно, в образовании и интерпретации посессивных конструкций есть свои нетривиальные правила, подробно обсуждать которые здесь было бы нецелесообразно (отчасти об этом см. [5]), но принцип построения этих правил ясен: они опираются на то обстоятельство, что в семантической структуре предметного имени есть предикатная часть и семантические переменные*.

Посессивные отношения — это важнейший тип семантико-синтаксических отношений, свойственных предметным именам. Другой тип таких отношений выражен в предложно-падежных конструкциях типа *кобура от пистолета, замок на двери, мужчина с газетой* и подобных.

С точки зрения семантики можно различать три типа таких конструкций. Во-первых, конструкции, описывающие отношение часть/целое — *стрелки часов* или *угол стола*. Это простые в плане лексикографического описания конструкции. Дело в том, что конституирующей лексикографической информацией о стрелках или угле будет то, что это часть некоторого объекта X. Таким образом, здесь синтаксическая конструкция описывает постоянное семантическое отношение, лексикографически заданное. Другой тип конструкций — тоже простой с теоретической точки зрения — это разнообразные варианты случайного соположения двух, вообще говоря, не связанных между собой объектов. Так, каждое из сочетаний типа *муха на полу, полотенце на стуле, береза с рябиной, коврик под диваном* описывает свой способ взаиморасположения двух произвольных объектов: в случае X на Y-е соприкасаются две рабочие поверхности объектов; в случае X с Y-м X и Y находятся близко (рядом), и т. п.

Третий семантический тип предложных конструкций не выражает отношения часть/целое и в то же время не обозначает простого соположения двух не связанных друг с другом объектов. Например, сочетание *ботинки со шнурками* интерпретируется обычно как ботинки, в которые вдеты шнурки (а не как ботинки, рядом с которыми шнурки, ср. *ботинки с сапогами, васильки с ромашками, береза с рябиной*). То же для сочетания *наволочка на подушке*, которое по преимуществу значит не то, что наволочка лежит на подушке, а то, что наволочка надета на нее. Таких примеров нетривиальной интерпретации предложно-падежных конструкций довольно много. Соответствующие пары лексем: *чашка/блюде, дверь/замок, пояс/пальто*,

* Исключения в данном случае подтверждают правило: действительно, в некоторых специальных прагматических условиях сочетание, например, *мой велосипед* может интерпретироваться не как 'велосипед, на котором я езжу', а, предположим, как 'велосипед, который я чиню' или еще каким-то другим способом. Это означает, что в данной прагматически отмеченной ситуации возникает некоторое новое окказиональное прагматическое отношение между двумя участниками ситуации, которое оказывается сильнее семантических лексикографически заданных постоянных связей, встроенных в лексему *велосипед*.

ключ/шкаф, болт/гайка, иглолка/нитка и мн. др. — находятся в некоторых отношениях, очень похожих на отношение часть/целое. Однако это не отношение часть/целое ни с точки зрения денотативной (блюде — не часть чашки), ни с точки зрения языковой, потому что партиитивные отношения в русском языке легко оформляются генитивной конструкцией (ср. *стрелки часов, угол стола, ручка двери*), а это отношение — никогда (нельзя сказать ни **шнурки ботинок*, ни **блюде чашки*, ни **наволочка подушки*).

Теоретическая проблема, связанная с описанием последнего типа отношений, может быть сформулирована следующим образом. Есть пары семантически связанных предметных лексем (т. е. таких, которые описывают объекты, являющиеся участниками одной и той же ситуации). При этом некоторые такие пары входят в предложно-падежные конструкции с предлогами *на, с, из* и др., причем последние в этих случаях приобретают нетривиальную семантическую интерпретацию. Те же пары свободно входят в конструкцию с предлогом *от*: X от Y-а (*ключ от шкафа*). И есть пары семантически связанных лексем другого типа: они если и входят в предложно-падежные конструкции, то только в конструкции с тривиальной интерпретацией*, но никогда — в конструкцию с предлогом *от*.

Ср., с одной стороны, пару *шнурки—ботинки*, образующую нетривиально интерпретируемые сочетания: *ботинки со шнурками, шнурки в ботинках, ботинки на шнурках, шнурки от ботинок* и т. п., и, с другой стороны, пары *молоток—гвоздь, коса—трава* с тривиально интерпретируемыми *молоток с гвоздем* (= 'молоток и гвоздь'), *коса на траве* (= 'коса, лежащая сверху травы'), а также запреты типа **гвоздь от молотка, *коса от травы, *топор от дерева* и др.

Если при решении данной задачи также исходить из того, что в лексикографическом представлении соответствующих лексем входит предикат (*молоток* — то, чем забивают гвозди, свай, колья ...; *шнурок* — то, чем зашнурован ботинок, корсет ...), то окажется, что обозначенные два типа предметных имен различаются тем, какого типа это предикат. Если это стативный предикат, то отношение между лексемами может быть описано конструкцией с предлогом *от*, а также другими предложными конструкциями с нетривиальной семантической интерпретацией. Предикаты нестативной, «динамической» природы описывают отношения в парах другого типа. Так, отношения в парах *иглолка—нитка, чашка—блюде, ботинок—шнурок* описываются стативными предикатами (в конечном счете сводимыми к предикатам пространственной локализации**, а в парах *молоток—гвоздь, пол—тряпка, коса—трава, кисть—потолок* и подобных отношения описываются динамическими предикатами, и это обстоятельство отражается в их языковом поведении).

Итак, толкования многих предметных лексем удобно строить таким образом, чтобы они содержали предикаты, семантически связанные с соответствующей предметной лексемой, и наследовали бы валентности этих предикатов. Это дает возможность описать некоторые семантико-синтаксические свойства предметных лексем. Семантическое же различие между предметной и не-предметной лексикой в таком случае оказывается состоящим не в том, что у предметных лексем нет переменных в толковании, а у предикатных — есть, а в том, что является вершиной толкования — предикат у предикатов, терм у не-предикатов.

* Под тривиальной интерпретацией имеется в виду такая, которую приобретает предложная конструкция, содержащая две «случайные», семантически не связанные предметные лексем.

** Ср. семантически близкое (хотя и не тождественное) стативное отношение 'часть/целое'.

рактика терминологических данных», шестая — «Коммуникативный формат ТБД». Видно, что авторы сконцентрировали свое внимание на трех основных направлениях: научно-теоретическом, методическом и организационном. Следует особо отметить нормализационную функцию фонда. Предполагается, что ТБД МФ РЯ будет влиять на все стили и сферы речи, закрепляя и развивая общеязыковую норму. Объединяются четыре нормализаторских течения:

— стандартизация научно-технической терминологии в изданиях Госстандарта;

— нормализация научной терминологии в изданиях Комитета научно-технической терминологии АН СССР;

— селекция лексических единиц информационно-поисковых языков, выражающаяся в выделении предпочтительных терминов (дескрипторов) и предпочтительных терминов (аскрипторов);

— индивидуальные усилия специалистов по культуре речи, выражающиеся в эпизодических публикациях и рекомендациях по применению того или иного термина.

Наиболее важна, по нашему мнению, нормализационная функция. Она проявляется при сборе и структурировании лексического фонда, а также при последующих обращениях к нему и заключается в стремлении к определенному единообразию описания терминологии:

— единообразию в определении состава полей формата записи и организации порядка их следования;

— единообразию в организации терминологической информации в пределах полей и способах ее кодирования;

— использованию единой классификации или совокупности совместных классификаций и средств для структурной организации терминологической информации.

Представляется, однако, что такое определение нормализационной функции не совсем полно, и вот почему. Проект не рассматривает эксплицитно противоречие между термином-логосом и термином-лексемом. Первый член понятийной системы данной предметной области существует независимо от словесной оболочки. Второй — слово естественного языка, функционирующее в разных сферах речевой деятельности (как известно, доля узко специальных терминов в любых текстах достаточно мала). Термин испытывает влияние тех же факторов, что и любое слово — возникает многозначность, синонимия, разные коннотационные оттенки и пр. В условиях научно-технического общения, когда за каждым термином, особенно в переводе с одного языка на другой, стоит не всегда четко осознаваемая предметная реальность, «словесные» характеристики термина играют большую роль в аспекте его нормализации. Все эти соображения заставляют терминологов различать нормализующий и описательный аспект, причем нормализующий относится и к термину как слову естественного языка, а не только к термину-логосу — возникает понятие языковой нормы, а отличие от предметной, которая определена выше. В фундаментальных трудах терминологов разделяются прескриптивный и дескриптивный аспекты; Х. Фельбер, например, обосновывает существование дескриптивной и прескриптивной терминологии [3]. Деление такого рода особенно заметно переводчику и/или редактору научно-технической переводной литературы: перевод в сильной степени эксплицитно различает между термином-словом и термином-понятием. Проблема перевода, кстати, тоже совершенно не затрагивается в проекте — и это зря. Отсутствие явного внимания к такому рода характеристикам термина приводит к неполной классификации источников терминологических данных. Так, в четвертой главе читаем: «В качестве источников ТБД могут использоваться также (до этого перечислены словари, классификаторы, стандарты и пр. — Ю. М.):

— монографии,
— вузовские и школьные учебники,
— рефераты книг и статей,
— описания изобретений к патентам и авторским свидетельствам,
— сборник экспресс-информации (то же, что и рефераты Ю. М.),
— текущие документы производства, предприятия и т. п. (стр. 32).

В этом списке мы не находим такого источника, как тексты книг и статей, только рефераты. Между тем, есть значительная разница между текстом журнальной статьи и ее рефератом. Реферат может, например, исключить непонятные ему термины или заменить их синонимами. Правда, в типологии источников появляется возможность выразить и каким-то образом обозначить «лексическую» характеристику термина, отнести его к каким-то разновидностям жанра, коммуникативной цели текста и т. п. Но все-таки в целом типология источников и система базовых признаков такой типологии не очень эксплицитно дескриптивны и прескриптивны ТБД.

Кстати сказать, в систему базовых признаков целесообразно было бы ввести признак типа происхождения термина. Несмотря на то, что речь идет о ТБД русского языка, многие входящие в этот язык термины являются кальками, и не всегда удачными, с других языков, чаще всего с английского (например, *хеширование* от *cash* или от *hush*). Список источников терминологии можно дополнить еще одним и очень важным — переводами научно-технической литературы и документации на русский. Переводы, кстати, есть разных типов: от нормативных, типов стандартов — до рабочих, которые могут давать лишь пробные, предварительные переводы новых терминов. «Тетради новых терминов» Всесоюзного центра переводов научно-технической литературы и документации являются реально одним из важных источников русской научно-технической терминологии. «Переводной словарь» упомянут в признаке «жанр текста». Однако переводной словарь может быть и нормативен, прескриптивен в энциклопедическом, например, издании, в то время как «Тетради новых терминов» и тексты научно-технических переводов — только дескриптивны. Если при всем этом принять во внимание долю переводной терминологии в общем количестве терминов, которая заведомо чрезвычайно велика, то можно представить важность таких источников и таких признаков в типологии источников банка терминологических данных.

В главах «Характеристики терминологических данных» и «Коммуникативный формат ТБД» приводятся представление формы термина, его грамматическая и семантическая характеристики, дополнительная информация о термине, форматы представления данных в ТБД, структура коммуникативного формата. Изложение материала четкое, снабжено примерами. Названы литература и источники формата ТБД. Говоря о публикации проекта в целом, следует отметить его безусловную полезность для информационных работников, лингвистов-прикладников и терминологов самого различного профиля.

СПИСОК ЛИТЕРАТУРЫ

1. Машинный фонд русского языка: идеи и суждения / Под ред. Ю. Н. Караулова. — М.: Наука, 1986. — 240 с.
2. Лингвистическая прагматика и общение с ЭВМ / Под ред. Ю. Н. Марчука. — М.: Наука, 1989. — 141 с.
3. F e l b e r H. Terminology manual. — Paris: INFOTERM, 1984. — 426 p.

Материал поступил в редакцию 03.04.91.

Сказанное, по-видимому, естественно отнести и к семантической структуре отглагольных имен ситуации (S_0 в терминологии модели «Смысл ↔ Текст»), ввиду того, что последние исследования их семантики показали, что между их семантической структурой и структурой инфинитива имеются существенные различия. Одним из способов отразить эту разницу было бы считать вершиной семантического представления S_0 семантические термины типа 'процесс', 'состояние', 'факт' и др.— ср., например, [6, 7], а также статью Е. В. Падучевой в настоящем сборнике.

СПИСОК ЛИТЕРАТУРЫ

1. Апресян Ю. Д. Лексическая семантика.— М.: Наука, 1974.

2. Мельчук И. А. Опыт теории лингвистических моделей «Смысл ↔ Текст». — М.: Наука, 1974.
3. Bach E. Nouns and noun phrases // *Universals in Linguistic Theory* / Ed. by E. Bach, R. T. Harms and oth. — Chicago, 1968.
4. Wierzbicka A. *Lexicography and conceptual analysis*. — Ann Arbor, 1985.
5. Рахилина Е. В. К семантике посессивного отношения в русском языке // *Функциональная грамматика русского языка*. — В печати.
6. Падучева Е. В. *Высказывание и его соотносительность с действительностью*. — М.: Наука, 1985.
7. Арутюнова Н. Д. *Типы языковых значений: Оценка. Событие. Факт*. — М.: Наука, 1988.

Материал поступил в редакцию 20.02.91.

НАШИ АВТОРЫ

НЕСТЕРОВ Петр Владимирович — доктор технических наук, профессор, директор ВИНТИ, Москва

БЕЛОНОГОВ Герольд Георгиевич — доктор технических наук, профессор, зав. отделом ВИНТИ

КУЗНЕЦОВ Борис Антонович — кандидат технических наук, зав. отделом ВИНТИ

ТЕРЕЩЕНКО Сергей Сергеевич — кандидат технических наук, зав. кафедрой ИПКИРа ГКНТ, Москва

ЕФРЕМОВ Игорь Александрович — младший научный сотрудник Московского инженерно-физического института

ПАДУЧЕВА Елена Викторовна — доктор филологических наук, ведущий научный сотрудник ВИНТИ

РАХИЛИНА Екатерина Владимировна — кандидат филологических наук, научный сотрудник ВИНТИ

МАРЧУК Юрий Николаевич — доктор филологических наук, профессор, зав. кафедрой прикладной лингвистики Московского государственного университета

Редактор Т. Н. Лаппалайнен

Технический редактор Л. В. Кутакова

Сдано в набор 02.05.91 Подписано в печать 25.06.91
 Формат бумаги 84X108^{1/16} Бум. кн.-журн. Литературная гарнитура Высокая печать
 Усл. печ. л. 3,36 Усл. кр.-отт. 3,91 Уч.-изд. л. 6,307 Тираж 4092 экз. Заказ 4166 Цена 1 р. 65 к.
 Адрес редакции: 125219, Москва, А-219, ул. Усиевича, 20а. Тел. 152-66-71
 Производственно-издательский комбинат ВИНТИ,
 140010, Люберцы, 10, Московской обл., Октябрьский пр., 403