

© 1994 г. Г.И. КУСТОВА, Е.В. ПАДУЧЕВА

СЛОВАРЬ КАК ЛЕКСИЧЕСКАЯ БАЗА ДАННЫХ

В статье содержится описание и лингвистическое обоснование экспертной системы "Лексикограф", разрабатываемой в ВИНТИ межинститутской исследовательской группой, в которую, кроме авторов, входят Е.В. Рахилина, Р.И. Розина, М.В. Филипенко, Н.М. Якубова, Т.Е. Янко (программное обеспечение системы осуществляется И.С. Красильщиком и Е.Н. Хасиной), см. нашу первую публикацию [1]¹.

В основе системы лежит идея представления словарной информации в виде лексической базы данных (БД) реляционного типа: информация о лексеме распределяется по признакам (полям), имеющим фиксированный набор значений; в результате пользователь может осуществлять поиск любой степени логической глубины в заданном множестве дескрипторов.

Система состоит из двух компонентов – собственно БД и экспертная часть. Основное внимание обращается на семантическую информацию о слове. Таким образом, речь идет, в сущности, о семантическом словаре, представленном в виде реляционной базы данных.

Для удобства разработчиков, БД поделена на несколько разделов. В данной работе речь идет о разделе "Глагол", который находится в стадии разработки. Раздел "Предметные имена" близок к завершению, см. [2] (в разработке этого раздела принимал участие В.А. Плунгян).

Словарная статья в Базе данных "Глагол" имеет следующие поля: 1) **заглавная лексема** (с примерами ее употребления в предложении); 2) **морфологическая характеристика** (по Грамматическому словарю [3]); 3) **актанты**; 4) **таксономическая категория**; 5) **толкование**; 6) **аспектуальная характеристика**; 7) **свод значений**.

Поле **актанты** и поле **толкование** в свою очередь делятся на подполя.

ТАКСОНОМИЧЕСКАЯ КАТЕГОРИЯ И ТОЛКОВАНИЕ

Узловым семантическим признаком слова является его **таксономическая категория** (Т-категория); различаются: действия, процессы, состояния, происшествия и т.д., приблизительно по З. Вендлеру [4]. Т-категория играет такую же роль для семантики слова, как часть речи – для грамматики.

Система Т-категорий имеет иерархическую структуру; например, внутри Т-категории процесс различаются субкатегории процесс **предельный** (*таять*) и процесс **непредельный** (*кипеть*); с другой стороны, категорию **действие в его целостности** (в которую входят только глаголы СВ) можно объединить с категорией **действие в развитии** в более общую категорию **действие**.

Все существенные семантические характеристики глагола – парадигма частных видовых значений, падежная рамка и мн. др. – так или иначе связаны с его категориальной принадлежностью и, в конечном счете, предопределены ею. Способом представления Т-категории глагольной лексемы, ее структурным воплощением в нашей БД, является **формат толкования**. Каждой категории соответствует особый

¹ В 1993 году работа поддерживалась Российским фондом фундаментальных исследований.

формат, и, соответственно, все глаголы с одинаковым форматом толкования принадлежат к одной и той же категории (или субкатегории).

Предпосылкой форматирования является расчленение толкования на отдельные синтаксически независимые компоненты – аналогично толкованиям в [5] и особенно в [6] (и в противоположность синтаксически связным толкованиям в модели "Смысл ↔ ↔ Текст"). Компоненты имеют предикативную форму, например: 'Субъект действует'; 'Идет процесс в Объекте'; 'Объект перемещается' и под. Форматирование толкования достигается за счет того, что каждый его компонент представляется как значение некоторого параметра (один из вариантов идеи форматированного толкования реализован в [6], где компоненты толкования глаголов речи идентифицируются как исходная предпосылка, диктум или иллюкутивная цель; ср. ~~также~~ частично форматированные толкования в [7], [8]). Примеры параметров: деятельность, результат, каузация, каузатор, процесс, предел, исходное состояние, новое состояние, экспозиция и под. С точки зрения БД, параметры – это названия подполей поля толкование. С точки зрения семантики глагола, параметр – это тип семантического компонента в семантическом разложении глагола и, соответственно, определенный аспект обозначаемой глаголом ситуации.

Толкование – это упорядоченный набор семантических компонентов, представленных как значения параметров. **Формат толкования** задается обязательным наличием в толковании определенных параметров.

Как уже говорилось, формат толкования глагола предопределен его принадлежностью к той или иной Т-категории. Так, для глагола Т-категории Действие обязательным (категориальным) будет набор параметров "деятельность" и "результат (соответствующий цели Субъекта)"; для глагола происшествия – "исходное состояние" и "новое состояние".

Кроме категориальных (форматообразующих) параметров, обязательных для толкования глагола данной Т-категории, в толковании могут быть факультативные параметры, т.е. такие, которые для каких-то слов или классов слов данной Т-категории принимают нулевое значение. Значение категориального параметра не может быть нулевым (т.е. компонент, характеризующий ситуацию по этому параметру, обязательно должен присутствовать в толковании); в самом деле, ситуацию, в которой отсутствует целесообразная деятельность и результат, нельзя назвать действием; ситуацию, в которой не происходит изменения состояния Объекта или Субъекта, нельзя назвать происшествием и т.д. Между тем параметры, факультативные для данного формата, у каких-то слов или подклассов слов данной Т-категории могут не участвовать в толковании. Например, параметры "каузация" и "процесс" являются факультативными для некоторых подклассов глаголов действия.

Так, в толковании конативов (типа *ловить*), в несов. виде, не участвуют параметры "каузация" и "процесс". Дело в том, что *ловить* значит 'прилагать усилия с целью'; т.е. не предполагается даже частичная каузация результата, что отличает конативы от обычных действий, типа *пришивать (пуговицу)* или *варить (суп)*.

Параметр "каузация" не принимает участия в толковании происшествий, обозначаемых некаузативными глаголами (*выздороветь, заснуть, забыть*). Для многих форматов факультативным является параметр "экспозиция".

Толкование включает компоненты двух типов – **метакомпоненты** и **квалификаторы**. Метакомпонент содержит тот же предикат, что и название параметра; добавляется только аргумент, например: 'Субъект действует', 'Идет процесс: в Объекте / в Субъекте', 'Объект/Субъект пребывает в состоянии', и под. В толковании той или иной глагольной леммы метакомпонент обычно сопровождается квалификатором, который раскрывает конкретное содержание деятельности, процесса, состояния и т.п. Например, квалификатор может фиксировать вид процесса в Объекте – 'Объект перемещается', 'Объект деформируется', 'Объект нагревается'; содержание деятельности Субъекта; участие Инструмента и т.д.

Такого рода многоступенчатость толкования – и, в силу этого, его избыточность – является намеренной и обусловлена задачами нашей БД: данная структура толкования возникла из необходимости обеспечить эффективный поиск классов лексем и представить в эксплицитном виде семантические соотношения между лексемами разных классов.

Строя толкование лексемы, мы не стремимся (по крайней мере на данном этапе) к исчерпывающей полноте: толкование лексемы в БД должно отражать только **повторяющиеся** семантические противопоставления, т.е. только то, что **объединяет** слова друг с другом. Что же касается семантических различий, которые уникальны для некоторой пары лексем, т.е. не являются "классообразующими", то они на данном этапе разработки БД в толковании этих лексем не эксплицируются. Так, глаголы *наполнить* и *заполнить*, в своих основных значениях, имеют одинаковые толкования; и вообще, их словарные статьи различаются только в поле **свод значений**. Аналогично для глаголов *приклеить* и *наклеить*. Глаголы типа *спилить* и *срубить* имеют совпадающие толкования и различаются только информацией в поле **АКТАНТЫ**: то, что на самом деле различает эти глаголы – способ действия, – в толковании этих глаголов не уточняется.

Ниже приведен пример словарной статьи.

1. ЛЕКСЕМА: *наклеить* 1

Почтальон наклеил марку на конверт

2. МОРФОЛОГИЯ: СВ 4а 0

3. АКТАНТЫ:

1) Сб	Агенса	Лицо
2) Об	Пациенса	Материальный предмет
3) Периф	Средства	Вещество; обычно не выражается
4) Периф	Места	Материальный предмет
5) (Периф)	(Инструмент)	(Материальный предмет)

4. Т-КАТЕГОРИЯ: действие {обычное}

5. ТОЛКОВАНИЕ:

экспозиция	: до t < МН Об не находился на Месте;	<презумпция>
	Об мог свободно перемещаться	<фон>
деятельность:	в t < МН Об действовал с Целью	<импликация>
	: воздействовал на Об с помощью Средства (и Инстр)	
каузация	: Деятельность КАУЗИРУЕТ Процесс	
процесс	: шел процесс с Об: синхронный: имеет Предел	
результат	: началось и в МН имеет место состояние	<ассерция>
	: Об находится на Месте;	
	Об имеет прочный контакт с поверхностью Места;	
	тем самым Об является деталью Места	
следствие:	Об не может свободно перемещаться	

6. АСПЕКТУАЛЬНАЯ ХАРАКТЕРИСТИКА:

НСВ *наклеивать*: действие {обычное}

7. СВОД ЗНАЧЕНИЙ: *наклеить* 2 – кумулятив от *клеить* 2,

глагол создания: 'клея 1, изготовить'.

Примечание. В словарных статьях используются следующие сокращения и условные обозначения: помета НСВ в поле 6 означает, что данный глагол входит в видовую пару; другие типы аспектуальных коррелятов глагола имеют специальные обозначения – Proc., Res., Iter. и под.; МН – момент наблюдения; Сб – Субъект, Об – Объект, Периф – актанта, не входящий в перспективу, (см. ниже). Факультативные актанта (и вообще факультативные компоненты толкования) помещаются в круглые скобки.

Кроме собственно семантической информации, толкование содержит пометы, касающиеся коммуникативного статуса компонентов: <презумпция>, <ассерция>, <имп-

ликация) (в обычном значении этих терминов; об импликация см. [9]). Дополнительно используются еще две пометы – (фон) и (импликатура). Эти пометы обозначают "ослабленный" коммуникативный статус: фон – это "ослабленная" презумпция, а импликатура – "ослабленная" импликация. Компонент с такой пометой имеет необязательный характер: соответствующее ему положение дел не входит непосредственно в состав ситуации, обозначаемой данным глаголом. Тем не менее оно тесно связано с этой ситуацией, почему и желательно его включение в толкование.

Фоновые компоненты (как и их "сильные" аналоги – пресуппозитивные компоненты) помещаются в поле "экспозиция". Примерами фоновых компонентов могут служить **фоновая деятельность** и **фоновая каузация**.

Фраза *Иван разбил чашку* предполагает, что Субъект что-то делал в тот момент, когда произошло описываемое глаголом событие; например, он мог идти и случайно задеть чашку, мог нести ее и нечаянно уронить, в результате чего чашка упала и разбилась, и т.д. Конкретный характер деятельности Субъекта, вызвавший непредвиденные последствия, конечно, не входит в содержание глагола *разбить* (ненамеренно), однако сам компонент 'Субъект действовал' должен быть включен в толкование; при этом он получит статус фонового.

Другой пример фонового компонента – фоновая каузация. Всякое изменение состояния происходит под действием определенной причины, хотя не всякий глагол изменения состояния является каузативным. Так, во фразах *Лужи быстро высохли*, *Трава пожухла* глагол не каузативный. Однако если причина происшедших изменений известна и важна для говорящего, она может быть эксплицирована при этих глаголах с помощью сирконстантов: *Лужи быстро высохли на солнце*; *Трава пожухла от жары*. Сочетаемость глагола с соответствующими обстоятельствами семантически мотивирована и должна предсказываться толкованием. В данном случае эта сочетаемость обеспечивается наличием в толковании фоновой каузации. У других подклассов глаголов изменения такой компонент отсутствует, а вместе с ним – и обусловленная этим компонентом сочетаемость. Ср., например, глаголы изменений, происходящих по законам природы: **Яблоки созрели от осени / от жары*, **Стемнело от вечера* (во фразах типа *Из-за жары яблоки созрели раньше времени* в сфере действия обстоятельства причины находится не глагол, а обстоятельство *раньше времени*).

Импликации – это не обязательные, а всего лишь вероятные следствия основной ситуации. Импликации, как и обычные следствия, помещаются в поле *результат*. Статус импликатуры имеет, например, компонент 'Субъект (или другое лицо) имеет Результат в своем распоряжении', входящий в толкование глагола *отрезать*, а также компонент 'ущерб Субъекту или другому лицу', входящий в толкование глагола *разбить* (ненамеренно). Импликатура второго порядка в семантике *разбить* – 'Субъект несет ответственность за причиненный ущерб'. Конечно, все эти семантические компоненты не являются обязательными для глаголов *разбить* и *отрезать*. Однако их включение в толкование этих глаголов – в "пониженном" статусе – представляется оправданным: полное семантическое описание должно предсказывать не только актанты глагола, но и менее обязательные элементы семантико-синтаксической структуры предложения – такие, как *мою* в контексте *Петя разбил мою любимую чашку* (где притяжательное местоимение обозначает субъекта ущерба), или как *себе* в контексте *Петя отрезал себе кусок торта* (где дополнение *себе* обозначает субъекта пользы). Именно компонент 'Субъект имеет Результат в своем распоряжении' в толковании глагола *отрезать*, *постелить*, *налить* и под. отвечает за возможность дополнения *себе* при этих глаголах и объясняет невозможность *себе* при других глаголах, у которых такой компонент отсутствует, ср.: *Петя отрезал себе торта*; *Петя постелил себе постель*; *Петя налил себе чаю*; но: **Петя забил себе гвоздь*; **Петя заправил себе постель*; **Петя посадил себе цветок*.

ИМПЛИКАТИВНЫЕ СВЯЗИ МЕЖДУ ПОЛЯМИ СЛОВАРНОЙ СТАТЬИ

В принципе, информация во всех полях словарной статьи должна предсказываться информацией в поле **ТОЛКОВАНИЕ**. В самом деле, Т-категория, как следует из сказанного выше, выводима из формата толкования. Аспектуальная информация в поле б – наличие видовой пары и допустимый набор частно-видовых значений – также во многих случаях предсказывается толкованием. Например, глаголы состояния, типа *знать*, как известно, не имеют актуально-длительного значения; глаголы постоянного свойства / соотношения (*весить*, *стоять*) не входят в стандартные видовые пары, и т.д. Есть, конечно, и ограничения, не выводимые из семантики; например, отсутствие парного НСВ у глаголов *очнуться*, *встрепенуться* – это лексическая случайность (lexical gap), см. [10]; то же можно сказать об отсутствии парного СВ у глаголов *утверждать*, *отрицать*.

Аспектуальная информация в поле **МОРФОЛОГИЯ** тоже выводима из толкования. В самом деле, из толкования, например, глагола *наклеить*, см. выше, однозначно вытекает аспектуальная принадлежность этого глагола к сов. виду. Уже первая попытка описания семантики вида методом толкований, осуществленная А. Вежбицкой, см. [11], показала, что толкование вида – это не что иное, как толкование лексического значения глагола данного вида. Перефразировки, предложенные Вежбицкой, являются толкованиями вида в том смысле, что они имеют структурные элементы, повторяющиеся в толкованиях всех лексем с одинаковым видовым значением. Именно на этом принципе основаны толкования видовых значений, разработанные М.Я. Гловинской, см. [12].

Тот же принцип неотделимости лексического значения от видового заложен и в нашу БД (точнее даже – лексическое значение неотделимо от видо-временного: глагол НСВ толкуется в настоящем времени, а глагол СВ – в прошедшем).

Впрочем, частичное преодоление этой неотделимости лексического значения глагола от вида достигается в нашей системе за счет формулирования правил, которые позволяют получить толкование аспектуального производного глагола по толкованию исходного глагола в видовой паре данного семантического типа, см. ниже.

Информация в поле **АКТАНТЫ** предопределена толкованием лишь частично, а частично содержит независимую семантическую и коммуникативную информацию. Каждый актант характеризуется по трем параметрам:

1) синтаксическая характеристика, т.е. поверхностный падеж; различаются: Субъект – подлежащее; Объект – прямое дополнение; и Периферийный актант (предложно-падежные формы косвенных дополнений и обстоятельств не фиксируются);

2) семантическая характеристика, т.е. глубинный падеж по Филлмору; различаются: Агенс, Пациенс, Место и под. роли;

3) таксономическая характеристика – ‘лицо’, ‘материальный предмет’, ‘вещество’, ‘числовой параметр’, ‘ситуация’ и под.

Избыточен только глубинный падеж: например, наличие в толковании компонента ‘Субъект действует’ означает, что Субъект данного глагола – Агенс. Синтаксическая характеристика и Т-категория актанта непосредственно из толкования не выводимы.

В толковании актант может быть назван своим поверхностным или глубинным падежом, а иногда и Т-категорией. Например, в толковании глагола *наклеить* актанты называются своими синтаксическими характеристиками (Субъект, Объект) и глубинными падежами (Средство, Место). Переменные для называния актантов в словарной статье не используются.

Сопоставление поверхностного падежа актанта с глубинным позволяет охарактеризовать глагол с точки зрения **коммуникативной перспективы** в смысле Якобсона-Филлмора (имеется в виду противопоставление центральных и периферийных падежей в якобсоновском ‘Учении о падеже’ и понятие перспективы по Филлмору [13]).

У обычного глагола Пациенс является синтаксическим Объектом (т.е. выражается

вин. падежом) и, тем самым, входит в перспективу. Существуют, однако, глаголы со **сдвинутой перспективой**. Таковы, например, глаголы полного охвата типа *наполнять, заваливать*, см. [14]. У них Пациенс выражен периферийным актантом, а Место, семантически периферийный актант, выражается прямым дополнением, т.е. входит в перспективу и является, тем самым, синтаксически центральным актантом. Ср. соотношение глубинных и поверхностных падежей у глагола с нормальной перспективой *наливать* и у глагола полного охвата *наполнять*:

<i>наливать воду в бассейн</i>	<i>наполнять бассейн водой</i>
Субъект – Агенс (лицо)	Субъект – Агенс (лицо)
Объект – Пациенс (вода)	Объект – Место (бассейн)
Периф – Место (бассейн)	Периф – Пациенс (вода)

Сдвиг перспективы порождает характерное для этих глаголов семантическое приращение – семантику "полноты охвата", ср. известные примеры: *загрузить картошку в машину* не то же, что *загрузить машину картошкой*, поскольку во втором случае, когда *загрузить* употребляется как глагол полного охвата, явно сказано, что машина загружена полностью, см. [15, с. 279]; [13, с. 513].

Принятая трехчленная структура описания актантов дает также возможность представить в эксплицитном виде многочисленные регулярные семантические соотношения между актантами, состоящие в диатетическом сдвиге, в результате которого синтаксический Объект меняет свою таксономическую или ролевую характеристику: *полоть сорняки* → *полоть грядки, мести сор* → *мести пол* (Пациенс → → Место); *пробить стену* → *пробить дыру в стене* (Пациенс → Результат); *привязать веревку к столбу* → *привязать лошадь веревкой к столбу* (Пациенс → Средство), *защитить рукав* – *защитить рубашку, закрыть дверь* – *закрыть квартиру, застегнуть молнию* – *застегнуть куртку, настроить струну* → *настроить гитару* (часть-целое), ср. описание соотношений этого рода в [15].

Следует подчеркнуть, что "прогностический потенциал" поля **ТОЛКОВАНИЕ** не означает, разумеется, избыточности прочих полей словарной статьи. Во-первых, из толкования выводится все-таки не вся лексикографическая существенная информация о лексеме; во-вторых, правила вывода, чаще всего, достаточно сложны, так что с практической точки зрения заведомо экономнее эксплицитировать "внешние" (аспектуальную, ролевую и т.п.) характеристики лексемы в соответствующих специальных полях. Потенциальная возможность вывода одного рода информации из другого, на которой мы настаиваем, носит, конечно, сугубо теоретический характер.

ПОЛЬЗОВАНИЕ БАЗОЙ ДАННЫХ

Пользование базой данных предполагается в нескольких режимах. Помимо семантической информации о каждом отдельном слове, пользователь может получать перечни всевозможных семантических классов глагольных лексем – в принципе, классы задаются любым признаком или любой конфигурацией (набором) признаков. Примеры классов: глаголы СВ, не входящие в видовую пару; физические действия, не допускающие инструмента (напр., *сорвать, жать* можно только руками или зубами) или, наоборот, требующие его (ср. *пахать, резать*). Выразимы на языке БД и традиционные семантические классы, такие, как глаголы движения, глаголы речи, глаголы эмоционального состояния и проч.

Среди классов глаголов, выразимых на языке БД, различаются **субкатегориальные** и **тематические**. Субкатегориальные классы – это подклассы внутри определенной Т-категории; в тематический же подкласс могут входить глаголы разных Т-категорий; ср. такие тематические классы, как глаголы изменения местонахождения, глаголы речи и под.

Субкатегориальные классы порождаются, в частности, разными значениями параметров "предел", "каузация" и "процесс". Так, Т-категория Процесс включает субкатегории, различающиеся значением параметра "предел": Процесс предельный (*сохнуть, таять*) vs. Процесс непредельный (*кипеть, гореть*);

Т-категория происшествие включает субкатегории, различающиеся значением параметра "каузация": Происшествие обычное, с Субъектом (*упасть, заболеть, забыть*) vs. Происшествие каузированное, с Объектом (*Ветер порвал провода = 'Ветер имел место, в результате провода были порваны'; Камень загородил вход в пещеру = 'Камень переместился, в результате вход оказался загорожен'*).

Т-категория Действие включает субкатегории:

– Действие обычное (есть процесс в Объекте, синхронный деятельности Субъекта; *красить стену, подметать пол, мыть кастрюлю*);

– Действие-конатив (нет процесса в Объекте, синхронного деятельности Субъекта; *ловить бабочку, решать задачу, уговаривать приятеля*);

– Действие сверхкратное (процесс в Объекте сверхкраткий и/или несинхронный деятельности Субъекта: *разбить бокал на счастье, бросить камень в пруд*);

– Действие абстрактное (деятельность Субъекта не специфицирована: *углубить яму*), и др.

Разные тематические классы глаголов порождаются, в частности, разными значениями параметров "деятельность" и "результат". Например, компонент 'прочный контакт' в поле *результат* задает класс глаголов прикрепления (т.е. каузации прочного контакта: *прибить, пришить, наклеить, привязать* и т.д.); компонент 'части Объекта (существуют)' или 'отверстие в Объекте (существует)', в том же поле, задает глаголы деформации (*разрезать, проколоть, пробить*); компонент 'Объект пребывает в функциональном состоянии: готов к использованию' задает класс глаголов каузации функционального состояния объекта (*настроить гитару*); *выстирать (рубашку)*; *вспахать (поле)* и т.д.); компонент 'Объект не существует' – глаголы уничтожения (*сломать (сарай), порвать (записку), сжечь (письмо)* и т.д.).

Среди множества семантических классов, задаваемых Базой данных системы "Лексикограф", для лингвиста представляют интерес так называемые **релевантные классы**. Релевантный класс – это такой класс (порождаемый набором или конфигурацией семантических компонентов), который характеризуется определенной общей особенностью поверхностного поведения входящих в него лексем. Например, релевантным классом является Т-категория: слова одной и той же Т-категории характеризуются одинаковой сочетаемостью с обстоятельствами цели, времени и многими другими общими свойствами. Обычно эта сочетаемость считается свободной; на самом деле она предопределена Т-категорией. Релевантный класс образуют также глаголы полного охвата; глаголы, допускающие генитивный субъект; глаголы, подчиняющие косвенный вопрос, и мн. др.

В оптимальном случае нахождение релевантного класса позволяет не только получить список глаголов с определенной особенностью поверхностного поведения, но **объяснить** наличие у данной группы глаголов данного свойства. Например, презумптивный статус компонента "деятельность" у конативов, ср. *решить (задачу)*, *уговорить* (факт, отмеченный в [16, с. 64]; так, *не решил* предполагает 'решал'; *не уговорил* предполагает 'уговаривал'), объясняется наличием в семантике этих глаголов компонента 'Субъекту удалось', который, в свою очередь, имплицитно подразумевает компонент 'Субъект пытался', т.е. 'Субъект действовал'.

Базой для формулирования запроса пользователя может быть не только информация о классах слов, но и о том или ином параметре – толкования или актанта – как о совокупности признаков. Например, возможны запросы: полный список Т-категорий; допустимые значения параметра "каузация"; Т-категории актанта Агенс (или Па-

циенс); и под. Полученная таким образом информация позволит не только проверять и уточнять традиционные лингвистические представления, но и решать новые задачи. Одна из таких задач – исследование необязательных, "слабых" компонентов семантики лексемы. До сих пор подобные исследования касались отдельных слов или небольших групп слов (см. о глаголах со слабым компонентом 'способ движения' в [17, с. 31] или о неустойчивом оценочном компоненте в [18, с. 24]) и тормозились отсутствием достаточно представительного материала. Последовательное отражение неассертивных компонентов (фоновых, имплицативных и под.) в соответствующих полях толкования позволит пользователю БД получить информацию о сотнях лексем и десятках семантических классов.

Еще один режим использования БД – проверка всевозможных гипотез о связи семантических характеристик слова с его поверхностными свойствами. Так, база данных дает возможность выяснить, верно ли, что в семантику глагола СВ всегда входит компонент 'наступление нового состояния', как предполагала А. Вежицка [11]. (Ответ отрицательный, ср. глаголы *произойти, осветить, защитить*, которые включают идею прекращения старого состояния, но не начала нового; так *Плащ защитил меня от дождя* значит 'защитал, пока дождь не кончился').

ЭКСПЕРТНАЯ ЧАСТЬ СИСТЕМЫ "ЛЕКСИКОГРАФ"

Второй компонент системы, лексическая база знаний (ЛБЗ), – это своего рода грамматика лексикона. ЛБЗ включает семантические правила и общие закономерности, которые позволяют манипулировать семантической и грамматической информацией, помещенной в БД. Мы исходим (опираясь, прежде всего, на работу [19]) из идеи о том, что, в принципе, любая особенность поверхностного поведения слова может быть выведена из его семантики. ЛБЗ должна содержать всевозможные обобщения этого типа, а также формальные правила, позволяющие делать такого рода выводы. Примером формального правила является правило, обеспечивающее толкованием глагол противоположного вида, входящий с данным в стандартную видовую пару.

Так, в предельной паре, чтобы получить по толкованию глагола СВ, например, *наклеить*, толкование парного глагола НСВ (*наклеивать*), нужно изменить в толковании СВ положение наблюдателя (НСВ, в его основном значении, актуально-длительном, предполагает синхронного наблюдателя) и изменить коммуникативные пометы; а именно: в поле "экспозиция" заменить 'до t, предшествующего МН' на 'до МН'; в поле "деятельность" заменить метакомпонент на 'в t, предшествующем МН, Сб действует с Целью', и в поле "результат" заменить параметр "результат" на параметр "Цель" с метакомпонентом 'в t после МН будет иметь место состояние'. При этом компонент "деятельность" получает коммуникативный статус (ассерция), а компонент "Цель" – статус (импликация). Предполагается, что порождающий потенциал системы будет увеличиваться по мере роста ее ЛБЗ.

На данном этапе ЛБЗ включает следующие разделы.

1) **Т-категории глагола.** Идея о связи видо-временных характеристик глагола с его семантикой высказывалась многими авторами; самую широкую популярность приобрела классификация Вендлера (ср. также подчиненную аспектуальным задачам семантическую классификацию глаголов Ю.С. Маслова). Однако классификация Вендлера не охватывает всей глагольной лексики и нуждается в уточнениях. Необходимо выявить полный набор Т-категорий русских глаголов, а также охарактеризовать исчерпывающим образом особенности поведения лексемы, вытекающие из ее принадлежности к данной Т-категории, ср. [20].

К настоящему моменту установлено, что принадлежность глагола к той или иной Т-категории предопределяет: набор основных актантов; возможность образования видовой пары и ее семантический тип; ограничения на набор частных видовых значений; сочетаемость с разными видами обстоятельств (цели, времени и под.); способность мотивировать маркированные способы действия. Так, глаголы, относящиеся

к Т-категории Деятельность (например, *играть на скрипке*), способны мотивировать делимитатив (*поиграть*) и производный инхоатив (*заиграть*).

Особую проблему составляет разграничение базовых и производных Т-категорий. Так, Т-категория делимитативов (обсуждавшаяся в работе [21]) производная, и потому не случайно, что ей не нашлось места в классификации Вендлера, рассчитанной на базовые категории.

2) **Семантическая сочетаемость.** Ограничения сочетаемости предопределяются не только Т-категорией, но и другими, более частными компонентами толкования. Например, у абстрактных глаголов физического действия типа *расширить* (с компонентом 'способ действия не специфицирован') затруднена сочетаемость с инструментом, [22]; так, недопустимо **расширить яму лопатой*, при правильном *копать землю лопатой*, хотя действие – расширение – производится, скорее всего, с участием инструмента. Предполагается, что все такого рода связи между семантикой и синтаксисом могут быть исчислены.

3) **Частные видовые значения.** В основном, ограничения на набор частно-видовых значений глагола (неполнота функциональной парадигмы по Ю.С. Маслову) выводятся из его Т-категории, но могут играть роль и другие компоненты толкования; например, невозможность или затрудненность употребления глагола в актуально-длительном значении обуславливается компонентом 'процесс в Объекте: сверхкраткий', как у *ударять* или 'процесс в Объекте: несинхронный деятельности Субъекта', как у глаголов *стрелять, взрывать, отравлять, убивать*.

4) **Семантическая деривация.** Многим глаголам в нашей БД свойственна регулярная многозначность (см. у Ю.Д. Апресяна [15]). Ставится задача исчислить регулярные типы семантической деривации, а также выявить **предрасположение** тех или иных классов слов к тому или иному типу семантической деривации.

Основной источник регулярной многозначности – метонимический перенос, ср.: *Хозяйка покрывает диван пледом – Плед покрывает диван; В своей комедии он высмеивает интеллигенцию – Его комедия высмеивает интеллигенцию; Его лишают стипендии – Это лишает его возможности уехать; Он подтверждает свое согласие – Это подтверждает его виновность.* Как оказывается, переносы такого рода дают регулярное соотношение Т-категорий исходного и производных значений слова. Исходное и производные значения глагола образуют регулярные пары и даже тройки, которые в совокупности составляют **категориальную парадигму** глагола. Потенциальные парадигмы Т-категорий одного глагола, в принципе, могут быть исчислены, равно как и степень их реализованности у отдельных глаголов или у глаголов определенного семантического класса. Так, глаголы полного охвата имеют трехчленную категориальную парадигму, см. [14]: действие (*сторож наполняет бассейн водой*), процесс (*Вода постепенно наполняет бассейн*) и состояние (*Вода наполняет бассейн до краев*).

Метафорические переносы обычно считаются непредсказуемыми. На самом деле, однако, Т-категория потенциального метафорического деривата во многих случаях может быть предсказана. Например, у глагола физического действия следует ожидать в первую очередь образования значения, относящегося к Т-категории "происшествие" (*порезал хлеб → порезал палец*). Что действительно является случайным, так это Т-категории актантов семантического деривата (ср. *мальчик подпрыгнул* и *цены подпрыгнули*), а также закрепление окказионального метафорического употребления в языке.

В принципе, зная исходный набор семантических компонентов и их соотношение, можно исчислить все типы метафорических переносов, все теоретически возможные способы получения новых значений путем устранения тех или иных компонентов и изменения коммуникативного статуса оставшихся (такая задача ставится, в частности, в [23]). Однако наличие определенных механизмов образования метафорических зна-

чений само по себе, конечно, не гарантирует того, что все эти типы семантических дериватов (или хотя бы некоторые из них) обнаружатся у конкретной леммы.

Полезным при описании семантической деривации оказывается также понятие тематического класса; ср. класс регулярных семантических дериватов, относящихся к тематическому классу 'создание Объекта': *выкопать картошку – выкопать яму; сварить яйца – сварить суп*.

Интересная лингвистическая проблема, к решению которой мы надеемся подойти в дальнейшем, – это проблема зависимости свойств леммы от положения того или иного компонента в ее семантической структуре, а не просто от наличия этого компонента. Само по себе наличие одинакового компонента у двух (или более) лемм еще отнюдь не гарантирует их принадлежности к одному и тому же релевантному семантическому классу. Последнее слово остается за позицией данного компонента в семантической структуре леммы и за его коммуникативным статусом. Так, наличие в семантике глагола компонента 'Объект перемещается в Место' еще не означает, что это глагол движения. Такой компонент встречается и у глаголов полного охвата (*наполнить бочку водой*), и у глаголов сверхкраткого действия (*бросить камень в пруд*), и у глаголов происшествия (*уронить коробку на пол*), и у многих других (*послать письмо в Казань*).

Отсюда возникает, с одной стороны, чисто техническая сложность – как корректно сформулировать запрос, чтобы получить нужный класс глаголов; а с другой стороны – теоретическая проблема: как идентифицируются носителями (и исследователями!) языка семантические классы глаголов; от чего зависит восприятие данного глагола как члена того или иного класса.

Таковы задачи, которые ставит перед собой система "Лексикограф". Решение даже части этих задач позволит выйти на новый уровень системности в описании лексики (см. [24]), а лексическая база данных должна стать, по замыслу ее разработчиков, не только способом хранения информации, но и полезным инструментом лингвистического исследования и получения нового знания.

СПИСОК ЛИТЕРАТУРЫ

1. Paducheva E.V., Rakhilina E.V. Predicting co-occurrence restrictions by using semantic classifications in the lexicon // COLING-90. Papers presented to the 13-th International conference on computational linguistics. V. 3. Helsinki, 1990.
2. Красильщик И.С., Рахилина Е.В. Предметные имена в системе "Лексикограф" // НТИ. Сер. 2. 1992. № 9.
3. Грамматический словарь русского языка. М., 1977.
4. Vendler Z. Linguistics in philosophy. Ithaca, N.Y., 1967.
5. Wierzbicka A. Lingua mentalis. Sydney, 1980.
6. Wierzbicka A. English speech act words. Sydney, 1987.
7. Иорданская Л.Н. Попытка лексикографического толкования группы русских слов со значением чувства // Машинный перевод и прикладная лингвистика. 1970. Вып. 13.
8. Зализняк Анна А. Семантика глагола *болиться* в русском языке // ИАН СЛЯ. 1983. № 1.
9. Карпунен Л. Логика английских конструкций с синтаксическим дополнением // Новое в зарубежной лингвистике. Вып. XVI. М., 1985.
10. Маслов Ю.С. Вид и лексическое значение глагола в современном русском литературном языке // ИАН СЛЯ. 1948. Вып. 6.
11. Wierzbicka A. On the semantics of verbal aspect in Polish // To Honour Roman Jakobson. The Hague; Paris, 1967.
12. Гловинская М.Я. Семантические типы видовых противопоставлений русского глагола. М., 1982.
13. Филлмор Ч. Дело о падеже открывается вновь // Новое в зарубежной лингвистике. Вып. X. М., 1981.
14. Падучева Е.В., Розина Р.И. Семантический класс глаголов полного охвата: толкование и лексико-синтаксические свойства // ВЯ. 1993. № 6.
15. Апресян Ю.Д. Лексическая семантика. М., 1974.
16. Апресян Ю.Д. Типы информации для поверхностно-семантического компонента модели "Смысл-Текст" // Wiener Slavistischer Almanach. Bd 1. Wien, 1980. S. 64.
17. Богуславский И.М. Исследования по синтаксической семантике. М., 1985. С. 31.

18. *Зализняк Анна А.* Исследования по семантике предикатов внутреннего состояния. München, 1992. С. 24.
19. *Wierzbicka A.* The semantics of grammar. Amsterdam, 1988.
20. *Падучева Е.В.* Таксономические категории глаголов imperfectiva tantum в русском языке // НТИ. Сер. 2. 1993. № 2.
21. *Flier M.S.* The scope of prefixal delimitation in Russian // The scope of Slavic aspect. Columbus, Ohio, 1985.
22. *Плунгян В.А., Рахилина Е.В.* Сирконстанты в толковании? // Metody formalne w opisie językow słowiańskich. Białystok, 1990.
23. *Туровский В.В.* О соотношения значений многозначного слова // Семиотика и информатика. Вып. 26. М., 1985.
24. *Apresjan Y.D.* Systemic lexicography // EURALEX-92. Papers submitted to the 5th EURALEX International congress of Lexicography, Helsinki, 1992.