

## СЕМАНТИКА ПРЕДЛОЖНО-ПАДЕЖНЫХ ФОРМ РУССКОГО ЯЗЫКА

## SEMANTICS OF PREPOSITIONAL-CASE FORMS OF THE RUSSIAN LANGUAGE

В.А. Тузов (*tuzovvitalij@mail.ru*)

Санкт-Петербургский государственный университет

Обсуждается проблема вычисления смысла предложно-падежных форм в рамках формального толкования слов русского языка, заданного семантическим словарем. Дополнение словаря базой данных, содержащей сведения о предметной области, позволяет автоматически вычислять значения всех предложно-падежных форм русского языка. Как правило, проблема сводится к выбору атрибута для объекта, присоединяющего предложно-падежную форму. Рассматривается возможная структура подобной базы данных.

## 1. О базе знаний

Функциональная модель языка [1], в свое время успешно использованная как средство формализации и описания искусственных языков, оказалась моделью, адекватной и естественным языкам. Доказательством этого факта является работающий семантический анализатор, правильно транслирующий практически любое предложение русского языка в формальный семантический язык.

Семантический анализатор является первым звеном двухчленной цепи: Русский Язык => Семантический Язык => База Знаний. Информация, поставляемая анализатором, имеет вид множества формул, очень похожих на обычные логические предикаты. Например, предложение русского языка *Иван белит стены* на семантическом языке имеет вид формулы:  $Caus(Иван, IncерCopul(стены, БЕЛЫЙ)) =$  Иван делает так, что стены начинают быть белыми. Однако эти формулы не являются предикатами.

Во-первых, каждая формула не имеет переменных: в нее входят только константы. Поэтому, мы имеем дело не с логикой предикатов, а с алгеброй высказываний. Во-вторых, нет необходимости в каких-либо правилах вывода. Сравнительно небольшой набор эквивалентных преобразований аналогичен преобразованиям в обычном алгебраическом языке [2]. Таким образом, мы имеем дело даже не с алгеброй высказываний, а с алгебраическим языком. Эквивалентные преобразования в этом языке могли бы стать объектом математического исследования, но даже и в этом нет никакой необходимости. Таким образом, структура данных семантического языка в принципе аналогична структуре данных обычного языка программирования.

Информация, поставляемая семантическим анализатором, может использоваться разными способами. Но, если базисные функции –  $Caus(x,y)$ ,  $Copul(x,y)$  и т. п. – являются функциями, то запись  $Caus(Иван, IncерCopul(стены, БЕЛЫЙ))$  является записью алгоритма. Поэтому выполнение этой записи является наиболее естественным способом построения отображения с семантического языка в базу знаний. Чтобы выполнить эту запись, необходимо построить алгоритм вычисления каждой базисной функции.

Областью определения базисных функций является множество объектов базы знаний. Общая структура базы знаний (БЗ) описывается классификатором семантического языка [2]. Основу БЗ составляют объекты, которые находятся в тех или иных отношениях друг с другом.

Рассмотрим процесс отображения в БЗ предложения *На окне стоит красный цветок*. В результате семантического анализа это предложение примет вид:

*стоит (@Где На (@Пред окне), @Им цветок(@Им красный))*

стоит X003.006: Oper01(Z1[X005.001], СТОЯНИЕ\$12013(Z2  
[Где:Х001.079]))  
На X001.079: НА\$123372~@Где Y1>  
Loc(Y1:,Z1[Пред:Х002.001])  
Окне X002.001: ОКНО\$123372 S1>Наб  
(ЗДАНИЕ\$1233(Z1,Z2,Z3), S1:ОТВЕРСТИЕ\$12/07)  
Цветок X005.001: ЦВЕТOK\$12232  
({Z0[КАКОЙ:Х004.001]}, ПРИСУЩИЙ:Z1)  
красный X004.001: Наб( [X005.001],  
ЦВЕТ\$12/012(КРАСНЫЙ\$12/1121/01))

По двум объектам – ОКНО\$123372 и ЦВЕТOK\$12232 – строится их образ в БЗ. Номер класса однозначно предписывает структуру объектов этого класса. Далее вычисляется функция Наб:

Наб(ЦВЕТОК\$12232, ЦВЕТ\$12/012(КРАСНЫЙ\$12/1121/01), которая по полю **цвет** объекта **цветок** присвоит значение КРАСНЫЙ\$12/1121/01. Функция Oper01 присвоит полю **местонахождение** этого объекта два значения: *способ нахождения* СТОЯНИЕ\$12013 и *место* – ОКНО\$123372.

Из этого примера видны основные моменты отображения семантической информации в БЗ. Во-первых, построение объектов. На этом этапе решается проблема их идентификации: быть может, этот объект уже есть в базе (при этом объект может иметь разные наименования, которые выражаются словом или словосочетанием, на него может указывать местоимение и т. п.). Во-вторых, вычисление значений связей, которые выявлены в процессе анализа. Связи могут быть двух типов. Первый тип связи указывает на отношения между объектами (например, отношение местонахождения объектов ЦВЕТОК\$12232, и ОКНО\$123372). Второй тип связан с конкретизацией объекта (например, присвоение полю ЦВЕТ\$12/012 значения *красный*). В обоих случаях происходит присваивание значения атрибуту объекта и главная проблема здесь – найти имя этого атрибута.

Первый тип связи предопределен семантико-грамматическим типом (СГТ) связи (например, @Где) (см. [3]), однозначно указывающим на атрибут, которому следует присвоить значение. Второй тип связи определяется базисными функциями, поэтому его реализация сводится к построению алгоритмов, вычисляющих значения базисных функций. Некоторые функции можно представить в виде таблицы, но для реализации таких функций как Rel табличного представления недостаточно.

Семантические функции, входящие в текст на семантическом языке, либо берутся непосредственно из семантического словаря, либо вычисляются в процессе связывания слов. Сравнительно небольшой набор тождеств (см. [2]) позволяет сократить число вхождений семантических функций. Например, тождество  $\text{Наб}(x, S1) \rightarrow f(S1: y, x1, \dots, xN) = f(x, y, x1, \dots, xN)$  позволяет убрать одно вхождение функции Наб. При помощи этого тождества предложения типа – *Сержант Иванов имеет 5 задержанных преступников* – преобразуются в предложения вида: *Сержант Иванов задержал 5 преступников*.

Рассмотрим ряд небольших фрагментов описаний функций.

#### Функция *Наб*.

##### Классы аргументов

Поле объекта и его значения

НЕЧТО\$1, ЦВЕТ\$12/012

ЦВЕТ\$12/012

\ Красный цветок

ЧЕЛОВЕК\$1241, ВЕЩЬ\$1213

СОБСТВЕННОСТЬ\$12411/0 \ Дом Ивана

ДОКУМЕНТ\$14002, ИНФОРМАЦИЯ\$1440

СОДЕРЖАНИЕ\$1440200 \ Книга анекдотов

ПОСУДА\$12131111, ЖИДКОСТЬ\$121112

СОДЕРЖИМОЕ\$12/013 \ Бутылка молока

#### Функция *Соруl*.

В большинстве случаев информация для вычисления функции Соруl содержится непосредственно в исходном семантическом словаре. Например:

АББАТИСА\$12413211

(S1>Соруl(S1: АББАТ\$12413211(!Род), ЖЕНЩИНА\$124131))

АБДИКАЦИЯ\$131334

(S0>Соруl(!Род, FinСоруl(!Род, МОНАРХ\$124134021))

АБОРИГЕН\$124113

(S1>Соруl(S1: ЖИТЕЛЬ\$1241 (СТРАНА: !Род), МЕСТНЫЙ\$1241))

АБРИКОСНЫЙ

(A1>Соруl(A1: ЦВЕТ\$12/012, ПРИСУЩИЙ: АБРИКОС\$122332))

АБРИКОСНЫЙ

(A1>Соруl(A1: ВЕТВЬ\$1223/2, ПРИСУЩИЙ: АБРИКОС\$12231))

и потому вычислить значение этой функции несложно. Немного сложнее вычисление функции Соруl в записях, которые порождаются выражениями вида *Директор фабрики Иванов, Иванов является директором фабрики*.

#### Функция *Rel*.

Эта функция в основном используется в описаниях прилагательных, образованных от существительных. Подавляющему числу таких прилагательных приписывается достаточно абстрактный смысл: *такой, который имеет какое-то отношение к мотивирующему его существительному*. Заводская труба – труба, которая имеет какое-то отношение к заводу. Таких прилагательных в словаре более 11 тысяч. Некоторые из них могут быть описаны с помощью функции Соруl. Например,

АБРИКОСНЫЙ (A1>Соруl(A1: ЦВЕТ\$12/012, ПРИСУЩИЙ:

АБРИКОСНЫЙ (A1>Copul(A1:ВЕТВЬ\$1223/2, ПРИСУЩИЙ:  
АБРИКОС\$122332))  
АБРИКОС\$12231))

Но разброс значений многих прилагательных достаточно велик, и потому размещение этих значений в семантическом словаре нецелесообразно, а иногда и невозможно. Например, прилагательное *заводской* в выражениях *заводской рабочий*, *заводская труба*, *заводской гудок*, *заводские ясли*, *заводская территория*, *заводской митинг* присваивает значение *завод* разным полям объектов-существительных. А словосочетание *карандашная зарисовка* показывает, что статическое описание взаимодействия прилагательных этого класса с существительным практически невозможно, так как прилагательное взаимодействует не со словом зарисовка, а с его аргументом. Система тождеств, позволяющая заменять прилагательные существительными (*карандашная зарисовка* = *зарисовка карандашом*, *восстановительные работы* = *работы по восстановлению*, *южный ветер* = *ветер с юга* и т. п. ), дает возможность свести описание функции Rel к вычислению смысла предложно-падежных форм.

Работа по реализации семантических функций еще не завершена, поэтому говорить о результатах целостного отображения семантической информации в базу знаний еще рано. Но независимо от того, какой окончательный вид примет это отображение, уже сейчас можно отчетливо представить основные контуры базы знаний, адекватно взаимодействующей с естественным языком [4]. Компьютерная база знаний – это база активных информационных процессов, получаемых из текстов на естественном языке. Взаимодействие двух информационных процессов порождает набор новых знаний. Эти взаимодействия – основа логического (хотя правильнее было бы сказать – семантического) компьютерного “мышления”. Исследование этих взаимодействий – одна из основных теоретических проблем информатики ближайшего будущего.

## 2. Смысл предложно-падежных форм

Чтобы уяснить постановку задачи рассмотрим ряд примеров: *книга Петра*, *книга анекдотов*, *книга издательства*, *книга времен Ивана Грозного*. Каждое из этих выражений имеет собственную интерпретацию: *книга Петра* = книга, принадлежащая Петру, *книга анекдотов* = книга, содержащая анекдоты, *книга издательства* = книга, изданная издательством, *книга времен Ивана Грозного* = книга, относящаяся ко временам Ивана Грозного. Или в более формальной интерпретации:

*книга Петра* = КНИГА(ВЛАДЕЛЕЦ: *Петр*)  
*книга анекдотов* = КНИГА(СОДЕРЖАНИЕ: *анекдоты*)  
*книга издательства* = КНИГА(ИЗДАНИЕ: *издательство*)  
*книга времен Ивана Грозного*  
= КНИГА(ВРЕМЕНА: *Иван Грозный*)

Родительный падеж, связывая два слова, приписывает значения разным атрибутам первого слова. Аналогичная ситуация возникает с любой предложно-падежной формой. Таким образом, проблема вычисления смысла предложно-падежной формы сводится к поиску конкретного атрибута присоединяющего слова.

Если посмотреть на связи, которые строит семантический анализатор:

КНИГА\$1213	РОД	ПЕТР\$12413/11000
КНИГА\$1213	РОД	АНЕКДОТ\$14322
КНИГА\$1213	РОД	ИЗДАТЕЛЬСТВО\$12341
КНИГА\$1213	РОД	ВРЕМЯ\$1613

то легко увидеть, что родительный падеж следует рассматривать как бинарную функцию, определенную на множестве номеров семантических классов, значением которой является имя атрибута первого аргумента:

РОД(\$1213,\$1241)	=	ВЛАДЕЛЕЦ
РОД(\$1213,\$1432)	=	СОДЕРЖАНИЕ
РОД(\$1213,\$12341)=	ИЗДАНИЕ	
РОД(\$1213,\$16)	=	ВРЕМЕНА

Однако, есть альтернативное решение. В семантическом словаре слово *книга* описывается одной альтернативой:

КНИГА \$1430(S1>Usor(S1:ВЕЩЬ\$1213(РОД:!Род),ЧТЕНИЕ\$144  
(ТЕМА:!Про!\обПред)))

Если добавить еще четыре альтернативы:

КНИГА \$1430(S1>Наб(ЧЕЛОВЕК\$1241~!Род,S1: КНИГА\$1430  
(ТЕМА:!Про!\обПред)))  
КНИГА\$1430(S1>Content(S1:КНИГА\$1430(ТЕМА:

!Про\!обПред), ИДЕЯНИЕС14~!Род))  
 КНИГА \$1430(S1>PerfCaus(ИЗДАТЕЛЬСТВО\$123~!Род,  
 IncepFunc(S1:КНИГА\$1430(ТЕМА:!Про\!обПред)))  
 КНИГА \$1430(S1>Temp(S1:КНИГА\$1430(ТЕМА:  
 !Про\!обПред), ВРЕМЯ\$16~!Род))

то семантический анализатор, обрабатывая выражение *книга Петра* выберет первую альтернативу, выражение *книга анекдотов* – вторую альтернативу и т.д. Тем самым проблема вычисления смысла родительного падежа может быть сведена к вычислению базисных функций. На первый взгляд такое решение выглядит достаточно простым, но необходимость многократного увеличения словаря делает его по крайней мере непрактичным.

В большинстве случаев смысл предложной формы вычисляется семантическим анализатором. Например, предложение – *компот из Франции хорош* – после анализа примет вид: @Крат хорош(@Им компот (@Откуда из(@Род Франции))), и нет необходимости в дальнейшем анализе предложной формы *из Франции*. Разбор предложения – *Они делают компот из Франции* – оставит предлог *из* не вычисленным: делают(@Им Они,@Вин компот,@Изо из(@Род Франции)), так как его смысл определяется глаголом *делают*. Близкое по внешней структуре предложение – *Они делают компот из персиков* – анализатор разберет следующим образом: делают(@Им Они,@Вин компот(@ОКакой из(@Род персиков))), и семантика предлога *из* определяется не словом *делают*, а словом *компот*, что несколько меняет процесс дальнейшего анализа, в результате которого в первом случае глагол *делают* присвоит атрибуту МАТЕРИАЛ объекта *компот* значение *Франция*, а во втором случае атрибуту СОСТАВ будет присвоено значение *персики*.

Общее представление о множестве атрибутов можно получить из следующих трех списков.

Общие свойства:

Азы Аналогия База Базис Возможность Выигрыш Высота Градус Доза Доля Закон Звено Значение Идеал Индекс Инфраструктура Итог Категория Качество Квалификация Квота Класс Композиция Компонент Кондиция Конституция Конструкция Конфигурация Корпус Критерий Кусок Логос Ляпус Макроструктура Манера Мера Модель Недооценка Необходимость Норма Нужды Образец Ополоски Орудие Основа Остов Отрывок Оценка Ошибка Пласт Подсистема Показатель Положение Полоса Полнос Порция Порядок Правило Превосходство Предпосылка Пример Принцип Прок Прототип Профит Распорядок Режим Результат Ритм Роль Самоцель Свойство Сегмент Сеть Система Следствие Слой Соединение Сор Способ Средство Стандарт Статут Степень Стереотип Стиль Стоимость Строение Строй Структура Субстрат Трафарет Узел Упор Управа Уровень Условие Устав Устой Формация Форма Формула Фрагмент Фракция Фундамент Функция Характеристика Характер Хартия Хлам Цена Часть Черта Член Шаблон Штамп Штука Экземпляр Экономия Элемент Эталон

Свойства физических объектов:

тип звук свет запах температура величина цвет состав Габарит высота Глубина Длина Толщина Ширина Вес Масса Объем Адрес Название Облик Поза Имидж Положение Крен Осадка Равновесие Уклон Изоляция Количество Число Часть Процент Множество Бок Сторона Верх Низ Зад Перед Периферия Середина Центр Поверхность Глянец Пятно Тавро Щель Углубление Галс

Свойства человека:

Поощрение-Наказание Успех-Неуспех Слава-Беславие Репутация Цивилизация Зажиточность Темперамент Лихой-Растерянный Трезвый-Пьяный Родной-Посторонний Женский-Мужской Бодрый-Усталый Активный-Пассивный Дородный-Плюгавый Местный-Чуждедальный Талант Право Практика Чувство Депрессия-Вдохновение Довольство-Недовольство Торжество Церемониал Покой-Тревога Радость-Печаль Веселье-Плач Страх-Бесстрашие Возмущение Обида Горе-Счастье Тягота-Легкость Благословение-Проклятие Восхищение-Отвращение Благодарность-Неблагодарность Любовь-Ненависть Отзывчивость-Бессердечность Уважение-Презрение Доброта-Жадность Дружба-Вражда Доверие-Недоверчивость Верность-Измена Чистоплюйство-Неразборчивость Лень Забытье Сон-Бдение Поза Невзгоды Опасность-Безопасность Мораль Искренность-Хитрость Скромность- Чванство Благородство Самообладание Упорство Добросовестность Непринужденность Покорность-Покорение Зависимость-Независимость Согласие-Разногласие Компромисс Настрой Легкомыслие-Серьезность Грех Обольщение Вина Обвинение-Оправдание Покаяние-Прощение Воля Намерение Желание Надежда Мечта Интерес-Скука Глупость-Здравомыслие Сомнение-Уверенность Преувеличение-Преумножение Разум Память Возможности Сословие Многолюдые Населенность Племенной Авангард-Арьергард Мобилизация Гражданка Коллективизм-Индивидуализм Отметка Ребенок-Старик Деторождение Детскость Любитель-Ненавистник Верующий Брюнет-Блондин Знарок-Неук Друг-Враг Озорник-Фифа Ловелас-Анахорет Лиходей-Джентльмен Умница-Балбес Работяга-Лентяй Красавец-Урод Молодец-Раззява Великан-Карлик

Ниже приведен небольшой фрагмент таблицы, позволяющий по типу связи и номеру класса присоединяющего слова определить имя атрибута, которому нужно присвоить значение – присоединяемое слово. Например, после вычисления выражения *цена пять рублей* атрибуту ВЕЛИЧИНА объекта *цена* будет присвоено значение – *пять рублей*. Число в последней колонке таблицы показывает число базовых слов, обладающих данным типом связи. Вычисление некоторых атрибутов сводится к вычислению базисных функций. Например, связь !Дат сводится к вычислению функции  $\text{Nab}(\text{ДАТ:}x,y)$ : благодарность Петру =  $\text{Nab}(\text{Петр}, \text{благодарность})$ .

Тип связи	Атрибут	Количество слов
Связь !Вин стоимость цена рост габарит глубина длина калибр объем площадь размер толщина формат ширина	ВЕЛИЧИНА	15
Связь !Вокруг ситуация атмосфера обстановка условия среда барьер забор изгородь ограда тын виток извилина	ОКРУЖЕНИЕ	20
Связь !Вопр ошибка судьба чувство счастье фантазия иллюзия заблуждение сомнение убеждение брехня клевета	СОДЕРЖАНИЕ	340
Связь !Где барахтанье бродяжничество бунт бытие взрыв висение восстание встреча вход дежурство строительство	МЕСТОНАХОЖДЕНИЕ 229	
Связь !Дат батя благодарность благословение близнец брехня буллит веление вендетта вера команда комментарий	Наб(ДАТ:х,у)	458
Связь !Для абажур абонемент бутылка дробь дровни еда емкость ерш жбан приспособление присыпка прицеп	Usor(х, ДЛЯ:у)	3635
Связь !Изо \$1/424 Части отрывок фрагмент \$12/133 Количество \$101/ Пища абрикосовка абсент ай анисовка аперитив арака азу антрекот \$12117 Парфюмерия амбра кольдкрем косметика лосьон \$121210 Повязки бинт заплата лонгет повязка \$121211 Ткани бинт заплата лонгет повязка рюш материя отрез ткань	Content(х,ИЗО:у)	3578
Связь !Из-под бадьа бак баллон банка блюдце бочка бурдюк футляр хребтук цибики цистерна чан	Perf Content(х,ИЗ-ПОД:у)	182
Связь !Инфин боязнь веление возможность гарантия глупость настроение необходимость обет мужество мысль надежда	СОДЕРЖАНИЕ	138
Связь !КакДолго ожидание перерыв тайм-аут	ПЕРИОД	3
Связь !Когда действие наблюдение ожидание положение работа саммит ситуация	ВРЕМЯ	13
Связь !Куда \$12/31 Направление галс галфвинд траверз фордевинд \$12321 Дороги авиалиния авиарейс траверз проспект бетонка колея коммуникация \$15121 Стрельба пальба стрельба \$154 Перемещение бегство бег вход въезд поездка полет прыжок побег виток волочение качение лазанье носка падение \$15501 Поворот вращение оборот поворот \$15513 Колебание	Direkt(х,КУДА:у) НАПРАВЛЕНИЕ НАПРАВЛЕНИЕ НАПРАВЛЕНИЕ НАПРАВЛЕНИЕ НАПРАВЛЕНИЕ НАПРАВЛЕНИЕ	410

	виляние качание колебание нутация шатание	
\$155222	Тяга	НАПРАВЛЕНИЕ
	притяжение тяготение	
\$16153	Опоздание	НАПРАВЛЕНИЕ
	опоздание опозданье спешка	
\$1214	Деньги	ПРЕДНАЗНАЧЕНИЕ
	деньга деньги десятирублевка дидрахма динарий ассигнация бабки баксы	
\$1241/4141	Стремление	ЖЕЛАНИЕ
	намерение стремление тяга	
\$1440	Информация	АДРЕСАТ
	декларация депеша заявка заявление каблограмма обращение сообщение телефонограмма	
\$151413	Критика	АДРЕСАТ
	антикритика критика	
\$1449	Документы	ПРЕДНАЗНАЧЕНИЕ
	абонемент билет бронь виза контрамарка купон курсовка талон	
\$1511/0	Нападение	ЖЕРТВА
	агрессия наступление	
\$152515	Погрузка	ЕМКОСТЬ
	навал нагрузка погрузка посадка	
\$15303	Вовлечение	СФЕРА
	вовлечение завлечение заманивание переманивание совлечение	
\$153021	Путешествие	МЕСТОПРЕБЫВАНИЕ
	круиз паломничество поход путешествие рейд турне турпоездка турпоход экспедиция	
\$1531061	Выборы	ДОЛЖНОСТЬ
	баллотирование выборы	

### **Список литературы**

1. Тузов В.А. Математическая модель языка. – Л.: Изд-во Ленингр. Ун-та, 1984. – 176 с.
2. Тузов В.А. Компьютерная семантика русского языка. – СПб.: Изд-во СПбГУ, 2004. – 400 с.
3. Тузов В.А. Компьютерная грамматика русского языка // Вестник С.-Петербур. ун-та. Сер. 10. СПб.: Изд-во С.-Петербур. ун-та, 2004. Вып. 1-2. С. 94–100.
4. Тузов В.А. Языки представления знаний. Л.: Изд-во Ленингр. Ун-та, 1990. 126 с.