

Контекстная грамматика предложных конструкций русского языка

В.П. Захаров¹, В.Д. Михайлова²

¹ Санкт-Петербургский государственный университет, Институт лингвистических исследований РАН

² Санкт-Петербургский государственный университет

v.zakharov@spbu.ru, BukvaCe@ya.ru

Аннотация

В работе представлена методика построения контекстной грамматики предложных конструкций русского языка (онтологии предлогов) с опорой на лексико-семантические и морфологические контекстные маркеры и один из вариантов её реализации — онтология первообразных предлогов русского языка, построенная на материалах «Синтаксического словаря» Г. Золотовой.

Данная грамматика представляет собой описание лексико-семантических характеристик знаменательных слов, в соединении с которыми реализуются значения предлогов.

Практическая и научная значимость данной работы связана с возможностью использования полученных данных для решения практических лингвистических задач в системах АОТ и для дальнейших исследований в области семантики предлогов.

Ключевые слова: предлоги, русский язык, семантика, синтаксические отношения, контекст, лексико-семантические маркеры.

1. Введение

Проблема автоматического извлечения смысла из текста включает в себя самые разные лингвистические подзадачи. К ним относятся задачи снятия лексической неоднозначности, снятия грамматической неоднозначности, семантической интерпретации синтаксических связей и др. При этом разработчики обычно концентрируются на знаменательных словах и зачастую обходят стороной предлоги и нередко не выделяют семантический анализ как отдельный уровень.

В то же время семантический анализ текста, будучи выделенным в отдельную процедуру или нет, является неотъемлемой частью задачи

понимания текста. Суть семантического анализа понимается разными исследователями по-разному. Многие ученые сходятся во мнении, что в сферу семантического анализа входит как минимум:

- построение семантической интерпретации слов и конструкций;
- установление «содержательных» семантических отношений между элементами текста, которые уже принципиально не ограничены размером одного слова (могут быть больше или меньше одного слова).

Разные авторы в своих работах используют различные уникальные определения семантического анализа текста, причем эти определения часто зависят от задачи. Так, по мнению авторов работы [1], результатом компьютерного семантического анализа связного текста должен быть формальный или адаптированный текст естественного языка (ЕЯ), который выражает его смысловое содержание. В работе [2] А.В. Сокирко под семантическим анализом текста на ЕЯ подразумевается поиск семантических отношений, где под семантическим отношением понимается некая универсальная связь, усматриваемая носителем языка в тексте. Эта связь бинарна и направлена, т.е. она идет от одного семантического узла к другому узлу. Автор предлагает использовать следующий формат записи семантического отношения: $R(A,B)$, где R — название семантического отношения, A — зависимый член отношения, B — управляющий член отношения. С одной стороны, такой подход сильно похож на синтаксис, с другой, понятно, что синтаксические отношения всегда нагружены семантикой.

В настоящее время широкое распространение получили так называемые грамматики конструкций, описывающие как лексическую сочетаемость, так и сочетаемость на уровне классов [3] [4] [5]. С точки зрения структурной организации, конструкция — это комбинация целевого слова и слотов, заполняемых регулярными контекстными маркерами. Контекстные маркеры могут быть разного уровня абстракции: лексемы, словоформы, части речи, граммы, синтаксические позиции, лексико-семантические классы. Контекстные маркеры разного типа в корпусах идентифицированы соответствующими тегами, доступными в многоуровневой разметке (в частности, наиболее в этом смысле богата разметка контекстов Национального корпуса русского языка). Имеется ряд корпусно-ориентированных проектов, в которых особое внимание уделяется формализации лексико-синтаксических связей единиц текста. Однако проектов, посвященных собственно предложным конструкциям, практически нет.

В традиционной лингвистике о предлогах и, в частности, о семантике предлогов написано немало. Но описание семантики предлогов в стиле обычных толковых словарей ориентировано на человека. Словари лишь некоторым образом систематизируют использование предлогов в речи. Для автоматического анализа такие описания бесполезны. В то же время без построения их формальной модели невозможно реализовать эффективные технологии автоматического анализа текста.

В свете традиционной лингвистики сочетаемость предлогов, как правило, описывается с точки зрения грамматики, однако семантике предлогов до настоящего момента должного внимания не уделялось. Особенно это

справедливо для русского языка, где корпусных исследований, посвященных предлогам, крайне мало.

Трудности описания предлогов также не в последнюю очередь связаны с вызывающими споры вопросами о наличии у предлогов лексического значения как такового, о границах между значениями предлогов-полисемантов, о грамматико-лексическом дуализме семантики предлогов.

В данной статье предпринимается попытка, взяв за основу авторитетное описание семантики предлогов, построить на его основе концептуально простую модель *минимальных лексико-семантических контекстов*, в которых реализуются предложные значения, предназначенную для целей автоматической обработки текста. Совокупность таких контекстов мы называем грамматикой предложных конструкций, или онтологией предлогов.

В современной лингвистике опора на контекст становится одним из ведущих принципов описания семантики. Здесь можно сослаться на описание семантики английских глаголов П. Хэнксом [6]. Можно также вспомнить высказывание Дж.Р. Фёрса «Вы поймете слово по его окружению» (“You shall know a word by the company it keeps”) [7, p. 179] и мн. др. Авторитетным описанием семантики предлогов, опирающимся на контекст, является «Синтаксический словарь русского языка» Г. Золотовой [8], на базе которого мы и строим свою грамматику.

Термин *минимальный контекст* в данной работе понимается как пара «хозяин» — «слуга», то есть пара знаменательных слов, вступающих в непосредственные синтаксические отношения с предлогом. О наличии «хозяина» и «слуги» мы говорим в случаях, когда предложно-падежная конструкция занимает в предложении присловную или квазиприсловную позицию (является частью сложного предиката). Целью данной статьи является описание фрагмента онтологии первообразных предлогов русского языка и проверка ее пригодности для автоматической обработки текста посредством семантического интерпретатора, построенного на базе данной онтологии.

Гипотезу данного исследования можно сформулировать следующим образом: если предложно-падежная конструкция занимает в предложении присловную или квазиприсловную позицию, значение предлога можно с высокой долей вероятности предсказать, опираясь на лексико-семантические характеристики минимального контекста.

2. Предлог как единица разных языковых уровней

«Предлог — это служебная часть речи, оформляющая подчинение одного знаменательного слова другому в словосочетании или в предложении и тем самым выражающая отношение друг к другу тех предметов и действий, состояний, признаков, которые этими словами называются» [9, с. 706]. В сфере АОТ предлогам посвящено значительно меньшее число научных работ, чем другим частям речи, они регулярно игнорируются в системах информационного поиска, автоматического реферирования и других программах автоматической обработки текста. Такое непочетное место предлоги занимают во многих флективных языках, где трудно отделить значение предлога от значения

падежа, однако, как будет показано ниже, предлоги зачастую имеют первостепенное значение для понимания высказывания.

Наличие лексического значения у служебных слов вообще и у предлогов в частности зачастую ставится под сомнение [10]. А. Пешковский, А. Потеня, А. Шахматов и другие выделяют у предлогов только грамматическое значение. В. Виноградов и его последователи утверждают, что у предлогов есть и лексическое значение. Другие лингвисты (например, А. Смирницкий) занимают пограничную позицию, утверждая, что специфика предлогов с точки зрения лексического значения заключается в том, что, у разных предлогов оно проявляется в разной степени. Иногда предлоги выступают десемантизированными «морфологическими показателями синтаксических связей» (термин В. Виноградова [11]) — «препозиционными флексиями» имен существительных или «послелогоми» глаголов (термины В. Богородицкого). В других случаях предлоги бывают семантически наполненными и могут иметь первостепенное значение для понимания высказывания. В таких предложениях, как *Часы висят над картиной* и *Часы висят под картиной* лексическое значение предлога становится очевидной.

Наиболее нейтральной позиции относительно наличия значения у предлога придерживается «Академическая грамматика»: «... в любом случае предлог имеет лексическое значение, различна лишь степень его абстрактности <...> семантически „пустых“ предлогов не существует» [12]. Именно этой позиции мы и будем придерживаться в дальнейшем, и даже в случае крайне грамматикализации будем выделять формальное — объектное — значение.

Предлог в сочетании с падежной формой образует предложно-падежную синтаксему — минимальную и нечленимую единицу синтаксиса. Для русского языка предлог, способный управлять разными падежами, нужно трактовать в лексиконе онтологии как грамматический омоним. Место предложно-падежной синтаксемы в синтаксической структуре предложения наиболее полно описано в «Синтаксическом словаре» Г. Золотовой.

Семантическая интерпретация предлогов должна рассматриваться как часть более общей задачи — задачи семантической интерпретации синтаксических связей с учетом семантических характеристик самого предлога. Хотим обратить внимание, что интерпретируется не связь между знаменательным словом и предлогом, а связь двух знаменательных слов, точнее, двух значений, уточняемая и конкретизируемая соединяющим их предлогом.

3. Предлог и его место в компьютерной онтологической системе

Необходимость внедрения семантического описания предлогов в вычислительные ресурсы, обусловленная задачами автоматической обработки текста, ставит перед лингвистами сложную задачу: преобразование существующих семантических описаний (со всеми их недостатками и противоречиями) в некоторую логическую структуру — формат, пригодный для компьютерной обработки.

При создании ресурса, представляющего программ обработки текста необходимые сведения о семантике предлогов, возникает проблема, в рамках

какого формализма (желательно универсального) давать семантическое описание предлогов. Однако ни одна универсальная онтология (типа СУС или SUMO) ничего для описания предлогов не предлагает.

Поэтому мы ставим задачу достаточно автономно — описать совокупность лексико-семантических контекстов, в которых реализуются предложные значения — онтологию предлогов. Такую онтологию можно описать на базе одной из достаточно подробно и широко проиллюстрированных классификаций предложных значений. Как мы уже говорили, мы опираемся на материал «Синтаксического словаря» Г. Золотовой. Мы обратились к этому словарю по трем причинам:

- «Синтаксический словарь» предлагает чрезвычайно богатый иллюстративный материал, на основе которого можно выделить частотные лексико-семантические характеристики «хозяев» и «слуг»: некоторые значения имеют более двухсот пятидесяти иллюстраций; а суммарное количество иллюстраций значений отдельных предлогов превышает полторы тысячи;
- в «Синтаксическом словаре» основной акцент сделан на синтаксических функциях предложно-падежных конструкций, поэтому отдельно выделены группы, где предложно-падежные синтаксемы имеют очевидного синтаксического «хозяина», что позволяет нам вынести за рамки работы большую часть рассуждений о выделении «левого контекста»;
- дефиниции предложных значений в словаре описаны скорее в лексикологических, чем в лексикографических терминах, что упрощает процедуру формализации лексико-семантических полей слов, при которых реализуются предложные значения.

4. Методология исследования

В данной работе рассматриваются только первообразные предлоги: с одной стороны потому, что только они фигурируют в «Синтаксическом словаре», с помощью материалов которого мы описываем онтологию, но также и потому, что они «являются наиболее употребительными словами русской речи, что следует интерпретировать как свидетельство центрального положения этих слов в языковой системе, а изучение этих слов, по всей видимости, — наиболее короткий путь, который приведет нас к пониманию языковой системы в целом» [13, с. 116], а значит — описание онтологии первообразных предлогов должно стать первым шагом для описания общей онтологии предлогов.

Так как наша работа основана в первую очередь «Синтаксическом словаре» Г. Золотовой, мы в дальнейшем будем ориентироваться на список первообразных предлогов, использованный в этом словаре: *без (безо), в (во), для, до, за, из (изо), из-за, из-под (из-подо), к (ко), между, на, над (надо), о (об, обо), от (ото), перед (передо, пред, предо), по, под (подо), при, про, против, с (со), сквозь, среди, у, через (чрез)*.

Следует отдельно оговорить, что ниже мы будем использовать термины 1) значение предлога, 2) отношение между «хозяином» и «служгой», выраженное предлогом (определение «Академической грамматики») и 3) ролевая позиция

«слуги» по отношению к «хозяину», выраженная предложом (определение В. Рубашкина [13]) как синонимы.

Этапы работы над формированием онтологии предлогов

Работа над описанием онтологии первообразных предлогов состояла из четырех основных этапов:

- унификация и обобщение материалов «Синтаксического словаря»;
- выделение конструкций, в состав которых входят предложно-падежные синтаксемы; семантическая разметка «хозяев» и «слуг»;
- частотный анализ, составление гистограмм распределения семантических тегов «хозяев» и «слуг»;
- выделений лексико-семантических характеристик минимальных контекстов и формулирование условий автоматического определения значения предлогов на языке лексико-семантических и морфологических маркеров;

5. Создание контекстной грамматики предложных конструкций (онтологии)

5.1. Унификация и обобщение материалов «Синтаксического словаря»

Мы не ставим своей целью разработку новой классификации предложных значений — более рафинированной по отношению к существующим. Ориентируясь на возможности автоматической обработки текста, мы вынуждены, наоборот, даже несколько огрубить и обобщить взятую нами за основу исключительно нюансированную классификацию Г. Золотовой (в которой некоторые предлоги разработаны до выделения 38 значений [8, с. 159–179, 296–323]). Поэтому в некоторых случаях мы позволили себе объединить сходные группы значений, имеющие очевидную общность и часто называемые у Золотовой одинаково, так как мы не видим возможности со всей полнотой и точностью передать в формальном виде их различия. Например, при описании синтаксем с предлогом *на* Золотова выделяет, среди прочего, пять медиативных значений. Мы понимаем индивидуальность каждого описанного значения, но видим и то, что их объединяет: они все указывают на средство или способ. Эта общность, как мы считаем, дает нам право объединить их в одно составное медиативное значение.

Тут мы следуем принципу минимальной достаточности смыслообразительных средств языка представления знаний. Этот принцип обеспечивает сбалансированность выразительных средств языка с возможностями алгоритмической обработки. «Богатство смысловых оттенков в описаниях предлогов имеет смысл различать в формальной модели только при условии, что оно будет адекватно поддерживаться изощренными алгоритмами анализа-синтеза текста и логического вывода. Между тем практический опыт показывает, что достижение такой изощренности — даже там, где это может принести какую-то пользу — дело не столь близкого будущего» [14, с. 294]. Пока же на пути формального описания предлогов делаются первые шаги,

приоритет должны иметь модели, использующие максимально обобщенный набор смыслов, представляющих варианты значения предлога.

5.2. Выделение конструкций и семантическая разметка предложно-падежных синтаксем

Существующие семантические модели различаются прежде всего по языку представления смысла (или на уровне отдельного знака, или связного текста). Среди наиболее известных следует назвать модель компонентного анализа, модель семантических классов, формально-логическую модель, реляционно-ситуационную модель, модель, основанную на ролевой структуре предложения, семантическую модель в рамках теории «Смысл ↔ Текст», модель концептуального и прецедентного анализа, язык UNL (см., в частности, в [15, с. 5–22]).

Но все они или неполны, или разработаны на материале английского языка (и некоторых других). Для русского языка наиболее полным таким языком, по крайней мере, в части предлогов, является реляционно-ситуационная модель, базирующаяся на теории коммуникативной грамматики Г.А. Золотовой [16]. Основополагающим термином теории коммуникативной грамматики является синтаксема. Под синтаксемой понимается минимальная, далее неделимая семантико-синтаксическая единица русского языка, выступающая одновременно как носитель элементарного (категориально-семантического) смысла и как конструктивный элемент более сложных синтаксических построений. Был создан словарь таких синтаксем, представляющий собой первое в языкознании последовательное функциональное описание элементарных единиц русского синтаксиса — синтаксем, из которых строятся словосочетания и предложения [8].

Для выделения лексико-семантических характеристик минимальных контекстов, в котором реализуются синтаксемы, требуется язык семантических классов, элементы значения которых будут присваиваться выделенным «хозяевам» и «слугам». Здесь мы воспользовались моделью семантических классов Е.В. Падучевой. Одной из причин обращения к ней является то, что она имеет компьютерную реализацию в виде семантической разметки «Национального корпуса русского языка» (НКРЯ) [17].

Семантическая разметка НКРЯ приписывает единицам текста один или несколько семантических и словообразовательных признаков, например, 'лицо', 'вещество', 'пространство', 'скорость', 'движение', 'обладание', 'свойство человека', 'диминутив', 'отглагольное имя' и т.п. Используется фасетная классификация, при которой одно слово может попадать в несколько классов. В основу семантической разметки положена система классификации русской лексики, принятая в базе данных «Лексикограф», которая разрабатывалась с 1992 г. в Отделе лингвистических исследований ВИНТИ РАН под рук. Е. В. Падучевой и Е. В. Рахилиной [18]. И указанные признаки, разбитые по нескольким категориям: разряд, таксономия, мереология, оценка, словообразование и т.п. — могут быть извлечены из корпуса в привязке к интересующим нас контекстам (см., например, рис. 1);

	ружьё	
Лемма	ружьё	
Грамматика	сущ., неод., с., ед., вин.	
Семантика основная	t tool weapon	
Доп признаки	numred	

Рис. 1. Разметка лексемы в терминах НКРЯ

Разработанный нами модуль (экстрактор) извлекает эту информацию из данных НКРЯ и определенным образом обрабатывает. Фрагмент выходных данных представлен в табл. 1.

Таблица 1. Результат работы автоматического экстрактора семантической разметки

выпротстал из-под одеяла	d:pref	r:concr, t:tool:bedding
летел из-под колёс	ca:noncaus, t:move	pc:tool:device, pt:part, r:concr, t:tool, top:disk — r:concr,
выбрался из-под одеяла	ca:noncaus, d:pref, der:v, t:move — ca:noncaus, d:pref,	r:concr, t:tool:bedding
вылез из-под одеяла	ca:noncaus, d:pref, der:v, t:move	r:concr, t:tool:bedding
достал из-под сиденья	ca:caus, d:pref, t:poss — ca:noncaus, d:pref	der:v, pc:tool:furn, pt:part, r:concr

5.3. Частотный анализ и составление гистограмм распределения семантических тегов «хозяев» и «слуг»

Получив таблицы с семантической разметкой минимального контекста для всех значений всех предлогов — суммарно это более **семи тысяч** конструкций «хозяин» — предлог — слуга» мы составили гистограммы частотного распределения семантических тегов «слуг» и «хозяев» для каждого предложного значения.

Например, для локативного значения предлога *над*, мы получили таблицу 2 семантически размеченных «хозяев» и «слуг».

Далее мы составили гистограммы частотного распределения семантических тегов «хозяев» (рис. 2) и «слуг» (рис. 3), в соединении с которыми реализуется локативное значение предлога *над*.

В таких гистограммах встречались не только чисто семантические теги (sc:plant, t:loc, qc:stuff и т.п.), но и словообразовательные (der:s, der:v и др.) и морфологические (ca:noncaus, ca:caus и др.). При анализе гистограмм мы регулярно игнорировали словообразовательные теги, так как в разных гистограммах они имеют примерно равную частотность, что говорит о том, что они являются параметрами, нерелевантными для нашего исследования.

Таблица 2. Фрагмент таблицы семантически размеченных контекстов, в которых реализуется локативное значение предлога *над*

Конструкция	Семантическая разметка «хозяина»	Семантическая разметка «слуги»
вился над кровлею	ca:noncaus, d:root — ca:noncaus, d:root	der:v, pc:constr, pt:part, r:concr
стоя над Невой	der:v	r:propn, t:topon
стоял над равниной	ca:noncaus, d:root, t:loc:body	der:a, r:concr, t:space, top:form
возвышалось над мехом	ca:noncaus, d:pref, der:v	pc:animal, pc:tool:device, pt:aggr, pt:part, r:concr, sc:hair, t:hair, t:tool — pt:aggr:aggrpl, r:concr, sc:tool:cloth, t:stuff
стояла над Академией	ca:noncaus, d:root, t:loc:body — ca:noncaus, d:root, t:loc	r:concr, t:org — r:concr, t:org
висящие над колодцем	ca:noncaus, d:root, t:loc — ca:noncaus, d:root	r:concr, t:constr, t:space, top:mineshaft — r:concr, t:constr, t:space, top:mineshaft
летит над полями	ca:noncaus, t:move	r:concr, t:space, top:horiz — der:shift, dt:space, pc:tool:cloth, pc:workart, pt:part, r:abstr, r:concr, t:plot, t:space, top:stripe, top:surface
бегут над морем	ca:noncaus, d:root, t:move — d:root	r:concr, t:reservoir, t:space, top:contain — der:shift, pt:qtm, r:abstr, t:quant:max
брожу над морем	ca:noncaus, d:root, t:move — ca:noncaus	r:concr, t:reservoir, t:space, top:contain — der:shift, pt:qtm, r:abstr, t:quant:max
ехал над рекою	ca:noncaus, d:root, t:move — d:root	r:concr, t:reservoir, t:space, top:contain — pt:qtm, r:abstr, t:quant:max
плывет над дубравой	ca:noncaus, d:root, t:move — d:root	der:s, r:concr, sc:plant, t:space
пролетая над рекой	ca:noncaus, d:pref, der:v, t:move	r:concr, t:reservoir, t:space, top:contain — pt:qtm, r:abstr, t:quant:max

Полный корпус размеченных минимальных контекстов всех значений всех первообразных предлогов и гистограммы частотного распределения семантических тегов можно назвать контекстной грамматикой предложных конструкций русского языка (онтологией).

5.4. Выделение лексических и семантических маркеров предложных значений

Анализируя семантические теги в гистограммах, описанных в предыдущем разделе, мы выделяли доминирующие по частотности семы¹ как семантические

¹ Здесь и далее для краткости вместо термина *семантический тег* используется

маркеры значения. С опорой на собственную языковую интуицию мы пополняли список правил близкими по смыслу тегами (например, если кандидатом в лексические маркеры была сема *t:book*, мы также считали кандидатом в маркеры сему *t:text*) и тегами, которые очевидно являются маркерами, но не попадали в список наиболее частотных из-за малого количества или разнородности иллюстраций.

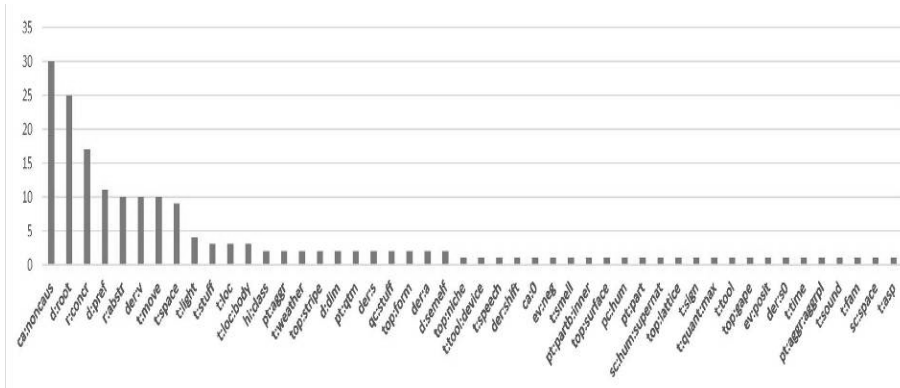


Рис. 2. Гистограмма частотного распределения семантических тегов слов-«хозяев», в соединении с которыми реализуется локативное значение предлога *над*

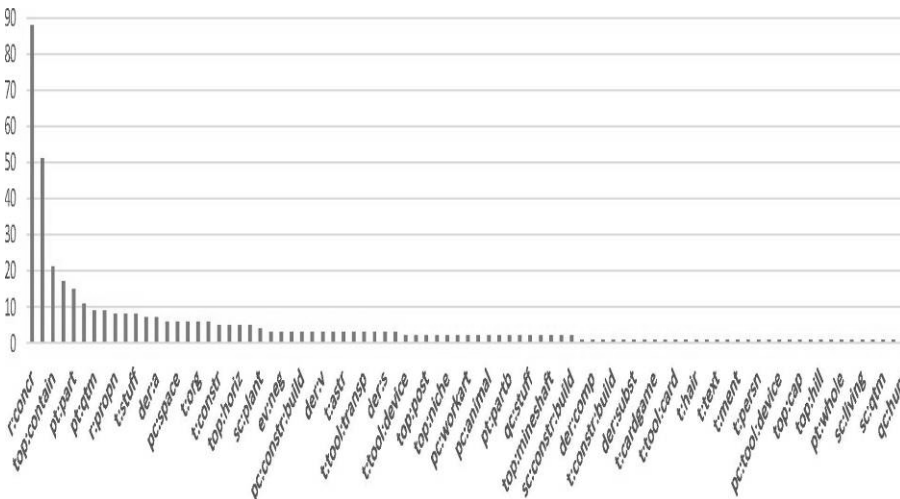


Рис.3. Гистограмма частотного распределения семантических тегов слов-«слуг», в соединении с которыми реализуется локативное значение предлога *над*

К сожалению, из-за скупости доступной нам семантической разметки (в НКРЯ достаточно подробно разработана только семантика существительных) многие семантические характеристики «хозяев» и «слуг» оказались

неформализованными, и, даже видя несомненную семантическую общность между словами — «хозяевами» или словами — «слугами», мы никак не могли выделить их характерные семантические теги. Например, при описании делиберативно-объектного значения предлога *над* мы обнаружили, что «хозяевами» регулярно выступают «глаголы насмешки»: *смеяться, насмехаться, издеваться, глумиться, шутить, хохотать, подтрунивать*. К сожалению, семантическая общность этих глаголов не выражена в разметке НКРЯ, и поэтому — за неимением другого выхода — в таких случаях мы были вынуждены выделять не семантические, а лексические маркеры.

Следует отдельно сказать, что для предлогов-грамматических омонимов, управляющих несколькими падежами, мы также выделяли *морфологические* маркеры «слуг», так как управлению разными падежными формами предлоги реализуют разные значения.

6. Онтологическое описание простых непроектируемых предлогов

В данном разделе представлен пример выделения лексико-семантических и морфологических маркеров контекстов, в которых реализуются предложные значения, т.е. онтология первообразных предлогов русского языка.

Предлог *над*

Синтаксемы с предлогом *над* реализуют три значения: локатив, объект-делибератив и субмиссив.

Этот предлог всегда управляет творительным падежом, поэтому мы не выделяем морфологические маркеры для «слуг».

Локатив — «обозначение места, пространства, среды, характеризуемой состоянием, либо происходящим в ее пределах действием, либо нахождением в ней предмета»: *стоять над морем, тучи над полем*. [8, с. 262].

«Хозяевами» могут быть глаголы пребывания, положения в пространстве: *повиснуть, нависнуть, стоять, возвышаться, висеть, сидеть* (семы t:loc, t:loc:body) глаголы движения: *лететь, бежать, бродить, пролетать, проноситься* (сема t:move) (в последнем случае конструкция приобретает транзитивное значение — обозначение пути движения). Также «хозяевами» могут быть глаголы, «сообщающие об осуществлении в названном пространстве звучащих и зримых проявлений бытия»: *сиять, шептать, загораться, жужжать, сверкать, светить* (семы t:light, t:sound). Также «хозяевами» могут быть отвлеченные имена, обозначающие действие, происходящее в названном пространстве: *восход, пенье, путь, ветер* или конкретные имена, обозначающие предмет, ориентированный в пространстве относительно названного в локативе лица или предмета: *небо, снег, тучи, решетки, облака, черточки*.

Объект-делибератив — обозначение предмета переживания (иногда с каузативным оттенком), проявления чувства, речемыслительной деятельности или физической деятельности (иногда с локативным оттенком): *трудились над полосой, корпел над книжкой, раздумывал над задачей, мудрил над колесом, работа над проектами, наблюдение над периодичностью, сжалась над*

Чаплицким, смилостивьтесь над голодающим, смеясь над собратом, шутит над султаном, глумиться над Герасимом, насмешка над землей, издевка над кузнецом, улыбка над глупостью, издевательств над людьми.

«Хозяевами» выступают глаголы мысли, умственной или физической работы и коррелирующие с ними существительные: *трудиться, корпеть, задуматься, биться, раздумывать, мудрить, работа, наблюдение* (сема t:ment). Все перечисленные; глаголы проявления горестного чувства и коррелирующие с ними существительные: *рыдать, плакать, стенать, слезы, стоны* (сема t:psych:emot). Также «хозяевами» выступают глаголы проявления отношения, основанного на сознании превосходства, социального или нравственного, и коррелирующие с ними имена: *сжалиться, смилостивиться, смеяться, насмехаться, шутить, подтрунивать, глумиться, измываться, издеваться, смех, насмешка, шутка, улыбка, издевательство* (все перечисленные слова можно считать лексическими маркерами).

Так как работать, размышлять, смеяться и горевать можно над совершенно разными вещами, «слугами» выступают самые разные существительные.

Субмиссив — обозначение объекта, над которым властвует или довлеет другой объект: *начальствовать над ними, поцарствовал над Москвой, властвует над реками, восторжествовало над слабостью, царствуешь над землей, преобладало над чувством, власть над душой, господство над трудом.*

«Хозяевами» выступают глаголы количественного, социального и нравственного превосходства (и коррелирующие с ними отвлеченные имена существительные): *господствовать, властвовать, начальствовать, преобладать, восторжествовать, царствовать, командовать, власть, господство, шефство, победа, суд, контроль, начальник, воевода, царь и некоторые другие* (все перечисленные слова можно считать лексическими маркерами).

7. Построение, тестирование и оценка семантического интерпретатора

Онтология предлогов, фрагменты которой были описаны выше и представляющая из себя набор лексических, семантических и морфологических маркеров минимального контекста, в котором реализуются предложные значения, может быть использована как основа для создания интерпретатора, определяющего значение предлога с опорой на минимальный размеченный контекст. Семантический интерпретатор использует описанную выше грамматику как правила определения значения предлога. Такой интерпретатор был нами разработан. Создание его само по себе является отдельной трудоемкой задачей. и подробно он будет описано в отдельной статье.

Также отдельная задача — его оценка. Можно лишь сказать, что оценка работы интерпретатора сводилась к сравнению значения предлога, приписанного нами вручную, с значением предлога, вычисленным интерпретатором. Мы считаем, что оценка работы такого интерпретатора является и косвенной оценкой адекватности созданной грамматики.

Заключение

В работе представлена методика описания онтологии предлогов с опорой на лексико-семантические и морфологические контекстные маркеры и один из вариантов её реализации — онтология первообразных предлогов русского языка, построенная на материалах «Синтаксического словаря» Г. Золотовой. Данная онтология представляет из себя описание лексико-семантических характеристик знаменательных слов, в соединении с которыми реализуются значения предлогов.

В процессе работы был произведен анализ материалов «Синтаксического словаря», из иллюстраций словаря были выделены конструкции, в состав которых входят предложно-падежные синтаксемы, знаменательным элементам этих конструкций была приписана многозначная семантическая разметка в терминах Национального корпуса русского языка, автоматическая извлекаемая из корпуса с помощью специально написанного экстрактора, составлены гистограммы частотного распределения семантических тегов знаменательных слов для каждого значения каждого первообразного предлога, выделены лексические и семантические маркеры минимальных контекстов, в которых реализуются значения предлогов, с помощью этих маркеров составлены правила автоматического определения значения предлогов для предлогов *до, из-за, из-под, над, от*; корректность правил проверена с помощью специально сконструированного семантического интерпретатора на материале тестового корпуса семантически и морфологически размеченных словосочетаний, содержащих предложно-падежные синтаксемы (корпус состоит из пятисот словосочетаний). Оценка работы интерпретатора выявила математическое ожидание успешной семантической разметки предлогов — 82,6%, что можно считать очень высоким показателем для анализатора, заведомо лишённого информации о более широком контексте и прагматической ситуации и ориентирующегося исключительно на минимальный лексико-семантический контекст.

Практическая и научная значимость данной работы связана с возможностью использования полученных данных для дальнейших исследований в области семантики предлогов и с возможностью использования интерпретатора и экстрактора, созданных нами, для решения практических лингвистических задач для более детальной разработки компьютерных средств семантического анализа.

Литература

- [1] Святогор Л., Гладун В. Семантический анализ текстов естественного языка: цели и средства. [Электронный ресурс]: http://www.foibg.com/ibs_isc/ibs-09/ibs-09-p06.pdf.
- [2] Сокирко А.В. Семантические словари в автоматической обработке текста: По материалам системы ДИАЛИНГ: диссертация ... канд. тех. наук: 05.13.17. Москва, 2001. 120 с.: ил.
- [3] Fillmore Ch.J. The Mechanisms of Construction Grammar // Proceedings of the Berkeley Linguistic Society. Vol. 14. 1988.

- [4] Goldberg A.E. *Constructions at Work: the Nature of Generalization in Language*. Oxford: Oxford University Press, 2006.
- [5] Кузнецова Ю.Л. Грамматика конструкций. Обзор // Научно-техническая информация. Серия 2, № 4, 2007.
- [6] Hanks, P. *Lexical analysis. Norms and Exploitations*. London, 2013.
- [7] Firth, J. R. A synopsis of linguistic theory 1930-1955, In: F. Palmer (Ed.), *Selected Papers of J. R. Firth 1952-1959*, London, Longman, 1957, pp. 168-205.
- [8] Золотова Г. А. *Синтаксический словарь: Репертуар элементарных единиц русского синтаксиса*. Изд. 4-е. М., Едиториал УРСС, 2011.
- [9] *Русская грамматика*. М., 1980. Т. I.
- [10] Коновалова Т. Е. Специфика предлогов широкой семантики и её отражение в словах // *Вопросы романо-германского языкознания*. — Саратов: Издательство Саратовского университета, 1988. С. 60—67.
- [11] Виноградов В. В. *Русский язык*. М., 1972.
- [12] *Русская грамматика*. М., 1980.
- [13] Солоницкий А. В. *Проблемы семантики русских первообразных предлогов*. — Владивосток: Издательство Дальневосточного университета, 2003.
- [14] Рубашкин В. Ш. *Онтологическая семантика. Знания. Онтологии. Онтологически ориентированные методы информационного анализа текстов*. М., 2012.
- [15] Мочалова А.В., Захаров В.П., Кузнецов В.А., Мочалов В.А. *Семантический анализ. Некоторые проблемы автоматической обработки текста*. Петрозаводск, 2016.
- [16] Золотова Г. А., Онипенко Н. К., Сидорова М. Ю. *Коммуникативная грамматика русского языка*. М., 2004.
- [17] Кустова Г. И., Ляшевская О. Н., Падучева Е. В., Рахилина Е. В. *Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы // Национальный корпус русского языка: 2003-2005. Результаты и перспективы*. М., 2005. С. 155—174.
- [18] Падучева Е.В. *Динамические модели в семантике лексики*. М.: Языки славянской культуры, 2004.

Context grammar of prepositional constructions in Russian

V.P. Zakharov ¹, V.D. Mikhailova²

¹ St. Petersburg State University, Institute for Linguistics Studies of the Russian Academy of Sciences

² St. Petersburg State University

The paper features a new method of building the grammar of prepositional constructions for the Russian language (the ontology of Russian prepositions) based on lexical, semantic and morphological contextual markers, as well as one of its possible realisations — the ontology of simple Russian prepositions, based on data of the Syntactic Dictionary by Galina Zolotova. The grammar consists of descriptions of lexical and semantic features of meaningful words that build up a context in which some meaning of a preposition is being actualised. The paper describes performance of an interpreter uncovering meanings of prepositions in “master” — preposition — “slave” constructions. The practical and scientific value of this work is connected with the possibility to use its results when computing linguistic tasks in NLP-systems and in further investigation of semantics of prepositions.

Keywords: prepositions, the Russian language, semantics, syntactic structures, context, lexical and syntactic markers.