

Поиск хозяина предложной группы в русском предложении¹

Search of the control word of a prepositional phrase in Russian

Кобзарева Т. Ю. (stamstam@mtu-net.ru)

Российский государственный гуманитарный университет

Рассмотрены проблемы поиска хозяина предложной группы в русских текстах разных жанров при автоматическом поверхностно-синтаксическом анализе. Описывается лингвистический базис и стратегия поиска хозяина предлога в рамках системы модульного анализа русского синтаксиса (MARS).

1. Общая стратегия решения проблемы

В настоящем исследовании рассматривается проблема поиска синтаксического хозяина предложных групп (ПГ) в обычном художественном тексте (не записи и не имитации устной речи) и в словарных определениях энциклопедических словарей.

Поиск хозяина предлога (р) — одна из самых трудных задач при построении синтаксических связей.

Для успешного ее решения необходимо определить особенности поведения ПГ в разных контекстах и принципиальную схему анализа.

Проблема рассматривается в рамках подхода, принятого в системе MARS, разрабатываемой в настоящее время в РГГУ. Ее специфика заключается в том, что собственно моделированию структуры предикативных ситуаций, составляющих предложение (S), предшествует сегментация — определение зон его линейной структуры, где могут находиться члены этих предикативных ситуаций.

Система состоит из 6 процедурно независимых модулей [2].

- 1) Постморфология — несловарные проблемы морфанализа: обработка имен собственных и названий, окказиональных аббревиатур, числительных [3];
- 2) Разрешение частичной омонимии — совпадения форм разных частей речи [4];
- 3) Предсегментация — построение **проективных фрагментов** (ПФ) именных и предложных групп (ИГ и ПГ) [5]. ПФ — часть S, границами которой служат связанные подчинительной

связью слова. Например, существительное (N) и его слуга — небособленное согласованное определение с вершиной A*– прилагательным, причастием, местоименным прилагательным или порядковым числительным (например, *лежащий на ложе в грозном полумраке прокуратор* — ПФ из прозы М. Булгакова (далее Б), границы задает связь *прокуратор* R *лежащий*); существительное и его слуга-существительное в роли небособленного приложения (*командир алы — сириец* (Б)); предлог (р) и его слуга-существительное (или его синтаксический субститут) (*на зеленой, никогда не езженной мостовой* (М)), конструкции с числительными, сложные сказуемые и др., т. е. связи, определяющие единицы линейной структуры, необходимые для сегментации.

- 4) Сегментация — построение **сегментов** [6], то есть простых-главных S, придаточных S, деепричастных оборотов, обособленных согласованных определений, обособленных приложений, уточняющих и всех других оборотов, обособляемых по правилам русской пунктуации.
- 5) Моделирование структуры синтагматических связей внутри сегментов.
- 6) Построение связей сегментов [7].

Поиск синтаксического хозяина ПГ происходит на 5-ом этапе

Сегментация и построение проективных фрагментов чрезвычайно сужают зоны, в пределах которых ведется поиск связей слов, что на практике облегчает решение задачи.

¹ Доклад подготовлен при частичной поддержке РФФИ — грант № 09-06-00275-а

Таким образом, при поиске хозяина ПГ границами поиска хозяина являются границы ПФ, (если мы находимся внутри ПФ), границы сегмента, если мы находимся вне ПФ (при этом слова внутри ПФ из рассмотрения исключаются) и, если ПГ находится внутри сегмента между двумя сочиненными словами, зону поиска определяет свойство сочинительной проективности [8].

2. Синтаксические особенности функционирования предложных групп

2.1. **р** по способности выступать слугами тех или иных частей речи делятся на три класса: 1) **р**, способные быть слугами только глагола (**V**), 2) **р**, которые могут выступать и слугами **V** и слугами прилагательного (**A**), и, наконец, 3) **р**, которые могут быть слугами и **V**, и **A**, и существительного (**N**) [1]. Естественно, что наибольшие трудности возникают при поиске хозяина предлогов третьего класса.

При этом ПГ, во-первых, может замещать некоторую валентность предикативной вершины сегмента (или ПФ) или выступать в роли обстоятельства этой вершины (т. е. быть актантом или сирконстантом вершины). В этих случаях ПГ подчинена вершине.

Во-вторых, ПГ может быть слугою актантов и сирконстантов или их слуг, чаще всего представленных **N**.

Последняя группа случаев, когда ПГ подчинена не вершине сегмента или ПФ, распадается на два вида ситуаций.

2.1.1 Хозяевами ПГ являются некоторые слова — не вершины ПФ и сегментов — и эти слова имеют некоторое предложное управление (предложную валентность). Если ПГ с вершиной — этим **р** находится в определенной позиции относительно слова — потенциального хозяина этого **р**, то оно и будет скорее всего ему подчинено. Например, *Влияние этого человека на ее сына ее пугало*. Возможно совпадение предложного управления вершины и некоторого слова в сегменте. При этом у каждого из них может быть свой слуга, например: *Его влияние на сына действовало на всех*. Но такая ситуация может и породить неоднозначность, например: *Негативное влияние этого человека на ее сына действовало очень заметно*.

2.1.2. Существуют структуры, в которых ПГ, независимо от возможностей вершины и других слов управлять этим **р**, выступают слугою некоторого слова, для поиска которого оказывается необходимым учитывать и слугу **р** и управляющее им слово, Например: *Она привезла туфли из кожи. Нужна трубочка длиной со спичку. Этим занимается комитет по правам человека. Он увидел человека в синем плаще*. Для определения синтаксиче-

ских хозяев ПГ = *со спичку \ по правам (человека) \ на концерт \ в синем плаще \ из кожи* необходимо определить условия, по которым ПГ в этих контекстах не могут быть объявлены слугою сказуемого и не могут заполнять валентность другого слова.

Или же, например, в каждом из предложений типа *Она из Питера привезла туфли из кожи. Положи билет на концерт на стол*. мы находим по две ПГ с одним и тем же предлогом: *из Питера, из кожи; на концерт, на стол*. При этом надо задать условия, определяющие, какая именно ПГ является определительной, а какая — подчинена предикату: *Она привезла туфли из Питера. Она только вчера привезла из мастерской туфли из Питера. Она привезла туфли из кожи. Она вынула туфли из Питера из шкафа*.

Подобные ситуации естественно объединяются в классы.

Чтобы в подобных случаях правильно определить хозяина ПГ или не потерять синтаксическую неоднозначность, задаются описания контекстных ситуаций таких классов с указанием порядка следования ИГ и ПГ в тексте и семантических классов или списков хозяев и слуг предлогов.

Например, класс структур типа *туфли из кожи, нож из стали, чашка из фарфора* и др.² описывается как [ИГ с вершиной **N** ∈ семантическому классу (СК) 'предмет неодушевленный' + *из* со слугою **N** ∈ СК 'вещество, материал'].

Процедуры анализа строятся с учетом того, что и в этих ситуациях могут возникать синтаксические неоднозначности. Например, *Эта фабрика шьет сумки из крокодиловой кожи* синтаксически неоднозначно из-за управления *шить*³.

3. Встречаемость разных ситуаций в текстах. Статистические наблюдения

Сужение зоны поиска хозяина в результате построения ПФ на этапе, предшествующем моделированию внутренней структуры сегмента и на этапе сегментации, приводит к тому, что часто в потенциально альтернативных случаях при поиске хозяина реально альтернативы не возникает.

Ниже приводятся результаты небольшого статистического обзора распределения типов контекст-

² Заметим, что приведенные примеры приобретают большую языковую достоверность при наличии у слуги предлога определения, например, *туфли из мягкой кожи, нож из дамасской стали, чашка из саксонского фарфора*.

³ Не очевидно, что подобную неоднозначность стоит вводить, т. е. учитывать содержательное различие этих интерпретаций.

ных ситуаций, возникающих при поиске хозяина ПГ, для художественной прозы и статей энциклопедического словаря (ЭС). По полученным результатам можно судить, насколько универсальна проблема поиска хозяина ПГ для текстов столь разных жанров, в частности, насколько эта проблема при анализе статей ЭС отличается от аналогичной задачи в обычных художественных текстах, на которые в первую очередь ориентирована система MARS.

Результаты представлены в Табл.1 и Табл.2.

Для анализа художественного текста были выбраны произведения с почти 100-летним разрывом во времени и совершенно разные по литературным критериям. Обращает на себя внимание близость результатов.

В Табл.1 рассмотрены следующие типы контекстных ситуаций.

1. В сегменте или в ПФ есть вершина, которая и является хозяином ПГ. *Это был повергнутый в пыль хаос чудейский.* (М) В зоне поиска хозяина, которая определяется границами ПФ *повергнутый в пыль хаос*, предикативная вершина — *повергнутый*.
2. Хозяин ПГ определяется тем, что в зоне поиска в определенной позиции по отношению к предлогу есть N, имеющее «преимущественное право» управление данным р: [*отверстие, специалист, участие, статус, ситуация...*] R в_{предл}: [*вступление, выборы, кандидат...*] R в_{вин}: [*взгляд, ответ, право, цена...R на_{вин}*]; [*борьба, диалог, дискуссия, договор, отношения, переговоры, связь... R с_{тв}*] и т. п.

Например: Я всегда смутно чувствовал особенное значение Финляндии для петербуржца и... То, что хозяином ПГ для петербуржца является не вершина сегмента чувствовал, определяется способностью слова значение управлять р=для_{род}.

Для подобных случаев невозможно обойтись без информации об управлении конкретными предлогами, например: песенка о_{предл} жертва за_{вин} игра в_{вин}, ударение на_{тв}, специалист по_{дат} сравнение с_{тв} и др.

3. В зоне поиска хозяина в результате построения ПФ и сегментов существует единственный претендент на роль хозяина.

- а) Семен Афанасьевич Венгеров, родственник мой по матери... (М) В зоне поиска — обособленном приложении родственник мой по матери — есть единственный претендент на роль хозяина ПГ.
- б) ...не менее интересно и значительно, чем майский парад на Марсовом поле. (М)
- в) Там, на Торговых, попадаются еврейские вывески с быком и коровой, женщины с выбивающимися из-под косынки накладными волосами и семящие ... (М)

Хозяин ПГ = *из-под косынки* однозначно определяется в результате построения ПФ ИГ *выбивающимися из-под косынки накладными волосами*, а в качестве хозяина ПГ = *с ... волосами* — *женщины* и хозяина ПГ = *с быком и коровой* — *вывески* однозначно определяются по условию сочинительной проективности. в результате построения в ходе сегментации сочинительной связи слов *вывески* и *женщины*

4. Хозяина определяет вхождение ПГ в специфическую структуру, для определения которой нужно учитывать и порядок слов, и семантику хозяев, и семантику ПГ. Например: Яростно прижимает к груди книгу с надписью... Или И вдруг два господина в цилиндрах, прекрасно одетые, лоснящиеся богатством,... (М)
5. Есть альтернатива и есть синтаксическая неоднозначность.

Например: а) *Винодел с дочками уехал* к бабушке. (М) Возможны две интерпретации: *винодел R с (дочками)* и *уехал R с б)* ..., но настоящим адом для большинства неловких, не слишком здоровых и нервических детей был ручной труд. (М) Возможны две интерпретации: *ад R для* и *был R для*.

Заметим, что в прозе Пелевина существенно меньше ситуаций 3-его типа: у Пелевина практически нет сегментов с именными вершинами, в частности, назывных предложений, которые так любит Мандельштам, и сегментов-приложений.

Жанр текста	Всего ПГ в выборке	1. VRp	2. N ^{спис} Rp	N ^{неглаг} Rp		
				3. Хозяин — не глагол и в зоне поиска хозяина ПГ нет альтернативы	4. В зоне поиска хозяина ПГ есть альтернатива, но нет синт. неоднозначн.	5. В зоне поиска хозяина ПГ есть альтернатива и есть синт. неоднозначн.
Худ. проза — Шум времени, О. Мандельштам	562 100 %	403 72 %	30 6 %	48 8 %	69 12 %	13 2 %
Худ. проза — Поколение «П», В. Пелевин	269 100 %	210 78 %	22 8 %	3 1 %	30 11 %	4 2 %

Табл. 1. Результаты статистического обзора ситуаций управления предлогами в «Шуме времени» О. Мандельштама (далее М) и «Поколения «П»» В. Пелевина (далее П)

Жанр текста	Всего ПГ в выборке	1а. VRp	1б. НМП = $N_{\text{ном}} R p$	2. $N^{\text{спис}} R p$	$N^{\text{неглаг}} R p$		
					3. Хозяин — не глагол и в зоне поиска хозяина ПГ нет альтернативы	4. В зоне поиска хозяина ПГ есть альтернатива, но нет синт. неоднозначн.	5. В зоне поиска хозяина ПГ есть альтернатива и есть синт. неоднозначн.
ЭС	266 100 %	104 39 %	92 34 %	45 17 %	7 3 %	13 5 %	5 2 %

Табл. 2. Статистика распределения ситуаций управления ПГ в статьях ЭС

В данные таблиц не включены прямая речь и сегменты с нулевым V-предикатом, такие, как: ...махнул рукой, как бы (*будучи) в обиде на такую деловитость.; Да у него самого под столом (*есть) такие надписи. У всякого бренда (*есть) своя легенда. (П).

Для определений словарных статей «Советского энциклопедического словаря» (ЭС), который после 1991 г. выходил под названием «Новый энциклопедический словарь», был проведен анализ специфики синтаксической и линейной структуры статей.

Синтаксическая структура определений, как правило, начинается с сегмента с именным сказуемым $N_{\text{им}}$ и осложнена чаще всего вставлением обособленных согласованных определений разного рода и придаточных с подчинительным союзом **который**.

Несмотря на тяготение к именным структурам, случаи глагольного управления предложом (VRp) в ЭС составляют почти половину. Во-первых, они входят в придаточные определительные с глагольными сказуемыми и деепричастные обороты, например: *Абсентеизм земледельческий — форма землевладения, при которой собственник земли, не участвуя в процессе производства, получает денежный доход в виде ренты или прибыли*. И, во-вторых, VRp появляются в обособленных и необособленных согласованных определениях, которые в именных структурах компенсируют отсутствие морфологического сказуемого. Например: *Абрис — схематический план, сделанный от руки, с обозначением данных полевых измерений, необходимых для построения точного плана или профиля*.

В ЭС часто встречаются необычные для нормальной повествовательной структуры конструкции: первое именованное сказуемое — $N_{\text{им}}$, и оно входит в цепочку сочиненных сказуемых, которые, начиная со второго, уже представлены обычными морфологическими сказуемыми. (Надо отметить и необычную при этом пунктуацию: зоны влияния этих сказуемых разделены точкой с запятой, что практически не встречается в других жанрах). Например: *Абонентское телеграфирование — разновидность телеграфной связи; служит для непосредственного двустороннего обмена телеграфными сообщениями между абонентами*.

Таким образом, в ЭС грамматически возможны все те же варианты соединений сегментов в одном

S, что и в художественном тексте, а значит, требуются те же процедуры сегментации.

4. Словарь исключений — «частных случаев» (ЧС)

Словарь ЧС описывает контекстные условия, определяющие безусловного хозяина ПГ с точностью до порядка слов и семантических классов (или списка лексем) хозяев и\или слуг конкретных предлогов.

Использование словаря ЧС удобно: оно позволяет, не меняя словаря основ, легко пополнять набор типовых структур, представляющих как большие классы слов, так и уникальные словосочетания.

К подобным конструкциям относятся хорошо известные комитативные группы, где слугами являются ПГ с вершиной *с*, управляющей $N_{\text{ив}}$. (*мы с приятелем, дом с мезонином, машина с желтыми фарами* и т. д.)

В словаре ЧС представлены ситуации для двух типов комитативных групп — существительных неодушевленных и одушевленных:

- *тележка с двигателем \ вагон с двигателем \ автомобиль с кузовом \ форма с рельфовыми печатающими элементами \ трансформатор с одной обмоткой* и т. п.
- *инженер И. Ф. Иванов с другом \ инженер нашего завода Ваня Петров со своей младшей сестрой \ Иван с другом \ мы с Ваней* и т. п.

В некоторых конструкциях и хозяин *р* и его слуга — большие группы слов. Иногда их удается задать семантическими классами (СК). Например:

- $N^j \dots + p = \text{для } R N^v$ (N^v — глагольное N), где $N^j \in \text{СК}$ 'устройство, приспособление' (*аппарат для проведения; прибор для автоматического удержания; экскаватор для переукладки* и др. (ЭС)).
- $[N^j - \text{'параметр} + \text{с } R N_{\text{вин}}^v]$, где $N_{\text{вин}}^v$ — 'предмет неодуш.' (например, *толщиной \ длиной со спичку*).
- $[N^v (+ \text{ИГ}_{\text{род}}^k) (+ \text{ИГ}_{\text{род}}^{k+1}) + \text{с помощью}_{\text{род}} \backslash \text{при помощи } R N^{k+2}]$ где N^{k+2} 'прибор, аппарат'

(например, *опыление ... с помощью пульверизатора \ вертолета*)

- $[N^j + p = \text{в виде } R N^k]$ где N^j и N^k — 'предмет неодуш.' (например, *застежка в виде броши*)

Значимы в силу высокой частотности две конструкции с предлогами *в_{предл}* и *без* с большими списками и хозяев и слуг, частые в текстах любых жанров:

- $N^j + [ПГ = \text{в } R N^k^*]$ (например, *старичок в соломенной шляпе \ дама в бальном платье \ подсудимый в наручниках*)
- $N^j^* + \text{без } R N^k^*$ (например, *старичок без пиджака*)

N^j^* — нарицательные имена человека, групп людей или изображений человека, различающиеся по полу и возрасту: *человек, мужчина, старик, старичок, старуха, старушка, дядя, тетя ...*; по роду деятельности: *король, царь, премьер (-министр), председатель, конгрессмен, заседатель, судья, терапевт, психиатр, отоларинголог...*; изображения человека: *фигура, манекен, статуя, ...*; собирательные одушевленные: *толпа, свита, солдатня, колонна (демонстрантов)...*

N^k^* — предметы, которые могут быть надеты на человека, служить одеждой, украшением и т. д.: *пальто, плащ, горжетка, кольчуга...*; *шляпа, кепка, платок, шлем, каска...*; *галстук, шарф, ...*; *бусы, кольцо, перстень, браслет, наручники, ошейник, очки, пенсне...*

В некоторых ситуациях для небольшого набора ПГ существует большой список хозяев, который не всегда можно задать семантическим классом.

- $[N^j = N^v + (ИГ^{k+1}_{\text{род}} (ИГ^{k+1}_{\text{род}})) + \text{в}_{\text{предл}} R ИГ^{k+2}]$ где вершина $ИГ^{k+2} = \text{цель}$ (например, *конструирование и постройка моделей автомобилей в технических или спортивных целях* (ЭС)).
- $[N^j — \text{имя родственника (дядя, тетя, бабушка, бабушка...)} (+N^j + k — \text{имя родственника в Род})] + [[\text{со стороны} \setminus \text{по линии } R \text{ матери} \setminus \text{отца}] \vee [\text{с материнской} \setminus \text{отцовской стороны}]]$
- $[N^j \dots + (N_{\text{ген}}) + ((ПГ) + \text{в } N^{\text{пим}} \setminus \text{несколько} \setminus \text{много} \setminus \text{этажа} \setminus \text{этажей})]$, где одно из значений N^j — строение: *строение \ дом \ башня \ небоскреб \ магазин \ театр...*
- $[N^j \dots ПГ = \text{по правам} + \text{человека}]$: где между N^j и *по* м. б. только ИГ в Род. п., ПГ, D, частицы.
- $N^j = \text{евросуд, конвенция, комитет, документы, госбюро, центр, обзор, законодательство, образование, просвещение, библиотека, уполномоченный, линия, суд, комиссар...}$
- $[N^j_{\text{одуш собирав}} + \text{во главе с } R N^k_{\text{Тв}}]$ где N^j — организация или собирательное одушевленное, а N^k — одушевленной нарицательное или собственное (например, *фирма \ школа \ группа \ класс во главе с Иваном Ивановичем*).

В настоящее время словарь ЧС этого алгоритма насчитывает 35 линейных конфигураций⁴. Список ЧС открыт. Многие структуры для их введения в словарь ЧС требуют наблюдений за их поведением в тексте.

5. Алгоритмическое решение задачи

Построен алгоритм, реализующий следующую принципиальную схему анализа.

Нормой в алгоритме считается 1-ая ситуация: ПГ объявляется слугою вершины сегмента, которая в простом-главном и придаточном S м. б. выражена и неморфологическим сказуемым, в частности — $N_{\text{им}}$. Если ПГ вложена в ПФ, то ее хозяином объявляется предикативное слово — вершина ситуации, границы которой заданы ПФ.

Чтобы объявить вершину сегмента или вершину предикативной ситуации ПФ хозяином, необходимо удостовериться, что в зоне влияния этой вершины нет ситуаций — исключений из этого правила, представленных в Табл.1 столбцами 2 и 4, когда управление ПГ «перехватывает» слово — не вершина.

Существенно, что в обычных текстах с цепочками именных и предложных групп вершин — $N_{\text{им}}$ часто появляется в простых назывных S. В ЭС $N_{\text{им}}$ обычно в выступает в роли именного сказуемого, причем в силу специфики задач текста это сказуемое распространено и косвенными падежами существительных и предложными группами. Например: *Аэрозольный генератор — машина для образования термомеханическим способом аэрозолей пестицидов и их распыления в целях уничтожения вредителей сельскохозяйственных культур, лесных и плодовых насаждений, а также для обеззараживания сельскохозяйственных хранилищ и животноводческих помещений.*

При поиске хозяина ПГ в таких сегментах с именным сказуемым приходится решать те же проблемы, что и в сегментах с глагольными вершинами, т. е. ситуации 2-ого, 3-его и 4-ого столбцов играют ту же роль, что в сегментах с глагольной вершиной и используется тот же лингвистический базис.

Соответственно, качество анализа определяется

- точностью отбора сегментов, для которых в результате сегментации не возникает альтернативы (ситуации 3-его столбца Табл.2),
- полнотой словаря специфических конструкций со слугами-ПГ (4-ый столбец) и

⁴ Программная версия алгоритма показывает хорошие результаты применения предлагаемой стратегии при обработке тестовых примеров, и на следующем этапе отладки системы, когда список ЧС достигнет определенной степени полноты, предполагаются эксперименты, которые позволят определить количественные оценки работы алгоритма.

- полнотой списков слов — хозяев конкретных предлогов (для ситуаций 2-ого столбца).

Первый шаг анализа — поиск ситуаций, когда хозяина можно найти без анализа управления. Для этого алгоритм определения безальтернативного управления ПГ ищет признаки контекстов, в которых хозяин ПГ определяется однозначно линейной структурой сегмента.

Второй шаг — поиск случаев-исключений. Если не найдена безальтернативная ситуация, алгоритм обращается к словарю ЧС и проверяет, не соответствует ли контекст анализируемой ПГ одной из структур этого словаря.

Третий шаг. Если нет ситуации безальтернативного управления и нет ситуации словаря ЧС, проверяется, что в сегменте или ПФ нет лексем, способных управлять конкретным предлогом, которые могут «перехватить» управление ПГ у вершины. В MARS в словаре основ управление предлогом лексически не конкретизировано, поэтому в специальной таблице для каждого *p*, управляющего определенным падежом и способного быть слугою и *V* и *N*, задается список *N*. В настоящее время такие списки заданы для 34 предлогов и суммарно насчитывают 428 существительных. При появлении в определенной позиции по отношению

к ПГ такого *N*, если нет конкурирующей валентности, мы отдаем предпочтение этому *N*. Если же эта ПГ — единственный вариант заполнения одинаковых валентностей двух слов, фиксируется неоднозначность.

6. Заключение

Предложена и алгоритмически реализована универсальная стратегия поиска хозяина ПГ, опирающаяся на использование зон поиска хозяина *p*, определяемых на этапе сегментации предложения, предшествующем этому поиску.

Анализ ситуаций, возникающих при поиске хозяина ПГ, доказывает возможность построения универсальной грамматики поверхностного синтаксиса для этой зоны автоматического анализа русского предложения.

Успешность работы алгоритма в сегментированном предложении определяется полнотой и точностью описаний, задающих модели предложного управления, в частности насыщением важного ее фрагмента, исключающего контекстные ситуации специфического управления предложными группами.

Литература

1. Мельчук И. А. Автоматический синтаксический анализ. Т. 1. — Новосибирск: Ред.-изд. отдел Сибирского отделения АН СССР, 1964.
2. Кобзарева Т. Ю. Иерархия задач поверхностно-синтаксического анализа русского предложения // НТИ, Сер.2, №1, 2007 — С 23–35.
3. Кобзарева Т. Ю. Морфализация in vivo // Труды Международной конференции Диалог'2004, — М.: Наука, 2004 — С. 286–291.
4. Кобзарева Т. Ю., Афанасьев Р. Н. Универсальный модуль предсинтаксического анализа омонимии частей речи в русском языке на основе словаря диагностических ситуаций // Труды международного семинара Диалог'2002, Протвино 2002. Т. 2. С. 258–268.
5. Кобзарева Т. Ю. Некоторые свойства линейной структуры именных и предложных групп (Поверхностно-синтаксический анализ русского предложения) // Вестник РГГУ, № 8/07, Серия «Языкознание» (Московский лингвистический журнал № 9/2) — М.: 2007. — С. 113–130.
6. Кобзарева Т. Ю. Принципы сегментационного анализа русского предложения. Московский лингвистический журнал // М.: 2004. Т. 8. №1 — С. 31–80
7. Кобзарева Т. Ю. Построение графа связей сегментов (поверхностно-синтаксический анализ русского предложения) // Компьютерная лингвистика и интеллектуальные технологии. Труды Международной конференции Диалог, М, Наука, 2008 — С. 192–198.
8. Кобзарева Т. Ю. Рекурсивность и проективность сочинительных связей в русском тексте // Труды Международной конференции Диалог 2006, Бекасово — М.: Наука, 2006. — С. 223–229.