

Русский фреймнет: к задаче создания корпусного словаря конструкций

Russian framenet: towards a corpus-based dictionary of constructions

Ляшевская О. Н. (olesar@mail.ru),
Кузнецова Ю. Л. (julia.kuznetsova@uit.no)

University of Tromsø, Тромсе, Норвегия

В российской компьютерной лингвистике до сих пор нет ресурсов, аналогичных FrameNet, VerbNet и PropBank, в которых можно было бы получить иллюстрации глагольного и именного управления, а также периферийных лексически ориентированных конструкций. В работе описываются задачи создания и структура FrameNet-ориентированной системы, предназначенной для исследования морфологических, синтаксических, семантических и лексических ограничений в русских конструкциях.

1. Введение

В последнее время благодаря появившимся поисковым системам (таким как Яндекс и Google) и корпусам текстов с возможностью поиска по ним (в частности, НКРЯ (<http://ruscorpora.ru>) и многим другим) возможности лингвиста значительно расширились. В лингвистике теперь стало можно ставить такие задачи, которые прежде решить было невозможно или для исследования которых требовалось необозримое количество времени и ресурсов. Например, теперь легко и быстро можно узнать, сочетается ли некоторый глагол с некоторым объектом, и с каким из двух объектов он встречается чаще, чем с другим.

Вместе с тем, при решении ряда задач наши поисковые возможности ограничены. С одной стороны, в запросах к поисковым системам и к большинству корпусов мы можем задавать только линейный порядок слов, поэтому лингвисту, исследующему конструкции, приходится перебирать все возможные комбинации элементов. С другой стороны, зачастую не хватает информации о синтаксической связи слов. Например, сложности возникают при поиске примеров конкретного варианта управления глагола или примеров конкретной конструкции (такой как конструкция с квазимперативом долженствования *Они едят в ресторанах, а я плати*). Как правило, в результатах поиска присутствует столько «шума», что требуется множество дополнительных усилий для того, чтобы отделить нужные примеры от примеров, случайно попавших в выданные результаты.

По-видимому, лучше всего указанную задачу решает синтаксически размеченный корпус НКРЯ

(<http://www.ruscorpora.ru/search-syntax.html>), в котором можно искать синтаксические зависимые, в том числе без учета порядка слов, но пока что этот ресурс не очень представительен в связи с малым размером. Кроме того, при создании этого корпуса не ставилась задача обеспечить достаточную выборку примеров для отдельных лексических единиц; в частности, даже частотные русские глаголы в нем представлены неравномерно. Например, обнаружив в синтаксическом корпусе, что в четырех из пяти имеющихся примеров на глагол *знать* этот глагол употреблен с прямым объектом, мы не можем сделать вывод ни о том, что это отражение наиболее частотной модели, ни что эта модель встречается в среднем в 80% случаев.

В российской компьютерной лингвистике до сих пор нет специализированных ресурсов, аналогичных FrameNet (Johnson, Fillmore et al. EE), VerbNet (Kipper et al. 2006) или PropBank (Palmer 2005). В этих источниках исследователи английского языка могут получить данные о типах глагольного управления, их распределении и вариативности в разных лексических единицах, просмотреть иллюстративный материал – причем на примерах из реальных текстов. Система FrameNet, кроме того, содержит аналогичные данные об именах существительных и прилагательных, и, что примечательно, в настоящее время эволюционирует в сторону словаря конструкций («New Constructicon», см. Fillmore 2008).

Таким образом, речь идет о создании русского фреймнет-ориентированного ресурса, спроектированного с учетом традиций отечественной лексической семантики и специфики русского языка, где информация о предложно-падежной реали-

зации управления предикатов и поверхностно-синтаксических свойствах других конструкций имеет особую ценность. Эта компьютерная система должна решать не только задачи словаря (ср. системе «Лексикограф», www.lexicograph.ru, бумажные словари Апресян, Палл 1982, Сазонова 2008, лексикографические проекты Азарова и др. 2004, Апресян 2008 и др.), но и представлять аннотированный корпусной материал. Этот ресурс реализует принципы гибридных систем, в которых авторитетные лексикографы видят будущее словарей (Atkins 1992, Kilgarriff et al. 2006): словарь в выходе в корпус. С одной стороны, это словарь – но с возможностью расширения иллюстраций за счет поиска в корпусе. С другой стороны, это корпус, но с экспертным отбором примеров. Наконец, это лексически ориентированный ресурс, где выборка примеров строится с ориентацией на конкретные лексемы, однако в каждом предложении разметке подлежат все предикатные слова и связанные с ними конструкции.

В результате появляется возможность проследить, как реализуются активные и пассивные валентности глаголов, реляционных имен типа *ненависть* или *пациент*, прилагательных (ср. *готов к выступлению*), наречий и т. п., как влияют на это «нелексические» грамматические конструкции, например, инфинитивная или компаративная, как это связано с порядком слов, пунктуацией и так далее. Тем самым, создаются предпосылки для прорыва в еще одной важной и малоисследованной области лингвистики – в изучении взаимодействия различных конструкций на пространстве текста.

Технологию создания и структуру такого корпусного словаря конструкций и описывает данная работа.¹

2. Формирование банка предложений

В первую очередь, необходимо собрать представительный массив данных для исследований управления и сферы действия предикатов (глаголы, имена существительные, имена прилагательные, неизменяемые части речи) и грамматики конструкций. Наиболее тщательно описанная область — глаголы, с них мы и предполагаем начать сбор данных. В дальнейшем банк примеров будет пополняться с целью сбалансированного представления имен существительных, а также прилагательных с предложным управлением. Наименее полно описаны конструкции «вне управления»: в настоящее время нет даже их сколько-нибудь полного инвентаря. Мы со-

бираемся выявить и разметить такие конструкции в собранных примерах, а затем, на последнем этапе, целенаправленно собрать материалы для составленного «словника» конструкций.

Таким образом, формирование банка предложений будет проходить в три стадии: «глаголы» — «имена» — «малый синтаксис» (прочие конструкции). Ниже мы опишем процедуру сбора данных для первой стадии.

Список «целевых» глаголов (1000 единиц) формируется по признаку их частотности, а также разнообразия и вариативности управления. На каждую лексическую единицу должно быть собрано 100 предложений НКРЯ (в дальнейшем, с учетом собранных данных, выборка может быть доведена до 200 примеров). Выборка должна быть сбалансирована по следующим показателям:

1) метатекстовые признаки:

- время создания текста (после 1950 г., первая половина XX в., XIX в., XVIII в.);
- (в идеале) не более 1 примера из 1 автора;
- жанр (художественная литература, публицистика, прочая нехудожественная литература, устная речь);

2) характеристики предложения:

- длина;
- сложность (простое, в т. ч. осложненное знаками препинания, сложносочиненное, сложноподчиненное, парцелляты);
- место предложения в тексте;

3) место предиката в предложении:

- начало — середина — конец.

Для особо частотных глаголов выборка может быть дополнительно сбалансирована относительно контекстных маркеров, а именно, слов открытых лексических классов, которые наиболее часто встречаются в предложении с данным глаголом, ср. *речь идет, подписать... договор* и т. д. Если их частотность в выборке будет не слишком велика, это послужит некоторой гарантией тематического разнообразия иллюстраций.

Банк предложений будет включать отобранные предложения и их ближайший контекст (3 предложения справа и слева от них в тексте).

3. Разметка предложения

Каждое предложение проходит несколько этапов разметки. Во-первых, в банк данных заносится информация о ближайшем контексте (ее можно просмотреть, нажав на знак $\leftarrow \dots \rightarrow$, см. пример ниже). Во-вторых, на основании мета-текстовых данных НКРЯ заполняется паспорт текста (\square автор, его пол и год рождения, время создания текста, жан-

¹ Проект выполняется в рамках программы фундаментальных исследований ОИФН РАН (2009–2011 гг.), направление «Лингвистические аспекты исследования текста».

ровая принадлежность). В-третьих, размечаются характеристики предложения: длина, сложность, место в тексте (начало, середина, конец), состав и основные характеристики лексических элементов и пунктуационных знаков.

(1) Он служил в конном корпусе Гая.

лемма: корпус	Словари	НКРЯ
слово: корпусе		
прописные: нет		
ударение: 2		
часть речи: S		
грамматика: inan,m,sg		
семантика: org		
модель управления	конструкции	
tid2485; sid000387; lid005		

[Анатолий Рыбаков. Тяжелый песок (1975–1977)] ←...→

Рис. 1. Пример предложения и паспорт слова корпус

Каждое слово в предложении также получает свой паспорт: поля паспорта заполняются автоматически, в частности, с использованием данных лексико-грамматической и лексико-синтаксической разметки НКРЯ. В дальнейшем эта информация может пополняться и редактироваться. Предусмотрена возможность посмотреть информацию о слове в словарях МАС и Ожегова-Шведовой (интернет-версии), а также получить справку об употреблении слова в НКРЯ (основной и синтаксический корпус).

Последний этап разметки предложения — определение элементов, которым будут должны быть приписаны модели управления (МУ) и словоцентричные конструкции. В примере (1) МУ приписываются глаголу *служить*, существительному *корпус*, прилагательному *конный* и предлогу *в*. Из словоцентричных конструкций следует отметить устойчивый оборот *конный корпус*.

4. Модели управления: словарь

Информация о моделях управления содержится в двух частях системы: в словаре (основные МУ) и при конкретных лексических элементах в предложении (реализация МУ в тексте). Словник словаря МУ имеет две версии: целевой (все единицы, для которых целенаправленно собирались данные) и полный (все предикаты, которые встретились в пред-

ложениях). Словарь также делится на частеречные разделы: глаголы, абстрактные и предметные имена, прилагательные, наречия, предлоги и прочее.

Пилотная версия глагольной части словаря (инвентарь МУ и их характеристики) базируется на работе Апресян, Палл 1982, в котором собрана информация о более чем 1300 глаголах. В дальнейшем список моделей управления может быть сокращен или расширен с учетом информации из других источников («Лексикограф», толковые словари, ТКС, НОСС, синтаксический корпус НКРЯ, RussNet, материалы проекта дизамбигуации глагольных значений (Толдова 2008) и др.), а также новых эмпирических данных. По мере необходимости будут формироваться словарные входы имен существительных и других частей речи.

В пределах словарной статьи все модели управления получают условное название (например, «служить в министерстве», «служить царю» и т.д.) и связаны между собой отношениями «синтаксической деривации» (Падучева 2004). При каждой МУ указывается число примеров, ассоциированных в банке предложений с данной моделью. Просмотреть эти примеры можно, перейдя по гиперссылке. Таким образом, мы получаем иерархию МУ от основных к производным и далее к частным МУ в примерах.

Способ представления МУ в словаре заимствован нами в основных чертах из системы «Лексикограф» (Падучева 2004). МУ включает в себя следующие сведения: сокращенное обозначение участника, стандартный способ морфосинтаксического выражения, синтаксический ранг, экспликация участника, семантические ограничения на заполнение валентности (см. рис. 2).²

«служить в министерстве» (примеров: 4)				
Имя	Экспликация	Морфосинтаксис	Ранг	Семантические ограничения
X	тот, кто служит	NPnom	Субъект	hum ▶
Y	место, где служат	в + NPloc	Периферия	org ▶

Рис. 2. Фрагмент статьи глагола *служить*, МУ «служить в министерстве»

² Здесь и далее используются сокращенные обозначения грамматических категорий и лексико-семантических признаков, принятые в основном корпусе НКРЯ, в частности, следующие пометы частей речи: А (имя прилагательное), S (имя существительное), V (глагол), SPRO (местоименно-существительное); лексико-семантические пометы inan («неодушевленное»), abstr («абстрактное»), hum («лицо»), org («организация»), food («пища»), famn («фамилия»), persn («имя»). Также используются принятые в грамматике составляющих обозначения групп: NP (именная группа), VP (глагольная группа), PP (предложная группа).

Экспликация — это адаптация понятия «семантическая роль» к разным типам предикатов. Как показал опыт развития FrameNet, традиционный список семантических ролей неизбежно разрастается при расширении круга размеченных глаголов и увеличении тематического разнообразия описываемых ими ситуаций, а применительно к именам существительным и прилагательным инвентарь ролей вообще не разработан. В силу этих обстоятельств было принято соглашение, что в поле «Экспликация» может быть указана семантическая роль или стоять иное пояснение, помогающее отождествить участника, например, в МУ имени *слуга* (ср. *слуга Петра*) участник Петр может быть обозначен как «хозяин» или «тот, кому служат»). Мы предполагаем, что в дальнейшем инвентарь этих маркеров будет пересмотрен и систематизирован в особую иерархическую структуру (ср. граф фреймовых элементов во FrameNet).

По желанию пользователя «имя» участника может быть представлено в двух вариантах (X, Y, Z... или 1, 2, 3...), а способы поверхностно-синтаксического выражения — в традициях грамматики зависимостей ($v + S_{loc}$, $v + N_{loc}$, ср. практику ТКС, Апресян, Палл 1982, системы «Лексикограф» и др.) или грамматики составляющих ($v + NP_{loc}$; PP, см. ниже). Поле синтаксического ранга (Субъект, Объект, Периферия, Инкорпорированный участник) присутствует только в МУ глаголов. Семантические ограничения на заполнение валентности записываются в стандарте тегов семантической разметки НКРЯ; в случае жестких лексических ограничений здесь может быть перечислен список конкретных лексем русского языка (ср. другое значение глагола *служить*, в котором объектная валентность может быть заполнена именами *служба, молебен, панихида, обедня*).

В качества образца МУ предметного имени на рис. 3. представлен фрагмент статьи слова *корпус*, где указана модель, релевантная для употребления этого существительного в примере (1).

«корпус Тухачевского» (примеров: 1) ▶			
Имя	Экспликация	Морфосинтаксис	Семантические ограничения
Y	тот, кто командует	NPgen	hum, famn ▶

Рис. 3. Фрагмент статьи имени корпус, МУ «корпус Тухачевского»

5. Модели управления: реализация в тексте

Задача разметки управления в предложении — определить конкретные способы реализации валентностей в тексте, а также указать сопутствующие сирконстанты и «новых» участников (например,

бенефицианта). В первую очередь, разметчик должен связать предикат с определенной МУ в словаре. Сделав это, он получает шаблон, включающий МУ из словаря и новые поля для заполнения, см. рис. 4.

Во вторую очередь, разметчик должен указать в предложении элементы, соответствующие каждому участнику. Шаблон заполняется в двух стилях: грамматики составляющих (ГС) и грамматики зависимостей (ГЗ). Например, заполняя шаблон глагола *служить* (см. пример выше) в стиле ГС, разметчик должен выделить в предложении словосочетание «в конном корпусе Гая» и связать его с полем <phrase>; а заполняя аналогичную позицию в стиле ГЗ, он должен выделить два элемента («в» и вершину именной группы, «корпусе») и связать их с полем <word>.

На следующем этапе в действие вступает программа автозаполнения, которая вносит информацию в остальные поля шаблона на основе имеющейся морфологической и семантической информации в паспорте лексемы. Затем разметчик может вручную откорректировать данные в шаблоне.

В случае, если в предложении присутствуют сирконстанты предиката или элементы, которые обозначают дополнительных (необязательных) участников ситуации, разметчик должен расширить шаблон и добавить данные о новых членах. Напротив, если кто-либо из обязательных участников не упомянут в предложении, следует указать причину его опущения. В частности, это может быть пассивизация; опущение субъекта в нефинитных формах; парцелляция; участник может быть известен из контекста (упомянут в предконтексте или постконтексте); управление может быть передано другому предикату (конструкция контроля) и т. д. Точно так же следует указывать конструкцию, «ответственную» за изменение синтаксического ранга участника.

Рассмотрим пример (2), в котором модели управления двух глаголов, *собрать* и *поесть*, реализуются нестандартно. Субъект ситуации, обозначаемой глаголом *собрать*, известен и упомянут как обращение в предшествующем контексте (*Оля*), однако он не может быть реализован при глаголе в форме императива. Пациент обозначается инфинитивной группой — это развитие переходной МУ (*собрать корзинку с едой*). Кроме того, в ситуации появляется бенефициант, который вводится дативной группой (дитранзитивная конструкция), а также указана цель действия (обстоятельство *в дорогу*).

Субъект ситуации, обозначаемой глаголом *поесть*, известен и упомянут в предложении (*нам*), однако он синтаксически зависит от другого глагола (*собрать*) и по правилам грамматики не может быть реализован при глаголе в форме инфинитива. Пациент у *поесть* генерический (любая еда) — это свойство конструкции «собрать поесть». Вообще, если

(1) Он служил в конном корпусе Гая. [Анатолий Рыбаков. Тяжелый песок (1975–1977)] ←...→ ▾

«служить в министерстве»		lid004	(примеров: 4) ▶	
Имя	Экспликация	Морфосинтаксис	Ранг	Семантические ограничения
X	тот, кто служит	NP _{nom}	Субъект	hum ▶
группа	<phrase>	<style NPnom>	<input type="text"/>	<semantic tags>
вершина	<word>	<style Snom>	<rank>	<semantic tags>
Y	место, где служат	в + NP _{loc}	Периферия	org ▶
группа	в конном корпусе Гая	PP: в + NPloc	Стандартный <input type="text"/>	org
вершина	в корпусе	в + Sloc	Периферия	org
+				

Рис. 4. Пример шаблона с частично заполненными полями

(2) Оля! Собери нам поесть в дорогу. [Фазиль Искандер. ... (...)] ←...→ ▾

«собрать поужинать»		собрать V,2p,act,imper,pf,sg	lid001	(примеров: 14) ▶	
Имя	Экспликация	Морфосинтаксис	Ранг	Семантические ограничения	
X	агенса	NP _{nom}	Субъект	hum ▶	
группа			Императив <input type="text"/>		
вершина			Нет	hum	
Y	то, что собирают	VP _{inf}	Периферия	inan ▶	
группа			Стандартный <input type="text"/>		
вершина			Периферия		
Z	бенефициант	NP _{dat}	Периферия	hum ▶	
группа	нам	NPdat	Дитранзитив <input type="text"/>		
вершина	нам	SPROdat	Периферия	hum persn	
W	цель		Сирконстант	▶	
группа	в дорогу	в + NPacc	Генерический <input type="text"/>		
вершина	в дорогу	в + Sacc	Нет	abstr	
+					

(2) Оля! Собери нам **поесть** в дорогу. [Фазиль Искандер. ... (...)] ←...→ ▾

«есть рыбу»		поесть V,act,inf,pf,v	lid003	(примеров: 2) ▶	
Имя	Экспликация	Морфосинтаксис	Ранг	Семантические ограничения	
X	агенса	NP _{nom}	Субъект	hum ▶	
группа			Инфинитив <input type="text"/>		
вершина			Нет	hum	
Y	пациенс	NP _{acc}	Периферия	food ▶	
группа			Генерический <input type="text"/>		
вершина			Нет	food	
+					

Рис. 5. Реализация МУ у глаголов собрать и поесть

бы пациенс ситуации *поест* был специфицирован, он бы был «захвачен» глаголом *собрать*, ср. *Вот, собрала вам пирожков поест*; однако чаще объект в этой конструкции опускается. Таким образом, конструкция «собрать поест», с одной стороны, входит как элемент в периферийную МУ глагола *собрать*, а с другой стороны, является частным случаем конструкции контроля объекта, обладая в то же время собственными синтаксическими свойствами.

6. Словарь конструкций

Понятие «конструкции» мы понимаем максимально широко, в традициях основного направления Грамматики Конструкций (Fillmore et al. 1988, Goldberg 1995, см. также обзор Кузнецова 2007). В частности, конструкциями мы называем:

- тривиальные синтаксические конструкции: A + S, на + СПРО и т. д.;
- общие синтаксические конструкции: сочинение и подчинение, перестановки порядка слов, конструкции повтора, пассив, генитив при отрицании, безличная конструкция, дитранзитивная конструкция (*испекла Пете пирог*), локативная трансформация (*грузить баржу лесом*), аппозитивная адъективная (*валялся пьяный*), компаративная (*гиены трусливее зайцев*) и т. д. – многие из этих конструкций основаны на «трансформациях» моделей управления;
- модели управления глаголов, предикатных имен (*ненависть*), реляционных предметных имен (*пациент*), имен прилагательных (ср. *готов к выступлению*), наречий, предлогов и т. д.;
- неоднословные лексические элементы (обороты): например, *иметь в виду*; *несмотря на*, вводные обороты, конструкции с лексическими функциями (*оказывать влияние*), а также их «малый синтаксис»;
- периферийные словоцентричные конструкции, например, сериальная (*сизу смотрю*), конструкции типа *гулять так гулять, взял да и помер* и т. д.

Как уже было указано ранее, мы хотим учесть опыт развития системы FrameNet и собрать данные о конструкциях последнего типа, т. е. ориентиро-

ванных вокруг отдельных лексических элементов или фразем. Источниками словаря являются Грамматика 1980, Шведова 1960, Золотова 1980, лингвистические описания отдельных конструкций, а главное, конструкции, обнаруженные при разметке предложений. В общих чертах структура словарного входа сходна со словарем МУ, представляя перечень структурных моделей конструкции. В каждой модели перечисляются элементы конструкции, как постоянные, так и переменные, и указаны их лексические, грамматические и семантические ограничения. В отличие от МУ, отдельное поле посвящено порядку слов и разрывности несущих элементов конструкции.

7. Заключение

Предлагаемая система должна иметь несколько индексов (словарь лексем, словарь типов моделей управления, словарь конструкций). Выбрав лексический вход, можно увидеть набор МУ данного слова, получить список конкретных поверхностно-синтаксических вариантов ее реализации и далее просмотреть аннотированный корпусной материал. Каждая строка в МУ также содержит гиперссылку, которая дает возможность получить список конкретных вариантов реализации данной валентности в предложениях.

Кроме того, корпусный словарь должен предоставлять возможности поиска. Пользователь может задать шаблон модели управления или конструкции, оговорить ограничения на лексические, грамматические и лексико-семантические признаки; ограничить длину предложения или его тип (сложноподчиненные и т. д.). В системе можно искать без учета порядка слов, или же, напротив, задать ограничения на линейный порядок элементов. В результатах запроса будет содержаться также информация о частотности того или иного явления в материалах корпусного словаря.

Естественное ограничение данной системы в том, что здесь нельзя проследить распределение моделей управления в пределах целого текста. Впрочем, в дальнейшем мы не исключаем возможности составления экспериментального корпуса со сплошной фреймнет-разметкой.

Литература

1. Азарова И. В., Синопальникова А. А., Яворская М. В. Принципы построения wordнетезауруса RussNet // Кобозева И. М., Нариньяни А. С., Селегей В. П. (ред.), Компьютерная лингвистика и интеллектуальные технологии: труды международной конференции «Диалог'2004». М.: 2004. С. 542–547.
2. Апресян Ю. Д. О проекте активного словаря (АС) русского языка // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 4–8 июня 2008 г.). Вып. 7 (14). М.: РГГУ, 2008. С. 23–31.
3. Апресян Ю. Д., Палл Э. Русский глагол — венгерский глагол. Управление и сочетаемость. Будапешт, 1982.
4. Грамматика 1980: Русская грамматика. М., 1980.
5. Золотова Г. А. Синтаксический словарь. Репертуар элементарных единиц русского синтаксиса. М., 1980.
6. Кузнецова Ю. Л. Грамматика Конструкций. Обзор // НТИ. Сер. 2, № , 2007?.
7. МАС: Евгеньева А. П. Словарь русского языка: В 4 т. 2-е изд. М., 1981–1984 г.
8. НОСС: Апресян Ю. Д. и др. Новый объяснительный словарь синонимов. М., 2003.
9. Ожегов С. И., Шведова Н. Ю. Толковый словарь русского языка. М., 1992.
10. Падучева Е. В. Динамические модели в семантике лексики. М., 2004.
11. Падучева Е. В., Кустова Г. И., Розина Р. «Лексикограф». <http://Lexicograph.ru>.
12. Сазонова И. К. Толково-грамматический словарь русских причастий. М., 2008.
13. Толдова С. Ю., Кустова Г. И., Ляшевская О. Н. Семантические фильтры для разрешения многозначности в Национальном корпусе русского языка: глаголы // Труды международной конференции «Диалог 2008». М., 2008. С. 522–529.
14. ТКС: Мельчук И. А., Жолковский А. К. Толково-комбинаторный словарь современного русского языка. Вена, 1984.
15. Шведова Н. Ю. Очерки по синтаксису русской разговорной речи. М., 1960.
16. Atkins B. T. Tools for computer-aided corpus lexicography: the Hector project // Papers in Computational Lexicography: Complex'92, F. Kiefer, G. Kiss and J. Pajsz (eds.) Hungarian Academy of Sciences, Budapest, 1992. P. 1–60.
17. Fillmore C., Kay P., O'Connor K. T. Regularity and idiomaticity in grammatical constructions: the case of let alone // Language, №64, 1988. С. 501–538.
18. Fillmore Ch. Border conflicts: FrameNet meets Construction Grammar // EURALEX 2008.
19. Goldberg A. Constructions: A Construction Grammar Approach to Argument Structure. Chicago: University of Chicago Press, 1995.
20. Johnson C., Fillmore C., Petruck M, Baker C., Ellsworth M., Ruppenhofer J. and Wood E. FrameNet: Theory and Practice. [Electronic resource]. Mode of access: <http://www.icsi.berkeley.edu/framenet>.
21. Kilgarriff A., Rundell M. and Uí Dhoonchadha E. Efficient Corpus Creation for Lexicography: building the New Corpus for Ireland' // Language Resources and Evaluation, 40, 2006. P. 127–152.
22. Kipper K., Korhonen A., Ryant N., Palmer M. Extending VerbNet with novel verb classes // Fifth International Conference on Language Resources and Evaluation (LREC 2006). Genoa, Italy. June, 2006. См. также <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>.
23. Palmer M., Gildea D., Kingsbury P. The Proposition Bank: A corpus annotated with semantic roles // Computational Linguistics Journal, 31:1, 2005. См. также <http://verbs.colorado.edu/~mpalmer/projects/ace.html>.
24. Ronald D., Jurafsky D., Menn L., Gahl S., Elder E., Riddoch C. Verb subcategorization frequency differences between business-news and balanced corpora: the role of verb sense // ACL Workshop on Comparing Corpora, 2000.