

Разработка квантитативной грамматики предложных конструкций

1. Предпосылки квантитативного описания предложных конструкций

1.1 Трактовка данных корпусной статистики

При разработке квантитативной грамматики предложных конструкций мы исходили из нескольких теоретических предпосылок. Предложная конструкция в русском языке предполагает не только использование предлогов, которые скорее относятся к грамматическому плану описания языка, но и падежных форм зависимых слов, обычно имен существительных, что также отсылает нас к грамматическому описанию. В трактовке особенностей грамматического значения мы отталкивались от позиции, сформулированной М.И.Стеблин-Каменским [Стеблин-Каменский 1974]: (а) степень детализации грамматических значений является существенно менее конкретной, чем лексических значений, (б) дифференциация грамматических значений не описывается явными логическими параметрами, что выражается, например, в приписывании одушевленности таким словам как *покойник*, *мертвец* или *кукла* и т. п. Довольно часто грамматические значения вообще слабо соотносятся с какой-либо осмысленной характеристикой, например, значение русских падежей. Такие грамматические категории, объединяющие группы значений, называют формальными. Однако важно, что даже в таких случаях возможно выделение некоторых оппозиций, как это было продемонстрировано Р.Якобсоном [Якобсон 1985] применительно как раз к значениям русских падежей и глагольным категориям. В описании оппозиций Р.Якобсон следовал собственной логике: возможности нейтрализации некоторого маркированного грамматического значения, так называемой признаковой категории. То есть грамматическая оппозиция не противопоставляет маркированный признак «А» его отрицанию «не А», как это обычно трактуют многие исследователи, а маркированному признаку «А» противопоставлен признак частичной нейтрализации противопоставления «не А или А». Для нас это дает довольно четкий принцип выделения грамматических оппозиций на основе корпусной статистики реализации однотипных грамматических значений: их соотношение будет стремиться к пропорции «1 : 1,5», т. е. признаковая категория встречается реже в текстовых корпусах, поскольку это же значение может быть выражено без прояснения вторым членом оппозиции. Описанные Р.Якобсоном оппозиции падежных значений по направленности, локализованности, периферийности хорошо подтверждаются в корпусной статистике. Мы используем идею «признаковых категорий» при анализе значений предложно-падежных конструкций, поскольку чисто логические сопоставления значений невозможны в силу грамматического характера этих конструкций, так называемой «непоследовательности» в употреблении предложно-падежных конструкций.

Корпусная статистика позволяет проследить зоны реализации грамматических и лексических значений. Общеизвестный тезис о том, что к сфере грамматического относится то, что повторяется в открытых классах слов, например падежные флексии существительных и т. п., позволяет оценить диапазоны корпусных значений собственно грамматического в интервале, превышающем 300 ipm (instances per million — число примеров на миллион употреблений в корпусе). Напротив наиболее частотные лексические значения являются базовыми понятиями [4; 5], они представлены в корпусе в диапазоне 100–50 ipm в зависимости от частеречной принадлежности лексемы: от существительных, глаголов, наречий к прилагательным. Зона среднечастотных лексических единиц может частично пересекаться с предыдущим интервалом 55–10 ipm. Низкочастотные лексические значения имеют единичные корпусные частоты или даже дробные. Лексико-грамматические явления характерны для подклассов слов в рамках частей речи, грамматической категории и т. п., в плане корпусной статистики они занимают переходный интервал от 300 до 40 ipm.

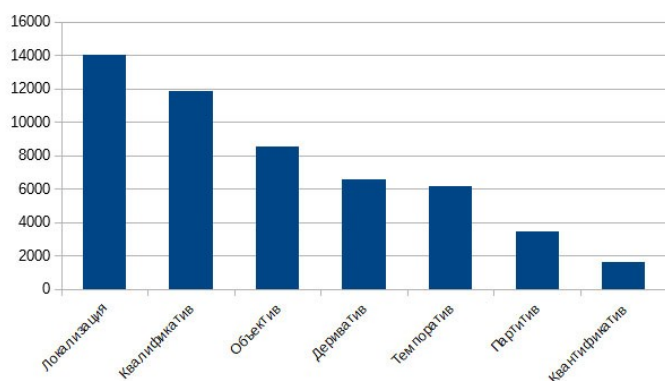
Значения корпусных частот для лексических единиц зависят от тематического наполнения корпуса, собственно грамматические явления должны быть стабильными, однако лексико-грамматические единицы могут варьироваться в зависимости от баланса текстов определенных функциональных стилей в корпусе.

Вычисление параметров корпусной статистики предполагает определенную методику экстраполяции, поскольку выявление типовых противопоставлений грамматических и лексических значений на миллионных выборках трудно осуществимо. Мы использовали методику анализа случайной выборочной совокупности контекстов из корпусов русских текстов общей тематики, в рамках выделенной совокупности производится дифференциация исследуемых параметров. Важным является случайный характер отбора контекстов в выборку, что не всегда реализовано в больших корпусах, например в НКРЯ. При возможности сохранить всю совокупность контекстов, удовлетворяющих критерию поиска, можно использовать генераторы случайных чисел, создавая собственную выборку нужного размера. Однако не всегда реализована и эта функция сохранения всех контекстов. Эксперименты с объемом выборочной совокупности показали, что для высоко- и среднечастотных явлений подходят выборки от 50 контекстов, для низкочастотных явлений объем выборки необходимо повышать до 100 или даже 1000 контекстов. Если доли грамматических и лексических значений в рамках случайной выборки превосходят 5%, то они хорошо коррелируют с данными, которые получаются при увеличении выборки на порядок. Доля частотности какого-либо параметра менее 5% в выборке является неустойчивой, не масштабируемой, для таких значений необходимо расширение выборочной совокупности, более того необходимо понимать, что эти значения будут отличаться от корпуса к корпусу. Важно подчеркнуть, что конкретное значение корпусной частотности является менее значимым, чем соотношение частот, поскольку в оценке частотности присутствует потенциальная 15% ошибка экстраполяции. Однако прогнозируемое полуторное различие частот как реализация значений признаковых и непризнаковых категорий.

1.2 Интерпретация компонентов предложно-падежной конструкции

В классической грамматике зависимостей предлоги трактуются как слова, синтаксически доминирующие над присоединяемой предложной формой имени, что хорошо иллюстрируется конструкциями с непервообразными предлогами, от которых можно задать вопрос к падежной форме: *идти в направлении* (чего?) *леса*. Конструкция с первообразными предлогами не всегда поддается вопросительной модификации: *варить на* (*чем?) — *на водяной бане*. Довольно хорошо подвергаются вопросительной модификации разные валентные употребления глаголов: *смотреть на* (кого?) *ребенка* — *смотреть за* (кем?) *ребенком*, но в этом случае нам помогает различие в глагольных значениях, которое относит первое употребление к типу активного перцептивного восприятия, которое предполагает валентность на «объект восприятия», а во втором случае это более сложное действие по обеспечению необходимых для жизни человека ресурсов заботиться о ком/чем, т. е. «объект действия». Для валентных структур русских глаголов важное значение имеет морфемная (обычно приставочная) модификация глагольных основ, например в случае повторения глагольного префикса совпадающим предлогом или его синонимичным заместителем: *войти в дом, зайти за дом, подойти к дому, выйти из дома* и т. д. Регулярно предложные конструкции проходят лексикализацию и приобретают наречный статус или фразеологизируются, функционируя как неделимые единства: *за бесценюк* ‘даром’, *за версту* ‘издалека, издали’ и т. п.

Поскольку не всегда возможно интерпретировать предложно-падежную конструкцию аддитивно «значение предлога» + «значение падежной формы», подход, предложенный Г.А.Золотовой



[Золотова 2011], при котором предложно-падежная конструкция рассматривается как синтаксема — минимальная грамматическая конструкция, выражающая определенное значение. Синтаксеммы могут быть независимыми, т. е. относительно самостоятельными, но чаще зависимыми, когда они присоединяются в виде целых блоков к знаменательным словам, чаще всего глаголам. Количество синтаксем в словаре Г.А.Золотовой составляет несколько десятков. Одна и та же синтаксема может быть выражена разными предлогами в сочетании с разными падежными формами, которые могут обеспечивать синонимичную замену при некоторых глаголах, но в общем случае они не синонимичны. Выделенные синтаксеммы в словаре Г.А.Золотовой по большей части совпадают с категориями семантических ролей (аргументов) функциональных грамматик: локатив, темпоратив, тематив, инструментив, медиатив и проч. Однако контекстный материал, приводимый в словаре, довольно ограниченный, что приводит к тому, что выделяются отдельные типы, которые с точки зрения корпусной частотности не носят грамматического характера.

2. Онтологическая схема квантитативного описания предложных конструкций

Идея синтаксеммы принята в нашем квантитативном описании как определенный средний уровень дифференциации значений предложно-падежных конструкций, при этом несколько однотипных синтаксем, обладающих общими свойствами, объединяются в семантическую рубрику, а с другой стороны дифференцируются на уровне более частных значений, так называемых субсинтаксем, у которых может быть более одной реализации. Таким образом онтологическая схема описания значений предложно-падежных конструкций имеет три уровня: семантические рубрики, синтаксеммы, субсинтаксеммы, для которых заданы определенные уровни корпусных частот, представленные в Таблице 1. В первом столбце приведены 7 семантических рубрик, в скобках указано число синтаксем и суммарное число реализаций субсинтаксем для каждой из них. Рубрики упорядочены сверху вниз в соответствии с убыванием их корпусной частотности, Вслед за названием рубрики идет перечень синтаксем в соответствии с убыванием их корпусной частотности, в скобках указано число реализаций для каждой синтаксеммы (см. описание рубрик в приложении)

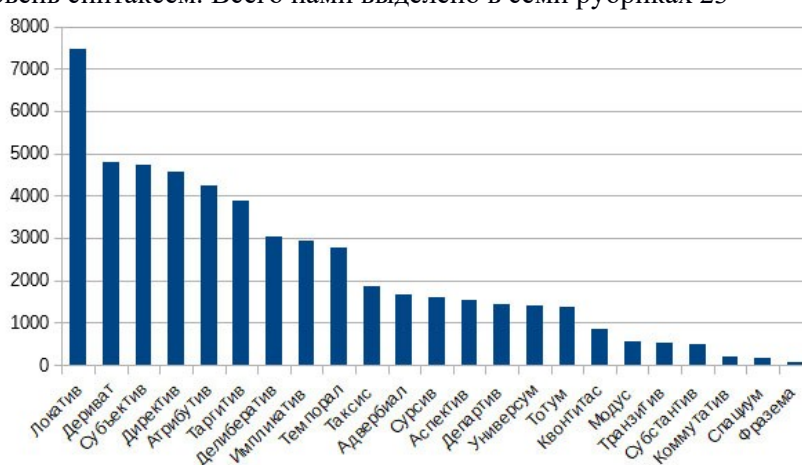
Таблица 1. Распределение корпусных частот по семантическим рубрикам, количество синтаксем и субсинтаксем.

Рубрика	Синтаксема	Синтаксема	Синтаксема	Синтаксема
Локализация (4, 19)	Локатив (6)	Директив (5)	Департив (5)	Транзитив (3)
14005 ipm	7465 ipm	4575 ipm	1432 ipm	533 ipm
Квалификатив (3, 9)	Субъектив (2)	Атрибутив (3)	Имплицатив (4)	
11866 ipm	4716 ipm	4222 ipm	2928 ipm	
Объектив (3, 16)	Таргитив (5)	Делибератив (8)	Сурсив (3)	
8515 ipm	3868 ipm	3037 ipm	1610ipm	
Дериватив (3, 7)	Дериват (3)	Адвербиал (2)	Фразема (2)	
6520 ipm	4780 ipm	1660 ipm	80 ipm	
Темпоратив (3, 11)	Темпорал (4)	Таксис (3)	Аспекттив (4)	
6155 ipm	2765 ipm	1872 ipm	1518 ipm	
Партитив (4, 4)	Универсум (1)	Тотум (1)	Субстантив (1)	Спациум (1)
3463 ipm	1413 ipm	1383 ipm	501 ipm	166 ipm
Квантификатив (3, 6)	Квонтитас (2)	Модус (3)	Коммутатив (1)	
1621 ipm	861 ipm	562 ipm	198 ipm	

На графике 1 приведено распределение корпусных частот для рубрик.

Предложно-падежная конструкция может синкретично передавать значения не одной рубрики (синтаксемы), а двух. Стандартно это происходит в случае переносных употреблений конструкций в некоторых контекстах. Регулярное пересечение синтаксем характерно для рубрики Дериватив, в которой описано, в частности, образование производных предлогов, некоторые из которых обладают практически грамматической частотностью и попадают в описание других рубрик. Аналогичная ситуация наблюдается между рубриками Квантификатив и Темпоратив. Среднее значение частоты для рубрик 7500 ipm. Превосходят среднее значение три рубрики: Локализация, Квалификатив и Объектив. Две рубрики существенно меньше среднего значения: Партидив и Квантификатив. Две рубрики попадают примерно в область средних значений: Дериватив и Темпоратив, они маркируют на графике «перегиб» кривой убывания параметров корпусной частотности.

Определение значения предложно-падежной конструкции с точки зрения ее «рубрикации» является так называемой «coarse-grane» (обобщенной) характеристикой. Переход к более детальной классификации значения — это уровень синтаксем. Всего нами выделено в семи рубриках 23



синтаксемы. Распределение их частот показано на графике 2. Среднее значение корпусной частоты для синтаксем равно 2267 ipm.

По распределению частот четко выделяется синтаксема Локатив, которую обычно довольно успешно идентифицируют при автоматическом анализе предложных конструкций в языках индоевропейской группы. Превышает среднее значение частотности две группы. Одна состоит из пяти синтаксем: Дериват, Субъектив, Директив, Атрибутив, Таргетив, вторая — из трех синтаксем: Делибератив, Импликатив и Темпорал. Ниже среднего значения частотности, но в зоне грамматических значений находятся семь синтаксем: Таксис, Адвербиал, Сурсив, Аспектив, Департив, Универсум, Тотум. Начиная с синтаксемы Квантитас идет зона лексико-грамматических значений, замыкает перечень синтаксема Фразема, имеющая лексический характер — формирование устойчивых (фразеологизированных) конструкций. В Таблице 2 представлено распределение наиболее частотных синтаксем по рубриками.

Рубрика	Суммарная частота (ipm)	Частотные синтаксемы (ipm)	Частотные синтаксемы (ipm)	Частотные синтаксемы (ipm)	Частотные синтаксемы (ipm)
Локализация	14005	7465	4575		
Квалификатив	11866		4716	4222	2928
Объектив	8515			3868	3037
Дериватив	6520		4783		
Темпоратив	6155				2765
Партидив	3463				

Рубрика	Суммарная частота (ipm)	Частотные синтаксемы (ipm)	Частотные синтаксемы (ipm)	Частотные синтаксемы (ipm)	Частотные синтаксемы (ipm)
Квантификатив	1621				

Для каждой из синтаксем в Приложении приведены списки наиболее частотных реализаций при помощи первообразных и непервообразных предлогов.

В таблице 3 приводится отсортированный сводный список реализаций первообразных предлогов в выделенных субсинтаксемах. Всего в перечне 121 пара «предлог-субсинтаксема», общая суммарная частота 43000 ipm, среднее значение корпусной частоты 360 ipm. Без учета 15 лексически связанных предложных конструкций с первообразными предлогами, для которых среднее значение существенно ниже 43 ipm, грамматическая корпусная частота равна 400 ipm.

Предложная конструкция в таблице задается частотным вариантом предлога: запись “о” подразумевает также его варианты “об” и “обо”. Цифра в виде подстрочного индекса указывает на присоединяемую предлогом падежную форму через порядковый номер в стандартной шестичленной парадигме. При наличии колебаний приводятся два варианта, например “по_{3/4}”. Полу жирным выделены частоты выше среднего.

Предложная конструкция	Субсинтаксема	Частота в ipm	Предложная конструкция	Субсинтаксема	Частота в ipm
<i>без₂</i>	Атрибутив_1	315	<i>за₄</i>	Имплицатив_3	250
	Имплицатив_4	203		Аспектв_2	230
	Субъектив_2	75		Темпорал_3	210
<i>в₄</i>	Директив_1	2500		Квонтитас_1	116
	Темпорал_2	920		Директив_4	115
	Дериват_1	650		L Имплицатив_1	110
	Квонтитас_1	70		Делибератив_6	60
	Коммутатив_1	67		Модус_1	29
	L Атрибутив_1	56		L Атрибутив_3	10
	L Делибератив_2	50		<i>за₅</i>	Локатив_4
	Фразема_1	3	Темпорал_4		110
<i>в₆</i>	Локатив_1	4000	Фразема_1		5
	Дериват_1	1210	<i>из-за₂</i>	Имплицатив_3	78
	Темпорал_1	1150		Департив_4	40
	Атрибутив_1	965	<i>из-под₂</i>	Департив_5	22
	Таргетив_3	850		Тотум_1	7
	Субстантив_1	150	<i>из₂</i>	Универсум_1	1172
	Фразема_1	55		Департив_1	660
	L Делибератив_2	50		Тотум_1	439
	Адвербиал_2	170		Субстантив_1	300
	<i>вне₂</i>	Имплицатив_4		14	Сурсив_3
<i>для₂</i>		Субъектив_1		890	Спациум_1
	Имплицатив_1	765		L Имплицатив_2	14

Предложная конструкция	Субсинтаксема	Частота в ipm		Предложная конструкция	Субсинтаксема	Частота в ipm
	Таргетив_4	720		κ_3	Таргетив_1	1200
$до_2$	Модус_1	298			Директив_3	575
	Таксис_3	266			Делибератив_3	460
	Аспектв_1	236			Имплицатив_1	395
	Директив_3	65			Аспектв_1	122
					Адвербиал_2	120

Предложная конструкция	Субсинтаксема	Частота в ipm		Предложная конструкция	Субсинтаксема	Частота в ipm
$между_5_{pl}$	Субъектив_2	156		$по_5$	Аспектв_4	180
	Локатив/Директив	61		$под_4$	Директив_5	70
$на_4$	Директив_2	1250		$под_5$	Локатив_5	145
	Делибератив_2	670			Атрибутив_1	130
	Квонтитас_1	258			Имплицатив_3	87
	L Темпорал_4	68			Атрибутив_3	78
	Коммутатив_1	34			Аспектв_1	25
	Аспектв_2	24		$при_6$	Таксис_2	675
			Спациум_1		10	
$на_6$	Локатив_2	2500		$про_4$	Делибератив_4	200
	L Делибератив_2	50		$против_2$	Делибератив_5	75
	L Субстантив_1	32			Локатив/Директив	14
	L Имплицатив_2	30		$ради_2$	Имплицатив_1	38
	L Квонтитас_1	19		c_2	Департив_2	410
	L Атрибутив_2	165			Аспектв_3	170
			Модус_3		77	
			L Спациум_1		19	
$над_5$	Локатив/Директив	82		c_5	Субъектив_2	1240
	Делибератив_7	47			Таргетив_2	1000
$о_6$	Делибератив_1	1455			Тотум_1	844
					Универсум_1	130
	Сурсив_1	800			L Субъектив_1	105
	Департив_3	300			Квонтитас_1	55
	Аспектв_3	110			L Субстантив_1	19
	Тотум_1	93		$сверх_2$	Модус_1	15
L Имплицатив_2	62		$среди_2$	Локатив/Директив	45	
Модус_3	10			$у_2$	Субъектив_1	2300
$перед_5$	Локатив/Директив	94			Сурсив_2	610
	Таксис_3	74				
$по_3$	Имплицатив_2	510				
	Транзитив_1	360				
	Локатив_6	110				

Предложная конструкция	Субсинтаксема	Частота в ipm		Предложная конструкция	Субсинтаксема	Частота в ipm
	L Атрибутив_2	70			Локатив_3	25
<i>по_{3/4}</i>	Квонтитас_1	146		<i>через₄</i>	Таксис_1	300
	L Коммутатив_1	27			Атрибутив_2	163
<i>по₄</i>	Аспектв_2	180			Транзитив_2	135

Девять конструкций с первообразными предлогами представляют суперпозицию двух предложных конструкций, их корпусные частоты невелики, однако они на порядок выше случайного следования двух предложных конструкций друг за другом. В Таблице 4 приводится их перечень.

Предложная конструкция	Субсинтаксема	Частота в ipm		Предложная конструкция	Субсинтаксема	Частота в ipm
<i>за₄... до₂</i>	Коммутатив_1	40		<i>с₂... до₂</i>	Модус_2	24
<i>из₂... в₄</i>	Фразема_2	15			Аспектв_2	22
<i>от₂... до₂</i>	Модус_2	109		<i>с₂... на₄</i>	Фразема_2	2
	Аспектв_2	14			Аспектв_2	24
				<i>через₄... после₂</i>	Коммутатив_1	30

В Таблице 5 приводится список реализаций субсинтаксем при помощи производных предлогов, приведенных в описаниях Приложения. Всего в таблице 54 пары «предлог-синтаксема», суммарная частота 2712 ipm, средняя частота 48 ipm; частоты выше среднего выделены полужирным.

Предложная конструкция	Субсинтаксема	Частота в ipm		Предложная конструкция	Субсинтаксема	Частота в ipm
<i>благодаря₃</i>	Атрибутив_2	71		<i>возле₂</i>	Локатив_3	66
<i>в адрес₂</i>	Таргетив_5	18			Атрибутив_3	29
<i>в виде₂</i>	Атрибутив_1	103		<i>исходя из₂</i>	Имплицатив_2	14
<i>в зависимости от₂</i>	Имплицатив_2	42			Имплицатив_2	18
<i>в качестве₂</i>	Атрибутив_1	25		<i>на основании₂</i>	Имплицатив_2	18
<i>в конце₂</i>	Аспектв_1	41			Имплицатив_2	26
<i>в начале₂</i>	Аспектв_3	27		<i>на основе₂</i>	Имплицатив_2	26
<i>в области₂</i>	Атрибутив_1	20			Имплицатив_2	26
<i>в отношении₂</i>	Таргетив_5	44		<i>насчет₂</i>	Делибератив_8	30
<i>в период₂</i>	Аспектв_2	30			Имплицатив_4	26
<i>в процессе₂</i>	Атрибутив_1	28		<i>независимо от₂</i>	Имплицатив_4	26
<i>в результате₂</i>	Имплицатив_3	50			Имплицатив_4	50
<i>в связи с₅</i>	Имплицатив_2	53		<i>около₂</i>	Квонтитас_2	142
<i>в случае₂</i>	Имплицатив_3	60			Локатив_3	16
<i>в составе₂</i>	Атрибутив_1	17		<i>по отношению к₃</i>	Таргетив_5	36
<i>в состоянии₂</i>	Атрибутив_1	50			Делибератив_8	40
<i>в сфере₂</i>	Атрибутив_1	55		<i>по поводу₂</i>	Имплицатив_3	10
<i>в течение₂</i>	Темпорал_3	140			Атрибутив_3	18
<i>в условиях₂</i>	Атрибутив_1	34		<i>подобно₃</i>	Атрибутив_3	29
<i>в форме₂</i>	Атрибутив_1	20			Имплицатив_3	8
				<i>после₂</i>	Таксис_1	500
					Атрибутив_2	19
				<i>посредством₂</i>	Атрибутив_2	19
					Атрибутив_2	27
				<i>при помощи₂</i>	Имплицатив_3	8
					Атрибутив_2	51
				<i>при условии₂</i>	Имплицатив_3	8
					Атрибутив_2	51
				<i>путем₂</i>	Атрибутив_2	51
					Атрибутив_2	51

Предложная конструкция	Субсинтаксема	Частота в ipm	Предложная конструкция	Субсинтаксема	Частота в ipm
<i>в ходе</i> ₂	Атрибутив_1	46	<i>с помощью</i> ₂	Атрибутив_2	98
<i>в целях</i> ₂	Импликатив_1	10	<i>с целью</i> ₂	Импликатив_1	24
<i>вместе с</i> ₅	Универсум_1	96	<i>сквозь</i> ₄	Транзитив_3	38
	Субъектив_2	80	<i>сквозь</i> ₄	Атрибутив_2	20
<i>вне зависимости от</i> ₂	Импликатив_4	5	<i>совместно с</i> ₅	Субъектив_2	11
<i>во время</i> ₂	Темпорал_4	138	<i>спустя</i> ₂	Таксис_1	57
<i>во главе с</i> ₅	Субъектив_2	15			
	Универсум_1	15			

2.1 Критерии выделения компонентов онтологической схемы анализа предложных конструкций

Дифференциация значений предложных конструкций позволила выработать трехуровневую онтологическую схему анализа предложных конструкций.

(1) Обобщенный уровень систематизации значений по наиболее общим параметрам семантических параметров предложения позволил выделить шесть базовых семантических рубрик: Локализация, Темпоратив, Объектив, Квалификатив, Квантификатив, Партитив и седьмую рубрику Дериватив, в которой предлог в сочетании с другими знаменательными лексемами (существительными, наречиями и глагольными деепричастиями) формирует новые предложные единицы или входит в структуру других составных единиц, таких как наречия или фразеологизмы. Последняя рубрика отчасти выпадает из собственно смысловой последовательности, однако игнорирование единиц такого типа приведет к неполноте даже обобщенной классификации, кроме того ограничит схему анализа закрытым перечнем базовых позиций, что абсолютно не соответствует природе языка, в котором одновременно присутствует «результат» и «процесс», т. е. эта область включает предложные конструкции, которые в перспективе могут менять внутреннюю конфигурацию семантических рубрик. Некоторые компоненты рубрики Дериватив уже грамматически вошли в структуру собственно семантических рубрик, поэтому между этой рубрикой и другими, например Темпоративом, есть зона пересечения в корпусной статистике. Корпусная статистика выделенных рубрик позволила сформулировать понимание статистики собственно грамматической реализации: от 14 тысяч до полутора тысяч ipm (instances per million — число примеров на миллион токенов в корпусе). Важным является то, что корпусная частота рубрик создает ряд геометрически убывающей прогрессии, хотя и с определенными нарушениями за счет совмещений некоторых реализаций: 14k, 12k, 9k, 7k, 6k, 3k, 2k ipm.

(2) Следующий уровень — более детализированных единиц в пределах рубрики, для которых мы использовали термин Г.А.Золотовой синтаксема. В нашем понимании это предложная конструкция, соотносимая с семантической рубрикой и представляющая определенное единство морфосинтаксической и лексической спецификации главного слов, присоединяющего предложную конструкцию, группы предлогов, каждый из которых присоединяет конкретную падежную форму существительного определенного семантического класса. Таким образом синтаксема включает несколько предложных конструкций, синонимичных или частично синонимичных, но не антонимичных. Лексические спецификации главных и зависимых слов в синтаксеме частично пересекаются. Всего нами выделено в семи рубриках 23 синтаксемы (от 3 до 4 для каждой рубрики). Среднее значение корпусной частоты для синтаксем равно 2267 ipm. Практически все синтаксемы по корпусной статистике находятся в зоне грамматических значений от 8k до 1k ipm, за исключением трех, которые попадают в зону лексико-грамматических значений. Синтаксемы в рамках рубрики также образуют ряд убывающей прогрессии.

(3) Практически каждая из синтаксем за исключением лексико-грамматических имеет ряд более детализированных вариантов — субсинтаксем, которые в свою очередь могут иметь один или несколько вариантов реализации предложной конструкцией.

Приведем наиболее очевидные примеры субсинтаксем для самой частотной рубрики Локализация. Самая частотная в рубрике синтаксема Локатив (7465 ipm). Она описывается в словаре Г.А.Золотовой перечнем предложных конструкций «от + род, между + род, против + род, среди + род, у + род, за + твор, между + твор, над + твор, перед + твор, под + твор, в + предл, на + предл, при + предл». При корпусном анализе наиболее частотной является конструкция предлог “в/во” в сочетании с формой предложного или местного падежа [3700 IPM]: *сидеть в саду, гулять в лесу*. Вторую позицию по частотности занимает конструкция с предлогом “на” в сочетании с формой предложного или местного падежа [1800 IPM]: *сидеть на стуле, дышать воздухом на веранде*. В [Herskovits 1985] разница в значениях связывается с определенной идеей «включения объектов» для первого предлога и «опоры» и «смежности объектов» для второго. Однако различие в значениях скорее имеет лингвистический характер, поскольку в отношении многих пространственных объектов неясно, есть ли там включение. Например «веранда» является трехмерным объектом, и человек находится внутри нее (*сидеть на веранде*), аналогично *стоять на улице, висеть на дереве* и т. п., причем в других индоевропейских языках и даже славянских для этих объектов используется конструкция с предлогом “в”. Это означает, что перед нами собственно внутриязыковая классификация пространственных объектов, не всегда логичная. Эти конструкции могут сочетаться в одном контексте: *сидеть на веранде в кресле, висеть в бильярдной на стене*, что обычно трактуется как доказательство того, что перед нами «разные» семантико-синтаксические роли. Очевидно, что локализация объектов носит «телескопический» характер типа настраиваемого фокуса: более грубо или более точно. То есть первая конструкция с предлогом “в” является субсинтаксемой «локатив_1» в силу большей корпусной частотности, а конструкция с предлогом “на” «локатив_2», поскольку она уступает первой по частотности. Если в приведенных примерах поменять местами предложные конструкции (*висеть на стене в бильярдной*), то второй локатив становится атрибутом (партитивом) первого. В качестве синтаксических «хозяев» для синтаксем локатив1 и 2 являются разнообразные типы глаголов: существования, принятия положения в пространстве, восприятия и проч. Проще указать на семантические типы, плохо сочетающиеся с этими синтаксемами, это глаголы направленного движения. Полные описания всех введенных синтаксем и субсинтаксем приведены в приложении.

Как видно из предыдущего примера, субсинтаксем высокочастотной синтаксем находятся в зоне грамматических значений по корпусной статистике, отсюда не вполне логичная дифференциация пространственных объектов, которые в лексических тезаурусах будут попадать в разные группы. Всего в 23 синтаксемах выделено 72 субсинтаксем, 173 пары «предлог-субсинтаксема», из них 86 пар для первообразных предлогов, их суммарная частота 35359 ipm, среднее значение 411 ipm; 47 пар с производными предлогами, суммарная частота 2455 ipm, средняя частота 53 ipm; лексически связанных конструкций с первообразными предлогами 15, суммарная частота 641 ipm, среднее значение равно 43 ipm.

2.2 Морфосинтаксические и семантические характеристики главных и зависимых лексем в синтаксемах семантических рубрик

В концепции Г.А.Золотовой [Золотова 2011] синтаксема — это синтаксический способ выражения функционального значения существительного или именного словосочетания в структуре простого предложения, поэтому в их перечень входят падежные формы как показатель способа выражения того или иного значения. При выделении синтаксем и субсинтаксем мы использовали несколько ключевых показателей: (1) морфосинтаксический и семантический класс главного слова; (2) семантический класс зависимого слова; (3) эквивалент предложной конструкции в виде падежной формы; (4) регулярность оппозиций в субсинтаксемах семантической рубрики.

Наиболее четкую организацию с точки зрения этих показателей имеет самая частотная рубрика Локализация. Мы приводили в предыдущем разделе пример с противопоставление первых двух частотных синтаксем самой частотной синтаксемы рубрики Локатив. Главными словами регулярно являются глаголы, глагольные дериваты-существительные, предметные существительные, при этом глаголы и глагольные дериваты встречаются примерно в 80% контекстов, а предметные существительные в 20%. Ясно, что это соотношение может меняться в зависимости от тематического наполнения текстов корпуса, однако доминирование глагольного характера среди главных слов будет сохраняться. Семантический класс глаголов — принятие положения в пространстве, хотя реже появляются обозначения действий. В качестве зависимых существительных используются обозначения пространственных объектов: природных и искусственных, при этом они классифицируются по лингвистическому принципу определения «объемности». Некоторые существительные чисто логически могут использоваться в синтаксеме Локатив_1 [3700 IPM] (*в шкафу*) и Локатив_2 [1800 IPM] (*на шкафу*), однако на уровне корпусной статистики первый вариант встречается в 56% контекстах лексемы *шкаф*, а второй менее 1%. Локатив_3 представлен сочетанием первообразного предлога “у” или производных предлогов “возле” и “около” с родительным зависимого существительного [350 IPM] *выскочить у забора/ возле забора*, при этом снимается противопоставление местоположения типа “в” и “на”, главные слова обозначают стационарное положение в пространстве. В случае этой субсинтаксемы нескольких предложных реализаций одной субсинтаксемы взаимозаменяемы и не могут стоять вместе в одном контексте. Локатив_4: предлог “за” в сочетании с творительным [260 IPM] по сочетаемости похож на предыдущий вариант, однако регулярно обозначает деятельность, хотя используется глагол принятия положения в пространстве: *стоять за стойкой, за дверью*. Локатив_5: предлог “под” в сочетании с творительным [145 IPM] накладывает определенные ограничения на пространство: *храниться под кроватью, синяк под левым глазом, тело под одеялом*. При оформлении субсинтаксем следующей синтаксемы (Директив) ее варианты Директив_1, Директив_2, Директив_4, Директив_5 используются те же самые предлоги, но в сочетании с винительным падежом, причем предпочтения семантического типа зависимого существительного совпадают с теми, которые были представлены в Локативе_1—Локатив_4. Меняется семантический класс главных слов для этой конструкции — это глаголы движения в широком понимании, включая глаголы перемещения объектов. Причем для этой главных менее частотно использование Локатива, чем Директива. Субсинтаксема Директив_3 оформляется иначе, чем Локатив_3 — это предлоги “к” или “до”, первый с дательным, второй с родительным. Частоты субсинтаксем образуют аналогичную убывающую прогрессию, которая была характерна для субсинтаксем Локатива. Следующая синтаксема Департив, которую мы ввели в отличие от [Золотова 2011], присоединяется к глаголам движения, обозначает движение от начальной точки, при этом Департивы (1–5) имеют такие же селективные ограничения на класс зависимых существительных, как Директивы (1–5): *прийти в сад — уйти из сада, прийти на веранду — уйти с веранды, подойти к дому — отойти от дома, дойти до дома — отойти от дома, залезть за портьеру — вылезти из-за портьеры, залезть под стол — вылезти из-под стола*. Такая симметрия и четкая организация присутствует не во всех рубриках. Что важно, что Директивы и Департивы присоединяются к приставочным дериватам глаголов. Обычно говорят, что предлоги «повторяют» префикс, если есть фактические эквиваленты, или дублируют их на уровне значения. Исходя из корпусной статистики получает несколько иная картина: в зависимости от значения префикса присоединяется группа Директивов, Департивов или Транзитивов. В таблице 6 приведено распределение синтаксем локализации в контекстах префиксальных образований глаголов и их дериватов (подчеркнуты наиболее частотные предлоги для данной приставки, полужирным шрифтом выделены частоты IPM, превышающие средние значения для группы синтаксем).

Синтаксемы Приставки	Локатив предлоги IPM		Директив предлоги IPM		Департив предлоги IPM		Транзитив предлоги IPM	
	в, на	25	в (на)	390				
в/во/въ	в, на	25	в (на)	390				

<i>вз/взо/взь/вс/воз/вос</i>	на	40	в	40	из-за	40		
<i>вы</i>			<u>на</u> , (к, в, за)	330	из	150	через	25
<i>за</i>	<u>на</u>	110	в	170	с, у	55		
<i>на</i>	в, (на, за)	265	<u>на</u> (к)	210	с, от	80		
<i>над/надо/надъ</i>	<u>на</u> , в, от, с	5						
<i>о/об/объ/обо</i>	в, на	185	в	40				
<i>от/ото/отъ</i>			в (к, на)	210	от (из)	80		
<i>пере</i>	в	25	в, (на, к)	120	из	25	через	40
<i>по</i>			<u>на</u> , в, (к)	520				
<i>под/подо/подъ</i>	на	20	<u>к</u>	420	с	50		
<i>пре</i>	в, на	25	в	20				
<i>пред/предо/предъ</i>	на	12	в	8				
<i>при</i>			в, <u>на</u>	270	из	60	через	60
<i>про</i>	в, на	230	в, к	170			через, по, сквозь	90
<i>раз/разо/разъ/рас</i>	в, на, по	165					по	25
<i>с/со/съ</i>	у	120	в	360	от	120		
<i>у</i>	в	35	в, к, на	170	из, от	70		

Такая четкая организация синтаксиса рубрики Локализации воспроизводится в других рубриках, например в Темпоративе, что иногда интерпретируют следующим образом «время структурируется по модели пространства». Отличия безусловно есть, но некоторые соотношения корпусных частот, противопоставление значений в трансформированном виде воспроизводит базовую модель рубрики Локализации.

Сравнение наборов предложных конструкций для рубрики Локализации с перечнем синтаксиса Локатив в [Золотова 2011] приведено в таблице 7. В первом столбце приведен перечень синтаксиса КГ (кватитативной грамматики) для семантической рубрики Локализация, во втором столбце приведены наборы предложных конструкций, соответствующие синтаксису, в нашей системе обозначений (предлог с подстрочным индексом, обозначающим позицию падежа в шестичленной парадигме), в третьем — частота синтаксиса в ipm. Предложные конструкции синтаксиса КГ упорядочены в соответствии с частотностью субсинтаксиса. В четвертом столбце приведено название синтаксиса у Г.А.Золотовой, для которой в пятом столбце указаны предложные конструкции. Неконгруэнтность наборов показана объединением ячеек.

Синтаксиса КГ	Набор предложных конструкций	Частота в ipm	Синтаксиса Г.А.Золотовой	Набор предложных конструкций
Локатив	<i>в₆, на₆, у₂, возле₂, около₂, за₅, под₅, по₃</i>	7465	Локатив	от + род, между + род, против + род, среди + род, у + род, за + твор, между + твор, над + твор, перед + твор, под + твор, в + предл, на + предл, при + предл
Директив	<i>в₄, на₄, к₃, до₂, за₄, под₄</i>	4575	Директив	до + род, из + род, из-за + род, из-под + род, от + род, против + род, с + род, к + дат, в + вин, за + вин, на + вин, под + вин, за + твор
Департив	<i>из₂, с₂, от₂, из-за₂, из-под₂</i>	1432		
Транзитив	<i>по₃, через₄, сквозь₄</i>	533	Транзитив	по + дат, сквозь + вин, через + вин, над + твор

Предложные конструкции *над*₅, *между*₅, *среди*₅, *перед*₅, *против*₂, приведенные в перечне Г.А.Золотовой, не разграничивают Локатив и Директив, т. е. их следует отнести к низкочастотной периферии рубрики, которая слабо структурирована. Суммарная корпусная частота этих конструкций (300 ipm) составляет менее 2% от корпусной частоты этих синтаксем.

Список источников

1. Стеблин-Каменский М.И. Спорное в языкознании. Л.: Издательство Ленинградского университета, 1974.
2. Якобсон Р.О. Избранные работы. М., 1985.
3. Золотова Г.А. Синтаксический словарь: Репертуар элементарных единиц русского синтаксиса. 4-е изд. М.: Едиториал УРСС, 2011.
4. Fellbaum С. WordNet: An Electronic Lexical Database. Cambridge: MIT Press, 1998.
5. Азарова, И. В., Браславский, П. И., Захаров, В. П., Киселев, Ю. А., Усталов, Д. А., & Хохлова, М. В. (2016). Интеграция тезаурусов RussNet и YARN. в Компьютерная лингвистика и вычислительные онтологии: сборник научных статей (стр. 7-13). НИУ ИТМО.
6. O'Hara T., Wiebe J. 2009. Exploiting Semantic Role Resources for Preposition Disambiguation // Computational Linguistics, vol. 35, no. 2, pp. 151–184.