

Б. Ю. Норман^{1,2}, М. Ю. Мухин²

¹ *Белорусский государственный университет, Минск, Беларусь*

² *Уральский федеральный университет им. Б. Н. Ельцина, Екатеринбург*

**Лексическая и грамматическая семантика:
корпусное статистическое исследование
лексико-семантических групп ***

Рассматривается взаимодействие лексической и грамматической семантики на примере словоизменения близких по смыслу русских имен существительных. Данное взаимодействие формирует понятие «грамматический профиль», которое в статье обосновывается на материале словоизменения нескольких лексических групп существительных. Это слова, обозначающие предметы одежды, инструменты, а также предприятия и организации. Приведены статистические сведения о распределении падежных форм существительных в зоне снятой омонимии Национального корпуса русского языка. Обнаружены закономерности в распределении падежных форм слов, входящих в одну лексико-семантическую группу и влиянии на это распределение типовой семантики группы. Рассмотренные группы слов различаются грамматическими профилями. На основании распределения падежных форм в структуре каждой группы выделено ядро – слова, для которых типовая семантика является предположительно более прототипической. В грамматических различиях, выявленных между словами одной семантической группы, реализуются аспекты композиционной семантики, которые далеко не всегда отражаются в словарных толкованиях. Сделаны выводы о дальнейших возможностях использования корпусных данных для современных исследований лексической и грамматической семантики.

Ключевые слова: семантика, грамматический профиль, словоизменительная парадигма, падеж, лексико-семантическая группа, имя существительное, корпусная лингвистика.

* Исследование выполнено при финансовой поддержке Российского научного фонда (проект № 16-18-02005).

Норман Борис Юстинович – доктор филологических наук, профессор кафедры теоретического и славянского языкознания Белорусского государственного университета (ул. К. Маркса, 31, Минск, 220030, Беларусь; boris.norman@gmail.com); ведущий научный сотрудник Уральского федерального университета им. Б. Н. Ельцина (ул. Мира, 19, Екатеринбург, 620002, Россия)

Мухин Михаил Юрьевич – доктор филологических наук, доцент, директор департамента лингвистики Уральского федерального университета им. Б. Н. Ельцина (ул. Мира, 19, Екатеринбург, 620002, Россия; mikhail.mukhin@urfu.ru)

Способность языка выполнять различные функции в самых разных коммуникативных условиях обеспечивается сложностью и гибкостью его структуры. В частности, в процессе функционирования языковой системы обнаруживается, что единицы и их классы, выделяемые сознанием на одном уровне, связаны многообразными внутренними связями с единицами и их классами на смежном уровне. Самый простой случай: мы регулярно убеждаемся в том, как лексика «прорастает» в грамматику, и наоборот. Как совершенно справедливо отмечала Н. Ю. Шведова, «грамматический строй языка пронизан лексикой. Нет почти ни одного грамматического правила, которое не требовало бы включения в свою формулировку “лексической части”» [Шведова, 1984, с. 11].

Неслучайно в последние десятилетия русское языкознание обогащается различными образцами новых грамматик и словарей, выявляющих эти «интегративные» отношения (см. [Зализняк, 1967; 1987; Мельчук, Жолковский, 1984; Золотова, 1988; Караулов, 1993; Активный словарь..., 2014] и др.). В указанных трудах с той или иной степенью глубины разрабатывается идея внутренней взаимосвязи лексического значения слова с набором его функций в речи.

Для многих лингвистов учет лексической базы (круга слов) служит условием описания особенностей словоизменительной парадигмы и, в частности, тех отклонений, которые наблюдаются в ее рамках. Вот пример экспликации такой связи: «Для многих существительных значение местного падежа оказывается основным значением: соответственно, у такого рода существительных обычной является именно форма 2-го предложного, которая успешно конкурирует с формой 1-го предложного и может даже вытеснить эту последнюю форму» [Успенский, 2006, р. 536] – ср.: *в плену*, но не *в плене*, *на мосту*, но не *на мосте* и т. п. В данном случае исследователь исходит из неявного определения некоторой лексической группы («многие существительные») и приходит к особенностям их словоизменения. Но не менее закономерен и обратный путь: от грамматики – к лексике. Это значит, что если в рамках речевого употребления слова обнаруживаются расхождения (варианты) в его формальной парадигме, то это верный признак расщепления слова на две лексемы. Скажем, русское слово *выбор* употребляется преимущественно в единственном числе: *Они сделали свой выбор* (не *выборы*). Однако параллельно ему существует слово *выборы*, которое в единственном числе не употребляется (это *pluralia tantum*) – и это другое, отдельное слово! Процесс лексической дивергенции, расщепления слова можно также проиллюстрировать русскими примерами типа *вкус – вкусы*, *вес – весы*, *капля – капли*, *право – права* и т. д.

Сказанное в очередной раз выносит на повестку дня лингвистических исследований вопрос о полной, неполной и избыточной парадигме. Когда речь заходит о неполной парадигме какого-то слова, то понятно, что она обнаруживается лишь на фоне полной парадигмы всего грамматического класса (части речи и т. п.). Не заполненную по тем или иным причинам «клеточку» матрицы носитель языка вправе соотнести с заполненными «клеточками» у других слов: системный характер языка позволяет ему сделать это (см.: [Норман, 2011; 2013]). Отдельная словоформа – это тот «черепок», по которому носитель языка реконструирует всю амфору.

Избыточная же парадигма только на первый взгляд нарушает принцип экономности языка. На деле за нею скрывается либо начальный этап лексико-семантической дивергенции (о чем уже было сказано), либо дополнительные смыслы прагматико-дискурсивного характера, такие как «архаичность», «научность», «разговорность» и т. п. В любом случае перед нами – застывший слепок эволюционирующей языковой системы. А закончится ли данный процесс появлением нового слова или оформлением особой функционально-стилистической категории – зависит от конкретной ситуации.

А. Л. Шарандин, разрабатывающий принципы «лексической грамматики», указывает на особую значимость дефектности парадигмы для этого вида лингвистического описания. Причем, процитируем, «в онтологическом аспекте семантическая дефектность не есть собственно дефектность. Уникальность знакового содержания, представленного лексическим значением, в том и состоит, что слово стремится иметь только свойственный ему набор грамматических форм» [Шарандин, 2009, с. 25]. Иное дело – та же дефектность в гносеологическом аспекте. Она оказывается важнейшим инструментом для изучения и описания лексической семантики: исследователь использует состав грамматических форм, свойственный одному слову, «в сравнении с набором (составом) форм, присущих другим языковым знакам. В результате обнаруживаются различия в этих наборах и возникает возможность, опираясь на формальный признак, выявить и описать лексическую семантику слов, имеющих различный состав парадигм и их форм» [Там же].

И вот тут не обойтись без данных лингвистической статистики. Известно, что в сознании человека присутствует интуитивное представление об относительной употребительности тех или иных языковых фактов. Оно становится явным, если попросить носителя языка расположить некоторую группу фактов (например, членов одной лексико-семантической группы) в порядке убывающей частотности.

Если взять два русских существительных – *река* и *рука*, то формально они очень похожи: принадлежат к одному типу склонения (3d и 3f, по грамматическому словарю А. А. Зализняка [Зализняк, 1987], с очень незначительными расхождениями в акцентуации) и при этом обладают полной парадигмой (или даже избыточной, с учетом устаревающих вариантов *рекою* и *рукою*). Но слово *рука*, согласно «Новому частотному словарю русского языка», употребляется в русской речи примерно в девять раз чаще, чем *река*, и это, несомненно, обусловлено его семантикой¹. Вообще названия частей тела входят в набор самых частотных существительных. Кроме того, формы слова *рука* входят во множество фразеологизмов (что в сегментации корпуса не отражается). Теми же причинами объясняется и неравномерное распределение употребляемых в речи словоформ. Прежде всего, формы множественного числа для существительного *река* составляют всего около 13 % всех словоупотреблений, в то время как для слова *рука* этот показатель, судя по зоне снятой грамматической омонимии Национального корпуса русского языка (НКРЯ)², равен 52,5 %! Различна и сочетаемость данных слов с предлогами. Скажем, для существительного *рука* характерны сочетания с предлогом *под* (*под руку*, *под руки*) или с предлогом *без* (*без руки*, *без рук*), а появление в русских текстах выражений типа *под реку* или *без реки* маловероятно. Река – пространственный объект, поэтому вполне возможны такие сочетания, как *вдоль реки*, *поперек реки*, *через реку*, *к реке*, *перед рекой*, *по ту сторону реки*, а для существительного *рука* подобные комбинации (*вдоль руки*, *поперек руки* и т. п.) остаются умозрительными. И даже внешне идентичные сочетания вроде *за реку* и *за руку* выполняют, конечно, разную семантико-синтаксическую функцию. Эти различия коренятся в лексической семантике сравниваемых слов. Река – природный объект, обладающий физическими параметрами (протяженность, ширина, извилистость и т. п.); в речи эта лексема выступает чаще всего как локатив. Рука – часть тела человека, и поведение данного слова в тексте подчиняется общим правилам функционирования соматизмов. Чаще всего *рука* в высказывании играет роль инструмента, но возможны для этого слова и иные семантико-синтаксиче-

¹ Ляшевская О. Н., Шаров С. А. Новый частотный словарь русского языка. URL: <http://dict.ruslang.ru/freq.php> (дата обращения 11.07.2018).

² Национальный корпус русского языка. URL: <http://ruscorpora.ru> (дата обращения 11.07.2018).

ские функции. Таким образом, даже на отдельном, случайно взятом примере можно показать связь лексического значения слова с закономерностями его синтаксического употребления.

В целом, возвращаясь к исходным положениям, можно сказать, что полнота словоизменительной парадигмы – это потенциальное, виртуальное свойство, вытекающее из системного характера языка. В сознании же конкретного человека слово существует не только в соответствии с идеальной моделью его словоизменения, но и в виде набора реально употребляемых словоформ. В таком случае возникает вопрос: происходит ли выбор словоформы в конкретный момент рече-производства с опорой на парадигму всей части речи как некую «матрицу» словоизменения или же по аналогии с конкретными образцами склонения, уже заложенными в памяти? Прочитав одного из создателей когнитивистской модели, основанной на языковом употреблении (*usage-based model*): «Существование в пределах грамматики общих схем и конкретных примеров предоставляет говорящему альтернативные пути для достижения составного регулярного выражения, имеющего статус единицы: можно обратиться к нему “напрямую”, но можно и вывести его по схеме» [Лангаккер, 1997, с. 164]. Причем признание реальными обоих механизмов речевой деятельности заметно повышает статус словоформы как «словоизменительного прототипа».

Ю. Н. Караулов, поставивший своей целью на материале ассоциативного словаря «восстановить всю систему грамматических форм, т. е. всю русскую грамматику» [Караулов, 1993, с. 9], пришел к выводу, что даже ограниченный список из 12 разнородных в частеречном отношении стимулов (*бабушка, бежать, белый, время, билет, видеть, это, сам, первый, сорок, ты, из*) «показал практически 100-процентное заполнение типовых клеток, т. е. засвидетельствовал наличие в этом небольшом фрагменте АВС (ассоциативно-вербальной сети. – Б. Н., М. М.) всех грамматических фактов для имени существительного в русском языке» [Там же, с. 18]. Однако при этом оказалось, что поведение испытуемых очень неравномерно «разносит» грамматику по отдельным словам-реакциям. Иными словами, «грамматика, которая находится в распоряжении стихийного носителя языка, вся сплошь лексикализована, привязана к отдельным лексемам, как бы распределена между ними» [Там же, с. 6–7]. Значит, парадигма отдельного слова имеет право в речи быть неполной, более того – это ее естественное состояние! Действительно, именно к такому выводу приходит Ю. Н. Караулов: свойством ассоциативно-вербальной сети «следовало бы считать принципиальную неполноту парадигмы для каждой лексемы... не существует такого ассоциативного словаря или тезауруса, который содержал бы в качестве реакций полную парадигму хотя бы одной лексемы; полная парадигма в АВС может быть представлена только в диссипированном виде» [Там же, с. 188–189].

Основу для нашего дальнейшего исследования будут составлять три тезиса, обладающих разной степенью научной силы и новизны.

Первый из них: в лексической семантике слова заложены возможности его речевого употребления. Эта гипотеза наиболее общая и потому малоинтересная. Она молчаливо принимается большинством лингвистов и по сути представляет собой трюизм. Впрочем, некоторые ученые считают все же необходимым подчеркнуть данную связь, ср.: «Разные группы существительных по-разному проявляют себя в синтаксисе, и это зависит прежде всего от их значения» [Золотова, 1982, с. 123].

Второй тезис: лексическая семантика слова определяет особенности реализации его морфологической парадигмы. Соответственно, лексико-семантическая группа слов должна характеризоваться некоторой общностью парадигмы (доминанты здесь выделяются на основании средних величин). Эта идея весьма актуальна для современной грамматики. Она, в частности, реализуется в разработке

так называемых грамматических профилей слова (см.: [Gries, Divjak, 2009; Janda, Lyashevskaya, 2011; Ляшевская, 2013] и др.).

Третий тезис: совокупность словоупотреблений лексемы (речевая парадигма слова) коррелирует с интуитивным структурированием лексико-семантических групп в сознании носителя языка. Можно, в частности, предположить, что ядру такой группы свойственно наибольшее соответствие «доминантному» грамматическому профилю. Эта гипотеза наиболее сильная, она нуждается в специальном доказательстве и проверке. В том числе она требует обращения к данным лингвистической статистики.

В качестве источников статистических данных и иллюстративного материала нами был использован Национальный корпус русского языка. Поиск производился в зоне снятой омонимии корпуса, которая на сегодняшний день (весна 2018 г.) составляет 6 003 398 слов. Предпосылкой данной работы является исследование, результаты которого изложены в статье [Норман, Мухин, 2015].

Для рассмотрения мы привлекли, с опорой на идеографическую классификацию «Большого толкового словаря русских существительных» под ред. Л. Г. Бабенко [2005], наиболее частотные слова трех идеографических групп имен существительных. Это классы «Одежда» (29 слов), «Инструмент» (25), а также слова из разных сфер (преимущественно сферы обслуживания, образования, культуры, медицины, экономики и т. п.) с общим значением «Предприятие, организация» (*гостиница, магазин, школа, университет, кинотеатр, больница* и др. – всего 27 слов). Из списка пришлось исключить такие лексемы, как *орудие, организация, рынок*, так как их многозначность коренным образом сказывается на результатах статистического анализа. Этим фактором, в частности, определен и выбор самих групп. Другим фактором является наличие интуитивных представлений о возможной количественной разнице в употреблении падежей, как было в случае с уже рассмотренными ранее формами слов *река* и *рука*.

Для каждой лексемы определено количество употреблений в шести основных падежных формах, которые в нотации корпуса обозначаются соответственно как Nom, Gen, Dat, Acc, Ins, Loc. Таким образом, основную базу исследования составили результаты $(29 + 25 + 27) \times 6 = 486$ запросов к корпусу.

Если взять количество словоупотреблений существительных в этих основных падежных формах за 100 %, то на диаграмме видно, что в процентном соотношении лидируют формы именительного, родительного и винительного падежей (рис. 1).

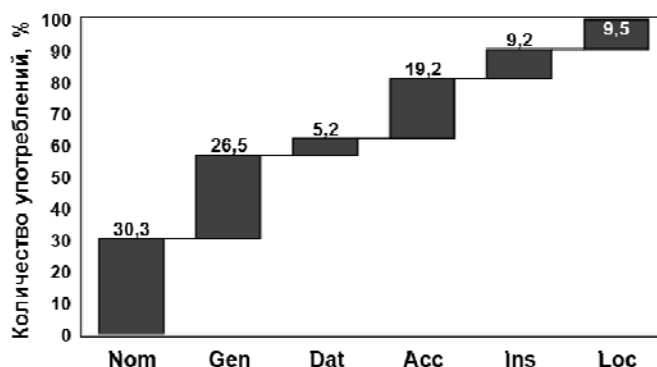


Рис. 1. Соотношение падежных форм существительных в зоне снятой омонимии НКРЯ
Fig. 1. The ratio of noun case forms in the disambiguation zone of the Russian National Corpus

Сходные статистические данные приводят и авторы проекта «Русская корпусная грамматика»³.

Несмотря на среднее статистическое доминирование трех падежей, интуитивное ожидание употребления слов конкретных идеографических групп предполагает другое распределение. Например, резонно прогнозировать, что частой и естественной формой для слов, относящихся к группе «Инструмент», является форма творительного падежа. Слова, называющие предметы одежды, кроме именительного падежа, часто употребляются в форме предложного падежа с предлогом *в*. Можно гипотетически говорить и о том, что у каждой идеографической группы, несмотря на возможную лексическую неоднородность, должен быть типовой грамматический профиль. Идею подобного рода высказывает О. Н. Ляшевская, создающая «Частотный лексико-грамматический словарь»: кроме грамматических профилей лексем, в словаре планируется давать «сведения для крупных лексико-семантических классов (в классификации НКРЯ), например, для глаголов движения, имен инструментов и т. п.» [Ляшевская, 2013]. К сожалению, заявленный словарь пока не вышел, и, кроме того, семантическая разметка корпуса базируется на очень крупных категориях. Это означает, что выявить лексико-грамматические особенности мелких семантических групп на фоне общей статистики будет не просто.

Следуя второму тезису о соотношении семантических групп и грамматических профилей слов, рассмотрим статистическую картину, которую дают избранные группы. Для слов группы «Одежда» (*костюм, одежда, платье, шапка, шляпа, пальто, рубашка, шинель, халат* и др. – расположены по убыванию частоты в НКРЯ) наиболее характерны формы винительного и предложного падежей (рис. 2).

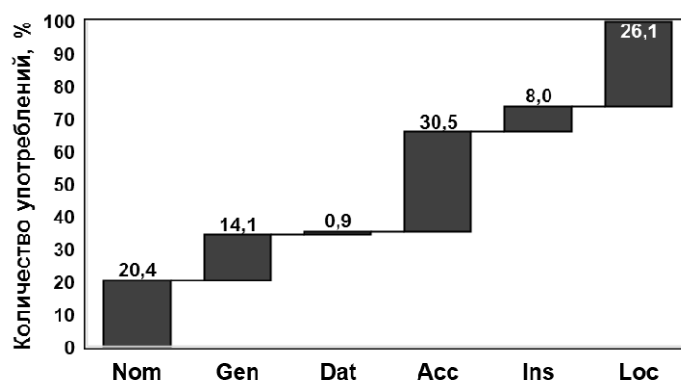


Рис. 2. Соотношение падежных форм существительных группы «Одежда»
Fig. 2. The ratio of noun case forms in the «Clothes» group

Как интерпретировать этот отход от стандартного распределения? Судя по конкретным контекстам, человек в своей «усредненной» речемыслительной деятельности воспринимает одежду как артефакт, служащий ему «объектом» деятельности (надевания, снятия, хранения, покупки и т. п.), – это объясняет высокий процент винительного падежа; затем – как «контейнер», «оболочку» для своего тела (это объясняет долю форм предложного падежа, в том числе

³ Воейкова М. Д. Именительный падеж // Русская корпусная грамматика. URL: http://rusgram.ru/Именительный_падеж (дата обращения 11.07.2018).

с предложом *в* – 17,4 из 26 %) и лишь затем – как «субъект» (состояния, размещения в пространстве, оценочной квалификации и т. п.) – об этом говорит доля именительного падежа.

Свой вариант падежного распределения имеет группа слов, обозначающих инструменты (*нож, топор, лопата, молоток, пила, ножницы, линейка* и др.; в группу включены также частотные слова *ложка* и *вилка*). Для этого класса, как и ожидалось, наиболее типична форма творительного падежа и, кроме него, винительного (рис. 3).

Приоритет инструментального значения очевиден в плане преобладания форм творительного падежа над всеми прочими, а если сложить доли «объектных» форм творительного и винительного падежей, то получится 61 %, что говорит об абсолютной объектной семантической доминанте группы. Основной ущерб в рассмотренных группах претерпевают именительный (на треть меньше в сравнении с общей статистикой), родительный (в два раза меньше) и дательный (в 4–5 раз меньше) падежи. Однако идея «общей объектности» (винительный падеж) для группы «Одежда» значительно важнее, чем для группы «Инструмент» (*надеть* или *повесить плащ* куда как весомее для языка, чем *взять молоток*; и так же более естественно и очевидно *ударить молотком*, чем *накрыться* или *размахивать плащом*, не говоря уже – *халатом, майкой* и т. п.). Наиболее разительно у этих групп отличаются показатели реализации творительного (8 vs. 37,9 %) и предложного (26 vs. 4,8 %) падежей в силу малой актуальности инструментальной функции одежды и семантических функций, связанных с формами предложного падежа, для разных инструментов.

В идеографической группе слов с семантикой «Предприятие, организация» наиболее существенно, в сравнении со стандартным распределением, «прибавляет» предложный падеж за счет убывания творительного и именительного. Винительный падеж держится около стандартных значений (рис. 4).

Преобладание локативного значения у слов этой группы вполне объяснимо: *остановиться в гостинице* или *лежать в больнице* – вполне естественные контексты. Организации предполагают наличие помещения, служащего пространственной координатой в обозначаемой ситуации. Что касается увеличенной доли родительного падежа, то здесь контекстная семантика достаточно разнородна и в значительной степени зависит от предлога, с которым сочетается падежная форма. Так, слово *гостиница*, как и другие, может встречаться в сочетаниях типа

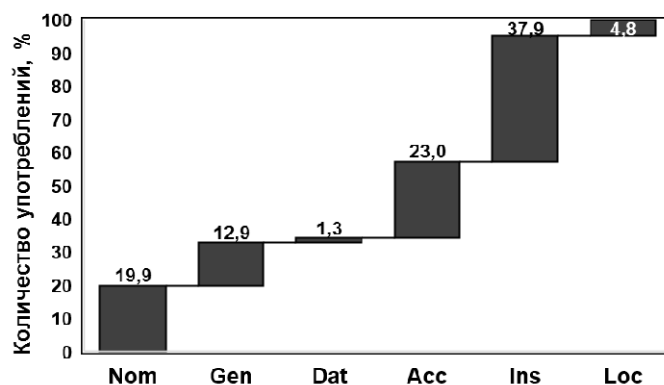


Рис. 3. Соотношение падежных форм существительных группы «Инструмент»
Fig. 3. The ratio of noun case forms in the «Instrument» group

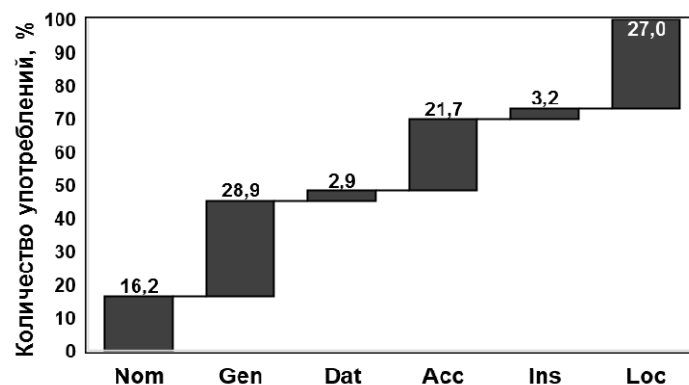


Рис. 4. Соотношение падежных форм существительных со значением «Предприятие, организация»
 Fig. 4. The ratio of noun case forms in the «Enterprise, organization» group

хозяин (директор) гостиницы, номер гостиницы, нет гостиницы и т. д. Однако значительная часть контекстов здесь выражает также пространственную семантику: *выехать из гостиницы, доехать до гостиницы, выписаться из больницы* и др. Таковы у этих слов и многие случаи употребления форм винительного падежа (*поехать или лечь в больницу, поселить в гостиницу*).

Третий тезис, заявленный в постановке проблемы, по сути, утверждает, что грамматический профиль слова соотносится с тем местом, которое оно занимает в составе идеографической группы. Вероятно, чем «прототипичнее» профиль, тем более «прототипично» и слово в плане выражения типовой семантики группы.

Так, если принять, что для группы «Одежда» наиболее характерны формы винительного и предложного падежей, то можно рассмотреть, для каких конкретных слов эти формы наиболее и наименее характерны (табл. 1).

Согласно полученным данным, больше всего форм предложного падежа у слов *халат, майка, джинсы, трусы, костюм*, меньше всего – у слов *галстук, шапка, шаль, шарф*. Оценивая эти крайние показатели, мы констатируем, что «контейнерное» значение, связанное с формой предложного падежа, характерно для основных предметов одежды, а чисто объектное значение (винительный падеж) – для «аксессуаров». Грубо говоря, халат – в большей степени одежда, чем шапка или шарф. Однако статистика не объясняет различий между *свитером* и *курткой* (винительный падеж), *джинсами* и *штанами* (предложный падеж). В этих случаях для измерения семантических расстояний предложенный метод не работает. Скорее всего, данных, извлеченных из зоны снятой омонимии корпуса, пока недостаточно для таких частных семантических суждений.

Рассмотрим, какова статистика для наиболее значимых падежных форм группы «Инструмент» (табл. 2).

И в этой парадигме ряд показателей можно связать с лексической семантикой. Есть «безусловные» инструменты вроде *плоскогубцев, лобзика, ножниц, молотка* (слова с выраженной долей форм творительного падежа), а есть инструменты, которыми не обрабатывают материал или поверхность: *тиски, верстак, линейка*. Сюда относится и *долото*, не которым, а по которому обычно ударяют, – например, *молотком*. Для слов *верстак* и, в особенности, *тиски* естественным образом характерна форма предложного падежа. Неудивительно, что эти существительные находятся в нижней части табл. 2.

Таблица 1

Доля основных падежных форм внутри группы «Одежда»
The share of the main case forms within the «Clothes» group

Элемент группы «Одежда» (в порядке убывания использования винительного падежа)	Доля винительного падежа, %	Элемент группы «Одежда» (в порядке убывания использования предложного падежа)	Доля предложного падежа, %
Шарф	25,5	Халат	48,4
Шаль	23,8	Майка	39,7
Колготки	20,0	Джинсы	35,6
Галстук	11,7	Трусы	34,8
Свитер	11,3	Костюм	32,8
Плащ	10,2	Шуба	32,6
Юбка	10,1	Свитер	32,3
Кепка	8,7	Куртка	32,0
Шляпа	8,1	Кофта	31,8
<...>		<...>	
Трусы	4,5	Штаны	22,3
Платье	3,6	Юбка	20,9
Майка	3,4	Колготки	20,0
Куртка	3,4	Галстук	18,7
Халат	3,2	Шапка	18,4
Костюм	2,5	Шаль	12,7
Пальто	2,0	Шарф	7,8

Таблица 2

Доля основных падежных форм внутри группы «Инструмент»
The share of the main case forms within the «Instrument» group

Элемент группы «Инструмент» (в порядке убывания использования творительного падежа)	Доля творительного падежа, %	Элемент группы «Инструмент» (в порядке убывания использования винительного падежа)	Доля винительного падежа, %
Плоскогубцы	85,7	Паяльник	43,8
Лобзик	70,0	Линейка	33,3
Ножницы	54,3	Щипцы	33,3
Напильник	50,0	Кусачки	33,3
Пинцет	50,0	Лопата	30,6
Молоток	49,5	Кувалда	30,0
Стамеска	44,4	Нож	29,8
Щипцы	44,4	Ложка	29,3
<...>		<...>	

Окончание табл. 2

Элемент группы «Инструмент» (в порядке убывания использования творительного падежа)	Доля творительного падежа, %	Элемент группы «Инструмент» (в порядке убывания использования винительного падежа)	Доля винительного падежа, %
Верстак	22,2	Плоскогубцы	14,3
Линейка	15,7	Тиски	13,3
Пила	13,2	Напильник	12,5
Долото	12,5	Лобзик	10,0
Тиски	6,7	Рубанок	8,3

Таким образом, полевая структура лексико-семантической группы слов получает некоторое подтверждение в виде статистики грамматических форм. Идея грамматической доминанты в данном случае коррелирует с понятием центра поля. Вместе с тем рассматривать падежные формы в качестве «диагностирующих» показателей, а именно – форму предложного падежа для группы «Одежда», а творительного – для группы «Инструмент» представляется рискованным, потому что реальная картина употребления этих форм в тексте достаточно сложна.

Группа слов, обозначающих предприятия и организации, является весьма пестрой с точки зрения сфер действительности. В табл. 3 мы упорядочили их список по убыванию наиболее значимых – предложного и родительного падежей.

Таблица 3

Доля основных падежных форм
внутри группы существительных со значением «Предприятие, организация»
The share of the main case forms within the «Enterprise, organization» group

Элемент группы «предприятие, организация» (в порядке убывания использования предложного падежа)	Доля предложного падежа, %	Элемент группы «предприятие, организация» (в порядке убывания использования родительного падежа)	Доля родительного падежа, %
Столовая	44,9	Университет	41,4
Бар	35,6	Отель	38,5
Госпиталь	34,4	Институт	37,9
Гостиница	32,5	Комбинат	37,8
Аптека	32,4	Кинотеатр	37,5
Больница	32,4	Банк	36,0
Баня	31,8	Театр	35,0
Буфет	31,0	Цирк	34,9
Фабрика	29,9	Вуз	34,2
Ресторан	29,8	Офис	32,9
Школа	29,6	Поликлиника	32,4
Цирк	29,2	Завод	32,2

Элемент группы «предприятие, организация» (в порядке убывания использования предложного падежа)	Доля предложного падежа, %	Элемент группы «предприятие, организация» (в порядке убывания использования родительного падежа)	Доля родительного падежа, %
Контора	27,7	Музей	31,7
Отель	26,7	Гостиница	31,1
Театр	26,5	Госпиталь	28,5
Библиотека	24,7	Магазин	28,1
Магазин	24,5	Библиотека	27,4
Офис	24,4	Фабрика	26,8
Поликлиника	23,9	Школа	24,9
Музей	22,8	Ресторан	24,1
Кинотеатр	22,5	Контора	23,6
Вуз	21,8	Бар	21,9
Университет	21,7	Больница	20,6
Институт	19,8	Буфет	17,7
Завод	19,0	Аптека	16,2
Комбинат	15,6	Столовая	14,6
Банк	14,1	Баня	13,4

Любопытно, что слова по представленности форм предложного и родительного падежей достаточно четко расходятся. Сочетаемость с предлогом *в* (*в столовой, баре, госпитале, гостинице, аптеке* и т. п.), вероятно, свидетельствует о более важном семантической признаке ‘помещение’ в структуре значения перечисленных слов. Для существительных *банк, комбинат, завод, институт, университет, вуз*, кроме пространственного значения, реализуемого совместно с предлогом *из*, важно также и атрибутивное или объектное: *директор банка, студент университета, строительство комбината, проходная завода, реструктуризация завода* и т. д. Разумеется, лексическое значение слова реализуется во всей совокупности употребляемых словоформ. Но даже сопоставление данных по двум падежам может иметь «диагностирующий» характер: мы получаем, что *столовая* – это помещение, а *университет* – нечто иное. По сути, в данных различиях реализуется один из аспектов композиционной семантики, утверждающей, что определенная часть значения слова находится «за его пределами».

Результаты, представленные в статье, показывают, как можно применять данные корпусных методов для изучения регулярных лексико-грамматических явлений. В то же время мы видим и определенные ограничения, накладываемые на применение статистических методов в семантике. Некоторые причины этих ограничений имеют универсально-языковой характер, некоторые являются переходящими и обусловлены современным состоянием лингвистических корпусов. Однако уже сегодняшние собрания позволяют делать значимые выводы об отражении связи лексики и грамматики.

Человек укладывает, упаковывает свой жизненный опыт в определенные языковые структуры. Одни из них более очевидны – это, скажем, падежные и предложно-падежные формы. Другие менее очевидны – это лексико-семантические

группы слов. Но и в том и в другом случае мы имеем дело с участием языка в когнитивной (познавательной) деятельности. Причем эти разные виды моделирования действительности внутренне предполагают друг друга и коррелируют друг с другом.

Слово существует только через совокупность своих словоформ и контекстов, в которых оно участвует. Мы показали, что особенности словоизменительной парадигмы слова воплощают в себе своеобразие его лексической семантики. Имеется в виду, что категориальная семантика (например, изменение по падежам и числам существительных определенного рода), представленная в языковом сознании, имеет для конкретной лексемы условный и схоластический характер. Слово в соответствии со своим лексическим значением выберет из этой парадигмы только те формы, которые ему необходимы для участия в коммуникативных актах. Вернейший показатель такой специализации – частота той или иной словоформы и ее доля в общем грамматическом «профиле» слова. Таким образом, второй из сформулированных выше тезисов можно считать полностью подтвержденным.

Что же касается третьего тезиса – о внутренней структуре лексико-семантической группы и ее ранжировании с помощью грамматических «профилей», – то он требует дополнительной проверки на более объемном статистическом материале.

Список литературы

- Активный словарь русского языка. Т. 1: А–Б / Отв. ред. Ю. Д. Апресян. М., 2014.
- Большой толковый словарь русских существительных: Идеографическое описание. Синонимы. Антонимы / Под ред. Л. Г. Бабенко. М.: АСТ-Пресс книга, 2005. 864 с.
- Зализняк А. А.* Русское именное словоизменение. М., 1967.
- Зализняк А. А.* Грамматический словарь русского языка. Словоизменение. 3-е изд. М., 1987.
- Золотова Г. А.* Коммуникативные аспекты русского синтаксиса. М., 1982.
- Золотова Г. А.* Синтаксический словарь. Репертуар элементарных единиц русского синтаксиса. М., 1988.
- Караулов Ю. Н.* Ассоциативная грамматика русского языка. М., 1993.
- Лангаккер Р. В.* Модель, основанная на речевом употреблении // Вестн. Моск. гос. ун-та. Сер. 9. Филология. 1997. № 4. С. 160–174.
- Ляшевская О. Н.* Частотный лексико-грамматический словарь: проспект проекта // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Междунар. конф. «Диалог». М., 2013. Т. 1. С. 478–489. URL: <http://www.dialog-21.ru/digests/dialog2013/materials/pdf/LyashevskayaON.pdf> (дата обращения 11.07.2018).
- Мельчук И. А., Жолковский А. К.* Толково-комбинаторный словарь современного русского языка. Опыт семантико-синтаксического описания русской лексики // Wiener Slawistischer Almanach. Sonderband 14. Wien, 1984.
- Норман Б. Ю.* Полнота/неполнота парадигмы как признак слова // Грамматика и лексика у словенским језицима. Нови Сад; Београд, 2011. С. 151–164.
- Норман Б. Ю.* Словоформа vs. парадигма (на материале русского языка) // Университетский научный журнал. 2013. № 4. С. 30–40.
- Норман Б. Ю., Мухин М. Ю.* О грамматических профилях лексико-грамматических групп: поиск внутренних связей между лексической и грамматической семантикой // SLAVIA: Časopis pro slovanskou filologii. 2015. Roč. 84, seš. 3. S. 348–359.

Успенский Б. А. Из наблюдений над русской грамматикой: Отношение падежных и числовых значений (Грамматическое выражение отношения части и целого) // *International Journal of Slavic Linguistics and Poetics*. 2006. Iss. XLIV–XLV. P. 503–544.

Шарандин А. Л. Методология лексической грамматики // *Взаимодействие лексики и грамматики в русском языке: проблемы, итоги и перспективы*. Тамбов, 2009. С. 13–28.

Шведова Н. Ю. Об активных потенциях, заключенных в слове // *Слово в грамматике и словаре*. М., 1984. С. 7–15.

Gries S. Th., Divjak D. Behavioral profiles: A corpus-based approach to cognitive semantic analysis // *New Directions in Cognitive Linguistics*. John Benjamins Publ. Comp., 2009. P. 57–75.

Janda L. A., Lyashevskaya O. Grammatical profiles and the interaction of the lexicon with aspect, tense, and mood in Russian // *Cognitive Linguistics*. 2011. Vol. 22, No 4. P. 719–763.

B. Yu. Norman¹, M. Yu. Mukhin²

¹ *Belarusian State University, Minsk, Belarus*

*Ural Federal University named after the first President of Russia B. N. Yeltsin
Ekaterinburg, Russian Federation, boris.norman@gmail.com*

² *Ural Federal University named after the first President of Russia B. N. Yeltsin
Ekaterinburg, Russian Federation, mikhail.mukhin@urfu.ru*

**Lexical and grammatical semantics:
a corpus-based statistical study of lexical semantic groups**

The paper explores the correlation between lexical and grammatical semantics showcased by the inflexion of close meaning nouns of the Russian language. The basic assumption is as follows: the lexical semantics of a word determines the features of its morphological paradigm, and the totality of lexeme variations in use (the speech paradigm of the word) correlates with the unmotivated structuring of lexical semantic groups in a native speaker's mind. The correlation between lexical and grammatical semantics conveys itself through the concept of the word's grammatical profile, herein exemplified by cases of inflexion of several lexical groups of nouns. These are the words that refer to clothing, tools, as well as companies and organisations. The material for statistical and contextual analysis was retrieved from the National Corpus of the Russian language.

Firstly, the statistical information is provided on the distribution of the case forms of nouns in removed homonymy zone of the National Corpus of the Russian language. Secondly, the regularities in the distribution of case forms of words are revealed that belong to the same lexical semantic group, and the ways this distribution is affected by the group's general semantics are found. The word groups concerned proved to differ in their grammatical profiles: the nouns of the «Clothes group» show peculiar statistical characteristics of the accusative and prepositional cases, the «Tools» group – in instrumental and accusative cases, the «Companies, organisations» group – in prepositional and accusative cases respectively. Thirdly, considering the distribution of case forms within each group, a «core group» of the words is identified where standard semantics is supposedly more prototypical. The analysis of grammatical differences found between the words of one semantic group, reveals certain aspects of compositional semantics, which are not always reflected in vocabulary definitions.

Conclusions are made about the further possibilities of using case data for current studies of the correlation between lexical and grammatical semantics.

Keywords: semantics, grammatical profile, inflectional paradigm, case, lexical semantic group, noun, corpus linguistics.

DOI 10.17223/18137083/64/17

References

- Aktivnyy slovar' russkogo yazyka. T. 1: A–B* [The active dictionary of the Russian language. Vol. 1: A–B]. Apresyan Yu. D. (Ed.). Moscow, 2014.
- Bol'shoy tolkovyy slovar' russkikh sushchestvitel'nykh: Ideograficheskoye opisaniye. Sinonimy. Antonimy* [The great explanatory dictionary of Russian nouns. Ideographic description. Synonyms. Antonyms]. Babenko L. G. (Ed.). Moscow, AST-Press kniga, 2005, 864 p.
- Gries S. Th., Divyak D. Behavioral profiles: A corpus-based approach to cognitive semantic analysis. In: *New Directions in Cognitive Linguistics*. John Benjamins Publ. Company, 2009, pp. 57–75.
- Janda L. A., Lyashevskaya O. Grammatical profiles and the interaction of the lexicon with aspect, tense, and mood in Russian. *Cognitive Linguistics*. 2011, vol. 22, no. 4, pp. 719–763.
- Karaulov Yu. N. *Assotsiativnaya grammatika russkogo yazyka* [Associative grammar of the Russian language]. Moscow, 1993.
- Langakker R. V. Model', osnovannaya na rechevom upotreblenii [A speech-based pattern]. *Moscow State University Bulletin. Series 9. Philology*. 1997, no. 4, pp. 160–174.
- Lyashevskaya O. N. Chastotnyy leksiko-grammaticheskiy slovar': prospekt proyekta [A frequency lexical-grammatical dictionary]. In: *Komp'yuternaya lingvistika i intellektual'nyye tekhnologii: Po materialam ezhegodnoy Mezhdunar. konf. "Dialog"* [Computer linguistics and intellectual technologies: Proceedings of the annual international conference "The dialogue"]. Moscow, 2013, vol. 1, pp. 478–489. URL: <http://www.dialog-21.ru/digests/dialog2013/materials/pdf/LyashevskayaON.pdf> (accessed 11.07.2018).
- Mel'chuk I. A., Zholkovskiy A. K. *Tolkovo-kombinatornyy slovar' sovremennog orusskogo yazyka. Opyt semantiko-sintaksicheskogo opisaniya russkoy leksiki* [Explanatory combinatorial dictionary of modern Russian. Semantic-syntactic studies of Russian vocabulary]. *Wiener Slavistischer Almanach*. Sonderband 14. Vienna, 1984.
- Norman B. Yu. Polnota/nepolnota paradigmy kak priznak slova [The complete/incomplete paradigm as a word feature]. In: *Gramatika i leksika u slovenskim jezicima* [Grammar and vocabulary of the Slovenian language]. Novi Sad, Belgrade, 2011, pp. 151–164.
- Norman B. Yu. Slovoforma vs. paradigma (na materiale russkogo yazyka) [Word form vs. paradigm]. *Humanities and Science University Journal*. 2013, no. 4, pp. 30–40.
- Norman B. Yu., Muhin M. Yu. Norman B. Yu., Mukhin M. Yu. O grammaticheskikh profilyakh leksikogrammaticheskikh grupp: poisk vnutrennikh svyazey mezhdru leksicheskoy i grammaticheskoy semantikoy [On the grammar aspects of lexico-grammatical groups: in search of internal ties between lexical and grammatical semantics]. *Slavia, časopis pro slovanskou filologii*. 2015, Ročník 84, sešit 3, pp. 348–359.
- Sharandin A. L. Metodologiya leksicheskoy grammatiki [The methodology of lexical grammar]. In: *Vzaimodeystvie leksiki i grammatiki v russkom yazyke: problemy, itogi i perspektivy* [The relation between vocabulary and grammar in the Russian language: Problems, conclusions, perspectives]. Tambov, 2009, pp. 13–28.
- Shvedova N. Yu. Ob aktivnykh potentsiyakh, zaklyuchennykh v slove [On active potentials within a word]. In: *Slovo v grammatike i slovare* [Word in grammar and in vocabulary]. Moscow, 1984, pp. 7–15.
- Uspenskiy B. A. Iz nablyudeny nad russkoy grammatikoy: Otnosheniye padezhnykhi chislovykh znacheniy (Grammaticheskoye vyrazheniye otnosheniya chasti i tselogo) [Some observations on the Russian grammar: The relations of case and number semantics (Grammatical expression of the part-to-whole)]. *International Journal of Slavic Linguistics and Poetics*. Iss. 44–45, 2002–2003. 2006, pp. 503–544.
- Zaliznyak A. A. *Russkoye imennoye slovoizmeneniye* [Russian nominal inflection]. Moscow, 1967.
- Zaliznyak A. A. *Grammaticheskiy slovar' russkogo yazyka. Slovoizmeneniye. 3-e izd.* [Grammar dictionary of the Russian language. Inflection. 3rd ed.]. Moscow, 1987.
- Zolotova G. A. *Kommunikativnyye aspekty russkogo sintaksisa* [Communicative aspects of the Russian syntax]. Moscow, 1982.
- Zolotova G. A. *Sintaksicheskii slovar'*. *Repertuar elementarnykh edinits russkogo sintaksisa* [Syntax dictionary: The repertoire of elementary units of Russian syntax]. Moscow, 1988.