

К СОЗДАНИЮ ЧАСТОТНОГО СЛОВАРЯ ЗНАЧЕНИЙ СЛОВ¹

Иомдин Б. Л. (iomdin@ruslang.ru)

Институт русского языка
им. В. В. Виноградова РАН, Москва, Россия;
Национальный исследовательский университет
Высшая школа экономики, Москва, Россия

Лопухина А. А. (nastya-merk@yandex.ru)

Институт русского языка
им. В. В. Виноградова РАН, Москва, Россия

Носырев Г. В. (grigorij-nosyrev@yandex.ru)

Яндекс, Москва, Россия

В докладе на материале русских существительных с предметными значениями обосновывается необходимость создания частотного словаря значений слов. Предлагаются методы приближенного определения частот, основанные на анализе данных опросов информантов и аннотировании наиболее частотных коллокаций в большом корпусе текстов (в настоящей работе был использован самый объемный на сегодняшний день корпус RuTenTen11, интегрированный в систему Sketch Engine). Такой словарь мог бы быть востребован в различных компьютерно-лингвистических приложениях (в частности, для вероятностного разрешения многозначности в отсутствие контекста), при создании обучающих ресурсов, в традиционной толковой лексикографии. Исследования наборов значений многозначных слов и их сравнительной частотности представляют и теоретический интерес для изучения эволюции лексической системы языка.

Ключевые слова: семантика, лексикография, многозначность, полисемия, омонимия, предметная лексика, опросы, эксперименты, частота, статистические методы

¹ Работа выполнена при частичной финансовой поддержке Программы фундаментальных исследований отделения историко-филологических наук РАН «Язык и литература в контексте культурной динамики», гранта РГНФ №13-04-00307а и гранта НШ-3899.2014.6 для поддержки научных исследований, проводимых ведущими научными школами РФ.

TOWARDS A WORD SENSE FREQUENCY DICTIONARY

Iomdin B. L. (iomdin@ruslang.ru)

Vinogradov Russian Language Institute
of the Russian Academy of Sciences, Moscow, Russia;
National Research University
Higher School of Economics, Moscow, Russia

Lopukhina A. A. (nastya-merk@yandex.ru)

Vinogradov Russian Language Institute
of the Russian Academy of Sciences, Moscow, Russia

Nosyrev G. V. (grigorij-nosyrev@yandex.ru)

Yandex, Moscow, Russia

Analyzing several Russian nouns denoting everyday life objects, we explain why a word sense frequency dictionary is necessary. Techniques of calculating the approximate frequencies are proposed, based on the analysis of native speaker surveys and the annotation of the most frequent collocations in a large text corpus (we used the huge RuTenTen11 corpus integrated into the Sketch Engine system). A word sense dictionary could be used in a variety of NLP tasks, in particular for a probabilistic word sense disambiguation without available context, in creating second language learning resources, as well as in academic lexicography. Besides, studies of sense sets of polysemous words and their comparative frequencies are important for the linguistic theory, because they shed light on the evolution of the lexical system.

Key words: semantics, lexicography, ambiguity, polysemy, homonymy, everyday life vocabulary, surveys, experiments, frequency, statistical techniques

1. Введение

В работе, опубликованной в материалах предыдущей конференции «Диалог», мы отмечали: «в последнее время большое значение, в частности, в компьютерной лингвистике, придается созданию частотных словарей и списков слов. К сожалению, большинство имеющихся частотных списков составляются из вокабул, но не отдельных лексем (то есть слов, взятых в определенном значении). Между тем очевидно, что разные лексемы одной и той же вокабулы частотны в очень разной степени; столь же очевидно, что составление частотных списков лексем представляет собой существенно более трудную и практически не автоматизируемую задачу» [Иомдин и др. 2013: 319].

Проблема отсутствия частотных словарей значений признается не только в лингвистике, но и, например, в сфере педагогики, в частности, при составлении списков слов для изучения в разных классах школ; ср. “Some problems are inevitable when word frequency is the primary source for identifying words. <...> A word that has different meanings is listed only once. For example, whether *bank* means financial institution, edge of a river, or angle of an airplane is not taken into account. *B-a-n-k* appears one time on the list, and its associated frequency represents all the different meanings. In other words, there is no way to get the frequency of the word *bank* meaning a financial institution” [Beck at al. 2013: 21]. Действительно, в существующих частотных словарях обычно не различаются не только разные лексемы, но и омонимы (ср., в частности, [Ляшевская и Шаров 2009]). Нет информации о частотности значений и в переводных словарях, что затрудняет работу с ними (особенно в случае сильно многозначных слов, когда недостаточно знакомый с языком читатель вынужден строить много равновероятных гипотез о значении незнакомого слова). Востребованность такой информации кажется бесспорной. Симптоматично, что система статистического машинного перевода Google Translate с 31.10.2012 стала визуализировать частотность различных вариантов перевода (но не значений переводимого слова!), хотя пока выдаваемые результаты достаточно спорны, во всяком случае с точки зрения адекватности переводных эквивалентов (ср. рис. 1–2):

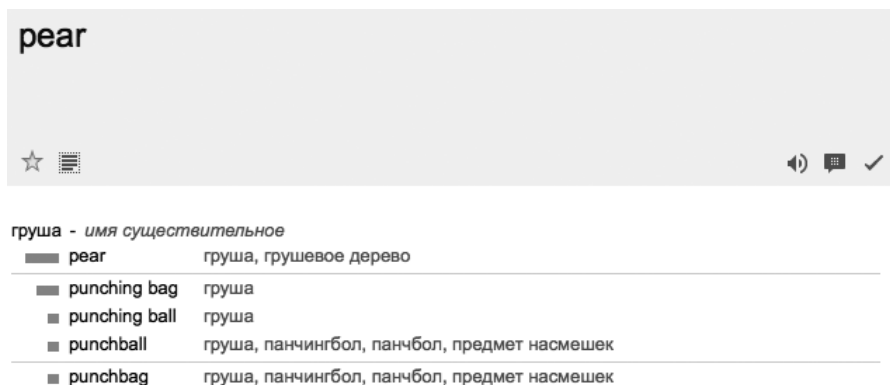


Рис. 1. Варианты перевода слова *груша* в Google Translate

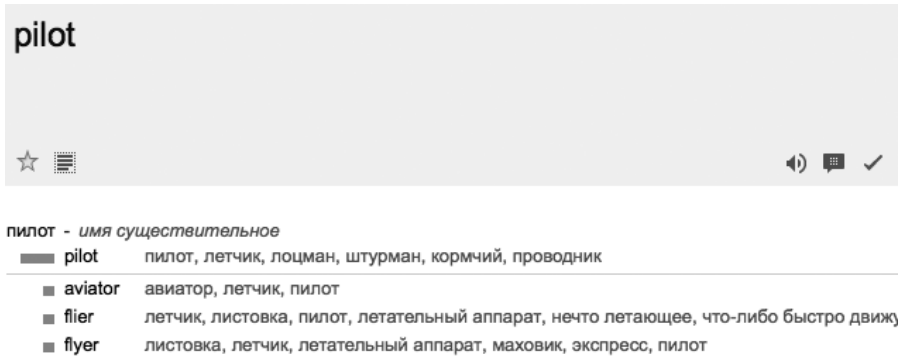


Рис. 2. Варианты перевода слова *пилот* в Google Translate

В настоящей работе делается попытка обосновать необходимость создания частотных словарей значений слов и предлагаются первые подходы к достижению этой цели.

2. Материал

Хотя задача создания частотных словарей значений предполагает саму возможность разделения слова на значения, на самом деле это сопряжено с большими трудностями. Известно, что носители языка часто дают не совпадающие ответы о наборах значений одних и тех же слов, разные лексикографы принимают разные решения на одном и том же материале, а в теоретической семантике высказываются мнения и о принципиальной недискретности полисемии (ср. обсуждение этих проблем, в частности, в работах [Апресян 1974/1995; Kilgarriff 1993; Pustejovsky 1996; Wilks 1998; Lin, Ahrens 2005; Зализняк Анна 2006; Иомдин 2014]). В практической лексикографии, однако, тем не менее делить слова на значения приходится. В частности, это очень существенно для переводных словарей, где объективным критерием выделения значений может считаться различие переводных эквивалентов. Кроме того, неопределенность полисемии неодинакова для разных частей речи и для разных семантических полей. Так, например, наборы значений и структура многозначности предлогов (*в, по* и др.) или глаголов положения в пространстве (*лежать, висеть* и др.) существенно различаются в разных словарях и для разных носителей, в то время как в предметной лексике значения часто выделяются более четко, поскольку возможно явное указание референтов, имеющих различные свойства (ср. *визитка* 'предмет одежды', 'сумочка', 'визитная карточка', *банан* 'фрукт', 'предмет одежды', 'водный аттракцион', 'оценка').

Для настоящей работы в качестве пробного материала мы отобрали семь достаточно частотных многозначных вокабул, имеющих конкретные предметные значения и описанных в Активном словаре русского языка (АС) [Апресян и др. 2014]: *альбом, билет, блок, вешалка, вилка, винт, горшок*, поставив

задачу выявить наиболее частотное значение каждого из этих слов. Для этого мы а) проанализировали существующие толковые словари, б) провели опрос информантов и в) исследовали наиболее частотные словосочетания с указанными словами на корпусном материале.

3. Исследование

3.1. Анализ толковых словарей

Для того чтобы определить набор и порядок значений исследуемых слов с точки зрения академической лексикографии, мы проанализировали их описание в восьми толковых словарях². Каждое значение мы называем при помощи короткого инварианта, составленного по материалам толкований из АС, а для тех значений, которые в АС отсутствуют, приводим упрощенное толкование из соответствующего словаря.

Для слова **альбом** в СУш, БАС 1, МАС и БТС приводятся два значения: 1) 'тетрадь или книга с чистыми листами' и 2) 'книга с репродукциями или фотографиями'. В СОШ, БАС 2, СЕф и АС, помимо этих двух, появляется третье значение — 'собрание музыкальных произведений' (в СЕф оно описывается в двух разных толкованиях: 'комплект из нескольких пластинок' и 'собрание песен, записанное на магнитную ленту').

Более интересной оказывается эволюция значений слова **билет**: в СУш приводится всего одно значение — 'документ, удостоверяющий право на услугу'; это же значение в БАС 1 идет под номером 3), а перед ним оказываются 1) 'документ, удостоверяющий членство в организации' и 2) 'денежный знак'. У данного слова в БАС 1 есть также значения 4) 'жребий, набор заданий на экзамене' и 5) 'документ, удостоверяющий личность'. В словарях, составленных позже (МАС, СОШ, БТС), значений всего четыре и они располагаются в следующем порядке: 1) 'документ, удостоверяющий право на услугу', 2) 'документ, удостоверяющий членство в организации', 3) 'денежный знак', 4) 'набор заданий на экзамене'. В АС значения 3) и 4) поменяны местами, т. е. более редким считается значение 'денежный знак'. В БАС 2 кроме этих четырех значений сохраняется 5) 'документ, удостоверяющий личность' с уточнением «в дореволюционной России».

Почти во всех словарях для слова **блок** выделяются два омонима: моносемичный со значением 'устройство для подъема грузов' и полисемичный, порядок и набор значений у которого варьируется от словаря к словарю. Первым значением второго омонима в СУш, БАС, МАС, СОШ, БТС и БАС 2 является значение 'группа организаций', в АС же данное значение — третье. Первыми, т. е. самыми актуальными, значениями в АС оказываются значения 1.1) 'большой

² Приведем список словарей в хронологическом порядке: СУш — 1935–1940 гг., БАС — 1948–1965 гг., МАС — 1981–1984 гг., СОШ — 1992 г., БТС — 1998 г., БАС 2 — 2004–2006 гг., СЕф — 2006 г. и АС — 2014 г.

сплошной брусок или плита' и 1.2) 'строительный элемент', которые, видимо, появились у слова **блок** достаточно давно (есть в МАС). Более новые значения — 2.1) 'несколько связанных помещений', 2.2) 'часть устройства' и 2.3) 'часть компьютерной программы' — либо зафиксированы в словарях, изданных позднее (значение 2.1 впервые встречается в БТС), либо пока не отмечены другими словарями (например, значение 2.3). В БТС в качестве значения 2) дается 'совокупность однородных предметов, понятий, явлений и т. п.', а в АС это большое значение разделено на два: 'совокупность однородных физических объектов' (*блок сигарет*) и 'совокупность информационных объектов' (*блок статей*) и стоит в слове **блок** последним. В БТС и БАС 2, кроме двух описанных омонимов, выделяется третий — в БТС он представлен одним значением 'защитный прием', а в БАС 2 к данному значению добавляется 'то же, что блокпост'.

Слово **вешалка** оказывается хорошим примером того, как особое употребление лексемы может развиваться в самостоятельное значение. Так, в СУш у слова два значения: 'приспособление с крючками' и 'петелька на одежде', они есть и во всех остальных рассмотренных словарях. В БАС у первого значения отмечено употребление 'планка с крючком, плечики', которое уже в СОШ дается как самостоятельное второе значение. В АС значения слова **вешалка** поданы в следующем порядке: 1.1) 'приспособление с крючками', 1.2) 'помещение гардероба', 2) 'планка с крючком, плечики' (с образным употреблением 'высокая худая женщина'), 3) 'петелька на одежде'.

Для слова **вилка** все словари выделяют как наиболее актуальное значение 'столовый прибор'. Вторым в СУш, БАС 1, МАС, СОШ, БТС, БАС 2 и СЕФ дается значение 'приспособление или устройство с раздвоенным концом'; в АС этот смысл выражен в значении 2) 'штепсель'. Более редкие значения — 'серия попыток' и 'положение в шахматной игре' (зафиксированы уже в БАС 1). Кроме того, достаточно новым можно считать значение 'крайние значения параметра'.

Первые два значения слова **винт** — 'предмет с резьбой для соединения деталей' и 'вращающаяся деталь' — описаны практически одинаково и в СУш, и в АС. Для данного слова изменения проявляются в развитии новых значений: 'элемент фигурного катания' и 'завинчивающаяся жестяная пробка' (БТС), 'спортивное упражнение, состоящее из вращения вокруг вертикальной оси тела' (БАС 2), 'жесткий магнитный диск' (СЕФ), 'наркотик' (АС) — а также в утрате неактуальных значений ('винтовка' — в СУш и 'карточная игра' — отдельное значение в СУш и БАС 1, омоним в МАС и БАС 2, в АС нет).

У слова **горшок** во всех словарях выделяются смыслы 'сосуд для приготовления пищи', 'сосуд для комнатных растений' и 'сосуд для естественных отправлений', причем в большей части словарей порядок именно такой. Различие состоит в статусе этих смыслов: в СУш, БАС 1 и МАС все это — употребления внутри одного значения, в БАС 2 последний смысл выделяется в отдельное значение, а в СЕФ и АС все представлены в виде разных значений. В СОШ у слова **горшок** два значения — 1) 'сосуд для приготовления пищи' и 2) 'сосуд для естественных отправлений' — а словосочетание *цветочный горшок* подается как фразема. Кроме этого, в БАС 2 есть устаревшее значение 'сосуд с горючими веществами, употреблявшийся в военном деле'.

Таким образом за 80 лет, прошедшие со времени выхода первого из рассмотренных толковых словарей, у всех семи слов в них менялись либо порядок значений, либо набор значений, либо статус приписываемых им смыслов (от употребления к значению или наоборот). В меньшей степени изменения коснулись слов *альбом* и *горшок*; больше всего изменений отражено в словарных статьях *блок* и *билет*.

3.2. Эксперимент с информантами

В феврале 2014 года мы провели эксперимент в интернете, состоящий из двух частей. Первая часть эксперимента выглядела следующим образом: информанты-носители русского языка должны были описать самое частое, по их мнению, значение каждого из данных выше семи слов и затем привести другие их значения (в том порядке, в котором они приходят в голову, без долгих размышлений). Значения разрешалось описывать разными способами: давать определение (можно краткое или неточное) или приводить примеры. При выполнении эксперимента не допускалось использование словарей и других источников. По завершении первой части информанту выдавалась вторая часть, где приводились краткие описания значений тех же семи слов, в основном адаптированные из АС. Для каждого значения каждого слова необходимо было оценить, часто ли оно, по мнению информанта, встречается в современных текстах на русском языке («часто», «средне», «редко», «никогда или почти никогда»).

В эксперименте приняло участие более 700 человек из более чем 100 населенных пунктов (около 30 % мужчин и около 70 % женщин), их средний возраст составил 33 года. 53 % информантов указали, что имеют законченное высшее образование, еще 16 % — степень кандидата наук и 2 % — степень доктора наук; среди остальных информантов было 17 % студентов, 7 % аспирантов, 3 % школьников и 2 % людей с законченным средним образованием.

Приведем некоторые результаты эксперимента.

В Таблице 1 в среднем столбце приведены данные о значениях, которые приводили информанты в первой части опроса, когда список значений еще не был им предъявлен («укажите самое частое, по вашему мнению, значение»), в правом столбце — процент информантов, оценивших соответствующее значение из выданного набора как «частое» во второй части опроса («часто ли это значение, по вашему мнению, встречается в современных текстах на русском языке»).

Таблица 1. Результаты опроса информантов

	привели первым	оценили как «частое»
Альбом		
Книга с репродукциями или фотографиями	42%	67%
Тетрадь или книга с чистыми листами	41%	56%
Собрание музыкальных произведений	8%	77%

	привели первым	оценили как «частое»
Билет		
Документ, удостоверяющий право на услугу	96 %	99 %
Набор заданий на экзамене	3 %	61 %
Документ, удостоверяющий членство где-л.	0 %	39 %
Денежный знак	0 %	2 %
Блок		
Строительный элемент	21 %	44 %
Упаковка сигарет	9 %	54 %
Устройство для подъема грузов	4 %	16 %
Часть устройства	3 %	43 %
Группа организаций	3 %	38 %
Спортивный прием	2 %	— ³
Вешалка		
Приспособление для хранения одежды	90 %	84 %
Помещение гардероба	<1 %	21 %
Худая женщина	<1 %	11 %
Неприятная ситуация	<1 %	— ⁴
Вилка		
Столовый прибор	95 %	99 %
Штепсель	2 %	81 %
Крайние значения параметра	0 %	13 %
Положение в шахматной игре	0 %	11 %
Серия попыток	0 %	1 %
Винт		
Предмет с резьбой для соединения деталей	68 %	90 %
Вращающаяся деталь	20 %	60 %
Наркотик	<1 %	16 %
Жесткий диск	<1 %	— ⁵
Горшок		
Сосуд для приготовления пищи	41 %	39 %
Сосуд для комнатных растений	16 %	89 %
Сосуд для естественных отпавлений	16 %	56 %

³ Это редкое значение, приведенное информантами, отсутствует в АС, поэтому не предлагалось в эксперименте изначально: ср. *Я поставил блок рукой, а левой стукнул ему в скулу* (С. Шаргунов).

⁴ Это новое сленговое значение, приведенное информантами, отсутствует в АС; ср. *С ребенком там совсем вешалки.. ни в магазин, ни в аптеку...* (Интернет-форум).

⁵ Этот омоним, использующийся в компьютерном сленге (сокращение от *винчестер*), отсутствует в АС; ср. *Т. е. для того чтобы записать по сетке видео на винт, подключенный к плееру, достаточно пары нажатий на пульте* (Компьютера.Лаб, 02.02.2010).

Как видно, значение, которое информанты выделяют как наиболее частотное до предъявления им полного списка значений, не всегда получает максимальную оценку при оценивании всего списка. Скажем, для слова *альбом* значение 'собрание музыкальных произведений' в первой части опроса назвали самым частотным лишь около 8% информантов, однако во второй части опроса именно это значение чаще всего оценили как «часто используемое» (77%); для слова *горшок* значение 'сосуд для комнатных растений' в первой части опроса назвали самым частотным лишь меньше 16% информантов, однако во второй части опроса именно это значение чаще всего оценили как «часто используемое» (89%).

Сложности предсказуемо возникли при наличии у слова большого числа трудно отделимых от друга и плохо формулируемых значений (ср. *блок*), а также в случаях, когда информанты давали слишком общее толкование, не позволявшее определить, какое именно значение имелось в виду (так, для слова *вешалка* в качестве первого значения чаще всего указывали 'то, на что вешают одежду', что не дает возможности различить значения 'приспособление с крючками' и 'планка или каркас с крючком, плечики').

Таким образом, однозначно определить наиболее частотные значения в представлении информантов удалось для трех слов: *билет* ('документ, удостоверяющий право на услугу'), *вилка* ('столовый прибор') и *винт* ('предмет с резьбой для соединения деталей').

3.3. Эксперимент с корпусом

Помимо опросов носителей, в современной лексикографической работе необходимо широкое использование корпусных данных. Самым объемным на сегодняшний день корпусом русского языка является собранный из текстов интернета корпус RuTenTen11, насчитывающий около 20 млрд словоупотреблений. Неоценимым удобством этого корпуса для лексикографов является его интеграция в систему Sketch Engine (www.sketchengine.co.uk, Kilgarriff et al. 2004), позволяющую получать списки наиболее частотных коллокаций с данным словом. Мы исследовали такие списки для указанных семи слов (первые десять самых частотных коллокаций в каждой синтаксической конструкции, предлагаемой системой Sketch Engine, в случае, если для этой конструкции система выдавала больше десяти коллокаций; в противном случае конструкция не учитывалась). Каждой коллокации два лексикографа-соавтора статьи независимо друг от друга сопоставили одно из значений, содержащихся в АС (случаи, где сделать это было трудно или невозможно, отмечались особо)⁶. Приведем пример такого аннотирования по одной конструкции для одного слова (всего для

⁶ Известно, что эксперты-носители языка по-разному оценивают конкретные употребления, что затрудняет работу над созданием аннотированных корпусов текстов (inter-judge variance). Так, в работе [Fellbaum et al. 1997] было показано, что на решения аннотаторов влияло расположение значений в списках, которые им выдавались; в работе [Snyder and Palmer 2004] несогласие аннотаторов связывается со слишком дробным делением на значения, принятым в тезаурусе WordNet [Июмдин 2014].

каждого слова использовались данные от 10 до 30 конструкций, то есть аннотировались от 100 до 300 частотных словосочетаний с каждым словом).

Таблица 2. Набор значений для слова вешалка (адаптированный из АС)

1	Приспособление с крючками (<i>У двери стоит вешалка</i>)
2	Помещение гардероба (<i>Театр начинается с вешалки</i>)
3	Планка или каркас с крючком, «плечики» (<i>Платье на вешалке</i>)
4	Худая женщина
5	Петелька на одежде (<i>пришить вешалку к куртке</i>)

Таблица 3. Пример работы аннотаторов

Самые частотные коллокации в конструкции с прямым дополнением:

	Значение (аннотатор 1)	Значение (аннотатор 2)
<i>прибивать</i>	1	1
<i>пришивать</i>	5	5
<i>смастерить</i>	1	1
<i>опрокинуть</i>	1	1
<i>уронить</i>	1,3	1,3
<i>повесить</i>	1,3	1,3
<i>закрепить</i>	1	1
<i>спланировать</i>	1	2
<i>крепить</i>	1	1
<i>прикреплять</i>	1	1

Затем для каждого значения было подсчитано суммарное количество отнесенных к нему коллокаций, помноженное на их общую частотность в корпусе. Если какая-то коллокация была отнесена к двум или нескольким значениям, ее частота учитывалась при подсчете частотности каждого из этих значений. Если результаты двух аннотаторов различались, подсчитывалось среднее арифметическое между получающимися частотами. См. формулы:

$$f_s(i) = \frac{1}{N_a} \sum_a \sum_w f(w, c = i_a)$$

$$f_s = \sum_w f_s(w)$$

$$F = \sum_s F_s$$

$$f(i) = \sum_s \left(\frac{f_s(i)}{f_s} \cdot \frac{F_s}{F} \right)$$

- N_a — Количество участвующих в разметке аннотаторов
 $f_s(w)$ — Частота коллокации w в синтаксической конструкции s
 $f_s(i)$ — Усредненная по аннотаторам частота класса i для синтаксической конструкции s
 f_s — Суммарная частота всех рассматриваемых коллокаций для данной синтаксической конструкции
 F_s — Частота синтаксической конструкции в корпусе
 F — Суммарная частота всех рассматриваемых синтаксических конструкций s

Таким образом были получены следующие результаты (число в процентах показывает, какая доля исследованных словосочетаний предположительно относится к соответствующему значению):

Таблица 4. Результаты аннотирования коллокаций

	F_s
Альбом	
Собрание музыкальных произведений	80%
Книга с репродукциями или фотографиями	38%
Тетрадь или книга с чистыми листами	5%
Билет	
Документ, удостоверяющий право на услугу	95%
Документ, удостоверяющий членство где-л.	4%
Набор заданий на экзамене	3%
Денежный знак	<1%
Блок⁷	
Часть устройства	41%
Устройство для подъема грузов	23%
Строительный элемент	22%
Большой сплошной брусок или плита	21%
Часть компьютерной программы	18%
Несколько связанных помещений	12%
Совокупность текстов	10%

	F_s
Группа организаций	9%
Упаковка сигарет	3%
Вешалка	
Приспособление с крючками	82%
Планка с крючком, «плечики»	39%
Помещение гардероба	14%
Петелька на одежде	2%
Вилка	
Столовый прибор	53%
Штепсель	25%
Крайние значения параметра	<1%
Положение в шахматной игре	<1%
Серия попыток	<1%
Винт	
Предмет с резьбой для соединения деталей	57%
Вращающаяся деталь	41%
Предмет в форме спирали	5%
Наркотик	<1%
Горшок	
Сосуд для комнатных растений	74%
Сосуд для приготовления пищи	34%
Сосуд для естественных отправлений	30%

⁷ Мы не приводим здесь данных об омониме *Блок* (фамилия), коллокации с которым также встречались среди частотных (*подржать Блоку, Блок и Маяковский и т. п.*)

Как видим, этот эксперимент позволяет достаточно надежно выявить наиболее частотные и наименее частотные значения для каждого из семи слов (если считать, что исследованный корпус адекватно отражает их распределение в языке)⁸.

4. Заключение

Итак, задача создания частотного словаря значений слов, во всяком случае в области предметной лексики, кажется теоретически выполнимой. Интересно, что если эксперимент, в котором информанты пытаются определить наиболее частотное значение спонтанно, далеко не всегда дает показательные результаты, то оценки частотности, данные носителями языка при предъявлении словарного списка значений, достаточно хорошо соотносятся с результатами экспертной оценки коллокаций, полученных в корпусе. Как кажется, сочетание этих методик может быть взято за основу при работе над частотным словарем значений. Первым шагом мог бы стать словарь, в котором для каждого многозначного слова было бы указано наиболее частотное значение, если оно надежно выделяется на основе экспериментов с носителями и корпусных исследований.

С практической точки зрения, такой словарь мог бы быть использован в различных компьютерно-лингвистических приложениях (в частности, для вероятностного разрешения многозначности в отсутствие контекста, см. также об этой проблеме работу [Иомдин 2014]), в создании лексических минимумов, разговорников, учебников и обучающих ресурсов, в традиционной толковой лексикографии (в частности, описанные методы уже используются при работе над Словарем бытовой терминологии (см. [Иомдин 2011, Иомдин и др. 2012, 2013]), в котором дается информация о частотности).

С теоретической точки зрения, регулярные исследования наборов значений многозначных слов и их сравнительной частотности представляют интерес для изучения эволюции лексической системы языка. Так, интересно, что сразу в нескольких рассмотренных словах на первый план выходят значения, утратившие связь с этимологически первоначальным смыслом [по Фасмер 1986]. Ср. *альбом*, восходящее к лат. *album* 'открытый лист для сбора подписей', букв. 'белый': в наиболее частотном сейчас «музыкальном» значении идея чистоты утрачивается, а разрабатывается идея совокупности; *блок*, восходящее к германскому корню **blok-* 'твердый ствол': в наиболее частотных значениях утрачивается идея твердости, но разрабатывается идея 'часть целого'; *горшок*, производное от *горн* 'плавильная печь': в наиболее частотном сейчас значении

⁸ Для более объективных результатов при работе над словарем необходимо учитывать данные разных источников, что пока осложняется отсутствием подобной Sketch Engine системы автоматического подбора частотных коллокаций, во всяком случае для русского языка. Кроме того, при наличии соответствующей разметки корпуса возможны и интересны также исследования особенностей социального, регионального, возрастного, гендерного распределения частотности значений.

‘сосуд для комнатных растений’ утрачивается связь с печью. Динамика частоты значений важна и для сравнительных исследований, прежде всего на материале региональных вариантов, диалектов и родственных языков.

Литература

1. *Апресян* 1974/1995 — Апресян Ю. Д. Лексическая семантика. Синонимические средства языка. М., 1974 (2-е изд.: М., 1995).
2. *Апресян* и др. 2014 (АС) — Активный словарь современного русского языка. А-Г / Под ред. Ю. Д. Апресяна. М.: «Языки славянских культур», 2014 (в печати).
3. *БАС 1* — Словарь современного русского литературного языка: В 17-ти т. / Изд-во АН СССР; Под ред. В. И. Чернышева. М., Л., 1948–1965.
4. *БАС 2* — Большой академический словарь русского языка / ИЛИ РАН; Под ред. К. С. Горбачевича. М, СПб.: «Наука», 2004–2006. Тт. 1–4.
5. *БТС* — Большой толковый словарь русского языка / Сост., гл. ред. С. А. Кузнецов. СПб.: Норинт, 1998.
6. *Зализняк* 2006 — Зализняк Анна А. Многозначность в языке и способы ее представления. М., 2006.
7. *Иомдин* 2011 — Иомдин Б. Л. Материалы к словарю-тезаурусу бытовой терминологии. СВИТЕР: образец словарной статьи // Слово и язык. Сборник статей к восьмидесятилетию академика Ю. Д. Апресяна. Отв. ред. И. М. Богуславский, Л. Л. Иомдин, Л. П. Крысин. М.: «Языки славянских культур», 2011. С. 392–406.
8. *Иомдин* и др. 2012 — Иомдин Б. Л., Лопухина А. А., Пиперски А. Ч., Киселева М. Ф., Носырев Г. В., Рикитянский А. М., Васильев П. К., Кадыкова А. Г., Матиссен-Рожкова В. И. Словарь бытовой терминологии: новые проблемы и новые методы // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2012» (Бекасово, 30 мая — 3 июня 2012 г.). Вып. 11 (18). М.: РГГУ, 2012. С. 213–226.
9. *Иомдин* и др. 2013 — Иомдин Б. Л., Лопухина А. А., Панина М. Ф., Носырев Г. В., Вилл М. В., Зайдельман Л. Я., Матиссен-Рожкова В. И., Винокуров Ф. Г., Выборнова А. Н. Маг вел мот: изменения в языке на материале бытовой терминологии // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2013» (Бекасово, 29 мая — 2 июня 2013 г.). Вып. 12 (19): в 2 т. Т. 1. М.: РГГУ, 2013, с. 311–324.
10. *Иомдин* 2014 — Многозначные слова в контексте и вне контекста // Вопросы языкознания, 2014, №4 (в печати).
11. *Ляшевская и Шаров* 2009 — О. Н. Ляшевская, С. А. Шаров, Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М.: Азбуковник, 2009.
12. *МАС* — Словарь русского языка: В 4-х т. /АН СССР, Ин-т рус. яз.; Под ред. А. П. Евгеньевой. М.: Русский язык, 1985–1988.

13. *СЕф* — Ефремова Т. Ф. Большой современный толковый словарь русского языка. В 3-х т. М., АСТ, Астрель, 2006.
14. *СОШ* — Ожегов С. И. и Н. Ю. Шведова. Толковый словарь русского языка. М.:Азъ, 1992.
15. *СУш* — Толковый словарь русского языка / Под ред. Д. Н. Ушакова. М.: Гос. ин-т Сов.энцикл.; ОГИЗ; Гос. изд-во иностр. и нац. словарей, 1934–1940.
16. *Фасмер* 1986 — Фасмер М. Этимологический словарь русского языка. Тома I–IV. М., 1986.
17. *Beck et al.* 2013 — Beck, Isabel L., Margaret G. McKeown, and Linda Kucan. *Bringing words to life: Robust vocabulary instruction.* Guilford Press, 2013.
18. *Fellbaum et al.* 1997 — Fellbaum C., Grabowski K., Shari L. *Analysis of a Hand-Tagging Task.* In: *Proceedings of ANLP-97 Workshop on Tagging Text with Lexical Semantics.* 1997.
19. *Kilgarriff* 1993 — Kilgarriff A. *Dictionary word sense distinctions: an enquiry into their nature // Computers and the humanities.* 1993. V. 26. No. 1–2.
20. *Kilgarriff et al.* 2004 — Kilgarriff A., Rychly, P., Smrz, P., & Tugwell, D. (2004). *ITRI-04-08 The Sketch Engine.* *Information Technology*, 105, 116.
21. *Lin, Ahrens* 2005 — Lin C. C., Ahrens K. *How many meanings does a word have? Meaning estimation in Chinese and English // J. W. Minett, W.S.-Y. Wang (eds). Language acquisition, change and emergence: Essays in evolutionary linguistics.* Hong Kong, 2005.
22. *Pustejovsky* 1996 — Pustejovsky J. *Lexical semantics: The problem of polysemy.* Oxford, 1996.
23. *Snyder and Palmer* 2004 — Snyder B., Palmer M. *The English all-words task // Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text.* 2004. P. 41–43.
24. *Wilks* 1998 — Wilks Y. *Senses and texts // Computational linguistics and Chinese language processing.* 1998. V. 3. No. 2.

References

1. *Apresjan Ju. D.* (1974/1995), *Lexical semantics. The synonymical means of language [Leksicheskaja semantika. Sinonimicheskie sredstva jazyka].* Moscow.
2. *Apresjan Ju. D.* (ed.). (2014), *Active Dictionary of Modern Russian. A-G [Aktivnyj slovar' sovremennogo russkogo jazyka. A-G].* *Jazyki slavjanskih kul'tur*, Moscow.
3. *Chernyšov V. I.* (ed.). (1948–1965), *Large Contemporary Dictionary of Standard Russian [Slovar' sovremennogo russkogo literaturnogo jazyka].* Moscow, St. Petersburg.
4. *Gorbachevich K. S.* (ed.). (2004–2006), *Large Academic Dictionary of Russian [Bol'shoj akademicheskij slovar' russkogo jazyka].* Moscow, St. Petersburg.
5. *Efremova T. F.* (2006), *Large Contemporary Explanatory Dictionary of Russian [Bol'shoj sovremennyj tolkovyj slovar' russkogo jazyka].* AST, Astrel', Moscow.
6. *Evgenyeva A. P.* (ed.). (1981–1984), *Russian Language Dictionary [Slovar' russkogo jazyka].* *Russkij jazyk*, Moscow.

7. *Iomdin B. L.* (2011), Materials for the thesaurus of Russian everyday life terminology. SWEATER: a sample dictionary entry [Materialy k slovarju-tezaurusu bytovoj terminologii. SVITER: obrazets slovarnoj stat'i]. Slovo i jazyk. Sbornik statej k vos'midesiatiletiju akademika Ju.D. Apresjana [The word and the language. A collection of papers to commemorate Academician Apresjan's 80th anniversary]. Jazyki slavjanskih kul'tur, Moscow, pp. 392–406.
8. *Iomdin B. L., Lopuhina A. A., Piperski A. Ch., Kisel'eva M. F., Nosyrev G. V., Rikitjanskij A. M., Vasil'jev P.K., Kadykova A. G., Matissen-Rozhkova V. I.* (2012), Thesaurus of Russian everyday life terminology: new problems and new techniques [Slovar' bytovoj terminologii: novye problemy i novye metody]. Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog 2012". Bekasovo, pp. 213–226.
9. *Iomdin B. L., Lopuhina A. A., Panina M. F., Nosyrev G. V., Vill M. V., Zajdel'man L. Ja., Vinokurov F. G., Matissen-Rozhkova V. I., Vybornova A. N.* (2013), Mag vel mot: language innovations in everyday life terminology [Mag vel mot: izmenenija v jazyke na materiale bytovoj terminologii]. Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog 2013". Bekasovo, pp. 311–324.
10. *Iomdin B. L.* (2014), Polysemous words within and without context [Mnogoznachnye slova v kontekste i vne konteksta]. Voprosy jazykoznanija [Issues in Linguistics]. Vol. 4. Moscow (in print).
11. *Kilgarriff A.* (1993), Dictionary word sense distinctions: an enquiry into their nature. In: Computers and the humanities. V. 26. No. 1–2.
12. *Lin C. C., Ahrens K.* (2005). How many meanings does a word have? Meaning estimation in Chinese and English. In: J. W. Minett, W. S.-Y. Wang (eds). Language acquisition, change and emergence: Essays in evolutionary linguistics. Hong Kong.
13. *Lyasevskaya O. N., Sharov S. A.* (2009) A Frequency Dictionary of Modern Russian (based on National Russian Corpus) [Chastotnyj slovar' sovremennogo russkogo jazyka]. Moscow: Azbukovnik.
14. *Kuznetsov S. A.* (ed.). (1998), Large Explanatory Dictionary of Russian [Bol'shoj tolkovyj slovar' russkogo jazyka]. Norint, St. Petersburg.
15. *Ozhegov S. I., Shvedova N. Yu.* (1992), Explanatory Dictionary of Russian [Tolkovyj slovar' russkogo jazyka]. Az, Moscow.
16. *Pustejovsky J.* (1996), Lexical semantics: The problem of polysemy. Oxford.
17. *Ushakov D. N.* (ed.). (1934–1940), Explanatory Dictionary of Russian [Tolkovyj slovar' russkogo jazyka]. OGIZ, Moscow.
18. *Vasmer M.* (1986), Etymological dictionary of Russian. I-IV [Etimologičeskij slovar' russkogo jazyka]. Moscow.
19. *Wilks Y.* (1998), Senses and texts. In: Computational linguistics and Chinese language processing. V. 3. No. 2.
20. *Zaluzniak Anna* (2006), Polysemy in language and its representation [Mnogoznachnost' v jazyke i sposoby ee predstavlenija]. Moscow.