

© 2014 г. Б.Л. ИОМДИН

МНОГОЗНАЧНЫЕ СЛОВА В КОНТЕКСТЕ И ВНЕ КОНТЕКСТА*

В статье затрагиваются актуальные вопросы выявления и описания лексической многозначности. Эта проблема выступает на первый план в современной компьютерной лингвистике. Существующие сейчас методы автоматического разрешения многозначности во многих отношениях различаются, однако их роднит общая черта – опора на контекст. Между тем изучение лексической многозначности вне контекста и исследование того, как многозначные слова представлены в сознании носителей языка, тоже представляет значительный научный и практический интерес. В статье кратко характеризуются основные подходы к разрешению лексической многозначности с учетом контекста в узком смысле (окружающий текст) и широком смысле (ситуационный контекст) и уточняется роль корпусных методов при выделении и описании значений слов. Приводятся экспериментальные данные, позволяющие выделять наборы лексем полисемичных слов, различающиеся для разных групп информантов, и ранжировать лексемы внутри одной вокабулы. Ставится проблема различения относительной и абсолютной многозначности и особенностей их лексикографического описания.

Ключевые слова: семантика, лексикография, компьютерная лингвистика, многозначность, полисемия, омонимия, разрешение многозначности, опросы, эксперименты, частота, норма, узус, идиолекты

The paper discusses various techniques of discovering and describing lexical ambiguity. This is one of the top issues in computational linguistics. A variety of techniques are used for word sense disambiguation, but all of them are based on context. Yet, studying how word senses work without context and what patterns of polysemous words could be found in speakers' minds also seems an interesting and important issue. The main approaches to WSD with or without context (in narrow and broad sense, including the situational context) are evaluated. The importance of corpora in discovering word senses is substantiated. New experimental data are presented, which allow defining subsets of senses for polysemous words for different speakers and rating the senses in the dictionary. Finally, the paper proposes to distinguish between absolute and relative polysemy and to search for ways of their adequate lexicographic description.

Keywords: semantics, lexicography, computational linguistics, ambiguity, polysemy, homonymy, WSD, surveys, experiments, frequency, norm, usage, idiolects

ВВЕДЕНИЕ

Многозначность была и остается одним из важнейших свойств лексической системы языка, а ее изучение представляет несомненный интерес как с теоретической, так и с практической точки зрения (в первую очередь в лексикографии). Она пронизывает все слои языка и особенно характерна для активной части лексики. Вспомним, что еще в работах Дж. Ципфа было показано существование корреляции между частотностью слова и количеством его значений [Zipf 1945].

* Работа выполнена при 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.

В справедливости этой закономерности легко убедиться на материале современной активной лексикографии. Так, в первых двух томах (А–Г) «Активного словаря русского языка» [Апресян и др. 2014] вокабул, у которых выделяется только одна лексема и нет омонимов, меньше четверти. При этом и у таких вокабул часто можно найти дополнительные значения или омонимы, не вошедшие в словарь из-за своей стилистической отмеченности, распространенности лишь в определенных регионах или достаточно низкой употребительности. Например: *афиша* 2 ‘(толстое) лицо’ (в уголовном жаргоне, см., например, [Грачев 2005]), ср. ниже пример (1); *байка*² ‘предмет верхней одежды из трикотажа’ (в вариантах русского языка в Белоруссии и Латвии, см. [Языки городов]), ср. (2)–(3); *балка*² ‘овраг, ложбина’ (см., например, МАС, СЕФ), ср. (4); *баян*² ‘старая, давно известная информация’ (в интернет-сленге), ср. (5); *баян*³ ‘шприц’ (в сленге наркоманов, см., например, [Никитина 2003]), ср. (6); *бомба* 2 ‘большая шпиргалка’ (в студенческом сленге, см., например, [Там же]), ср. (7); редк. *босоножка* 2 ‘босая девочка или женщина’ (см., например, СОШ, СШ, МАС), ср. (8); уходящ. *босоножка* 3 ‘танцовщица, выступающая босой’ (см., например, СУШ), ср. (9); разг. спец. *бумажник*² ‘рабочий в бумажной промышленности’ (см., например, СОШ, СШ, СЕФ), ср. (10); спец. *бухта*² ‘канат, сложенный кольцами’ (см., например, МАС), ср. (11); *бык* 2 ‘агрессивный человек, предпочитающий разрешать конфликты грубым физическим насилием’ (в сленге; ср. *бычить*, *быковать*), ср. (12); спец. *бык*² ‘промежуточная опора моста’ (см., например, СОШ, СШ, МАС), ср. (13); *весить* 2 ‘иметь определенный объем’ (о файлах, в интернет-сленге), ср. (14); *годный* 2 ‘хороший’ (в молодежном сленге), ср. (15), и др. Все эти лексемы, впрочем, могут встретиться и в современных неспециальных текстах:

- (1) *Марина: Она проникает к тебе в дом, афиша у нее броская... Певцов: Какая афиша? Марина: По-вашему – лицо* [Э. Брагинский. Авантюристка (1989)].
- (2) *В секунд-хэнде на улице Рокоссовского немногочисленно. Мужчины выбирают джинсы, молодые люди байки* [«Комсомольская правда в Беларуси», 12.09.2011].
- (3) *Мы надевали белые высокие кроссовки с развязанными шнурками, джинсы и байки типа «кенгуру» с капюшоном* [«Час», Рига, 19.04.2002].
- (4) *Так я перешел через балку, услышал дважды свист, принесенный ветром: егеря, выходя на патрулирование в параллельный маршрут, пересвистывались* [А. Иличевский. Перс (2009)].
- (5) *Ребят ну реально может хватит присылать картинки и публиковать это? в основном это баян и люди их видели* (Интернет-форум)¹.
- (6) *Как известно уже, увы, большинству носителей великого и могучего, «баян» – это шприц, а Ширянов – от слова «ширяться»* (Труд-7, 21.10.2006).
- (7) *Но иной раз видишь, как в лаборатории, спрятавшись за приборы, девушки вяжут; как на экзамене студент нагло достает шпиргалку или даже «бомбу» (готовый ответ на отдельном листке)* [«Студенческий меридиан» (1984)].
- (8) *Никак не могли взять в толк, почему сам капитан Бох снисходит до разговора с этой смуглой босоножкой, нищенкой, утопленницей, случайно оставшейся в живых, безродной девочкой* [Ю. Буйда. Город палачей (2003)].
- (9) *А еще я помню гастроль в Харбине Айседоры Дункан. Знаменитая «босоножка» уже пережила в ту пору расцвет своей карьеры* [Л. Лопато. Волшебное зеркало воспоминаний (2002–2003)].
- (10) *21 января 2005 – газете «Сегежский бумажник» 65 лет* (сайт газеты Открытого акционерного общества «Сегежский ЦБК»).
- (11) *Весь Левин сюжет легко свертывается кольцами, образуя как бы бухту каната или спящую змею* [«Звезда» (2002)].
- (12) *Так, выступая перед собравшимися, заместитель председателя Николаевской областной организации ВО «Свобода» Наталья Грудина назвала действующую власть «властью гопников и быков»* (Портал украинского информационного агентства «Преступности НЕТ»).
- (13) *Ослепительно-белые льдины проплывали по черной, будто маслянистой воде, с легким шорохом касались берегов, потрескивали и шумели, налетая на быки мостов* [Е. Хасецкая. Синие стрекоты Вавилона (1997)].

¹ Здесь и далее в текстах из Интернета сохраняется авторская орфография и пунктуация.

- (14) *Внимание: загруженный файл должен «весить» не более 100 мб* («Комсомольская правда», 11.08.2009).
- (15) *Смотрю Whitechapel. Нравится, годный сериал. Посмотрел первую серию, твердая четверка с плюсом* (Интернет-форум).

Задача настоящей статьи – рассмотреть актуальные проблемы выявления и лексикографического описания лексической многозначности. Поскольку в последние десятилетия проблема многозначности наиболее остро стоит в компьютерной лингвистике, в первом разделе статьи мы кратко резюмируем существующие подходы к ее решению в этой области. Во втором разделе мы описываем проведенные нами эксперименты, результаты которых могут способствовать выделению наборов лексем полисемичных слов, различающихся для разных групп носителей языка, и ранжировать лексемы внутри одной вокабулы, и сопоставляем их с данными толковых словарей. В третьем разделе обсуждается роль корпусных исследований при выделении и описании значений слов. Наконец, в заключении ставится проблема различения относительной и абсолютной многозначности и ее лексикографического описания.

1. МНОГОЗНАЧНЫЕ СЛОВА КАК ПРОБЛЕМА КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ

С развитием компьютерной лингвистики многозначность становится одной из главных проблем для систем автоматической обработки текстов. Проблеме автоматического разрешения лексической многозначности (word sense disambiguation, WSD) посвящены сотни работ, с 1998 г. регулярно проводятся посвященные ей международные рабочие семинары (SensEval, позднее SemEval), на которых тестируются автоматические и полуавтоматические системы идентификации лексических значений. Стоит сразу подчеркнуть, что в сфере компьютерной лингвистики различие между полисемией и омонимией, последовательно проводимое в теоретической семантике и (до последнего времени) в лексикографии (ср., однако, работу [Качурин 2014], где в частности на основе экспериментов с информантами предлагается ввести в лексикографическую практику «шкалу семантической связи между двумя значениями, крайними точками которой являются омонимия и полисемия» [Качурин 2014:16]), нерелевантно, поскольку для решения задачи определения значения слова в тексте наличие или отсутствие смысловой связи между возможными кандидатами несущественно. В настоящей статье мы также не затрагиваем проблему различения полисемии и омонимии и оперируем более общим термином «многозначность»² (ср., например, [Рахилина и др. 2006]).

В цели настоящей статьи не входит подробное описание методов, используемых в компьютерной лингвистике для решения проблемы автоматического разрешения лексической многозначности. Их развитие можно проследить, в частности, по обзорам [Ide, Véronis 1998; Agirre, Edmonds (eds) 2006; Navigli 2009]. Важно отметить, что эти методы основаны на использовании толковых словарей, тезаурусов или корпусов текстов.

² Более того, при автоматической обработке текста во многих случаях несущественно и различие между омонимами (разными вокабулами с совпадающей словарной формой) и омоформами (совпадающими формами разных вокабул). С учетом этого в первых двух томах «Активного словаря русского языка» остается еще меньше вокабул, не представляющих трудностей при автоматическом разрешении многозначности, так как многие из них имеют омоформы. Ср.: *байка* (*байка*¹, *байка*² или форма слова *байк* 'велосипед или мотоцикл'), *банту* (форма слов *бант* или *банту* 'народы в Африке'), *бегам* (наречие или форма слова *бег*), *бывало* (вводное слово или форма глагола *быть*), *вдов* (форма существительного *вдова* или прилагательного *вдовый*), *века* (форма слов *век* или *веко*), *видно* (вводное наречие или форма прилагательного *видный*), *винил* (существительное или форма глагола *винить*), *виски* (форма слов *виски* или *висок*), *вою* (форма существительного *вой* или глагола *выть*), *гною* (форма существительного *гной* или глагола *гноить*) и др. Неоднозначность такого рода может возникать и в реальной речевой практике, особенно с недостаточным контекстом или вне контекста.

Так, М. Леск еще в 1986 г. предложил использовать электронную версию словаря Oxford advanced learner's dictionary of Current English для автоматического определения значения слова в заданном словосочетании. Например, его программа определяла, что в сочетании *pine cone* слово *pine* с наибольшей вероятностью используется в значении 'сосна' (а не 'ананас', 'лимонка', 'томиться', 'тосковать' и др.), а слово *cone* – в значении 'шишка' (а не 'конус', 'рожок', 'пучок' и др.), поскольку в словарных толкованиях этих лексем больше всего пересечений (*pine* 1 'kind of evergreen trees...' и *cone* 3 'fruit of certain evergreen trees') [Lesk 1986]³. В целом ряде разработок используется улучшенный алгоритм Леска: из русскоязычных работ см. хотя бы [Гельбух, Сидоров 2004; Бабанов 2012]. Сходным образом действуют методы, использующие тезаурусы (чаще всего тезаурус Роже, сеть WordNet, Википедию): [Yarowski 1992; Jarmasz, Szpakowicz 2001; Mohammad, Hirst 2006; Лукашевич, Добров 2007; Lapata, Keller 2007; Турдаков 2010; Tejada-Cárcamo et al. 2010] и др.

Другое важное направление, оформившееся несколько позднее, – использование корпусов текстов: разрешение лексической многозначности осуществляется с опорой на сочетаемость. Огромные объемы современных корпусов позволяют делать статистически значимые наблюдения о совместной встречаемости слов в разных значениях. При этом лучшие результаты дают аннотированные корпуса, в которых у всех или у какой-то части слов размечены значения (полуавтоматически или вручную). Однако создание таких корпусов сопряжено с огромными трудозатратами. Ср. подробное описание лексико-семантической разметки в НКРЯ в работах [Кобрицов 2004; Кустова и др. 2005; Рахилина и др. 2006; 2009]. В [Апресян и др. 2005] описан опыт разработки семантической классификации и ее использования для создания глубоко аннотированного корпуса SynTagRus.

В последнее время перспективными оказываются разработки, совмещающие использование словарей и корпусов текстов: [Кобрицов и др. 2007; Roth, Schulte im Walde 2008]. Тем не менее проблема автоматического разрешения неоднозначности пока отнюдь не потеряла актуальность, и это легко может заметить, например, любой пользователь поисковых систем. Основные трудности, на которые наталкиваются разработчики, – отсутствие единообразного описания значений и недостаточная системность существующих лексикографических источников (убедительно продемонстрированная в работах Ю.Д. Апресяна о системной лексикографии, см. хотя бы введение к [Апресян и др. 2010]), ср. также примеры несистемности словарных описаний в области предметной лексики [Иомдин 2009; 2011]), принципиальная недискретность полисемии и неоднородность решений, принимаемых разными лексикографами⁴ (ср. обсуждение этих проблем в работах [Апресян 1974; Kilgariff 1993; Pustejovsky 1996; Lin, Ahrens 2005; Зализняк 2006]). Для настоящей статьи важно отметить еще одну проблему: частое несогласие экспертов (носителей языка) при выделении значений слов и интерпретации конкретных употреблений, затрудняющее 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, Palmer 2004] несогласие аннотаторов связывается со слишком дробным делением на значения, принятым в тезаурусе WordNet.

³ Заметим, что те же идеи лежат в основе «закона семантического согласования», описанного, в частности, в [Гак 1972] и в строгой форме сформулированного в [Апресян 1974].

⁴ «Lexicographers traditionally distinguish “lumpers” and “splitters” among colleagues: those who tend to break up senses further and those who go for large, homonymic, senses, of which Wierzbicka would be the extreme case» [Wilks 1998: 276] («Лексикографы традиционно делят своих коллег на “любителей объединять” и “любителей разделять”: тех, кто стремится к более дробным значениям, и тех, кто описывает широкие, омонимичные значения – автором наиболее крайнего проявления последнего подхода следует считать Вежбицкую»; здесь и далее перевод мой. – Б.И.).

2. МНОГОЗНАЧНЫЕ СЛОВА В ЭКСПЕРИМЕНТЕ

Как разнообразны бы ни были методы автоматического разрешения многозначности, их объединяет одно: опора на контекст. Ср.: «Context is the only means to identify the meaning of a polysemous word. Therefore, all work on sense disambiguation relies on the context of the target word to provide information to be used for its disambiguation» («Контекст – это единственное средство идентификации значения многозначного слова. Таким образом, вся деятельность по разрешению многозначности опирается на контекст анализируемого слова, в котором и содержится информация, используемая для определения его значения») [Ide, Véronis 1998: 18]. Более того, и в некоторых работах по теоретической семантике высказываются мнения, что значения существуют исключительно в контексте, ср.: «Do meanings also exist outside the transactional contexts in which they are used? I would argue that they do not» («Существуют ли значения также и вне операционных контекстов, в которых они используются? Я готов утверждать, что не существуют») [Hank 2013: 73].

Однако представляется, что задача определения наиболее вероятного значения многозначного слова вне контекста также имеет смысл. Помимо возможных практических применений (скажем, для ранжирования результатов поиска при вводе в поисковую систему однословного многозначного запроса: например, при вводе запроса «банки» одни пользователи могли бы прежде всего получить информацию о банковских организациях, а другие – о стеклянных емкостях), она представляет и теоретический интерес – в частности, для поиска ответа на вопрос, как соотносятся разные значения слова в мозгу разных носителей языка и насколько возможно предсказать их реакцию на соответствующие вербальные стимулы.

Исследований такого рода меньше, чем работ в области автоматического определения значения в контексте, хотя соответствующие эксперименты и проводятся (ср., например, [Foraker, Murphy 2012]). Приведем характерную цитату из недавней монографии О. Квонг: «Believe it or not, there is literally almost no study in the vast volume of psychology literature which tells us how humans disambiguate word senses. Rather, indirect evidence is found within the broader topic of human lexical processing, which mainly deals with the organization of words and the representation of meanings in the mental lexicon, and the mechanisms of lexical access, especially for words with multiple meanings [Kwong 2012: 53] («Трудно поверить, но в литературе по психологии не существует практически ни одного исследования, которое бы показало, как человек определяет значение многозначного слова. Есть лишь косвенные свидетельства в более широкой сфере лексической обработки текстов человеком, в основном связанные с организацией слов и представлением значений в ментальном лексиконе и механизмами лексического доступа, в том числе для многозначных слов»)). В психолингвистике, правда, широко распространены эксперименты по выявлению свободных ассоциаций, однако они редко соотносятся со списками значений и данными толковых словарей. Так, в эксперименте [Durkin, Manning 1989] носители английского языка должны были предложить значения 175 многозначных английских слов; авторы ранжировали ответы информантов по частотности и определяли доминирующие значения для каждого слова.

В таких экспериментах, впрочем, следует учитывать эффект прайминга (поскольку этот термин несколько по-разному употребляется в психологии, психолингвистике и когнитивной лингвистике, ср. [Seidenberg et al. 1982; Rayner, Frazier 1989; Vue et al. 2003; Ноеу 2005; Kwong 2012; Заботкина, Боярская 2011; 2012], уточним, что мы имеем в виду влияние ситуативного или текстуального контекста на актуализацию того или иного значения). Так, на конкурсе «Русский медвежонок – языкознание для всех» в 2005 г. школьникам 10 и 11 классов была предложена следующая задача: «Вот список слов: *джаз, рок, поп, блюз*. Сколько слов из него есть в “Толковом словаре живого великорусского языка” В.И. Даля, изданном в 60-е годы XIX века?» (автор задачи И.А. Рубанов). В конкурсе участвовали 182 577 школьников 10 и 11 классов, и верный ответ дали только 43 % из них; по-видимому, музыкальный контекст помешал осталь-

ным участникам заметить омонимы слов *рок* и *поп*, в целом более частотные в русских текстах [Русский медвежонок 2006]. Еще пример: в личной онлайн-переписке автора зафиксирован следующий диалог:

- (16) – Я пока без голоса, но лечу.
– Выздоровливай! А куда летишь?

(в предшествующем контексте речь шла о поездках, поэтому не вполне стандартное употребление глагола *лечить* (*лечить голос*) было воспринято как форма глагола *лететь*).

Важно также осознавать, что наборы значений многозначных слов могут существенно различаться в словарях, в лингвистических работах и в сознании носителей языка; см. обширные экспериментальные данные в работе [Lin, Ahrens 2005], где предлагается выделять соответственно «dictionary meanings», «linguistic senses» и «semantic intuition» (ср. также [Jorgensen 1990]).

В июле 2013 г. на XV Летней лингвистической школе в Дубне нами был проведен сходный эксперимент⁵. 82 участника (школьники, студенты, преподаватели, средний возраст 21 год) получили следующее задание: «Даны слова: *автомат, банан, бой, бумага, бюро, вид, визитка, действие, деталь, дипломат, картонка, книжка, крем, майка, мешок, пакет, пилот, таблетка, трубка, ящик*. Любым способом укажите для каждого слова его значения, нумеруя их в том порядке, в котором они вам приходят в голову». Ниже мы приведем некоторые результаты эксперимента и сравним их с данными толковых словарей (МАС, БТС, СОШ, СЕФ).

АВТОМАТ. БТС: «1. Самодействующее устройство (аппарат, машина, прибор), производящее работу по заданной программе без непосредственного участия человека. *Автоматы с газированной водой. Игровые автоматы. Касса-автомат. Телефон-автомат. Размен монет производится в автомате, автоматом.* // Разг. О телефоне-автомате. *Звонить из автомата. А. глотает, проглотил монетки* (о неисправном телефоне). 2. Ручное автоматическое скорострельное оружие. *Стрелять из автомата. А. Калашникова* (по имени его изобретателя М. Т. Калашникова). 3. О человеке, действующем по выработанному шаблону, безучастно и механически совершающем что-л. *Это не человек, а А.*». Значение 3 не выделил никто. Значение ‘телефон-автомат’ выделили 5 % респондентов. Значение ‘зачет автоматом’, выделенное у 30 % респондентов, присутствует только в СЕФ; значение ‘автоматическая коробка передач (в автомобиле)’, выделенное у 24 % респондентов, в исследованных нами словарях отсутствует.

БАНАН. БТС: «1. Высокое тропическое растение с огромными листьями и кистями продолговатых сладких мучнистых плодов желтого цвета; плод такого растения. *Гроздь бананов. Сушеные бананы.* 2. только мн. Разг. Брюки свободного покроя, собранные у пояса и суживающиеся книзу. *Сшить бананы.* 3. Жарг. Неудовлетворительная оценка; двойка». Значение 1 выделили все респонденты, значение 3 – 12 %, значение 2 не выделил никто (отметим, что в СЕФ значение 2 отсутствует, но присутствует значение «разг. Юбка из разноцветных клиньев, по форме напоминающая банан I 2»; его также не выделил ни один из респондентов). Кроме того, 41 % респондентов выделил значение ‘водный аттракцион – плавсредство в форме банана’, 7 % – значение ‘наушники в форме банана’, 5 % – значение ‘оценка «единица»’. Отметим, что значение ‘единица’ (в школьном сленге, см., например, [Елистратов 2000; Watzke, Sweigert 1997]) возникло, по всей видимости, раньше, чем значение ‘двойка’ (поскольку именно единица напоминает по форме банан), однако по прагматическим причинам (редкое использование оценки «1» в школе и ее фактическая замена оценкой «2») оно уступает первенство значению ‘двойка’ (см. подробнее об этом, а также о разных значениях словосочетания *схлопотать банан* в школьном и компьютерном сленге в [Иомдин, Пиперски 2010]).

⁵ Автор благодарит А.А. Лопухину и П.К. Васильева за помощь в проведении эксперимента и обработке его результатов.

ВИЗИТКА. БТС: «1. Однобортный короткий сюртук с закругленными, расходящимися спереди лапами (в 19 – начале 20 в. предназначался для утренних визитов). 2. Разг. = Визитная карточка. *Заказать визитки. Вручить визитку*». Значение 2 выделили (и привели первым в списке) все респонденты, значение 1 выделили только 6 %. Еще один респондент выделил значение, приводимое в СЕф: «Небольшая кожаная – обычно мужская – сумочка на ремешке для документов, денег, сигарет, ключей и т. п.; барсетка» (подробнее об эволюции значений слова *визитка* см. в [Июдин и др. 2013]). Кроме того, 8 % респондентов выделили значение, отсутствующее в словарях: «короткая сценка, представление команды, класса и т. п.»⁶; ср. (17–18):

- (17) *Первый этап конкурса – видеовизитка. Для этого заблаговременно для конкурсанток устроили фотосессию. Сама же визитка-представление – плод творчества женщин, где каждая – кто в прозе, а кто и в стихах – рассказала о себе самое интересное* («Липецкая газета: итоги недели», 27.08.2012).
- (18) *Найти свой образ и стиль, подготовить визитку-представление и творческий номер конкурсантам помогли профессионалы: режиссер, постановщик дефиле, хореограф, музыкант и стилист* («Тюменские известия», 2.04.2009).

МАЙКА. БТС: «Трикотажная рубашка без рукавов и воротника с большим вырезом». Все участники эксперимента выделили значение «предмет одежды». При этом 35 % описывали предмет нижнего белья с низким круглым вырезом, без воротника, застежки и рукавов (ср. то же значение у разг.-сниж. (*майка*)-*алкоголичка*), 24 % – предмет спортивной одежды с вырезом, короткими или длинными рукавами (ср. то же значение у слова *футболка*) и еще 8% выделили у слова два соответствующих значения («бельевая майка» и «майка-футболка»). Кроме того, 30 % респондентов выделили отсутствующее в словарях значение «вид полиэтиленового пакета» (ср. толкование из [Винокуров 2013]: «...иногда являющийся частью перфорированного рулона или пачки полиэтиленовый полупрозрачный предмет прямоугольной формы с двумя ручками, визуально напоминающий майку, имеющий нежесткую конструкцию, предназначенный для переноски чего-л. в руке и не рассчитанный на многократное использование»)⁷; ср. (19).

- (19) *Французский актер Жан Луи Увар вроде бы ничего «особенного» не делал. Но обычный полиэтиленовый пакет-майка вдруг оживал в его руках и начинал танцевать* («Московский комсомолец», Санкт-Петербург, 18.09.2013).

ПАКЕТ. БТС: «1. Упакованный в бумажную или иную обертку какой-л. предмет (предметы); сверток. *Передать п. Сложить вещи в п. Индивидуальный перевязочный п.* (упакованная готовая повязка для наложения на рану при оказании первой медицинской помощи). // Бумажный кулек для упаковки каких-л. предметов, продуктов и т.п. *Продажа овощей в пакетах. Молоко в пакетах.* 2. Конверт с письмом официально-делового содержания. *Секретный п. Контрольный п. (акций)* (финанс.; доля от общего количества акций, владелец которой может оказывать решающее влияние на деятельность акционерного общества)⁸. 3. Комплект документов, официальных бумаг. *П. мирных инициатив. П. законопроектов. П. требований забастовщиков.* 4. Спец. Стопка ящиков или одинаковых деталей, строительных материалов и т.п., уложенных на специальный поддон для погрузки, перевозки и т.п. *П. труб. П. пиломатериалов*». В эксперименте бумагу как материал, из которого изготавливают пакеты, упомянули лишь 8% участников.

⁶ По-видимому, это значение восходит к названию конкурса в популярной телепередаче «КВН».

⁷ На интернет-форуме словаря [Языки городов] слова *майка* и *маечка* в этом значении отмечены как региональные, однако большинство участников нашего эксперимента – ученики московских школ.

⁸ Отнесение словосочетания *контрольный пакет (акций)* к значению 2 вызывает недоумение.

9% респондентов выделили отсутствующее в рассмотренных словарях значение 'набор' (ср.: *пакет услуг*). Специальное значение 4 не выделил никто, при этом 5% выделили другое специальное значение 'блок информации, передаваемой по компьютерной сети' (ср.: *пакет данных*).

ТАБЛЕТКА. БТС: «1. Твердая лекарственная форма: лепешечка из прессованного дозированного порошка. *Т. аспирина. Т. от кашля. Проглотить таблетку.* 2. Разг. Женская шляпка твердой овальной формы с плоским дном. *Эlegantная т. Т. мне очень идет!*». Значение 1 выделили все респонденты, значение 2 – 7%. Еще три значения слова *таблетка*, описанные в СЕф («Тонкая пластинка – обычно из слоновой кости – для записи чего-либо на память»; «Дошечка, пластинка – обычно из обожженной глины – для письма»; «Небольшая, обычно двухсторонняя, икона, написанная на холсте, покрытом левкасом – смесью алебаstra или мела с клеем, предназначенной для грунтовки дерева»), не выделил никто из респондентов. Напротив, среди выделенных ими значений, отсутствующих в рассмотренных словарях, – «магнитный ключ круглой формы» (10%), «пиратская программа, предназначенная для взлома другой лицензионной программы» (интернет-сленг, 7%), «планшетный компьютер» (6%); ср. (20–22):

- (20) Попов «снял» его своим электронным ключом – *таблеткой*, такой же, какой открывают замки-домофоны («Известия», 23.05.2007).
- (21) Надо скачать *таблетку* (ссылка находится на том же сайте выше над системными требованиями к игре). Затем, после установки игры, скопировать файлы из *таблетки* (их там 3) в папку установленной игры на компе (Интернет-форум).
- (22) Планшетные компьютеры (они же планшетники, *таблетки*, айпады) весьма удачно обосновались на рынке мобильных устройств в 2010 году («Московский комсомолец», 25.10.2011).

Значения, отсутствующие в словарях, были обнаружены и у слов *бой*, *книжка*, *мешок*, *пилот*, *ящик*.

Проведенный нами эксперимент, безусловно, не претендует на статистическую достоверность и требует тщательной перепроверки на больших и репрезентативных выборках. Тем не менее и его результаты позволяют сформулировать некоторые наблюдения.

Во-первых, хотя у каждого слова было обнаружено как минимум три значения, у большинства слов одно из значений «лидировало» с большим отрывом (было выделено более чем двумя третями респондентов). Данные такого рода экспериментов позволили бы выдвигать гипотезы о наиболее частотных значениях слов и использовать их, в частности, при принятии вероятностных решений в системах автоматической обработки текстов, в том числе и вне контекста, гораздо менее трудозатратным способом, чем алгоритмы, основанные на аннотировании огромных массивов текстов и не работающие в отсутствие контекста.

Во-вторых, более точные предположения о наиболее частотных значениях можно делать, если принимать во внимание информацию о возрасте, поле, социальном статусе и сфере занятий респондентов. Даже в этом небольшом эксперименте было хорошо заметно, что более новые значения (*банан* 'плавсредство', *визитка* 'представление', *майка* 'пакет' и др.) существенно чаще выделяют школьники, чем преподаватели. Респонденты разного возраста по-разному ранжируют значения: так, для слова *бюро* чем старше был респондент, тем чаще он первым выбирал значение 'предмет мебели', а не 'организация'. Сходные наблюдения мы получаем при проведении «скрытых» экспериментов, в частности языковых игр. Так, в игре «Шляпа», где игрок должен за очень короткое время дать отгадывающему партнеру толкования возможно большего количества слов, регулярно отмечаем случаи, когда загадывающий игрок более старшего возраста воспринимает слово не в том значении, которое актуально для отгадывающего, принадлежащего к другому поколению (ср. объяснение слова *карточка* преподавателем школьнику: *фотографическая... фото... что получается, когда снимаешь... что я держу в руках* и (в другой игре) *продовольственная... раньше по ним давали хлеб*;

в обоих случаях слово не было отгадано) [Иомдин 2014]. Разная актуализация значений возможна и в контексте. Так, наши эксперименты показывают, что словосочетание *распечатать письмо* понимается по-разному представителями разных поколений; ср. (23) и (24):

- (23) *Да как же вы осмелились распечатать письмо такой уполномоченной особы?* [Н. Гоголь. Ревизор (1836)].
- (24) *Членам университетского интернет-сообщества было предложено подписаться под текстом с помощью комментариев, а также распечатать письмо и собирать «живые» подписи на лекциях и в общежитии* [«Московские новости», 13.04.2011].
- (25) *В общем... я кое-что вытянул из посудомойки* (А. Кристи. Лощина, пер. А. Вашенко).

Пример (25), предъявленный школьникам и студентам, однозначно интерпретировался в значении 'Я взял некоторые предметы из посудомоечной машины', тогда как в романе Кристи речь идет о допросе служанки.

Ср. также статью [Кронгауз 2013] об изменении значений слов *двушка*, *трешка* и *пятак* (наши опросы школьников и студентов подтверждают, что слово *двушка* в значении 'двухкопеечная монета', отмечаемом словарями (а не 'двухкомнатная квартира', словарями пока игнорируемом), им неизвестно).

Наконец, эксперименты показывают, что данные толковых словарей плохо соотносятся с современным узусом. Этот результат нельзя назвать неожиданным, поскольку хорошо известно, что словари (особенно бумажные) не успевают за появлением новых значений. Тем не менее столь же очевидно, что лексикографы должны использовать разные методы для оперативного пополнения словарей. Экспериментальные методы здесь, безусловно, должны дополняться тщательным исследованием в корпусах текстов (именно такой подход предложен нами в работе [Иомдин и др. 2014], где описывается необходимость создания частотного словаря значений слов и предлагаются методы приближенного определения частот, основанные на анализе данных опросов информантов и аннотировании наиболее частотных коллокаций в большом корпусе текстов).

3. МНОГОЗНАЧНЫЕ СЛОВА В КОРПУСЕ ТЕКСТОВ

Если выявление неологизмов в условиях существования текстов в электронной форме технически достаточно легко осуществимо (регулярно выпускаются словари новых слов, разрабатываются словари неологизмов на основе корпусов, в том числе НКРЯ и его подкорпусов, создаются программы по ежедневному мониторингу социальных сетей на предмет выявления неологизмов по ключевым запросам (например, «coined the (term|word) jargon for», «new word», «slang (expression|phrase) for» и т. п. [Cook 2012]), то выделение у слов новых значений значительно сложнее. Тем не менее в европейской лексикографии и в этой области ожидаются существенные продвижения: ср. работы [Lau et al. 2012; Cook et al. 2013], в которых описывается успешное автоматическое выделение новых значений путем сравнения корпусов текстов разных лет и статистического анализа контекстов (ср.: search 3 [uncountable] «the process of searching for information on the Internet, or the business and technology that supports this» (включено в «Macmillan dictionary» в 2013 г.; такие употребления начинают появляться и по-русски, ср. (26)), *vest* в выражениях *suicide vest*, *explosive vest* (на русский язык *vest* в этом значении, по-видимому, следует переводить словом *пояс*, ср.: *пояс шахида*) и др.).

- (26) *Сегодня представители поиска много говорят о микроформатах, зачем они нужны и почему им уделяется так много внимания?* (из интервью ведущего инженера и руководителя отдела разработки программ компании Google portalу SearchEngines.ru, 11.11.2013).

Появление больших корпусов текстов с широкими возможностями поиска, в частности кластеризации коллокаций [Kilgarriff 1993; 1997], дают лексикографам богатейшие возможности для проверки своей интуиции, определения актуальности описанных

в словарях значений и выявления новых значений. Приведем примеры использования самого объемного на сегодняшний день корпуса русского языка RuTenTen11, насчитывающего около 15 млрд словоупотреблений и интегрированного в систему автоматического подбора коллокаций [Sketch Engine] в составе полусотни корпусов текстов разных языков [Kilgariff et al. 2004].

Слово *вышка* в СУш и МАС имеет два значения: «1. Отдельная надстройка наверху здания. 2. Узкое высокое строение или площадка на высоких столбах и т.п., имеющие специальное назначение. *Наблюдательная вышка. Буровая вышка. Прыжки в воду с вышки*». В СОШ, СЕф, БТС к ним добавлено третье: «3. Высшая мера наказания, смертная казнь. *Дали вышку кому-л.*» (с пометами прост. или разг.). Система Sketch Engine выдает, в частности, следующие группы частотных словосочетаний в разных конструкциях (всего система выдает 25 групп конструкций, из которых мы приводим лишь 7; для каждой конструкции приводим не более десяти словосочетаний с самой высокой частотностью; частотность указана в абсолютных числах).

Глаголы, при которых в корпусе RuTenTen11 слово *вышка* чаще всего выступает как подлежащее:

<i>стоять</i>	274
<i>находиться</i>	196
<i>возвышаться</i>	70
<i>светить</i>	41
<i>торчать</i>	41
<i>располагаться</i>	36
<i>Чикаило</i>	33
<i>виднеться</i>	33
<i>устанавливаться</i>	29
<i>строиться</i>	21

Глаголы, при которых в корпусе RuTenTen11 слово *вышка* чаще всего выступает как прямое дополнение:

<i>построить</i>	106
<i>устанавливать</i>	101
<i>поставить</i>	83
<i>установить</i>	65
<i>поставлять</i>	51
<i>строить</i>	42
<i>соорудить</i>	34
<i>сносить</i>	24
<i>уничтожать</i>	23
<i>сооружать</i>	20

Глаголы, при которых в корпусе RuTenTen11 слово *вышка* чаще всего выступает как дополнение с предлогом *в*:

<i>поступать</i>	99
<i>поступление</i>	72
<i>поступить</i>	39
<i>врезаться</i>	27
<i>путевка</i>	12

Глаголы, при которых в корпусе RuTenTen11 слово *вышка* чаще всего выступает как дополнение с предлогом *к*:

<i>направиться</i>	25
<i>приговаривать</i>	12
<i>приговорить</i>	11
<i>подъехать</i>	8
<i>зашагать</i>	7
<i>подъезжать</i>	7
<i>побежать</i>	7

Прилагательные, в корпусе RuTenTen11 чаще всего выступающие при слове *вышка* как определение:

<i>нефтяной</i>	1344
<i>буровой</i>	673
<i>смотровой</i>	658
<i>наблюдательный</i>	639
<i>парашютный</i>	608
<i>телевизионный</i>	549
<i>сторожевой</i>	461

Существительные в корпусе RuTenTen11, при которых слово *вышка* чаще всего выступает как дополнение с предлогом *по*:

<i>экзамен</i>	30
<i>лекция</i>	25
<i>триангуляция</i>	21
<i>местоположение</i>	13
<i>препод</i>	12

Существительные в корпусе RuTenTen11, при которых слово *вышка* чаще всего выступает как дополнение с предлогом *о*:

<i>диплом</i>	36
<i>корочка</i>	7

Анализ приведенных контекстов показывает, что в корпусе RuTenTen11 слово *вышка* чаще всего выступает в значении, приводимом словарями на втором месте ('высокое строение'). Значение 'высшая мера наказания' представлено также достаточно частотными контекстами (*кому-л.*) *светит вышка*, *вышка Чикатило* (второе слово явно было неверно интерпретировано системой как глагол), *приговорить / приговаривать к вышке*. Некоторые контексты демонстрируют наличие у слова *вышка* современных разговорных значений, не отмеченных словарями: 'высшая математика' (*экзамен по вышке, лекция по вышке, препод по вышке*); 'высшее образование' (*диплом о вышке, корочка о вышке*); 'Высшая школа экономики' (*поступать / поступить в вышку, поступление в вышку*).

Еще один способ выявления дополнительных значений в системе Sketch Engine – сравнение сочетаемости пар слов, в частности синонимов (процедура Sketch Diff). Так, при сравнительном анализе сочетаемости слов *варежка* и *перчатка* выдаются, в частности, следующие данные:

Существительные, в сочинительных конструкциях с которыми встречается слово *рукавица* и не встречается слово *варежка*:

	<i>варежка</i>	<i>рукавица</i>
<i>крага</i>	0	100
<i>респиратор</i>	0	112
<i>спецодежда</i>	0	168
<i>нарукавник</i>	0	21
<i>унт</i>	0	23
<i>спецобувь</i>	0	17
<i>наплечник</i>	0	18
<i>наколенник</i>	0	18

Существительные, при которых в конструкции с предлогом *из* встречается слово *рукавица* и не встречается слово *варежка*:

	<i>варежка</i>	<i>рукавица</i>
<i>двуunitки</i>	0	9
<i>спилка</i>	0	6
<i>брзент</i>	0	31
<i>мешковина</i>	0	6
<i>шкурка</i>	0	6
<i>шкура</i>	0	17

Существительные, при которых в конструкции с предлогом *из* встречается слово *варежка* и не встречается слово *рукавица*:

	<i>варежка</i>	<i>рукавица</i>
<i>пух</i>	14	0
<i>замша</i>	6	0
<i>норка</i>	12	0
<i>флиса</i>	7	0
<i>мохер</i>	9	0
<i>микроволокна</i>	10	0
<i>микрофибры</i>	16	0

Глаголы, при которых слово *варежка* выступает как прямое дополнение, а слово *рукавица* – нет:

	<i>варежка</i>	<i>рукавица</i>
<i>захлопывать</i>	7	0
<i>разинуть</i>	16	0
<i>раззявить</i>	10	0
<i>раззявливать</i>	12	0
<i>разевать</i>	75	0

Анализ приведенных контекстов позволяет уточнить значение слова *рукавица* (частое использование в технических целях, не описанное словарями; ср. толкование МАС и БТС: «Род перчаток с одним пальцем (для большого пальца руки); варежки») и выделить у слова *варежка* разг.-сниж. значение ‘рот’, отсутствующее в рассмотренных нами словарях и внесенное автором настоящей статьи в [Апресян и др. 2014].

Очевидно, что не все вышеприведенные значения известны всем носителям литературного языка. Помимо стилистически отмеченных значений (жаргонных, сленговых, просторечных и т. п.) и значений, распространенных лишь в определенных регионах (см. выше пример слова *байка*; целый корпус таких примеров собран в словаре [Языки городов] и обсуждается в работах В.И. Беликова [Беликов 2004; 2009]), интересно также отметить случаи расхождения терминологического и бытового словоупотребления. В работе [Иомдин 2012] были приведены некоторые примеры такого рода (*курок, мимоза, паранойя, шизофрения, плечо, предплечье, целлофан* и др.) и было предложено регулярно различать в лексикографических описаниях общезыковые и специальные значения, руководствуясь не только материалом словарей и корпусов, но и экспериментальными данными.

Развитие больших аннотированных корпусов позволяет уточнять наборы значений слов не только в различных социальных группах или определенных предметных областях (ср. раннюю работу [Gale et al. 1992], в которой даже утверждалось, что в рамках одного текста каждое слово употребляется только в одном значении, и ее критику в [Wilks 1998]), но и в отдельных идиолектах. Эти данные также могли бы использоваться при разрешении неоднозначности (в условиях, когда отсутствует контекст, но известен говорящий, как в вышеупомянутой ситуации с анализом однословных поисковых запросов).

Так, анализ употреблений слова *вышка* в произведениях А. Солженицына, представленных в НКРЯ, показывает, что из 35 вхождений в 33 случаях это слово употреблено им в значении ‘сторожевая вышка’, ср. (27), в одном случае – в значении ‘надстрой-ка’ (28) и еще в одном случае – в значении ‘высшая мера наказания’ с комментарием, показывающим, что для Солженицына это значение менее естественно (29):

- (27) Два больших прожектора били по зоне наперекрест с дальних угловых **вышек** [А. Солженицын. Один день Ивана Денисовича (1961)].
- (28) Усеченная пирамида самой церкви, переходное здание с **вышкой** и две круглых крепостных башни [А. Солженицын. Захар-Калита (1965)].
- (29) Очень часто не помилование получали они, а **вышку** (так называют арестанты «высшую меру», они не терпят высоких слов и все называют как-нибудь поглубей и покороче) [А. Солженицын. Архипелаг ГУЛаг (1958–1973)].

Напротив, в проиндексированном поисковыми системами блоге А. Навального из 17 вхождений слова *вышка* в 13 случаях оно употреблено в значении 'Высшая школа экономики' (данные на октябрь 2013 г.).

Корпусные исследования позволяют выявлять и уникальные авторские значения. Так, слово *своеобычный* толкуется в словарях как 'то же, что своенравный', 'то же, что своеобразный' или в обоих указанных значениях (СОШ, МАС, БТС, СЕФ). Из 119 вхождений слов *своеобычный* и *своеобычно* в основном корпусе НКРЯ этим значениям соответствуют тексты всех авторов, кроме Л. Леонова (30–31) и А. и Б. Стругацких (32):

- (30) *Как и в вечер фирсовского появления на Благуше, цветные искорки необсохшей измороси переливались на плечах у Векшина, стынул перед ним в стакане **своеобычный**, без сладостей, чай* [Л. Леонов. Вор (1927)].
- (31) *Таиска постаралась представить, что сейчас творится на Енге, занятой немцами, померкла и, **своеобычно** оправив платок на голове, искоса взглянула на Полю* [Л. Леонов. Русский лес (1950–1953)].
- (32) *И когда мне навстречу, шаркая подбитыми кожей валенками, прошел Выбегалло, я испытал к нему даже нечто вроде симпатии, потому что у него была **своеобычная** пишенная каша в бороде, потому что он ковырял в зубах длинным тонким гвоздем и, проходя мимо, не поздоровался* [А. Стругацкий, Б. Стругацкий. Понедельник начинается в субботу (1964)].

Анализ всех произведений А. и Б. Стругацких, доступных в электронной форме, показывает, что слово *своеобычный* всегда употребляется ими в значении 'обычный, характерный для кого-л., свойственный кому-л.' (...но тут пришел доктор Р. Квадрига и заказал *своеобычную* бутылку рома; Лицо у Комова было непривычно озабоченное, ... без этой *своеобычной*, оскомину набившей ледяной сосредоточенности; ...он без всяких предисловий и *своеобычных* в таких случаях... шуточек доложил Тойво, что...). Можно предположить, что так употребляется это слово и в других произведениях Л. Леонова.

Стало быть, значение слова *вышка* в текстах А. Солженицына и А. Навального и значение слова *своеобычный* в текстах А. и Б. Стругацких и Л. Леонова с высокой вероятностью определяется и вне контекста. Привлечение данных авторской лексикографии может дать много интересных результатов такого рода.

ЗАКЛЮЧЕНИЕ

Языки члелят мир по-разному, и между лексическими системами разных языков нет и не может быть взаимно-однозначного соответствия. Даже слова одного происхождения в родственных языках или в случаях заимствования чаще всего имеют разные наборы значений. То, что такие расхождения возможны и в рамках одного языка, хорошо известно в диалектологии, а в последнее время такие примеры накапливаются и при обсуждении региональных вариантов литературного языка. Целью настоящей статьи было показать, что наборы значений многозначных слов могут различаться и у носителей одного и того же идиома и зависеть от социального, возрастного, профессионального статуса говорящих и от их личного речевого опыта. Если приводить аналогию с фонемами, можно сказать, что многие лексемы многозначных вокабул находятся в отношениях свободного варьирования (у одних говорящих они входят в активный словарный запас, у других – в пассивный, у третьих отсутствуют вовсе) и по-разному ранжированы в разных подъязыках и идиолектах, а некоторые лексемы находятся в отношениях, близких к отношениям дополнительной дистрибуции (так, по-видимому, обстоит дело со значениями слова *банан* ('двойка' vs. 'единица'), *фуфайка* ('поддевка' vs. 'свитер' vs. 'ватник' vs. 'футболка', ср. [Иомдин 2011]), *мимоза* ('*Mimosa pudica*' vs. '*Asacia dealbata*', ср. [Иомдин 2012]). Задача современной лексикографии – выявлять, адекватно оценивать и описывать явления такого рода. При этом интроспекцию, на которую обычно опираются классические лексикографы, необходимо сочетать с экспери-

ментальными данными, корпусными исследованиями и результатами, полученными в области компьютерной лингвистики⁹.

Данные о различиях в наборах значений слов в современном языке необходимы для создания учебников для разного уровня владения языком, словарей наиболее частотных слов в разных предметных областях (ср. проект словаря бытовой терминологии [Иомдин 2011; Иомдин и др. 2012; 2013], проект частотного словаря значений слов [Иомдин и др. 2014]), динамических словарей на основе аннотированных корпусов с настраиваемыми параметрами (подбор материала словарных статей с учетом данных о возрасте, социальном статусе, поле и др.), персонализации поиска в информационных базах данных (в том числе в поисковых системах, ср. [Персонализированный поиск Яндекс; Технология «Крипта»] и др.).

Кроме того, кажется, что накопление и анализ информации о различиях в употреблении многозначных слов людьми, говорящими на одном языке, важно и интересно не только для практических применений, но и для развития лексической семантики, исследования языковой вариативности, психо- и социолингвистики.

СПИСОК ЛИТЕРАТУРЫ

- Апресян 1974 – Ю.Д. Апресян. Лексическая семантика. Синонимические средства языка. М., 1974 (2-е изд.: М., 1995).
- Апресян и др. 2005 – Ю.Д. Апресян, И.М. Богуславский, Б.Л. Иомдин, Л.Л. Иомдин, А.В. Санников, В.З. Санников, В.Г. Сизов, Л.Л. Цинман. Синтаксически и семантически аннотированный корпус русского языка: современное состояние и перспективы // Национальный корпус русского языка: 2003–2005. М., 2005.
- Апресян и др. 2010 – В.Ю. Апресян, Ю.Д. Апресян, Е.Э. Бабаева, О.Ю. Богуславская, И.В. Галактионова, М.Я. Гловинская, Б.Л. Иомдин, Т.В. Крылова, И.Б. Левонтина, А.В. Птенцова, А.В. Санников, Е.В. Урысон. Проспект активного словаря русского языка / Отв. ред. акад. Ю.Д. Апресян. М., 2010.
- Апресян и др. 2014 – В.Ю. Апресян, Ю.Д. Апресян, Е.Э. Бабаева, О.Ю. Богуславская, И.В. Галактионова, М.Я. Гловинская, Б.Л. Иомдин, Т.В. Крылова, И.Б. Левонтина, А.А. Лопухина, А.В. Птенцова, А.В. Санников, Е.В. Урысон. Активный словарь русского языка. Т. 1–2: А–Г / Отв. ред. акад. Ю.Д. Апресян. М., 2014 (в печати).
- Бабанов 2012 – А.А. Бабанов. Применение классического словаря к задаче автоматизированного лингвистического анализа текста // Информационные системы для научных исследований: Сборник научных статей. Труды XV Всероссийской объединенной конференции «Интернет и современное общество». Санкт-Петербург, 10–12 октября 2012 г. СПб., 2012.
- Беликов 2004 – В.И. Беликов. Сравнение Петербурга с Москвой и другие соображения по социальной лексикографии // Русский язык сегодня. Вып. 3: Проблемы русской лексикографии. М., 2004.
- Беликов 2009 – В.И. Беликов. Стереотипы в понимании литературной нормы // Стереотипы в языке, коммуникации и культуре. М., 2009.
- БТС – Большой толковый словарь русского языка / Сост., гл. ред. С.А. Кузнецов. СПб., 1998.
- Винокуров 2013 – Ф.Г. Винокуров. Проблемы описания предметной лексики (на материале названий сумок в русском языке). Дипломная работа. М., 2013.
- Гак 1972 – В.Г. Гак. К проблеме семантической синтагматики // Проблемы структурной лингвистики 1971. М., 1972.
- Гельбук, Сидоров 2004 – А.Ф. Гельбук, Г.О. Сидоров. Метод автоматического разрешения неоднозначности значений слов в словарных толкованиях // Научно-техническая информация. Серия 2: Информационные процессы и системы. 2004. № 3.

⁹ Ср. предложение, высказанное в монографии [Kwong 2012]: «...take a step back to re-examine the computational strategies (by machines) and the cognitive strategies (by humans) for WSD in parallel, and to explore alternative classification senses which might shed light on their differential information susceptibility» («вернуться на шаг назад и пересмотреть компьютерные стратегии (используемые машинами) и когнитивные стратегии (применяемые людьми) для задачи разрешения многозначности и исследовать альтернативные способы классификаций значений, которые могли бы пролить свет на различия в восприятии информации») [Kwong 2012: vii–viii].

- Грачев 2005 – М.А. Грачев. От Ваньки Каина до мафии: прошлое и настоящее уголовного жаргона. М., 2005.
- Елистратов 2000 – В.С. Елистратов. Словарь русского арготизма: Материалы 1980–1990 гг.: Около 9 000 слов, 3 000 идиоматических выражений. М., 2000.
- Заботкина, Боярская 2011 – В.И. Заботкина, Е.Л. Боярская. Роль прайминга в разрешении полисемии // Когнитивные исследования языка. Вып. 9: Взаимодействие когнитивных и языковых структур. М.; Тамбов, 2011.
- Заботкина, Боярская 2012 – В.И. Заботкина, Е.Л. Боярская. Роль когнитивного контекста в разрешении многозначности: опыт концептуального моделирования // Когнитивные исследования языка. Вып. 12: Теоретические аспекты языковой репрезентации. М.; Тамбов, 2012.
- Зализняк 2006 – Анна А. Зализняк. Многозначность в языке и способы ее представления. М., 2006.
- Иомдин 2009 – Б.Л. Иомдин. Терминология быта. Поиски нормы // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2009» (Бекасово, 27–31 мая 2009 г.). Вып. 8 (15). М., 2009.
- Иомдин 2011 – Б.Л. Иомдин. 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., 2011.
- Иомдин 2012 – Б.Л. Иомдин. О «неправильном» использовании терминов: может ли язык ошибаться? // Смыслы, тексты и другие захватывающие сюжеты: Сборник статей в честь 80-летия И.А. Мельчука / Под ред. Ю.Д. Апресяна, И.М. Богуславского, Л. Ваннера, Л.Л. Иомдина, Я. Миличевич, М.-К. Л'Омм, А. Польгера. М., 2012.
- Иомдин 2014 – Б.Л. Иомдин. Наивные представления о значениях слов в русском языке // Антропологический форум. 2014. № 21 (в печати).
- Иомдин, Пиперски 2010 – Б.Л. Иомдин, А.Ч. Пиперски. Прагматика еды: коннотации в русской и немецкой пищевой лексике // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2010» (Бекасово, 26–30 мая 2010 г.). Вып. 9 (16). М., 2010.
- Иомдин и др. 2012 – Б.Л. Иомдин, А.А. Лопухина, А.Ч. Пиперски, М.Ф. Киселева, Г.В. Носырев, А.М. Рикитянский, П.К. Васильев, А.Г. Кадыкова, В.И. Матиссен-Рожкова. 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., 2012.
- Иомдин и др. 2013 – Б.Л. Иомдин, А.А. Лопухина, М.Ф. Панина, Г.В. Носырев, М.В. Вилл, Л.Я. Зайдельман, В.И. Матиссен-Рожкова, Ф.Г. Винокуров, А.В. Выборнова. 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., 2013.
- Иомдин и др. 2014 – Б.Л. Иомдин, А.А. Лопухина, Г.В. Носырев. К созданию частотного словаря значений слов // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2014» (Бекасово, 4–8 июня 2014 г.). Вып. 13 (20): В 2 т. Т. 1. М., 2014.
- Качурин 2014 – Д.В. Качурин. Проблема разграничения омонимии и полисемии применительно к практике составления толковых словарей. Дис. ... канд. филол. наук. М., 2014.
- Кобрицов 2004 – Б.П. Кобрицов. Модели многозначности русской предметной лексики: глобальные и локальные правила разрешения омонимии. М., 2004.
- Кобрицов и др. 2007 – Б.П. Кобрицов, О.Н. Ляшевская, С.Ю. Толдова. Снятие семантической многозначности глаголов с использованием моделей управления, извлеченных из электронных толковых словарей // Электронная публикация. <http://download.yandex.ru/IMAT2007/kobricov.pdf>.
- Кронгауз 2013 – М.А. Кронгауз. Актуальный счет, или Двушка, трешка, пятак... // Russian linguistics. 2013. V. 37. № 2.
- Кустова и др. 2005 – Г.И. Кустова, О.Н. Ляшевская, Е.В. Падучева, Е.В. Рахилина. Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы // Национальный корпус русского языка: 2003–2005. Результаты и перспективы. М., 2005.
- Лукашевич, Добров 2007 – Н.В. Лукашевич, Б.В. Добров. Разрешение лексической многозначности на основе тезауруса предметной области. Компьютерная лингвистика и интеллектуальные технологии // Труды международной конференции «Диалог 2007» (Бекасово, 30 мая – 3 июня 2007 г.). М., 2007.

- МАС – Словарь русского языка. Т. 1–4 / Под ред. А.П. Евгеньевой. М., 1981–1984.
- Митрофанова и др. 2012 – О.А. Митрофанова, О.Н. Ляшевская, М.А. Грачкова, А.С. Шиморина, А.С. Шурыгина, С.В. Романов. Эксперименты по автоматическому разрешению лексико-семантической неоднозначности и выделению конструкций (на материале Национального корпуса русского языка) // Структурная и прикладная лингвистика. 2012. № 9.
- Никитина 2003 – Т.Г. Никитина. Молодежный сленг. Толковый словарь. М., 2003.
- НКРЯ – Национальный корпус русского языка // <http://www.ruscorpora.ru>.
- Персонализированный поиск Яндексa – <http://api.yandex.ru/personalized-search>.
- Рахилина и др. 2006 – Е.В. Рахилина, Б.П. Кобрицов, Г.И. Кустова, О.Н. Ляшевская, О.Ю. Шеманаева. Многозначность как прикладная проблема: лексико-семантическая разметка в Национальном корпусе русского языка // Компьютерная лингвистика и интеллектуальные технологии: Труды Международной конференции «Диалог 2006». М., 2006.
- Рахилина и др. 2009 – Е.В. Рахилина, Г.И. Кустова, О.Н. Ляшевская, Т.И. Резникова, О.Ю. Шеманаева. Задачи и принципы семантической разметки лексики в НКРЯ // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб., 2009.
- Русский медвежонок 2006 – Русский медвежонок – 2005. Задачи, решения, информация, статистика / Сост. Е.В. Муравенко, И.С. Рубанов, И.Ф. Богдалов. Киров, 2006.
- СЕФ – Т.Ф. Ефремова. Большой современный толковый словарь русского языка: В 3 т. М., 2006.
- СОШ – С.И. Ожегов, Н.Ю. Шведова. Толковый словарь русского языка. М., 1992.
- СУШ – Толковый словарь русского языка / Под ред. Д. Н. Ушакова. М., 1934–1940.
- СШ – Толковый словарь русского языка с включением сведений о происхождении слов / Отв. ред. Н.Ю. Шведова. М., 2007.
- Технология Крипта // <http://company.yandex.ru/technologies/crypta>.
- Турдаков 2010 – Д.Ю. Турдаков. Методы разрешения лексической многозначности // Программирование. 2010. Т. 36. № 6.
- Языки городов // <http://www.lingvo.ru/goroda>.
- Agirre, Edmonds (eds) 2006 – E. Agirre, P.G. Edmonds (eds). Word sense disambiguation: Algorithms and applications. Dordrecht, 2006.
- Cook 2012 – P. Cook. Using social media to find English lexical blends // Proceedings of the 15th EURALEX International congress (EURALEX 2012). Oslo, 2012.
- Cook et al. 2013 – P. Cook, J.H. Lau, M. Rundell, D. McCarthy, T. Baldwin. A lexicographic appraisal of an automatic approach for detecting new word-senses // I. Kosem, J. Kallas, P. Gantar, S. Krek, M. Langemets, M. Tuulik (eds). Electronic lexicography in the 21st century: Thinking outside the paper. Proceedings of the eLex 2013 conference, 17–19 october 2013, Tallinn (Estonia). Ljubljana; Tallinn, 2013.
- Durkin, Manning 1989 – K. Durkin, J. Manning. Polysemy and the subjective lexicon: Semantic relatedness and the salience of intraword senses // Journal of psycholinguistic research. 1989. V. 18. № 6.
- Fellbaum et al. 1997 – C. Fellbaum, K. Grabowski, L. Shari. Analysis of a hand-tagging task // Proceedings of ANLP-97 Workshop on tagging text with lexical semantics. Washington, 1997.
- Foraker, Murphy 2012 – S. Foraker, G.L. Murphy. Polysemy in sentence comprehension: Effects of meaning dominance // Journal of memory and language. 2012. V. 67. № 4.
- Gale et al. 1992 – W. Gale, K. Church, D. Yarowsky. One sense per discourse // Proceedings 4th DARPA Speech and natural language workshop. San Mateo, 1992.
- Hank 2013 – P. Hank. Lexical analysis: Norms and exploitations. Boston, 2013.
- Hoey 2005 – M. Hoey. Lexical priming. Oxford, 2005.
- Ide, Véronis 1998 – N. Ide, J. Véronis. Introduction to the special issue on word sense disambiguation: The state of the art // Computational linguistics. 1998. V. 24. № 1.
- Jarmasz, Szpakowicz 2001 – M. Jarmasz, S. Szpakowicz. Roget's thesaurus: A lexical resource to treasure // Proceedings of the NAACL WordNet and other lexical resources workshop. Pittsburgh, 2001.
- Jorgensen 1990 – J.C. Jorgensen. The psychological reality of word senses // Journal of psycholinguistic research. 1990. V. 19. № 3.
- Kilgarriff 1993 – A. Kilgarriff. Dictionary word sense distinctions: An enquiry into their nature // Computers and the humanities. 1993. V. 26. № 1–2.
- Kilgarriff 1997 – A. Kilgarriff. «I don't believe in word senses» // Computers and the humanities. 1997. V. 31. № 2.
- Kilgarriff et al. 2004 – A. Kilgarriff, P. Rychly, P. Smrz, D. Tugwell. The sketch engine // Proceedings of the Eleventh EURALEX International congress. Lorient, 2004.

- Kwong 2012 – *O.Y. Kwong*. New perspectives on computational and cognitive strategies for word sense disambiguation. New York, 2012.
- Lapata, Keller 2007 – *M. Lapata, F. Keller*. An information retrieval approach to sense ranking // Proceedings of the Human language technology conference of the North American chapter of the association for computational linguistics (HLT-NAACL). Rochester (NY), 2007.
- Lau et al. 2012 – *J.H. Lau, P. Cook, D. McCarthy, D. Newman, T. Baldwin*. Word sense induction for novel sense detection // Proceedings of the 13th Conference of the European chapter of the association for computational linguistics. Stroudsburg, 2012.
- Lesk 1986 – *M. Lesk*. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone // Proceedings of SIGDOC-86: 5th International conference on systems documentation. Toronto, 1986.
- Lin, Ahrens 2005 – *C.C. Lin, K. Ahrens*. 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.
- Mohammad, Hirst 2006 – *S. Mohammad, G. Hirst*. Determining word sense dominance using a thesaurus // Proceedings of the 11th Conference on European chapter of the Association for computational linguistics (EACL). Italy, 2006.
- Navigli 2009 – *R. Navigli*. Word sense disambiguation: A survey // ACM computing surveys. 2009. V. 41. № 2.
- Pustejovsky 1996 – *J. Pustejovsky*. Lexical semantics: The problem of polysemy. Oxford, 1996.
- Rayner, Frazier 1989 – *K. Rayner, L. Frazier*. Selection mechanisms in reading lexically ambiguous words // Journal of experimental psychology: Learning, memory, and cognition. 1989. V. 15. № 5.
- Roth, Schulte im Walde 2008 – *M. Roth, S. Schulte im Walde*. Corpus co-occurrence, dictionary and Wikipedia entries as resources for semantic relatedness information // Proceedings of the 6th conference on language resources and evaluation. Marrakesh, 2008.
- Seidenberg et al. 1982 – *M.S. Seidenberg, M.K. Tanenhaus, J.M. Leiman, M. Bienkowski*. Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing // Cognitive psychology. 1982. V. 14. № 4.
- Sketch Engine // <http://www.sketchengine.co.uk>.
- Snyder, Palmer 2004 – *B. Snyder, M. Palmer*. The English all-words task // Senseval-3: Third international workshop on the evaluation of systems for the semantic analysis of text. Barcelona, 2004.
- Tejada-Cárcamo et al. 2010 – *J. Tejada-Cárcamo, H. Calvo, A. Gelbukh, K. Hara*. Unsupervised WSD by finding the predominant sense using context as a dynamic thesaurus // Journal of computer science and technology. 2010. V. 25. № 5.
- Vue et al. 2003 – *H. Vue, G. Kellas, E. Petersen, K. Metcalf*. Situation-evoking stimuli, domain of reference, and the incremental interpretation of lexical ambiguity // Memory and cognition. 2003. V. 31. № 8.
- Watzke, Sweigert 1997 – *J. Watzke, J. Sweigert Jr.* The Russian reference grammar: Core grammar in functional context. Dubuque, 1997.
- Wilks 1998 – *Y. Wilks*. Senses and texts // Computational linguistics and Chinese language processing. 1998. V. 3. № 2.
- WordNet // <http://wordnet.princeton.edu>.
- Yarowsky 1992 – *D. Yarowsky*. Word-sense disambiguation using statistical models of Roget's categories trained on large corpora // Proceedings of the 14 International conference on computational linguistics (COLING-92). Nantes, 1992.
- Zipf 1945 – *G.K. Zipf*. The meaning frequency relationship of words // Journal of general psychology. 1945. V. 33.

Сведения об авторе:

Борис Леонидович Иомдин
Институт русского языка им. В.В. Виноградова РАН,
НИУ «Высшая школа экономики»
iomdin@ruslang.ru

Статья поступила в редакцию 3.12.2013.