УДК 802.0:801.541.2

Е. Л. Боярская

КОНЦЕПТУАЛЬНОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССА РАЗРЕШЕНИЯ ПОЛИСЕМИИ С ИСПОЛЬЗОВАНИЕМ КОРПУСА

Утверждается, что лингвистический анализ, выполняемый на основе корпусов естественных языков, позволяет проводить аналитическую работу высокой степени точности и репрезентативности. Доказывается, что использование корпусов открывает новые возможности исследования когнитивных механизмов формирования структуры многозначного слов, а также идентификации его отдельных значений. Предлагается вероятностная модель (алгоритм) концептуальных операций, в результате которых становится возможной идентификация соответствующего значения многозначного слова.

Any linguistic analysis when done on the basis of a natural language corpus produces results of much higher accuracy and a higher degree of representation. The use of corpora opens up new possibilities for investigating the mechanisms of the formation of the cognitive structure of a polysemous word, as well as facilitates the identification of individual senses. The author proposes a probabilistic conceptual model (algorithm) of cognitive operations underlying the process of word sense disambiguation.

Ключевые слова: полисемия, разрешение полисемии, концептуальное моделирование, корпус.

Key words: polysemy, word sense disambiguation, conceptual modelling, corpus.

За последние несколько десятилетий в рамках науки о языке сформировалось особое направление, получившее название «эмпирическая лингвистика», к которому обычно причисляют и когнитивную лин-

---

Вестник Балтийского федерального университета им. И. Канта. 2013. Вып. 2. С. 35 – 44.

гвистику. Проведение когнитивных исследований в области лингвистики невозможно представить без применения эмпирических методов, концептуального моделирования и использования баз данных корпусов различного типа в качестве материала.

Как известно, корпусная лингвистика — особое направление лингвистики, цель которого — детальный анализ данных естественных языков, зафиксированных и представленных определенным образом.

Понятие «корпус» с момента своего появления в 1960—1980-х гг. прошлого столетия претерпело значительные изменения. Изначально любое множество текстов можно было назвать корпусом. Однако корпусы, используемые для целей лингвистического исследования, обладают рядом особых характеристик, среди которых обычно выделяют следующие:

• стандартизированное представление данных, позволяющее применять компьютерные методы обработки (sampling);

• репрезентативность — наличия перечня параметров, в соответствии с которыми и был оставлен тот или иной корпус (representativeness);

• конечный размер текстов (finite size);

• размещение информации на электронном носителе (machine-readable forms) [13, p. 6].

В настоящее время под корпусом понимают собрание текстов на определенном языке, представленных в электронной форме, специальным образом обработанное — аннотированное таким образом, чтобы исследователь мог быстро и в полном объеме найти в корпусе интересующую его информацию. В зависимости от типа аннотации возможен поиск по грамматическим характеристикам слов и предложений (морфологическим, синтаксическим, семантическим параметрам), а также по разнообразным данным о самих текстах, входящих в корпус: по автору, дате создания текста, жанру, тематике и т. п. [6].

Со времени возникновения корпусов первого поколения со стандартом около 1 млн словоупотреблений — the Brown Corpus (1961—1964), Lancaster-Oslo / Bergen LOB (1970—1978), London-Lund Corpus (1975), Машинного фонда русского языка (1985), Уппсальского корпуса русского языка (1980) — появилось новое поколение корпусов с сотнями миллионов словоупотреблений, наиболее известные из которых — British National Corpus (1995 г., 100 млн словоупотреблений), Национальный корпус русского языка (140 млн), the COBUILD project / the Bank of English (525 млн), Gigaword corpora (1 млрд) и многие другие.

Корпусная лингвистика, прошедшая несколько этапов развития, с течением времени выработала ряд критериев, которые выделяют ее из ряда других направлений науки о языке:

— Лингвистический анализ выполняется на основе корпуса или корпусов естественных языков, зафиксированных и формализованных фрагментов письменной и устной речи. Формализация фрагментов позволяет осуществлять компьютерный анализ определенного рода информации, извлекая ее из баз корпусов в соответствии с заданными критериями поиска.

— Данные корпусов представляют собой материал высокой степени репрезентативности, дающий полную информацию, удовлетворяющую целям исследования.

— Анализ, проводимый с использованием корпусов, — системное исследование, позволяющее получить максимально исчерпывающие результаты. Корпус служит не только базой данных примеров, которые выбираются *ad libitum,* он обеспечивает исключительную возможность осуществлять анализ явлений как более, так и менее частотных, тем самым формируя целостную картину.

— Использование данных корпусов позволяет не только выделять некие феномены категориального характера, но и проводить статистические исследования — анализировать частотность употребления, сравнивать процентное соотношение, определять вероятностный фактор, а также осуществлять собственно статистический анализ. Более того, применение корпуса дает возможность связать воедино область семантики и грамматики, прежде всего синтаксиса.

— Лингвистический анализ выполняется на основе созданных системой частотных списков, которые могут состоять из отдельных лексических единиц, их частей (например морфем), грамматических моделей, сочетаемости и т.д.

— Слово приводится и анализируется в естественном контексте, в составе ему присущих коллокаций [13, p. 4—5].

При этом нельзя не упомянуть о двух важных моментах, связанных с использованием корпуса: степени гранулярности, то есть уровня детализации данных при проведении исследования, а также соотношении качественных и количественных параметров.

Что касается гранулярности, то можно выделить несколько подходов к степени детализации информации:

• использование леммы с целью абстрагироваться от конкретных словоформ и сделать выводы более общего характера;

• использование одной конкретной словоформы в качестве отправной точки;

• использование более сложных, флективных форм леммы с целью определения системных отношений между индивидуальными формами слова, их окружением и т.д. [13].

Что касается соотношения количественных и качественных параметров, то в корпусной лингвистике количественный анализ имеет важное значение, так как способствует повышению валидности полученных результатов.

Появление корпусов дало новый импульс развитию лингвистических исследований в том числе и изучению многозначности, предоставив возможность фиксировать большой объем данных в максимально широком, а главное, *естественном* контексте. Активное применение современных компьютерных технологий позволяет представить весь объем данных, ассоциируемых с различными значениями многозначного слова, в сжатой форме, благодаря чему исследователи практически мгновенно могут извлекать сведения об интересующих единицах в необходимом объеме. Использование корпуса, фиксирующего естествен-

ный контекст употребления отдельных значений многозначных единиц, повышает степень валидности выводов исследования, минимизируя долю субъективности трактовок и неполноты анализируемого материала, прежде всего за счет значительного количества данных и их гранулярности.

Перейдем к методике концептуального моделирования, применение которой на материале корпусов открывает новые перспективы идентификации значений многозначного слова. Концептуальное моделирование позволяет создать некую модель объекта, явления или процесса, обеспечивая изучение их свойств, характеристик, последовательности протекания процессов, их анализа и прогнозирования. В когнитивной лингвистике метод концептуального моделирования выступает в роли метода эмпирического познания ментальных процессов.

Естественный язык, не обладая функцией порождения мысли, оказывается средством выявления структуры мысли носителя языка; он, по сути, дает ключ к *реконструкции сознания*. Реальная действительность отражается в сознании в процессе мышления как накопленное знание об этой действительности, репрезентируемое в материальных формах языка [8].

В рамках когнитивной лингвистики многозначное слово трактуется не как некий список фиксированных значений, как это делается в традиционной лексикографии, а как единица сознания, характеризующая комплексный характер человеческой деятельности, опыта и его осмысления. Язык определяет, объективирует то, как *увиден и понят* мир разумом, как он преломлен и категоризирован сознанием. Каждая языковая единица рассматривается как результат действия когнитивных процессов [4, с. 38].

Метод концептуального моделирования дает возможность представить процесс формирования комплексной ментальной репрезентации, а также воссоздать до определенной степени механизм идентификации значений многозначного слова.

Использование данных British National Corpus, анализ случаев употребления как одного и того же, так и разных значений многозначного слова в естественном и максимально широком контексте позволило сделать некоторые предположения касательно особенностей протекания когнитивных процессов.

Как выявил проведенный эмпирический анализ, каждое из значений многозначного слова ассоциировано с определенной комбинацией когнитивных контекстов, а полисемант в целом представляет собой комплексную ментальную репрезентацию [2]. Когнитивные контексты, ассоциированные с отдельными значениями многозначного слова, не статичные образования; они отражают динамический характер познания в процессе восприятия реальности, приобретения нового опыта и его оценки.

Анализ данных показывает, что когнитивный контекст, трактуемый широко, не является гомогенным по своему составу, он содержит определенные концептуальные элементы. Так, например, можно выделить имплицитно фиксируемую когнитивную информацию универсального характера (например, о пространственных отношениях, физическом устройстве мира, движении и его характере, времени и т.д.). Это то концептуальное содержание, которое остается неизменным вне зави-

симости от любых прагматических факторов, некая константа, присутствующая *by default* в когнитивном контексте. Однако без данного типа концептуальной информации знания, ассоциированные с определенным фрагментом объективной действительности, не могут отражать ее целостно, а лишь фиксируют значимые, но все-таки фрагменты этой реальности. Так, например, любой концептуальный сценарий подразумевает не только фиксацию информации о типичных, конвенциональных ситуациях, типах действий субъекта, их последовательности, характере и т.д. Сценарий разворачивается на определенном фоне, информация о котором может играть важную роль в дальнейшем развитии данного сценария, выполняя функцию не только «упаковки» знания, но и — потенциально — его генерирования, а в случае многозначности — способствовать более точной идентификации смысла [2].

Когнитивный контекст отдельных значений многозначного слова может содержать культурно-специфический компонент.

Вероятно, можно вести речь и о наличии *индивидуально-личностного элемента когнитивного контекста*, так как не принимать во внимание субъективные факторы прагматического (социального) характера при восприятии когнитивного контекста, ассоциированного с определенным смыслом многозначного слова, нельзя. Даже используя многозначную единицу в одном и том же значении, разные носители языка, не осознавая этого, оперируют несколько различным объемом информации, фиксированной данным смыслом. В момент формирования когнитивного контекста употребления данного смысла каждый индивид обладает *своим* представлением о мире, порядке его устройства, его ценностных аспектах, которые были заложены всем его предыдущим опытом.

Таким образом, многозначное слово может присутствовать в ментальном лексиконе, комплексной ментальной репрезентации — наборе когнитивных контекстов, ассоциированных со смыслами многозначного слова, фиксируя при этом значительный объем концептуальной информации, относящейся к самым разным областям. Модель распознавания значения многозначного слова может быть представлена следующим образом:

- предъявление слова, употребленного в одном из своих смыслов;
- активизация набора когнитивных контекстов, ассоциированных со словом;
- идентификация соответствующего когнитивного контекста (или его элемента);
- идентификация соответствующего значения многозначного слова.

Необходимо заметить, что указанная выше последовательность когнитивных операций, связанных с формированием, переработкой, хранением и извлечением концептуальной информации, имеет гипотетический характер и является результатом применения метода концептуального моделирования.

Ниже приводятся примеры, отобранные из British National Corpus и подтверждающие высказанные выше гипотезы. Именно возможности корпуса, фиксирующего фрагменты живой речи, а также корпусной лингвистики, рассматривающей единицы текста в глобальной перспек-

тиве, фокусирующей внимание на как можно более широком взгляде на текст, не ограниченном никакими догмами, прибегающей к вероятностным методам и обработке эмпирических данных [7], позволяют нам приблизиться к решению некоторых из краеугольных проблем лингвистики, таких, как многозначность.

Обратимся к анализу путей разрешения многозначности на примере существительного *bird,* ранее использованном нами для демонстрации потенциала концептуального моделирования процессов формирования нового значения и его репрезентации [3]. Данная лексическая единица имеет несколько значений, зафиксированных в большинстве лексикографических источников:

Warm-blooded egg-laying vertebrates characterized by feathers and forelimbs modified as wings:
- S: (n) the flesh of a bird or fowl (wild or domestic) used as food.
- S: (n) informal terms for a (young) woman.
- S: (n) a cry or noise made to express displeasure or contempt.
- S: (n) badminton equipment consisting of a ball of cork or rubber with a crown of feathers (WordNet).

Рассмотрим несколько примеров употребления данной лексической единицы в некоторых из вышеуказанных значений. Начнем с группы примеров, в которых слово *bird* потреблено в его исходном вышеуказанном значении:

1. a. Glancing up, I saw a beautiful yellow **bird** perched on a telegraph wire, looking like a prize long-tailed canary.
b. A **bird** shuffled along the perch.
c. **A bird** tweetered in the grey light through the blinds.
d. It's 10 minutes today: less than a cat puts in on **a bird**.
e. **A bird** had gained entry through one of the broken windows and flown helplessly around until it collided with her.
f. It must be great to be **a bird** — you know, just flying over people and buildings and that.
g. I expect he was after a squirrel or a bird up there; he's a regular hunter.

Первая группа примеров демонстрирует употребление рассматриваемого существительного в его исходном, прототипическом значении. При этом следует оговориться, что понятие «прототипичность» трактуется поразному в когнитивной и корпусной лингвистике: в когнитивной лингвистике прототип — наилучший представитель некой категории, в то время как в корпусной лингвистике это самая частотная единица [12, p. 159].

В примере 1а ментальная репрезентация содержит концептуальную информацию, которая ассоциативно и категориально связана с анализируемой лексической единицей — указание на месторасположение по направлению вверх от говорящего (glancing up, perched on a telegraph wire), внешние характеристики (beautiful and yellow), отсылку к определенной концептуальной категории (long-tailed canary). В сознании активизируются когнитивные контексты, свойственные лишь одному значению данного слова из всех перечисленных выше — исходному значению полисеманта. Даже при наличии минимального объема кон-

цептуальной информации — «glancing up, I saw a beautiful and yellow bird…» — определение значения данной лексической единицы не представляется сложным.

В примере 1b существительное *bird* также употреблено в своем исходном значении, как и в случае 1a. Однако указание на характер передвижения (shuffled) формирует совсем иной когнитивный контекст, тоже ассоциативно связанный с птицей, но более крупной, тяжелой, возможно, даже неспособной летать, то есть совершать действие, являющееся одной из основных черт прототипического представления о птице как таковой. Итак, значение одно, оно имеет сходные описания в лексикографических источниках, но объем концептуальной информации, фиксированный данным значение, когнитивные контексты, ассоциированные с ним, различны.

Сходный по характеру когнитивный процесс инференции с элементами специализации — сужения первоначального объема концептуальной информации — наблюдается и в примерах 1c, 1e и 1g, в которых активизируемые в сознании когнитивные контексты способствуют более точной идентификации смыслов, по сути, сужая концептуальную информацию о птицах вообще до представителей определенного класса (типа) птиц.

В примерах 1d и 1f активизируются когнитивные контексты, ассоциируемые с собственно прототипическим значением без сужения объема концептуальной информации, то есть «warm-blooded egg-laying vertebrates characterized by feathers and forelimbs modified as wings».

Перейдем к рассмотрению второй группы примеров:

2. a. Victoria Wood's quite an attractive **bird** for a fat lady.
   b. Richie left with **a bird** yesterday.
   c. Oh look ain't that, that same **bird.**
   d. He had risen in the City to a financial position which permitted him to park his BMW beside the flash docklands flat, close to his rough but pure roots, in which he kept his nice French **bird**.

Во второй группе примеров активизируется комплекс когнитивных контекстов, связанных с категорией ЧЕЛОВЕК — одушевленный объект женского рода (Victoria Wood, lady), а также фиксирующих некоторые физические (fat) и оценочные (attractive) характеристики объекта. Активизация данных областей в сознании помогает идентифицировать метафорическое, прагматически окрашенное значение — «привлекательная женщина».

В примере 2b происходит активизация когнитивного контекста, отражающего вполне стереотипную ситуацию — мужчина покидает ресторан/кафе/вечеринку в компании привлекательной девушки. В данном случае когнитивный контекст настолько стереотипен, что его реконструкция не вызовет особых затруднений, равно как и идентификация метафорического значения существительного *bird*.

В примере 2c собственно лингвистический контекст употребления слова *bird* настолько узок, что не дает возможности однозначной иден-

тификации смысла. Однако почерпнув из корпуса информацию о говорящем — молодом человеке 19 лет, беседующем со своими сверстниками, по фонетическим характеристикам его речи можно предположить, что в данном случае анализируемая лексическая единица использована в значении «молодая привлекательная женщина». Данный вывод можно сделать по стереотипным представлениям о молодых людях определенного возраста, особенностях их поведения, возможных темах разговоров, то есть исходя из имеющегося у нас конкретного когнитивного контекста, являющегося частью комплексной ментальной репрезентации полисеманта, иными словами, на основе ряда прагматических стратификационных характеристик, которые в данном случае и позволили идентифицировать значение слова.

В примере 2d активизируется когнитивный контекст, фиксирующий стереотипную информацию об успешном в финансовом и карьерном плане чиновнике и «наборе» того, чем такой мужчина должен обладать: престижная и высокооплачиваемая работа (had risen in the City to a financial position), элитное жилье (flash docklands flat), дорогая машина у дома (to park his BMW), а также привлекательная девушка в качестве дополнения к вышеперечисленному. Когнитивный контекст, который уже перестал быть культурно-специфическим, по-видимому, может вызвать сходные ассоциации у многих людей, которые неоднократно наблюдали такого рода ситуацию. Поэтому из всех значений многозначного существительного сознание активизирует именно это.

Третья группа примеров демонстрирует еще одно метафорическое значение полисеманта, которое не зафиксировано большинством традиционных лексикографических источников, однако присутствует в корпусе, — «летательный аппарат, самолет»:

3. a. How soon could Kirov get his **bird** off the ground, do you think?
b. That's a *** there is still no other **bird** that can take off!

Ментальная репрезентация в примере 3а содержит разные когнитивные контексты — эпоха (Киров), место (Россия) и действие над объектом — «поднять что-либо в воздух» (get smth. off the ground). В результате в сознании активизируется когнитивный контекст, ассоциируемый с метафорическим значением слова — «летательный аппарат, самолет, ракета». Наши фоновые знания о времени события, состоянии научно-технического прогресса сужают объем значения слова-мишени до значения «самолет», так как конструкции ракет в то время еще не были разработаны.

Во втором примере данной группы активизируется совсем иной когнитивный контекст — невозможность вылететь (take off), покинуть аэропорт и вызванное этим раздражение, выраженное обсценно. Стереотипная ситуация, знакомая всем, кто часто совершает перелеты. Данный когнитивный контекст способствует идентификации единственно подходящего значения.

Интересным представляется следующий пример:

4. Ninety seven there it is now showing, Lot ninety seven, singing **bird** in a cage I have three hundred offered for this, three hundred pounds and twenty,

three fifty, three eighty, four hundred and twenty, fifty four eighty five hundred fifty six hundred and fifty, seven hundred fifty eight hundred eight hundred pounds seated now, any more at eight hundred and fifty nine hundred and fifty nine fifty nine fifty, one thousand new bidder thousand one hundred one thousand one hundred pounds, any more?

В этом случае ментальная репрезентация многозначного слова фиксирует информацию о типичной конвенциональной ситуации — проведение торгов на аукционе. В данном контексте присутствует указание на все необходимые элементы стереотипной ситуации — продавец (ведущий аукционных торгов), покупатели (bidder), деньги (перечисление предлагаемых сумм) и, наконец, товар-лот (bird in a cage), обозначенный словом-полисемантом. Понятно, что речь вряд ли идет о продаже живой птицы, так как стереотипное представление о данной ситуации — проведении аукциона — не ассоциировано в первую очередь с продажей живых птиц. Речь, вероятно, идет о неком предмете — картине, статуэтке и т.д. Отсутствие категориальной отсылки делает невозможной более точную идентификацию смысла, создавая тем самым эффект неоднозначности.

В результате осуществленного концептуального моделирования процесса разрешения полисемии можно сделать вывод о важной роли когнитивных контекстов разного типа, являющихся частями комплексной ментальной репрезентации, ассоциированной с определенным значением слова, и способствующих идентификации соответствующего значения.

Таким образом, проведенные исследования продемонстрировали преимущества сочетания методов эмпирической лингвистики, концептуального моделирования и данных корпусов для осуществления анализа такого сложного феномена, как многозначность.

Работа выполнена в рамках гранта РФФИ 11-06-00301 «Когнитивный анализ семантики слова (компьютерно-корпусный подход)».

Список литературы

1. *Захаров В.П.* Корпусная лингвистика. СПб., 2006.

2. *Заботкина В.И., Боярская Е.Л.* Роль когнитивного контекста в разрешении многозначности: опыт когнитивного моделирования // Когнитивные исследования языка. М. ; Тамбов, 2012. Вып. 12 : Теоретические аспекты языковой репрезентации. С. 624—635.

3. *Заботкина В.И., Боярская Е.Л.* Когнитивное моделирование процесса разрешения полисемии // Когнитивные исследования языка. М.; Тамбов, 2012. Вып. 11: Международный конгресс по когнитивной лингвистике 10—12 октября 2012 года. С. 52—55.

4. *Кубрякова Е.С.* Части речи с когнитивной точки зрения. М., 1997.

5. *Кубрякова Е.С.* Язык и знание. На пути получения знаний о языке: части речи с когнитивной точки зрения. Роль языка в познании мира. М., 2004.

6. *Программы* фундаментальных исследований Президиума РАН. URL: http://www. corpling-ran.ru/links.html.

7. *Рыков В.В.* Курс лекций по корпусной лингвистике. URL: http://rykov-cl. narod.ru/c.html.

8. *Фесенко Т. А.* Концептуальное моделирование как метод познания ментальной реальности человека. Язык, сознание, коммуникация. М., 2000.

9. *Шаров С. А.* Параметры описания текстов корпуса. URL: http://bokrcorpora.narod.ru/header.html.

10. *Baker P., Hardie A., McEnery T.* A Glossary of corpus linguistics. Edinburgh, 2006.

11. *Biber D., Conrad S., Reppen R.* Corpus linguistics: investigating language structure and use. URL: http://books.google.com/books?id= 2h5F7TXa6psC.

12. *Gilguin G.* The place of prototipicality in corpus linguistics // Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis. Berlin, 2006.

13. *Gries A.* Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis. Berlin, 2006.

14. *McEnery T., Wilson A.* Corpus linguistics. URL: http://www.lancs.ac. uk/fss/courses/ling/corpus/Corpus1/1FRA1.HTM.

Об авторе

Боярская Елена Леонидовна — канд. филол. наук, доц., Балтийский федеральный университет им. И. Канта, Калининград.

E-mail: depti313@gmail.com

Author

Boyarskaya Elena — PhD, Ass. Prof., I. Kant Baltic Federal University, Kaliningrad.

E-mail: depti313@gmail.com