***S. Jones.* Antonymy: A corpus-based perspective.** London; New York: Routledge, 2002. – xvi, 193 p.

Книга Стивена Джонса, вышедшая в серии "Routledge advances in corpus linguistics" ("Публикации по корпусной лингвистике издательства Раутледж"), замечательна тем, что стремится навести мосты между стремительно развивающейся в последние годы корпусной лингвистикой и традиционной лингвистической теорией. Она призвана показать, какие преимущества дает использование текстовых корпусов в изучении такой "изъезженной" темы, как антонимия. В этой связи книга может быть интересна как специалистам в области корпусной лингвистики, так и исследователям лексической семантики, лингвистики текста и лексикографам.

Исходной идеей данной работы послужило наблюдение Дж. Джастесона и С. Катца, что "прилагательные не случайно имеют тенденцию сочетаться со своими антонимами в пределах одного предложения чаще, чем можно было бы ожидать" [Justeson, Katz 1991: 142]. Исследование С. Джонса базируется на 280-миллионном неаннотированном корпусе английских газетных текстов ("The Independent", 1988–1996 гг.), из которого программными средствами выделен подкорпус

предложений, содержащих одновременно оба члена антонимической пары. Цель монографии состоит в том, чтобы выяснить, зачем авторы употребляют слова "с противоположным значением" вместе, и создать новую классификацию антонимов, основанную не на хорошо известных синтагматических характеристиках (таких как градуируемость, тип общей семантической части, лексическая – морфологическая антонимия и др. [Leech 1974; Lyons 1977; Cruse 1986]; см. также [Апресян 1974]), а на том, что С. Джонс называет "парадигматическими свойствами" связанной антонимической пары[1].

Книга состоит из 11 глав. В первых двух главах дается общее представление об антонимии и представлена краткая история ее изучения. В третьей главе говорится о том, как создавалась рабочая база данных и каким образом классифицировались вошедшие в нее предложения. Четвертая, пятая и шестая главы – фактографические, в них описываются восемь классов ролей, которые играют антонимические пары в организации предложения. В седьмой главе представлен статистический анализ совместной встречаемости антонимов на фоне их общей частотности, и определяются наиболее эндемичные, т.е. "преданные друг другу", антонимические пары (например, наречие *indirectly* в каждых 10 случаях из 28 употребляется вместе с *directly*). В восьмой главе утверждается, что, употребляя антонимы совместно, пишущие обычно придерживаются определенного "стандартного порядка" их появления, и исследуются семантические и иные факторы, обуславливающие этот порядок для каждой пары. В девятой главе изучается вопрос о том, насколько признаки частеречной принадлежности и градуируемости влияют на функционирование антонимов в тексте (в терминах нововыявленных антонимических ролей). Десятая глава посвящена выявлению новых антонимических пар в тексте на основе выявленных ранее типичных контекстов семантически связанных употреблений. В одиннадцатой главе подводятся итоги исследования.

С точки зрения заявленной "corpus-based" перспективы, особый интерес представляют главы 3-я (методы отбора материала) и 10-я (автоматическое обнаружение антонимических пар в корпусе), на которых мы бы хотели остановиться подробнее. При этом нам кажется важным проанализировать их содержание сквозь призму следующих проблем:

1) Корпусная лингвистика предполагает разделение труда между человеком-исследователем и компьютером. Где проходят эти границы?

2) Насколько корпусная лингвистика приближается к точным наукам? Какова степень вмешательства интуиции исследователя в корпусное исследование?

Процедура отбора языкового материала, излагаемая в 3-й главе, состоит из нескольких этапов. Во-первых, создается словарь "бесспорных" антонимов (56 пар), сбалансированный по разным признакам. В нем представлены как высоко-, так и низкочастотные антонимы, включены представители разных частей речи (прилагательные, наречия, глаголы и существительные), лексические и морфологические антонимы (ср. *new/old* 'новый/старый' и *honest/dishonest* 'честный/нечестный'), градуируемые и неградуируемые (ср. *high/low* 'высокий/низкий' и *male/female* 'мужчина/женщина'). Во-вторых, на основании этого словаря компьютерная программа выбирает из большого корпуса все предложения, содержащие антонимические коллокации (более 55 000 примеров); к ним добавляются предложения, содержащие некоторое слово и его коррелят с приставкой *un-* (*X / un-X;* ср. *clever / un-clever* 'умный/неумный'; еще более 10 000 примеров). В-третьих, из полученного подкорпуса отбирается эталонная база данных, насчитывающая 3 000 предложений. Это наиболее трудоемкий и длительный этап отбора материала, осуществляемый по большей части вручную: в самом деле, только исследователь может установить, что коллокация антонимов в предложении не случайна и что между ними существует определенная семантическая связь. Кроме того, эталонная база примеров отбирается в соответствии со сложной системой критериев (sampling method):

– должно быть не более 60% примеров на антонимы-прилагательные, и не менее чем по 10% примеров на существительные, глаголы и наречия;

– должно быть не менее 250 примеров (8,3%) на неградуируемые антонимы и не менее 250 примеров на морфологические антонимы;

– иерархия частотности в общем случае не нарушается (более частотные антонимы остаются более частотными), но шкала частот сознательно сжимается, чтобы при статистических расчетах можно было принять во внима-

[1] Исследование С. Джонса выполнено в рамках проекта ACRONYM Ливерпульского университета [Renouf 1996], нацеленного на автоматическую идентификацию гипо-гиперонимических, синонимических и тому подобных отношений в тексте.

ние даже самые редкие пары (рекордсмен – пара *dishonest/honest* – представлена только 12 предложениями).

К сожалению, автор не поясняет, почему были выбраны именно такие количественные соотношения, как осталось без объяснения и то, для чего вообще понадобилось отбирать 3 000 эталонных примеров (впрочем, последнюю цифру объяснить легче – как кажется, это разумный компромисс между тем количеством предложений, которое можно обработать вручную, и количеством примеров, достоверным для статистического анализа). Наибольший недостаток процедуры отбора материала, по нашему мнению, состоит в том, что в программе поиска коллокаций не был реализован морфологический компонент, тем самым, например, ни формы множественного числа существительных, ни формы прошедшего времени глаголов не попали в выборку.

С. Джонс критикует "до-корпусную" лингвистику за то, что в ней слишком многое зависит от интуиции исследователя: лингвисты, как правило, подбирают такие языковые примеры, которые подтверждают их исходные теоретические установки, но могут не заметить другие языковые факты, не объясняемые их теорией. Корпусный подход, по мысли С. Джонса, дает исследователю более объективный материал. Более того, исходный корпус примеров должен сам подсказать исследователю способ его классификации и даже последовательность анализа полученных классов.

Вторая часть 3-й главы так и называется – "классификация базы данных". На примере 23 предложений из базы С. Джонс показывает, каков был ход его рассуждений при признании примеров подобными или различными (на этом этапе участие интуиции приветствуется). Выделяется восемь основных классов:

1) "вспомогательная" (ancillary) антонимия: *stamps are* ***popular*** *but collecting is* ***unpopular*** 'марки популярны, но их коллекционирование непопулярно';

2) "сочиненная" (coordinated) антонимия: *the company's policy is to recruit* ***skilled*** *and* ***unskilled*** *workers* 'стратегия компании состоит в том, чтобы привлекать квалифицированных и неквалифицированных работников';

3) "сравнительная" (comparative) антонимия: *those who* ***succeed*** *more than they* ***fail*** 'те, кто достигает успеха чаще, чем терпит неудачу';

4) "различительная" (distinguished) антонимия: *the gap between* ***rich*** *and* ***poor*** *has widened* 'пропасть между богатыми и бедными увеличилась';

5) "транзитивная" (transitional) антонимия: *how easy to slip from the* ***legal*** *to the* ***illegal*** *trade* 'как легко соскользнуть с легальной торговли на нелегальную';

6) "негативная" (negated) антонимия: *to facilitate the re-establishment of* ***peace***, *not* ***war*** 'способствовать установлению мира, а не войны';

7) "экстремальная" (extreme) антонимия: *except when the soil is too* ***wet*** *or too* ***dry*** 'за исключением тех случаев, когда почва слишком влажная или слишком сухая';

8) "идиоматическая" (idiomatic) антонимия: *to blow* ***hot*** *and* ***cold*** 'постоянно менять свои взгляды' (букв. 'веять жарой и холодом').

В то же время, характерно, что С. Джонс не делает обобщений по поводу того, какие именно типы критериев он использует при классификации. Со своей стороны заметим, что здесь применяются и содержательные критерии, такие как семантическое отношение между членами антонимической пары и "риторический эффект", и некоторые нестрогие синтаксические критерии, а именно, исследователь нащупывает наиболее частотные языковые конструкции и сигнальные слова, характерные для каждого из выделенных классов, например, *X, not Y* 'X, а не Y' ("негативная" антонимия), *the gap* 'расхождение (между)'. Безусловным достижением анализа С. Джонса является отказ от использования чисто поверхностных синтаксических критериев, хотя, как показывают главы 4–6, при выделении подклассов и мелких "миноритарных" классов он руководствуется в большей мере формальными признаками; ср., в частности, неоднородный в семантическом отношении класс "косая черта" (oblique stroke), в котором примеры "объединены по признаку пунктуации, а не по функции": *some mistrust of the* ***alive/dead*** *distinction* 'некоторое недоверие к различию между живым/мертвым', *Sussex's* ***new/old*** *boy Adrian Jones* 'новый старый игрок из Суссекса Адриан Джонс' и др. (с. 96).

Накопленные сведения о словах и конструкциях, сигнализирующих об антонимической коллокации, находят применение в 10-й главе. Здесь делается попытка обнаружить устойчивые антонимические пары, еще не зафиксированные словарями – так называемые "антонимы завтрашнего дня". С помощью трех продуктивных диагностических конструкций (*both X and Y, between X and Y, whether X or Y*) выявляются потенциальные антонимы для трех достаточно частотных ключевых слов (в позиции X): *good* 'хороший', *natural* 'естественный' и *style* 'стиль, способ выражения'. В частности, у слова

good, помимо очевидного *bad* 'плохой', обнаруживается коррелят *evil* ('добро/зло') (37,3% коллокаций), а также ряд слов *wicked* 'злой', *nasty* 'отвратительный', *pathetic* 'жалкий', *poor* 'скудный', *lousy* 'вшивый' и др., контрастирующих по значению, но встреченных менее двух раз. В то же время, из списка коллокаций вычеркиваются пары, связанные особыми не-антонимическими отношениями, ср. *good and excellent* 'хороший и превосходный', а также окказиональные сочетания типа *good and true (story)* 'хороший и правдивый (рассказ)'. Конечно, было бы интересно узнать, какие результаты дал бы этот подход на более представительном корпусе и при использовании всех "сигнальных" конструкций, но a priori он кажется чрезвычайно плодотворным.

В отношении разделения труда между человеком и машиной исследование С. Джонса показывает, что компьютерные программы совершенно незаменимы для статистической обработки результатов. Однако не "интеллектуализованная" программа не может сама по себе отобрать необходимый исследователю материал: она только составляет исходный подкорпус примеров (по заданному исследователем словарю) и может привлекаться для поиска предложений, содержащих сигнальные слова и конструкции (которые затем нуждаются в дополнительной "человеческой" экспертизе).

Соответственно, отбор примеров на базе корпуса не свободен от интуиции лингвиста. Человек-исследователь участвует в создании эталонного словаря и в экспертизе эталонной базы данных, а критерии выборки зависят от его априорных представлений о том, как должны выглядеть эталонный словарь и эталонная база данных.

Наиболее независимы от интуиции статистические методы исследования. Стремясь никак не проявить собственные предпочтения относительно оценки важности выделенных классов, С. Джонс строит их описание в последующих главах строго в соответствии с результатами статистического анализа. Наибольшего внимания удостаиваются два самых частотных класса – "вспомогательная" (38,7% примеров) и "сочиненная" антонимия (38,4%). Удивительным образом, "сравнительная" антонимия встречается всего в 6,8% примеров, хотя "измерение одного антонима относительно другого всегда интуитивно признавалось одним из свойств антонимии" (с. 76). Из этого частотного различия С. Джонс делает важный вывод о природе антонимических коллокаций: "Пишущие предпочитают использовать антонимию как сигнал, будь то указание на полную шкалу значений признака, или подчеркивание контраста между двумя другими понятиями. Простая констатация, что X отличается от Y, имеет малый эффект при использовании в языке" (с. 103).

Вместе с тем, хотелось бы заметить, что "статистический" порядок изложения не всегда удобен для восприятия анализируемого материала. Например, близкие и логически связанные классы "вспомогательной" (ancillary) и "специфицирующей" (specification) антонимии рассматриваются в разных главах (последний "миноритарный" класс представлен в конструкциях с числительными: *there were 51 **male** and 140 **female** prisoners* 'там находились: 51 заключенный мужского и 140 заключенных женского пола'), а при анализе факторов, обуславливающих порядок антонимов в тексте (гл. 8), семантические признаки ("позитивность", "маркированность" и др.) перемежаются с признаками, характеризующими фонологическую структуру и частотность антонимов.

Кроме того, нельзя отрицать, что исходная статистика зависит от степени дробности классов, определяемой, опять-таки, по интуитивным критериям, а значит, "статистический" принцип изложения также не избавлен до конца от влияния исследователя.

В заключение хотелось бы отметить, что перед нами не просто семантическое исследование, использующее в качестве инструмента анализа текстовый корпус, но исследование, сознательно ограниченное его рамками. Поэтому, например, в книге отсутствуют отрицательные языковые примеры, привычные для традиционных работ по семантике. Утверждается, что использование корпуса избавляет исследователя от необходимости вводить отрицательный языковой материал: само отсутствие таких примеров в корпусе подтверждает правоту автора. Действительно, почти ни один пример в книге не вызывает нареканий у читателя (напомним, что цитаты берутся из газеты "The Independent", где публикуются хорошо отредактированные тексты). Однако, что же будет делать исследователь с менее "темперированным" корпусом, составленным, например, из текстов Интернета? Аналогичная проблема может возникнуть при использовании корпуса художественных произведений: например, С. Джонс критикует корпус произведений Агаты Кристи из [Mettinger 1994], в котором многие предложения "кажутся стилизованными и слишком книжными" (с. 22).

Как говорит в предисловии к книге редактор серии М. Хей (M. Hoey), существует опасность, что корпусная лингвистика сосредоточится на детальном описании самих кор-

пусов и станет, тем самым, наукой о корпусах. Но это было бы так же нелепо, как "микроскопная биология" или "лопатное садоводство". Как известно, изобретение микроскопа привело к возникновению микробиологии, но сначала микроскоп использовали для любования узорами на крыльях бабочек. По-видимому, требуется некоторое время, чтобы лингвистика научилась использовать текстовые корпуса как надежный инструмент языкового анализа.

СПИСОК ЛИТЕРАТУРЫ

Cruse 1986 – *D.A. Cruse*. Lexical semantics. Cambridge, 1986.

Justeson, Katz 1991 – *J.S. Justeson, S.M. Katz*. Redefining antonymy: the textual structure of a semantic relation // Literary and linguistic computing. 7. 1991.

Leech 1974 – *G. Leech*. Semantics. Harmondsworth, 1974.

Lyons 1977 – *J. Lyons*. Semantics. V. 2. Cambridge, 1977.

Mettinger 1994 – *A. Mettinger*. Aspects of semantic opposition in English. Oxford, 1994.

Renouf 1996 – *A. Renouf*. The ACRONYM project: discovering the textual thesaurus // J.M. Aarts, P. De Haan, N.H.J. Oostdijk (eds.). Synchronic corpus linguistics – Papers from the Sixteenth international conference on English language research on computerised corpora. Amsterdam, 1996.

Апресян 1974 – *Ю.Д. Апресян*. Лексическая семантика. М., 1974.

О.Н. Ляшевская