

УДК 81.139

Суворина Е.В.

Московский городской педагогический университет

**ИСПОЛЬЗОВАНИЕ СЕТЕВОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ
BNCweb ПРИ ОПИСАНИИ ЛИНГВО-КОГНИТИВНЫХ
ОСОБЕННОСТЕЙ СЛОВ «EMOTION» И «FEELING»
В СОВРЕМЕННОМ АНГЛИЙСКОМ ЯЗЫКЕ***

E. Suvorina

Moscow City Pedagogical University

**BNCweb-BASED LINGVO-COGNITIVE ANALYSIS OF THE WORDS EMOTION
AND FEELING IN THE CONTEMPORARY ENGLISH LANGUAGE**

Аннотация. Предлагается метод дискурсивного анализа данных аннотируемого лингвистического корпуса с использованием сетевого программного обеспечения BNCweb, позволяющий описывать лингво-когнитивные особенности семантически трудно дифференцируемых лексических единиц путем построения и сопоставления их лексических профилей. Основным отличием предлагаемого подхода является отказ от использования первичных частотных данных в пользу обобщенной статистической меры близости, вычисляемой методом максимума правдоподобия.

Ключевые слова: когнитивная лингвистика, эмоция, чувство, аннотируемый лингвистический корпус, сетевое программное обеспечение BNCweb, метод максимума правдоподобия, лексический профиль.

Abstract. An advanced technique for the discourse analysis of the annotated corpus via the BNCweb within the framework of cognitive linguistics is proposed. The analysis of near-synonyms is based on the lexical profile evaluation, improved, in turn, by the usage of the log-likelihood similarity measure instead of raw frequency data. The technique allows reproducible results and can be applied to the very hard cases of the pairs of words considered as absolute synonyms in contemporary dictionaries.

Key words: cognition, emotion, feeling, annotated linguistic corpus data, the BNCweb software, log-likelihood similarity measure, lexical profile.

Изучая концептуальную структуру языка, современный лингвист стоит перед непростым выбором метода исследования, так как диапазон методов, применимых к исследованию концептуального структурирования языка, достаточно широк. Основными методами когнитивной лингвистики являются: *интроспекция, аудио- и видеографический анализ, корпусной анализ и экспериментальный метод*. Л. Телми отмечает, что каждый метод имеет ресурсы и ограничения, и именно поэтому не существует метода, имеющего преимущество над другими и являющегося «золотым» стандартом исследования [12].

Если следовать основным критериям современного научного исследования (таким, как *объективность, доказательность и воспроизводимость результатов*), особую ценность представляет изучение данных, представленных в лингвистических аннотированных (размеченных) корпусах, или **автоматизированная экспертиза корпуса**, выбранного для лингвистического анализа. Основным преимуществом такого метода исследования является возможность обработки огромных массивов языковых данных, позволяющая делать объективные, статистически подтвержденные выводы, а также возможность формулировать различные запросы, составлять кросстабуляции и получать результаты обработки данных в кратчайшие сроки. При этом гарантирована воспроизводимость полученных результатов и полностью исключён субъективизм и личные пристрастия исследователя.

Предметом исследований в рамках корпусной лингвистики могут стать как отдельные лексические единицы, так и их сочетания (конкордансы), а также тексты. Основными фак-

* © Суворина Е.В.

торами, определяющими то, насколько полно можно использовать возможности обработки дискурсивных данных, является выбор языкового корпуса и наличие компьютерных программ, способных обработать отобранные для лингвистического анализа данные.

Британский национальный корпус (БНК), а также сетевое программное обеспечение BNCweb – идеальный пример использования языкового корпуса в лингвистических исследованиях. БНК является образцовым лингвистическим корпусом из множества существующих в настоящий момент. Отобранный в корпусе материал сопровождается детальным описанием с указанием критериев, по которым создавался корпус. БНК состоит более чем из 4.000 текстов (файлов), каждый из которых имеет идентификатор – имя файла. Каждый файл – пример устной или письменной речи современного английского языка. Размер примеров варьируется от нескольких десятков до нескольких сотен слов. Письменный компонент корпуса охватывает около 90 млн. слов.

Очевидно, что любые электронные варианты корпусов бесполезны без надлежащего интерфейса пользователя. Официальный дистрибутив БНК распространяется с программой Xaira, web-альтернативой которой служит BNCweb. Последняя его версия – CQP редакция – формулирует задачи системы BNCweb как «дружественный к пользователю, технически богатый инструмент анализа корпуса» [9, XIII]. Использование подобного сетевого обеспечения даёт возможность на практике гуманитариям с минимальной пользовательской подготовкой эффективно получать информацию, работая в режиме диалога с системой, без изучения сложных языков программирования или синтаксиса языков и систем управления базами данных. Программное обеспечение BNCweb позволяет исследователю максимально использовать данные, представленные в БНК, компилировать результаты запросов, а также представлять результаты запросов, выполненные по разным сечениям корпуса в виде классических многовходовых таблиц сопряжённости,

пригодных для многофакторного дисперсионного анализа, руководствуясь которым можно оценить не только влияние каждого фактора в отдельности, но и их взаимодействие [4, 246-260].

Используя возможности, предоставляемые сетевым программным обеспечением BNCweb, мы постараемся проанализировать, **что** говорит сам язык о структурах знаний, стоящих за обозначениями слов *emotion* и *feeling* в современном английском языке. В данной статье мы попытаемся установить, вслед за Дж. Лакоффом [11], **как** эмоции и чувства отражаются в «зеркале» языка [1; 3]. В связи с этим отметим, что если первые десятилетия когнитивизма были отведены описанию когниции различного рода мыслительных процессов, осуществляемых человеком, то позднее (вторая половина 80-х гг.), в результате этих исследований, когнитологи пришли к пониманию того факта, что когниция и эмоции связаны между собой неразрывной связью [6] и что в этом смысле и не может быть «бесстрастного», не связанного с эмоциями, восприятия мира. Поэтому сегодня речь идёт не столько о том, какие структуры знаний закрепились постепенно в конкретных языках в понятиях «эмоция», «чувство» (и т. п.), а какие из них (и как) – отражают опыт обыденного сознания человека. Для понимания этого факта и следует обратиться к достоверным, статистически выверенным дискурсивным данным, с помощью чего мы сможем отследить ту степень «фильтрации», которую они прошли со временем, прежде чем сформироваться и продолжить формироваться в более устойчивые структуры знаний, стоящие за словами *emotion* и *feeling* в современном английском языке.

В исследованиях, использующих данные лингвистического корпуса, для описания множества слов, в окружении которых находится анализируемое, принято использовать термины: **лексический профиль слова** (lexical profile) и **конструкционный профиль слова** (constructional profile). Лингвистический анализ лексического профиля слова подразумевает нахождение взаимосвязи

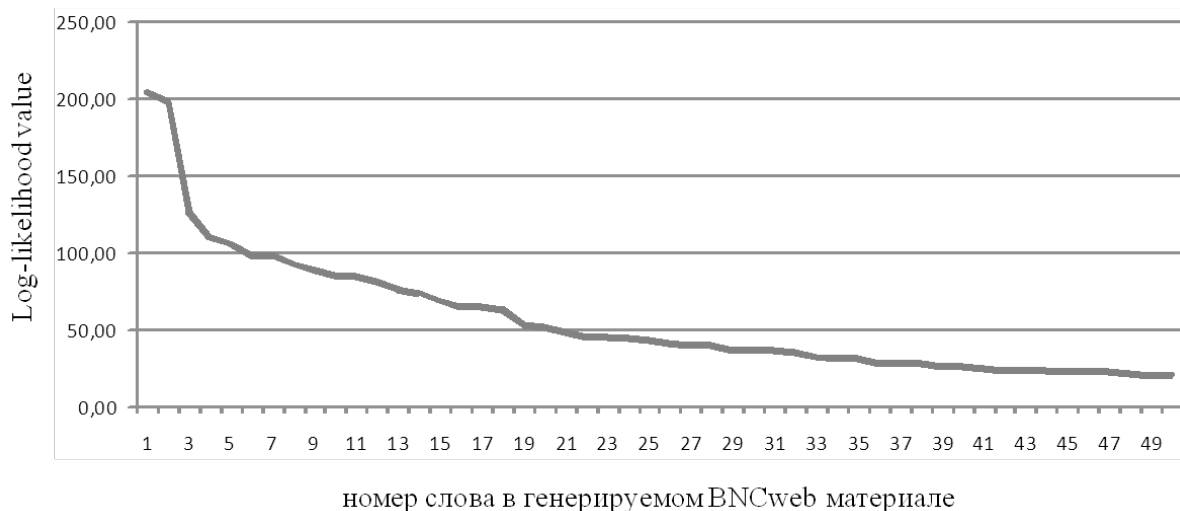


Рис. 1. Зависимость величины критерия максимума правдоподобия от номера слова при построении профиля слова *feeling*.

между значением слова и конструкциями, в которых оно употреблено в корпусе. Анализ конструкционного профиля подразумевает описание грамматических конструкций, в которых участвует данное слово [10]. В рамках данной статьи мы делаем попытку частично описать лексические профили английских слов *emotion* и *feeling*, определяющих лингвокогнитивные особенности этих языковых единиц, с целью их дальнейшей дифференциации. При формировании лексического профиля мы будем включать не вообще все слова, употребляемые вместе с анализируемым, а только те из них, статистические меры близости которых к анализируемому (вычисляемые программным обеспечением BNCweb) удовлетворяют заданным критериям. Принципиальным отличием настоящей работы от методики, использованной в современных исследованиях [7; 10], является отказ от частотных данных (также доступных в среде BNCweb) в пользу более сложных статистических мер близости. Как отмечалось исследователями, «сырые» частотные данные (raw data) нередко оказываются искажёнными (biased), особенно в области низких частот, и их корректное применение требует явного учёта истинного числа включений анализируемого слова [9]. То есть корректная работа с частотами требует на практике

нетривиального учёта 2-3 параметров, а переход на статистические меры позволяет избежать этих проблем. В настоящей работе из всего предлагаемого системой BNCweb списка мер нами был выбран **метод максимума правдоподобия** (log-likelihood method). Если рассмотреть зависимость величины λ – безразмерного значения правдоподобия [9] – от номера слова в множестве, отсортированном по убыванию λ , то окажется, что эта зависимость имеет устойчиво воспроизводимый вид. Вначале идут слова с высоким λ , значения λ быстро убывают и после небольшого участка плато падают до малых значений и далее практически не меняются. На рис. 1 показана зависимость величины критерия максимума правдоподобия (ось ординат) от номера слова в корпусе (ось абсцисс) при построении профиля слова *feeling*.

Лексическим профилем анализируемого слова мы будем называть подмножество, такое, что для всех слов, принадлежащих профилю, значение критерия максимума правдоподобия (λ) превышает критическое. Порог $\lambda_{\text{крит}}$ будем выбирать по выходу зависимости λ на плато по аналогии с методом L-кривой (L-curve technique) [8] (см. рис. 1).

Лексический профиль слова будет содержать конструкции типа: имя существительное + *emotion/feeling* (any noun + *emotion/*

feeling), глагол + *emotion/feeling* (any verb + *emotion/feeling*), а также имя прилагательное + *emotion/feeling* (any adjective + *emotion/feeling*). Разумеется, профиль может быть построен и для других частей речи, но в данной работе мы проанализируем эти три типа конструкций, исключая служебные слова, артикли и пр.

Отдельный важный аспект построения множества профиля анализируемого слова – выбор длины окна (смещение, измеренное в словах, относительно вхождения анализируемого слова – truncation boundaries). Система BNCweb, используемая в настоящей работе, позволяет строить подмножество корпуса и выполнять необходимые статистические расчёты в диапазоне ± 10 слов от заданного. Мы ограничились диапазоном ± 3 слова. Данный выбор обусловлен следующими соображениями: с ростом размера диапазона возникает риск включения в множество слов, не относящихся к анализируемому, а сокращение размера окна до 1 слова, следующего строго перед/за исследуемым, придаёт избыточный вес служебным словам, артиклям, притяжательным или указательным местоимениям и т. д.

В основе нашего анализа лежит композитная статистическая гипотеза, состоящая из двух положений: 1) психоэмоциональное состояние человека, вызванное внешними воздействиями, влияет на когнитивные процессы в мозге, стимулируя его к использованию (выбору) тех или иных слов; 2) при описании чего бы то ни было человек, используя те или иные слова, включает в контекст (с ненулевой вероятностью) описание сопутствующих или стимулирующих внешних факторов. Тогда, анализируя профили слов, можно рассмотреть связь внешнего, выраженного профилем слова, с процессом выбора анализируемого слова из когнитивного банка.

Дополнительное ограничение заключалось в выборе подмножества корпуса, содержащего только письменные (written) файлы. Соображения в пользу такого выбора строятся на допущении, что при письме человек более склонен подбирать слова и формулировки, зачастую довольно тщательно, с многократным редактированием, чем в устной речи. Другим аргументом является достаточно значительная ёмкость письменной компоненты корпуса (90% от общего количества файлов, представленных в БНК).

Таблица 1

Лексический профиль слова *FEELING*, полученный по всему корпусу

Части речи	Слова	Вероятный аналог в русском языке	Значения критерия максимума правдоподобия (λ)
Имена существительные	helplessness	беспомощность	204,53
	unease	беспокойство	197,69
	dread	страх	126,88
	guilt	вина	110,52
	satisfaction	удовлетворение	105,40
	panic	паника	98,76
Глаголы	to have		756,36
	to be		261,00
	to get		96,67
	to experience	испытывать	66,118
	to dispel	развевать	62,20
	to persist	сохраняться	47,57
Имена прилагательные	strong	сильный	460,54
	sinking	тонуший	408,44
	strange	странный	363,16
	bad	плохой	225,29
	uncomfortable	некомфортный	211,01

Таблица 2

Лексический профиль слова *EMOTION*, полученный по всему корпусу

Части речи	Слова	Вероятный аналог в русском языке	Значения критерия максимума правдоподобия (λ)
Имена существительные	voice	голос	128,17
	intensity	интенсивность	59,41
	expression	выражение	57,33
	depth	глубина	56,37
Глаголы	to show (showed)	показывать	160,35
	to be choked with	подавлен	73,82
	to show (showing)	показывая	63,05
	to show	показать	62,88
	to feel	чувствовать	57,33
Имена прилагательные	suppressed	подавленный	75,07
	expressed	выраженный	66,25
	strong	сильный	64,96
	powerful	мощный	64,24
	human	человеческий	59,50

Представим результаты построения лексических профилей анализируемых слов без разбиения на эпохи и возрастные когорты. Для слова *feeling* результаты приведены в табл. 1.

Аналогично результаты построения лексического профиля для слова *emotion* приведены в табл. 2. Включение слова «emotion» в список имён существительных вызвано попаданием в корпус научной литературы по психологии и смежным областям знаний, что показал анализ предложений, входящих в выборку. Именно поэтому мы не будем в дальнейшем принимать во внимание это слово при анализе, оставив его для адекватной визуализации динамики изменения значений λ .

Итак, дадим интерпретацию представленных в табл. 1-2 показателей. Начнём с сопоставления **лексических профилей** анализируемых слов. Необходимо учитывать, что эмоциональная составляющая всегда есть часть восприятия мира и что подлинная иерархия семантических отношений задаётся на более высоком уровне абстракции,

т. е. связями лексем *emotion* и *feeling* со словом *perception* (восприятие) и также, по-видимому, – со словом *consciousness* (сознание). Анализируя полученные результаты выборки, мы исходим из базового предположения о том, что человеческое сознание включает все воспринятое и осмысленное человеком, и что восприятие – это и процесс, и результат процесса взаимодействия человека с миром, осуществляемого по всем каналам поступления информации (всеми органами его чувств).

Прямое сопоставление полученных данных показывает, что лексические профили анализируемых слов отличны, т. е. перед нами две абсолютно разные лексические единицы, которые не могут быть взаимозаменяемыми. При этом в современных лексикографических справочниках слова *emotion* и *feeling*, как правило, представляются как взаимозаменяемые лексические единицы [5].

Представленные в табл. 1 данные свидетельствуют о том, что лексический профиль слова *feeling* составляют имена существительные с ярко выраженным негативным

оттенком значения (*беспомощность, беспокойство, страх, вина, паника*), т. е. отрицательным знаком эмоционального реагирования. Первое «положительное» относительно переживаний (ощущений) слово *satisfaction* (*удовлетворение*) появляется только на пятом месте, уступая по критерию значения максимума правдоподобия имени существительному *helplessness* (*беспомощность*) почти в два раза. В конструкциях «*any noun + emotion*» лексический профиль слова *emotion* формируют абстрактные имена существительные (*голос, интенсивность, выражение, глубина*), относящиеся скорее к характеристике изменения вегетативных показателей и психомоторике эмоционального реагирования человека.

Принимая во внимание вышесказанное, мы можем сделать вывод о том, что в сознании носителя современного английского языка, лексическая единица *emotion* занимает особое место. Слова *emotion* и *feeling* находятся на совершенно разных уровнях абстракции в процессе концептуализации мира.

В глагольных конструкциях («*any verb + feeling*») лидирующее место по значению λ занимают глаголы с общим значением «обладание» (*to have, to be, to get*). Лексический профиль слова «*emotion*» в конструкциях того же типа («*any verb+emotion*») представлен, прежде всего, глаголом *to show*. Основная масса примеров, представленных в БНК, свидетельствует об употреблении этого глагола в конструкциях с отрицательной частицей «но», т. е. в значении «скрывать эмоции, не показывать эмоции». Например: *Cranog Jones showed no emotion as the judge passed sentence* (K1L 794)* [9, 57-59]. Таким образом, дискурсивные данные, представленные в БНК, удивительным образом подтверждают одну из основных особенностей национального характера англичан – «эмоции даны, чтобы их контролировать» [2, 16].

Особый интерес при сравнении лексических профилей анализируемых слов пред-

ставляют списки имён прилагательных в конструкции типа *any adjective + emotion/feeling*. Так, в лексическом профиле слова *feeling* преобладают имена прилагательные, имеющие тенденцию к описанию интенсивности чувств (*strong*), а также описания негативных переживаний (*strange, bad, sinking*). Вероятнее всего, имя прилагательное *sinking* (*тонущий*) следует понимать в смысле ощущений, близких к ситуации тонущего человека (ср.: в русском языке – *хвататься за соломинку*). Например: 1) *I had the sinking feeling of failure* (AOU2570). 2) *The sinking feeling I get when I fold it up again and stuff the envelope under my bed makes me realize how much I wanted it to come up with some answers* (ADG118). Первое прилагательное с положительной окраской (*wonderful*) появляется в лексическом профиле лексемы *feeling* только на 12 месте в диапазоне глубокого плато на λ, уступая по величине лидерам более чем в четыре раза.

Дескриптивная лексика, формирующая лексический профиль слова «*emotion*», равно как и глагольная, прежде всего, сигнализирует о дихотомии «выражать эмоции – подавлять эмоции» (*suppressed, expressed*). Однако, как и в случае с глаголами, лидирует значение *suppressed* (*подавленный*). Имена прилагательные *strong, powerful* (*сильный, мощный*) характеризуют интенсивность переживаний. Следует отметить, что в именно конструкции «*any adjective + emotion/feeling*» (по компоненту *strong*) лексические профили анализируемых единиц совпадают. Дескриптив «*human*» особо подчёркивает тот факт, что само понятие «эмоция» является прерогативой именно человека.

Сравнение лексических профилей исследуемых слов позволяет сделать следующие выводы. Полученные данные по всему письменному корпусу показали, что даже на такой ограниченной выборке параметров для подбора дискурсивных данных в Британском национальном корпусе, лексические профили слов *feeling* и *emotion* обнаруживают как области их сходства, так и предполагают дифференцирующие их черты, что и позволяет относить эти лексические единицы к разряду

* Здесь и далее ссылки на источники примеров приведены в формате системы BNCweb (имя файла БНК + s-номер единицы текста).

синонимичных. Однако приведённые выше примеры свидетельствуют о том, что когнитивные процессы в сознании социализированного человека приводят к осознанному выбору только одного из этих слов при разных обстоятельствах. А это, в свою очередь, доказывает, что в речи отражены процессы мышления человека, во всяком случае, при *осреднении* по достаточно большому числу реализаций (большой выборке). Безусловно, всегда можно найти контрпример к полученным результатам, но, как и вообще в статистике, количество таких контрпримеров возрастает гораздо медленнее, чем увеличивается объём анализируемого корпуса. В то же время количество примеров, подтверждающих статистические выводы, растёт пропорционально объёму корпуса.

Очевидно, что описанные выше лингвокогнитивные особенности слов *emotion* и *feeling* лишь частично формируют лексические профили анализируемых слов согласно заданной выборке. Представленные результаты свидетельствуют о необходимости дальнейшего изучения и составления тонкой структуры лексических профилей выбранных для лингвистического анализа слов, а именно: рассмотрение предложных сочетаний, конструкций с анализируемыми словами в форме множественного числа и т. д. Результаты, полученные в ходе дальнейшего анализа, будут представлены нами в следующих публикациях.

ЛИТЕРАТУРА:

1. Кубрякова Е.С. О месте когнитивной лингвистики среди других наук когнитивного смысла и о её роли в исследовании процессов категоризации и концептуализации мира // Когнитивные исследования языка. – Вып. VII. Типы категорий в языке: сб. науч. тр./ Гл. ред. серии Е.С. Кубрякова, отв. ред. вып. Н.Н. Болдырев. – М.: Ин-т языкознания РАН; Тамбов: Издательский дом ТГУ им. Г.Р. Державина, 2010. – С. 13-18.
2. Паксман Дж. Англия: Портрет народа. – СПб.: Амфора. ТИД Амфора, 2009. – 380 с.
3. Рахилина Е.В. Когнитивный анализ предметных имен: семантика и сочетаемость. – М.: Русские словари, 2000. – 416 с.
4. Сидоренко Е.В. Методы математической обработки в психологии. – СПб.: ООО «Речь», 2001. – 347 с.
5. Суворина Е.В. Феномен «эмоция» в современной лексикографии // Аракинские чтения. Актуальные проблемы лингвистики и методики преподавания английского языка: Сб. науч. ст. – М.: МГПУ, 2006. – С. 60-68.
6. Daneš F. Cognition and emotion in discourse interaction. A preliminary survey of the field // Preprints of the plenary session papers of the 13-th international congress of linguists. – Berlin, 1987. – P. 272-291.
7. Delgado Ana R. Spanish basic emotion words are consistently ordered // Qual Quant (43). (DOI 10/1007/s11135 – 007 – 9121 – 3), 2009. – P. 509-517.
8. Hansen P.C., O'Leary D.P. The use of the L-curve in the regularization of discrete ill-posed problems // SIAM J., Vol.14 (6), 1993. – P. 1487 – 1503 p.
9. Hoffmann S., Evert S., Smith N., Lee D., Prytz Y. Corpus Linguistics with BNCweb – a Practical Guide. – Frankfurt am Main: Peter Lang, 2008. – 288 p.
10. Janda L., Solovyev V. What constructional profiles reveal about synonymy: A case study of Russian words for SADNESS and HAPPINESS // Cognitive Linguistics. Vol. 20(2). – 2009. – P. 367-393.
11. Lakoff G. Women, fire and dangerous things: What categories reveal about the mind. – Chicago, 1987. – 614 p.
12. Talmy L. Foreword // Methods in Cognitive Linguistics/ ed. by M. Gonzalez - Marquez. – Amsterdam: John Benjamins, 2007. – P. XI-XXII.