

Суворина Екатерина Владимировна

старший преподаватель кафедры фонетики
английского языка
Института иностранных языков
Московского городского
педагогического университета
тел.: (495) 607-47-74

**СЕМАНТИЧЕСКАЯ
ДИФФЕРЕНЦИАЦИЯ СЛОВ
С ПОМОЩЬЮ ПОСТРОЕНИЯ
ИХ ЛЕКСИЧЕСКИХ ПРОФИЛЕЙ**

Аннотация:

На примере словарно синонимичных слов "emotion" и "feeling" показано, как автоматизированная экспертиза лингвистического корпуса доказывает и определяет разницу в концептуальном содержании лексических единиц. Впервые приводятся полные лексические профили слов "emotion" и "feeling" и детальный анализ группы адъективных коллокаций.

Ключевые слова:

автоматизированная экспертиза лингвистического корпуса, сетевое программное обеспечение BNCweb, выборочное наблюдение, статистические таблицы, лексический профиль.

Suvorina Ekaterina Vladimirovna

senior lecturer of the chair of phonetics of
the English language,
Institute of Foreign Languages of
Moscow City
Pedagogical University
tel.: (495) 607-47-74

**ELABORATION OF THE SEMANTIC
DIFFERENCE BETWEEN WORDS
BY MEANS OF THE LEXICAL
PROFILES BUILT BY UNSUPERVISED
BNCWEB SYSTEM**

The summary:

The article describes how one can define the difference between concepts of any certain pair of words using the unsupervised elaboration of lexical profiles built by BNCWeb system. The example under analysis is the pair "emotion" vs. "feeling". For the first time we present the entire lexical profiles of the words and the in-depth analysis of the adjective collocations.

Keywords:

the unsupervised assessment of a linguistic corpus, the BNCweb software, sampling, statistical tables, lexical profile.

Современная корпусная лингвистика использует статистический анализ коллокаций изучаемых слов в языковом корпусе для дифференциации словарно синонимических единиц с дальнейшей возможностью построения их естественных и адаптивных лексических профилей [1]. Полученные результаты, являясь статистическими средними по большому ансамблю, выражают максимально правдоподобное употребление слов в языке. Для современного английского языка возможность проведения такого анализа предоставляется сетевым программным обеспечением BNCweb, созданным на базе Британского Национального Корпуса (БНК) [2].

В основе автоматизированной экспертизы корпуса (исследования корпусных данных, обработанных с помощью программного обеспечения) лежат принципы статистического исследования. Напомним, что любое статистическое исследование массовых явлений включает такие этапы, как: сбор статистической информации и ее первичная обработка; сведение и группировка результатов наблюдения в конкретные совокупности; анализ и обобщение информации. Важно отметить, что на первом этапе (этапе сбора информации) формируются первичные статистические данные, или исходная информация, представляющая собой фундамент исследования. «Получение первичной информации и итоговых данных должно быть тщательно продуманным и четко организованным: если материал для исследования некачественный, то теоретические и практические выводы, сформулированные по его результатам, будут ошибочными» [3, с. 13]. Надо признать, что традиционно лингвисты работали над сбором и обработкой материала вручную, тем самым повышая вероятность ошибочных выводов и относительный вес субъективных правил отбора данных.

Результаты обработки корпуса БНК системой BNCweb возможно представить в виде таблиц сводной количественной характеристики статистической совокупности. Как известно, такие таблицы представляют собой форму наиболее рационального, наглядного и систематизированного изложения результатов сводки и обработки изучаемого материала, позволяют компактно и выразительно отражать цифровые характеристики исследуемого материала, показывают состояние изучаемого явления, служат основой для сравнения, сопоставления, анализа, а также позволяют формулировать соответствующие выводы [4, с. 35].

Результаты поиска по корпусу построены на принципах статистической выборки. Выборочным наблюдением называется такое наблюдение, при котором характеристика всей сово-

купности единиц дается по некоторой их части, отобранной в случайном порядке. Выборочный метод применяется в тех случаях, когда проведение сплошного наблюдения невозможно или экономически нецелесообразно. Качество результатов выборочного наблюдения зависит от того, насколько выборка репрезентативна (представительна). Для репрезентативности выборки необходимо соблюдать принцип случайности отбора единиц, предполагающий, что на включение или исключение объекта из выборки не может повлиять ничего, кроме случая [5, с. 66]. Эти правила реализованы в системе BNCWeb.

Метод анализа дискурсивных данных аннотируемого лингвистического корпуса с использованием сетевого программного обеспечения BNCweb, позволяет описывать лингвокогнитивные особенности семантически трудно дифференцируемых лексических единиц (традиционно считающихся взаимозаменяемыми) путем построения и сопоставления их лексических профилей.

Лингвистический анализ лексического профиля слова подразумевает нахождение взаимосвязи между значением слова и конструкциями, в которых оно употреблено в корпусе. При этом вместо первичных частотных данных используются обобщенная статистическая мера близости слов, вычисленная методом максимума правдоподобия. О принципах и особенностях формирования лексического профиля слова [6].

Продемонстрируем возможности системы BNCWeb для описания дифференцирующих признаков и лингвокогнитивных особенностей английских слов «emotion» и «feeling». 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.

Для анализа использовались следующие конструкции: глагол + emotion(s)/feeling(s) (any verb+emotion(s)/feeling(s)), глагол + emotion/feeling + предлог (of) (any verb + emotion/feeling + preposition OF), имя прилагательное + emotion(s)/feeling(s) (any adjective + emotion(s)/feeling(s)), emotion + предлог (of) + имя существительное (emotion + preposition OF + any noun), feeling + предлог (of) + имя существительное (feeling + preposition OF + any noun). Выбор указанных конструкций для анализа был определен результатами дискурсивного анализа, проведенного методом сплошной выборки на материале современной художественной литературы английских и американских авторов общим объемом более 20 000 страниц. По результатам нашего наблюдения, в художественных произведениях исследуемые единицы употреблены наиболее часто в указанных выше конструкциях.

Для облегчения интерпретации мы включили в анализ номинативную лексику (группа «any noun + feeling/emotion»), окружающую целевое слово в заданном диапазоне (± 3 слова).

Выбранные для автоматизированной экспертизы Британского Национального Корпуса конструкции сформировали параметры поискового запроса BNCweb для составления лексических профилей слов «emotion» и «feeling».

В рамках данной статьи мы считаем целесообразным представить конечный результат сопоставления полученных лексических профилей целевых слов (таблица 1).

Таблица 1 – Сопоставление лексических профилей слов EMOTION и FEELING

| тип конструкции | ЛП FEELING | Перевод | ЛП EMOTION | Перевод |
|-------------------------------|---|--|--|---|
| verb+feeling/emotion | to have to get | иметь получать | to show to express to convey | показывать выражать передавать |
| verb+feelings/emotions | to mix to express to provoke to arouse | смешивать выражать вызывать пробуждать | to express to seethe to suppress | выражать быть охваченным сдерживать, подавлять |
| verb + a feeling/emotion + of | to give to create to have to overcome to experience | давать вызывать иметь подавить испытывать | to evoke to betray to convey | пробуждать выдавать передавать |
| adjective+ feeling/emotion | strong sinking strange bad uncomfortable uneasy | сильный слабый, неприятный странный плохой неловкий беспокойный, тревожный | suppressed expressed strong intense human pent-up powerful | подавленный выраженный сильный сильный, глубокий человеческий сдерживаемый мощный |

Продолжение таблицы 1

| | | | | |
|------------------------------|--|---|--|--|
| adjective +feelings/emotions | mixed strong personal true negative ambivalent unwanted | смешанные сильные личные истинные негативные противоречивые нежелательные | conflicting mixed negative strong human pent-up deepest uncomfortable suppressed | противоречивые смешанные отрицательные сильные человеческие сдерживаемые глубочайшие стесненные подавленные |
| noun+feeling/emotion | helplessness unease dread guilt satisfaction panic | беспомощность беспокойство страх вина удовлетворение паника | voice intensity expression depth | голос интенсивность выражение глубина |
| noun+a feeling/emotion+of | helplessness dread guilt unease satisfaction elation panic relief well-being security | беспомощность ужас вина беспокойство удовлетворение восторг паника, тревога облегчение благополучие безопасность | compassion fear loss anguish jealousy bitterness creativity gloom curiosity happiness friendship horror | сострадание страх потеря тоска ревность горечь, обида творчество уныние любопытство счастье дружба ужас |

Преимущество представления данных в виде статистических таблиц наглядно демонстрируется очевидным различием профилей анализируемых слов. Рассмотрим детально группу адъективных коллокаций. Список имен прилагательных, употребляемых с анализируемыми единицами, довольно многообразен. Британский Национальный Корпус приводит более 1 500 вариантов употребления имен прилагательных со словом «feeling», и 245 вариантов использования адъективной лексики со словом «emotion». Для анализа и обобщения полученных результатов мы используем лишь те пары слов, статистическая мера близости которых (определяемая как числовое значение логарифмической функции правдоподобия) удовлетворяет критериям, сформулированным ранее в работе [7, с. 73–76].

В лексическом профиле слова «feeling» преобладают имена прилагательные, имеющие отношение к описанию интенсивности чувств (strong), а также описанию исключительно негативных переживаний (strange, bad, sinking, uncomfortable, uneasy).

Дескриптивная лексика, формирующая лексический профиль слова «emotion», прежде всего, сигнализирует о дихотомии «выражать эмоции – подавлять эмоции» (suppressed, expressed). Однако лидирует значение «подавленный» (suppressed, pent-up). Имена прилагательные со значением «сильный, глубокий, мощный» (strong, intense, powerful) характеризуют интенсивность переживаний. Дескриптив human особо подчеркивает тот факт, что само понятие «эмоция» является прерогативой человека.

Данные адъективных коллокаций в группе «any adjective+feeling/emotion», приведенные в таблице 1, демонстрируют четкое разведение лексических профилей анализируемых слов.

Области пересечения лексических профилей обнаруживаются в анализе адъективной лексики при рассмотрении группы, где целевое слово употреблено во множественном числе («any adjective + feelings/emotions»). В указанной группе присутствует максимальное количество совпадений, полученных коллокаций. Выделим основные группы значений: «смешанный» (mixed feelings/emotions), «интенсивный» (strong, deepest feelings/emotions), «противоречивый» (ambivalent, conflicting feelings/emotions), «подавленный» (unwanted, uncomfortable, suppressed feelings/emotions). Такое пресечение возникает именно тогда, когда анализируемые единицы (в терминах описания типов семантических отношений) являются согипонимами, то есть рядоположными понятиями, входящими в одно концептуальное пространство.

Анализ адъективных коллокаций лексических профилей анализируемых единиц показал, что наличие слов «emotion» и «feeling» в современном английском языке оправдано их различным концептуальным содержанием. Построение лексических профилей даже для словарно синонимичных слов, выполненное автоматизированной экспертизой корпуса, позволяет провести их семантическую дифференциацию.

Проведенный анализ блестяще иллюстрирует тезис Т.В. Черниговской о том, что «человеческий язык является эффективным средством противостоять сенсорному хаосу, который постоянно атакует нас: именно язык обеспечивает номинацию ментальных репрезентаций сенсорного опыта и, таким образом, «объективизирует» индивидуальные впечатления, обеспечивая описание мира и коммуникацию. Именно язык, будучи базирующимся на генетически обусловленных алгоритмах культурным феноменом, соединяет объекты внешнего мира с нейрофизиологическими событиями в мозгу, используя конвенциональные семиотические механизмы» [8, с. 118–119].

Представленные результаты свидетельствуют о необходимости дальнейшего расширения критериев при построении лексического профиля. Такой анализ проводимого исследования подразумевает учет возрастных когорт и временных эпох, а также анализ дополнительной статистической информации, предоставляемой системой BNCweb [9].

Ссылки:

1. Суворина Е.В. Методика построения адаптивного лексического профиля слова (на материале Британского Национального Корпуса) // Интеллектуальный потенциал XXI века: ступени познания: сборник материалов VII Международной студенческой научно-практической конференции / под общ. ред. С.С. Чернова. Новосибирск, 2011. С. 181–187.
2. Hoffmann S., Evert S., Smith N., Lee D., Prytz Y. Corpus Linguistics with BNCweb – a Practical Guide. Frankfurt am Main: Peter Lang, 2008.
3. Акулич М.В. Статистика в таблицах, формулах и схемах. СПб., 2009.
4. Там же.
5. Там же.
6. См. подробнее: Суворина Е.В. Использование сетевого программного обеспечения BNCweb при описании лингво-когнитивных особенностей слов «emotion» и «feeling» в современном английском языке // Вестник МГОУ. Журнал Московского государственного областного университета. Серия «Лингвистика». 2011. № 1 (6). С. 73–79.
7. Там же.
8. Черниговская Т.В. Мозг и язык: врожденные модули или обучающаяся сеть? // Мозг: фундаментальные и прикладные проблемы / под ред. акад. А.И. Григорьева. М., 2010. С. 117–127.
9. Суворина Е.В. Построение лексического профиля слова по различным статистическим критериям (на материале Британского Национального Корпуса) // Вестник МГОУ. Журнал Московского государственного областного университета. Серия «Лингвистика». 2011. № 5 (6). С. 163–167.

References (transliterated):

1. Suvorina E.V. Metodika postroeniya adaptivnogo leksicheskogo profilya slova (na materiale Britanskogo Natsional'nogo Korpusa) // Intellektual'niy potentsial XXI veka: stupeni poznaniya: sbornik materialov VII Mezhdunarodnoy studencheskoy nauchno-prakticheskoy konferentsii / under general ed. of S.S. Chernova. Novosibirsk, 2011. P. 181–187.
2. Hoffmann S., Evert S., Smith N., Lee D., Prytz Y. Corpus Linguistics with BNCweb – a Practical Guide. Frankfurt am Main: Peter Lang, 2008.
3. Akulich M.V. Statistika v tablitsakh, formulakh i skhemakh. SPb., 2009.
4. Ibid.
5. Ibid.
6. See for details: Suvorina E.V. Ispol'zovanie setevogo programmnoogo obespecheniya BNCweb pri opisaniy lingvo-kognitivnykh osobennostey slov “emotion” i “feeling” v sovremennom angliyskom yazyke // Vestnik MGOU. Zhurnal Moskovskogo gosudarstvennogo oblastnogo universiteta. Series “Lingvistika”. 2011. No. 1 (6). P. 73–79.
7. Ibid.
8. Chernigovskaya T.V. Mozg i yazyk: vrozhdennyye moduli ili obuchayushchayasya set'? // Mozg: fundamental'nye i prikladnye problemy / ed. by acad. A.I. Grigor'ev. M., 2010. P. 117–127.
9. Suvorina E.V. Postroenie leksicheskogo profilya slova po razlichnym statisticheskim kriteriyam (na materiale Britanskogo Natsional'nogo Korpusa) // Vestnik MGOU. Zhurnal Moskovskogo gosudarstvennogo oblastnogo universiteta. Series “Lingvistika”. 2011. No. 5 (6). P. 163–167.