

ФИЛОЛОГИЯ

Суворина Екатерина Владимировна

канд. филол. наук, доцент

ГБОУ ВПО «Московский городской педагогический университет»

г. Москва

**ИСПОЛЬЗОВАНИЕ КВАЛИМЕТРИЧЕСКИХ МЕТОДОВ
В ЛИНГВОКОГНИТИВНЫХ ИССЛЕДОВАНИЯХ**

Аннотация: в статье описаны принципы применения новейших квалиметрических методов математической статистики объектов произвольной природы для решения задач когнитивной лингвистики в рамках методического аппарата корпусной лингвистики. Рассмотрена постановка и решение прямой задачи когнитивной лингвистики – определения семантической близости слов по их адаптивным лексическим профилям, а также обратной задачи – определения корректности выбора проецирующего множества для построения адаптивного лексического профиля слова.

Ключевые слова: корпусная лингвистика, корпусный менеджер, лингвокогнитивное исследование, адаптивный лексический профиль слова, математическая статистика, квалиметрические методы.

Последнее десятилетие ознаменовалось взрывным ростом количества работ, использующих в лингвистических исследованиях те или иные методы математической статистики. Исторически первые работы использовали достаточно сложные математические методы: латентный семантический анализ, разные виды сингулярных разложений, многомерное шкалирование. Недостатком этих методов являются избыточные требования к математической квалификации исследователей, а также трудности в интерпретации полученных результатов. Само по себе привлечение математических методов в решении задач когнитивной лингвистики преследует, в качестве важнейшей цели, исключение субъективных факторов в исследованиях и достижение объективных и воспроизводимых результатов.

Коренной перелом в когнитивных исследованиях связан с привлечением программно-методического аппарата корпусной лингвистики и появлением на рубеже веков так называемых корпусных менеджеров, позволяющих исследователю использовать запросы к корпусу естественного языка, реализуемые по сети *Интернет* в реальном масштабе времени.

Использование корпусного менеджера BNCWeb [6] позволяет строить естественные или адаптивные лексические профили слов английского языка с использованием статистического критерия максимума правдоподобия [2, 3, 4, 5]. Такой подход позволяет получать объективные и воспроизводимые результаты, исключая при этом сплошной просмотр всего корпуса. Полученные результаты являются наиболее правдоподобными при осреднении по ансамблю текстовых источников, формирующих используемый корпус современного английского языка.

Последние достижения математической статистики связаны с критическим анализом классических корреляционных и кластерных алгоритмов. Альтернатива, предложенная на сегодняшний день, заключается во введении достаточно простых количественных методов расчета близости объектов абсолютно произвольной природы, результаты которых остроумно интерпретируются максимально простым качественным образом [1].

Для полноты изложения приведем необходимые формулы и определения из [1]. Исходим из того, что в нашем распоряжении имеется два объекта, характеризующихся некоторой количественной характеристикой. На первом этапе исследователь должен выбрать k – число качественных уровней представления исходной величины. Имеющиеся количественные значения признаков объекта (x_i) переводятся в качественный аналог K_i по следующим формулам:

$$d = \frac{x_{max} - x_{min}}{k - 1}, \quad (1)$$

$$K_i = \frac{x_i - x_{min}}{d} + 1. \quad (2)$$

На втором этапе вычисляется коэффициент парной конкордации (согласованности) признаков объектов:

$$W = 1 - \frac{\sum_{j=1}^n |x_{1j} - x_{2j}|}{n(k-1)}, \quad (3)$$

где n – количество признаков, используемых для определения согласованности, k – заданное число качественных уровней, x_{1j} – качественные значения первого объекта, x_{2j} – качественные значения второго объекта.

Полученное значение лежит строго в интервале от 0 до 1. Интерпретация полученного значения коэффициента согласованности, согласно многим экспериментам выглядит следующим образом: $W=1$ – это тождество, абсолютная близость; $0.75 \leq W < 1$ – отличная близость; $0.7 \leq W < 0.75$ – хорошая близость; $0.65 \leq W < 0.7$ – удовлетворительная, и при меньших значениях близость отсутствует.

Используя полученные значения коэффициентов согласованности W можно выполнить кластерный анализ выбранных объектов, определив множества наиболее близких из них и, в отличие от классических подходов иерархической кластеризации, рассчитать качества найденных кластеров [1, с. 247].

Ранее, в работе [4] для решения задач когнитивной лингвистики было введено понятие адаптивного лексического профиля. *Адаптивным лексическим профилем* (АЛП) называется количественная мера совместной встречаемости целевого (анализируемого) слова и заданного множества слов-коллокатов в пределах некоторого окна.

Предлагается использовать следующий алгоритм построения АЛП. В отличие от естественного задается некоторое множество слов, с помощью системы BNCweb в режиме анализа коллокаций вычисляются значения критерия максимума правдоподобия между целевым словом и последовательно каждым из слов, входящим в заданное множество. В результате анализируемое слово проектируется на заданное множество.

В работах [2, 3, 4, 5] решалась задача описания лингвокогнитивных особенностей слов *emotion(s)* и *feeling(s)* в современном английском языке. В качестве проецирующего множества была выбрана классификация машинного языка EARL [7] (см. табл. 1).

Классификация эмоций языка EARL

<i>negative & forceful</i>	anger annoyance contempt disgust irritation	<i>positive & lively</i>	amusement delight elation excitement happiness joy pleasure
<i>negative & not in control</i>	anxiety embarrassment fear helplessness powerlessness worry	<i>caring</i>	affection empathy friendliness love
<i>negative thoughts</i>	doubt envy frustration guilt shame	<i>positive thoughts</i>	courage hope pride satisfaction trust
<i>negative & passive</i>	boredom despair disappointment hurt sadness	<i>quiet positive</i>	calm content relaxation relief serenity
<i>agitation</i>	shock stress tension	<i>reactive</i>	interest politeness surprise

С помощью корпусного менеджера BNCweb [6] были рассчитаны значения критерия максимума правдоподобия для каждого из слов, представленных в таблице 1, и анализируемых слов *emotion* и *feeling*. В результате были получены два массива x_{1j} , x_{2j} размерности 48. Дальнейшие вычисления по формулам (1–3) показали, что варьирование числа качественных групп k перестает менять значение W при $k \geq 10$. Достигнутое значение коэффициента семантической близости слов *emotion* и *feeling* оказалось равным 0,68. Согласно [1, с. 178] это – удовлетворительная близость или так называемые «троюродные родственники» [1].

Решение обратной задачи строится следующим образом. Для всех слов, представленных в таблице 1, рассчитываются значения критерия максимума

правдоподобия с помощью системы BNCweb в режиме анализа коллокаций. Полученная матрица «расстояний» обрабатывается по методике, представленной в [1], с определением истинных кластеров, их плотности и расстояния между кластерами. Если проекционное множество построено корректно, то группы классификаций из таблицы 1 будут совпадать с найденными кластерами и их «качество» будет достаточно высоким. Размытость кластеров, малое расстояние между ними, будут свидетельствовать о неверном ошибочном формировании проекционного множества. Эта же методика может использоваться для построения ментальных карт и при лингвистическом анализе так называемых близких синонимов в современном английском языке.

Уникальные возможности современных корпусных менеджеров позволяют также уточнять внешние условия максимальной/минимальной кластеризации слов при варьировании типа текста, учета возрастных характеристик, гендерных особенностей носителя языка. Предлагаемый подход при использовании корпусного менеджера BNCweb требует вспомогательных расчетов, например, в системе EXCEL. При использовании штатного программного обеспечения Хаира, созданного на базе Британского Национального Корпуса [6], вся процедура анализа может быть полностью автоматизирована.

Список литературы

1. Васильев, В.И., Красильников, В.В., Плаксий, С.И., Тягунова, Т.Н. Статистический анализ многомерных объектов произвольной природы [Текст] / В.И. Васильев и др. – М.: ИКАР, 2004. – 382 с.
2. Суворина, Е.В. Использование сетевого программного обеспечения BNCweb при описании лингвокогнитивных особенностей слов «emotion» и «feeling» в современном английском языке [Текст] / Е.В. Суворина // Вестник Московского государственного областного университета. Серия «Лингвистика». – 2011. – №1(6). – М.: Изд-во МГОУ. – С.73–79.
3. Суворина, Е.В. Построение лексического профиля слова по различным статистическим критериям (на материале Британского Национального Корпуса) [Текст] / Е.В. Суворина // Вестник Московского государственного областного

университета. Серия «Лингвистика». – 2011. – №5(6). – М.: Изд-во МГОУ. – С.163–167.

4. Суворина, Е.В. Результаты построения адаптивного лексического профиля слова (на материале Британского Национального Корпуса) [Текст] / Е.В. Суворина // III Международная заочная научно-практическая конференция «Филология и лингвистика: современные тренды и перспективы исследования»: сборник материалов конференции (1 февраля 2012 г.). – Краснодар, 2012. – С. 107–110.

5. Суворина, Е.В. Семантическая дифференциация слов с помощью построения их лексических профилей [Текст] / Е.В. Суворина // Теория и практика общественного развития. Научный журнал. – 2012. – №2(12). – Краснодар: ООО Издательский дом «ХОРС». – С.365–368.

6. Corpus Linguistics with BNCweb – a Practical Guide. Hoffmann Sebastian. Evert Stefan, Smith Nicolas, Lee David, Prytz Ylva Berglund. – Frankfurt am Main: Peter Lang, 2008. – 290 p.

7. Schroder, M., Pirker, H. First Suggestions for an Emotion Annotation and Representation Language. [Электронный ресурс] / Режим доступа: http://www.ofai.at/~hannes.pirker/papers/schroder-pirker-lamolle_LREC2006.pdf (дата обращения: 23.09.2011).