

## КРИПТОКЛАССНЫЕ ДАННЫЕ ДЛЯ ОПРЕДЕЛЕНИЯ МЕРЫ МЕЖЪЯЗЫКОВОЙ ЭКВИВАЛЕНТНОСТИ

О. В. Дони́на

*Воронежский государственный университет*

Поступила в редакцию 10 августа 2014 г.

**Аннотация:** в статье предлагается методика сопоставления эквивалентных слов в различных языках при помощи методов криптоклассного анализа и лингвостатистики.

**Ключевые слова:** корпусная лингвистика, лингвостатистика, эквивалентность, языковая категоризация, именной криптокласс.

**Abstract:** in this paper, we propose the method of comparing equivalent words in different languages using cryptotype analysis and linguistic statistics.

**Key words:** corpus linguistics, linguistic statistics, linguistic categorization, noun cryptotype.

В фокусе нашего исследования находится криптоклассное сопоставление эквивалентных слов в английском и русском языках на материале данных информационного ресурса «Криптоклассы английского языка» [1] и национального корпуса русского языка (НКРЯ) [2].

В рамках данной статьи мы попытаемся предложить методику сопоставления эквивалентных слов в различных языках при помощи методов криптоклассного анализа и лингвостатистики на примере лексемы русского языка *выгода* и ее английских эквивалентов.

Одним из основных понятий теории вероятностей и математической статистики является понятие корреляции. В лингвостатистике для определения корреляционной зависимости между двумя случайными величинами используют коэффициент корреляции К. Пирсона, основанный на значениях коррелирующих величин. Его разработали Карл Пирсон, Фрэнсис Эджуорт и Рафаэль Уэлдон [3].

С помощью коэффициента линейной корреляции (или коэффициента корреляции Пирсона) можно получить представление о тесноте корреляции случайных величин. Коэффициент линейной корреляции определяет характер и тесноту статистической связи.

Линейная корреляция предполагает линейную зависимость и является скалярной величиной. Численное значение коэффициента корреляции указывает на тесноту статистической связи, причем значения, мало отличающиеся от единицы, указывают на связь, близкую по характеру к функциональной, а значения, приближающиеся к нулю, свидетельствуют о слабой связи или даже указывают на отсутствие корреляционной связи (табл. 1). В данном исследовании мы

пользуемся общей классификацией корреляционных связей, предложенной в работе Э. В. Ивантера и А. В. Коросова [4].

Т а б л и ц а 1

*Общая классификация корреляционных связей*

Интервал значений коэффициента корреляции	Интерпретация
0–0,19	Очень слабая корреляция
0,2–0,29	Слабая корреляция
0,3–0,49	Умеренная корреляция
0,5–0,69	Средняя корреляция
0,7–1	Сильная / тесная корреляция

Определим характер и силу связи между русской лексемой *выгода* и ее английскими эквивалентами. По данным «Англо-русского синонимического словаря» [5] имя *выгода* имеет четыре английских аналога: *advantage, benefit; profit, gain* (те же данные получаем при обращении к электронному словарю ABBYY Lingvo Online Dictionary [6]).

Подтвердим это примерами из параллельного подкорпуса национального корпуса русского языка [2]:

(1) *I would have returned them in any event – as a matter of honor, not to let them take **advantage** of your hospitality for – but never mind. [Walter M. Miller, Jr. A Canticle For Leibowitz (1960)]*

*Я бы вернул их в любом случае: это дело чести – не дать им извлечь грубую выгоду из вашего гостеприимства. Но оставим это. [Уолтер Миллер. Страсти по Лейбовицу]*

(2) *His corporation had sent him to Harvard Business School to study the minds of the movers and shakers who were screwing up our economy for their own immediate **benefit**, taking money earmarked for research and development and new machinery and so on, and putting*

it into monumental retirement plans and year-end bonuses for themselves. [Kurt Vonnegut. *Hocus Pocus* (1990)]

Корпорация послала его на экономический факультет Гарвардского университета, чтобы он разобрался, как и чем думают плуты и проныры, которые пустили под откос нашу экономику ради своих сиюминутных выгод, тащат деньги, предназначенные на научные исследования и культуру и новую технику и прочее, и устраивают на эти деньги роскошные самоторши, раздают самим себе громадные пенсии и премии в конце года. [Курт Воннегут. *Фокус-покус*]

(3) He's a contractor and builder, fifty-five years old, makes a business of buying lots, putting houses on them, selling the house at a **profit**, buying other lots. [Erle Stanley Gardner. *The Case of the Daring Divorcee* (1964)]

Ему пятьдесят пять лет, он строитель-подрядчик. Покупает участки земли, ставит на них дома и с выгодой перепродает. [Эрл Стэнли Гарднер. *Дело смелой разведки*]

(4) But for those Georgians who had turned Republican for personal **gain**, there was no excuse. [Margaret Mitchell. *Gone with the Wind, Part 2* (1936)]

А вот тем уроженцам Джорджии, которые стали республиканцами выгоды ради, уже не было оправдания. [Маргарет Митчелл. *Унесенные ветром. Ч. 2*]

Также мы провели опрос среди 20 респондентов в возрасте от 18 до 30 лет, в их число входили русскоговорящие участники, которые изучают или преподают английский язык в России или проживают и обучаются на территории США более 2 лет, а также англоязычные респонденты, обучающиеся русскому языку. В ходе эксперимента мы просили участников ранжировать английские лексемы по смысловой близости с русской лексемой *выгода*. Полученные данные представлены в табл. 2.

Т а б л и ц а 2

Теснота статистической связи лексемы **выгода** и ее английских эквивалентов по данным эксперимента

Имя	Степень тесноты связи с лексемой <i>выгода</i> , %
Benefit	33,13
Profit	26,99
Advantage	20,86
Gain	19,02

После этого мы перешли непосредственно к данным криптоклассного анализа, сопоставив исследу-

емые имена по двум показателям: индексу разнообразия (ИРа – отношение количества классификаторов, с которыми имя встречается в корпусе, к общему количеству классификаторов криптокласса) и показателю криптоклассной активности (ПоКа – доля активности имени в образовании словосочетаний с классификаторами одного криптокласса на фоне других проекций) [7]. Результаты криптоклассного анализа вхождения исследуемых имен в шесть выделенных криптоклассов (Res Liquidae (эталон – ‘вода’), Res Acutae (эталон – ‘шип’), Res Filiformes (эталон – ‘нить’), Res Rotundae (эталон – ‘мяч’), Res Parvae (эталон – ‘камень’), Res Longae Penetrantes (эталон – ‘копьё’)) представлены в табл. 3.

Т а б л и ц а 3

Данные криптоклассного анализа

Имя	Показатели	Res Liquidae	Res Acutae	Res Filiformes	Res Rotundae	Res Parvae	Res Longae Penetrantes
Выгода	ИРа	0,2	0	0,2	0,167	0,5	0
	ПоКа	0,0976	0	0,2378	0,0061	0,6585	0
Advantage	ИРа	0,077	0	0,125	0	0,667	0
	ПоКа	0,0003	0	0,0002	0	0,9995	0
Benefit	ИРа	0,308	0,100	0	0	0,333	0
	ПоКа	0,2195	0,0098	0	0	0,7707	0
Profit	ИРа	0,385	0	0,125	0	0,444	0
	ПоКа	0,1845	0	0,0037	0	0,8118	0
Gain	ИРа	0	0,100	0	0	0,444	0
	ПоКа	0	0,3333	0	0	0,6667	0

Последовательно сопоставляя каждый английский эквивалент с лексемой *выгода*, мы высчитали коэффициент корреляции Пирсона по формуле

$$r_{xy} = \frac{cov_{xy}}{\sigma_x \sigma_y} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}$$

где  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ ,  $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$  – среднее значение выборки.

В табл. 4 приведены полученные эмпирические значения по индексу разнообразия сочетаемости лексем (ИРа).

Т а б л и ц а 4

Эмпирические значения коэффициента корреляции Пирсона по индексу разнообразия (ИРа)

Имя	r (ИРа)	Тип связи
Advantage	0,916819	Тесная/сильная
Profit	0,822767	Тесная/сильная
Gain	0,768952	Тесная/сильная
Benefit	0,666461	Средняя

Из приведенной таблицы видно, что по данному показателю имя *advantage* имеет наиболее тесную связь с русской лексемой *выгода*.

Аналогичные расчеты были проведены и по показателю криптоклассной активности имен (ПоКа) (табл. 5)

Т а б л и ц а 5

Эмпирические значения коэффициента корреляции Пирсона по показателю криптоклассной активности имен (ПоКа)

Имя	r (ПоКа)	Тип связи
Advantage	0,933642	Тесная/сильная
Profit	0,923739	Тесная/сильная
Benefit	0,91021	Тесная/сильная
Gain	0,643993	Средняя

В данном случае лексемы *advantage*, *profit* и *benefit* отличаются высокими коэффициентами линейной корреляции (> 0,9). Учитывая, что, согласно гипотезе криптоклассной организации именной лексики, показатель криптоклассной активности имени существительного отражает современное функционирование имен в языке [8], можно сделать вывод, что три вышперечисленных имени (*advantage*, *profit* и *benefit*) наиболее часто употребляются в английском языке со значением, которое носители русского языка приписывают лексеме *выгода*.

Как видим, полученные при расчете коэффициента линейной корреляции эмпирические значения (см. табл. 5) расходятся с данными опроса (см. табл. 2). Большинство наших респондентов отдавали предпочтение лексеме *benefit*, а *advantage* размещали лишь на 3-ю позицию. Для того чтобы вычислить степень согласованности экспертов, мы посчитали коэффициент конкордации рангов Кендалла [9; 10], который позволяет дать оценку исследуемому явлению в виде обобщенного мнения специалистов (экспертов) по изучаемому вопросу или проблеме. Этот общий коэффициент ранговой корреляции для группы, состоящей из *m* экспертов, вычисляется по формуле

$$W = \frac{12 \cdot S}{m^2(n^3 - n)},$$

где *S* – сумма квадратов отклонений всех оценок рангов каждого объекта экспертизы от среднего значения; *m* – число экспертов; *n* – число объектов экспертизы.

Коэффициент конкордации изменяется в диапазоне  $0 > W > 1$ , причем 0 соответствует несогласованности, а 1 соответствует полной согласованности. Если значение коэффициента конкордации превышает 0,40–0,50, то качество оценки считается удовлетворительным, если  $W = 0,70–0,80$  – высоким. В нашем случае коэффициент конкордации рангов составил  $W$  (*выгода*) = 0,313086, что указывает на очень слабую согласованность респондентов и дает право не учитывать данные эксперимента при формулировке выводов.

Таким образом, мы можем говорить о том, что на примере лексемы *выгода* и ее английских эквивалентов мы проиллюстрировали способ сопоставления эквивалентных слов в различных языках при помощи методов криптоклассного анализа и лингвостатистики. В результате исследования мы выяснили, что наиболее близким аналогом для лексемы русского языка *выгода* в американском варианте английского языка выступает имя *advantage*.

#### ЛИТЕРАТУРА

1. Информационный ресурс «Криптоклассы английского языка». – Режим доступа: <http://www.rgph.vsu.ru/coel.php>
2. Национальный корпус русского языка (НКРЯ). – Режим доступа: <http://www.ruscorpora.ru/>
3. *Pearson K.* (June 20, 1895) «Notes on regression and inheritance in the case of two parents», *Proceedings of the Royal Society of London*, 58: 240–242.
4. *Ивантер Э. В.* Основы биометрии : введение в статистический анализ биологических явлений и процессов / Э. В. Ивантер, А. В. Коросов. – Петрозаводск, Изд-во ПетрГУ, 1992. – 168 с.
5. *Литвинов П. П.* Англо-русский и русско-английский синонимический словарь / П. П. Литвинов. – М. : Астрель : АСТ, 2002.
6. *ABBYU Lingvo Online Dictionary.* – Mode of access: [www.lingvo-online.ru](http://www.lingvo-online.ru)
7. *Борискина О. О.* Криптоклассы английского языка / О. О. Борискина. – Воронеж : Истоки, 2011. – 333 с.
8. *Борискина О. О.* Классифицирующие структуры в криптоклассном исследовании / О. О. Борискина // Вестник Воронеж. гос. ун-та. Сер.: Лингвистика и межкультурная коммуникация. – 2010. – № 2. – С. 28–33.
9. *Kendall M.* (1938). «A New Measure of Rank Correlation». *Biometrika* 30 (1–2): 81–89.
10. *Kendall M.* (1948) *Rank Correlation Methods*, Charles Griffin & Company Limited.

Воронежский государственный университет

Донина О. В., аспирант кафедры английского языка в профессиональной международной деятельности  
E-mail: [olga-donina@mail.ru](mailto:olga-donina@mail.ru)

Voronezh State University

Donina O. V., Post-graduate Student of the English Language in the Professional International Activity Department  
E-mail: [olga-donina@mail.ru](mailto:olga-donina@mail.ru)