

Использование параллельного корпуса для количественного изучения лингвоспецифичной лексики

Для анализа я выбрала слова «тоска» и «глобус».

Я предположила, что слово «тоска» лингвоспецифичное, так как слово может использоваться для обозначения скуки, печали, уныния, грусти и т. д. Моя гипотеза подтвердилась, так как на 24 вхождения нашлось 11 моделей перевода (melancholy, gloom, boredom, depression и др.), но самой частотной моделью оказалось melancholy.

Слово «глобус» произошло от латыни, и во многих европейских языках существует однокоренное слово того же значения (globe (eng), Globus (de), globe (fr) и др.), соответственно, я предположила, что это слово не может быть специфичным. Действительно, на 16 переводов слова «глобус» пришлось всего 2 модели (globe, sphere).

	тоска	глобус
F(Mmax)/NumM	8/11=0,72	10/2=5
F(O)/NumM	24/11=2,18	16/2=8
F(Mmax)/F(Msec)	8/4=2	10/4=2,5
F(Mmax)/F(O)	8/24=0,33	10/16=0,63

Таким образом, можно сделать вывод, что слово «тоска» является лингвоспецифичным несмотря на то, что можно легко выделить самую частотную модель. Так или иначе, оно обладает большим количеством разных переводов с относительно небольшим процентом каждой модели из них от общего числа переводов. Особенно это видно в сравнении с данными, полученными по слову «глобус», для которого было найдено всего 2 модели перевода. Меры разброса по слову «глобус» больше чем по слову «тоска» в 1,3-6,9 раз.