

## ДЗ 7

### *Использование параллельного корпуса для количественного изучения лингвоспецифичной лексики*

Я выбрала два русских слова, гипотетические лингвоспецифичное (жуткий) и неспецифичное (разделять). Чтобы подтвердить свою гипотезу, я обратилась к Национальному корпусу русского языка.

В параллельном корпусе английского и русского языка я проверила каждое из слов на частоту вхождений, количество моделей перевода, абсолютную частоту самой частотной модели перевода и абсолютную частоту второй самой частотной модели перевода. Полученные данные я занесла в таблицу.

Параметр поиска	Лингвоспецифичное слово (жуткий)	Неспецифичное слово (разделять)
Частота вхождений	65	27
Количество моделей перевода	46	10
Абсолютная частота самой частотной модели перевода	7	15
Абсолютная частота второй самой частотной модели перевода	4	2

Слово “жуткий” имеет 46 моделей перевода, в то время как “разделять” всего 10. Абсолютная частота самой частичной модели перевода словоформы “жуткий” (horrible) занимает всего 17,5 процента от общего числа возможных соответствий, а ее частота несильно отличается от частот остальных. Слово “разделять”, напротив, имеет большую разницу абсолютных частот самой частотной и второй по частотности модели перевода. Также его абсолютная частота самой частотной модели (share) занимает около 56 процентов от общего числа возможных соответствий. Уже из этого можно сделать вывод, что словоформа “жуткий” является линвоспецифичным, а “разделять” неспецифичным, но, чтобы полностью удостовериться в этом, я обратилась к известным мне формулам разброса, с помощью которых можно оценить степень специфичности того или иного слова.

Я взяла две меры разброса: отношение абсолютной частоты самой частотной модели перевода ( $F(M_{max})$ ) к количеству различных моделей (NumM) и отношение абсолютной частоты самой частотной модели перевода к частоте второй ( $F(M_{max})/F(M_{sec})$ ). По первой из них получаем отношение 7:46 (0,15) у слова “жуткий” и 15:27 (0,6) у “разделять”, что говорит о небольшом проценте количества употребления самой частотной модели от общего числа возможных соответствий у лингвоспецифичного первого слова.

У неспецифичного второго слова самая частотная модель занимает больше половины всех возможных вариантов. По второй мере разброса мы можем отследить разницу абсолютных частот первого самого частотного слова и второго: 7:4 (1,75) у словоформы “жуткий” и 15:2 (7,5) у “разделять”. Частота самой частотной модели лингвоспецифичного слова почти не отличается от частот остальных, а у неспецифичного она в 7 раз больше остальных. Теперь можно с уверенностью сказать, что “жуткий” – лингвоспецифичное слово, “разделять” – неспецифичное.

Таким образом, НКРЯ позволяет лингвистам с легкостью определять специфику слов, предоставляя огромное количество данных. Также я пришла к выводу, что нельзя с одного взгляда, используя исключительно свой опыт и знания, различить лингвоспецифично ли слово или нет, так как изначально я решила отнести словоформу “разделять” к специфичным, зная как минимум 17 возможных переводов, однако оказалось, что модель перевода “share” используется во много раз чаще других вариантов (по данным из НКРЯ), поэтому оно не может быть лингвоспецифичным.