

Использование параллельного корпуса для количественного изучения лингвоспецифичной лексики

Проведенное с помощью русско-английского параллельного корпуса Национального корпуса русского языка исследование помогло мне подтвердить собственные догадки о специфичности слов *лапти* и *дуб*

Лексическая единица *лапти*, как я и предполагала, оказалось лингвоспецифичным словом: в переведенных с русского англоязычных текстах оно встретилось всего 7 раз, и практически везде в разных «обликах»: *bark shoes*, *plaited shoes*, *bast shoes*, просто *boots* – в одном из контекстов слово вовсе опущено. То есть, в силу нетипичности для английских реалий оно не может быть переведено однозначно.

Специфичность слова <i>лапти</i>			
Общее количество вхождений F(O)	7	Средняя частота вхождений на одну модель	1,2
Общее количество различных моделей перевода (NumM)	6		
Абсолютная частота самой частотной модели перевода (F(Mmax))	2	Отношение абсолютной частоты самой распространенной модели перевода к частоте второй	1
Частота второй частотной модели (F(Msec))	1		
		Отношение абсолютной частоты самой частотной модели перевода к общему кол-ву вхождений	0,3

Таким образом, слово, о котором идет речь, вошло в небольшое количество контекстов в английском корпусе и в большинстве случаев переведено по-разному (причем самая частотная из всех моделей перевода занимает небольшой процент от общего числа соответствий, и ее частота не отличается сильно от частот остальных), действительно относится к лингвоспецифичной лексике.

Второе выбранное мной для исследования слово – *дуб* – я изначально определила неспецифичным. Как выяснилось, при большем числе вхождений (по сравнению с частотой употребления первого слова) моделей перевода с русского на английский меньше. Две самые популярные – *oak* и *oak-tree*, которые встречаются в несколько раз чаще остальных:

Специфичность слова <i>дуб</i>			
Общее количество вхождений F(O)	83	Общая частота вхождений на одну модель	14
Общее количество различных моделей перевода (NumM)	6		
Абсолютная частота самой частотной модели перевода (F(Mmax))	49	Отношение самой распространенной модели перевода к частоте второй	8,2
Частота второй частотной модели (F(Mmsec))	16		
		Отношение абсолютной частоты самой частотной модели перевода к общему кол-ву вхождений	0,6

Я считаю, что единица *дуб* обладает меньшей специфичностью, чем единица *лапти*: на одну модель перевода первой приходится 14 вхождений, в то время, как на модель перевода второй – всего 1,2. Я бы не назвала слово *дуб* абсолютно неспецифичным, если не сравнивать его с другими, так как в таком случае количество моделей перевода было бы гораздо меньше.