

Использование параллельного корпуса для количественного изучения лингвоспецифичной лексики

«Предполагается, что у «лингвоспецифичного» слова должен наблюдаться большой список возможных моделей перевода, и на каждую будет в среднем приходиться сравнительно немного контекстов, а самая частотная из всех моделей будет занимать небольшой процент от общего числа возможных соответствий, ее частота не будет сильно отличаться от частот остальных». Данную гипотезу можно проверить с помощью данных параллельного корпуса.

Я выбрала гипотетически лингвоспецифичное слово «простор», на которое было найдено 18 вхождений в параллельном русско-итальянском корпусе. Среди них было обнаружено четыре модели перевода: spazio (пространство), libertà (свобода), vastità (ширь) и aria aperta (открытый воздух).

Самая частотная модель перевода, слово spazio, встретилось 7 раз, libertà – 4 раза, vastità – 2 раза, aria aperta – 1 раз.

Далее я оценила степень лингвоспецифичности слова «простор» с помощью несколько возможных мер разброса моделей перевода:

- отношение абсолютной частоты самой частотной модели перевода ($F(M_{\max})$) к количеству различных эквивалентов ($NumM$): **1,75**
- средняя частота вхождений на один эквивалент ($F(O)/NumM$, где $F(O)$ — частотность данного слова в оригинале): **4,5**
- отношение абсолютной частоты самой частотной модели перевода к частоте второй ($F(M_{\max})/F(M_{\sec})$): **1,75**
- отношение абсолютной частоты самой частотной модели перевода к общему количеству вхождений ($F(M_{\max})/F(O)$): **0,39**

Далее я выбрала гипотетически неспецифичное слово «любовь», которое имеет 343 вхождения, и проделала с ним те же расчёты. По данным НКРЯ (параллельный русско-итальянский корпус) это слово имеет следующие модели перевода: amore (встретилось 132 раза) и affetto (встретилось 5 раз). Таким образом, самая частотная модель перевода – amore.

- отношение абсолютной частоты самой частотной модели перевода ($F(M_{\max})$) к количеству различных эквивалентов ($NumM$): **66**
- средняя частота вхождений на один эквивалент ($F(O)/NumM$, где $F(O)$ — частотность данного слова в оригинале): **171,5**
- отношение абсолютной частоты самой частотной модели перевода к частоте второй ($F(M_{\max})/F(M_{\sec})$): **26,4**

- отношение абсолютной частоты самой частотной модели перевода к общему количеству вхождений ($F(M_{max})/F(O)$): **0,38**

Все расчеты можно увидеть в нижеприведенной таблице (а также в приложенном документе Excel):

	простор	любовь
$F(M_{max})$	7	132
NumM	4	2
$F(M_{max})/NumM$	1,75	66
$F(O)$	18	343
$F(O)/NumM$	4,5	171,5
$F(M_{sec})$	4	5
$F(M_{max})/F(M_{sec})$	1,75	26,4
$F(M_{max})/F(O)$	0,39	0,38

Вывод: на основании проведенного исследования мне удалось выяснить, что у лингвоспецифичного слова «простор» обнаруживается достаточно большое количество моделей перевода, которые не очень точно отображают суть слова в оригинале. У самой частотной из моделей перевода количество вхождение практически не отличается от частоты других моделей перевода.

Абсолютное большинство показателей в расчетах у слова «простор» численно меньше, чем у нелингвоспецифичного слова «любовь», имеющее большее число вхождений, из которых тем не менее можно выделить лишь две модели перевода, одна из которых по количеству вхождений значительно превосходит вторую.

Таким образом, всё это свидетельствует о том, что слово «простор» действительно является лингвоспецифичным, в то время как слово «любовь» таковым не является.