

Использование параллельного корпуса для количественного изучения лингвоспецифичной лексики.

Я выбрала два слова, которые я посчитала лингвоспецифичным и неспецифичным. Лингвоспецифичное – «*грубо*», неспецифичное – «*ложка*».

1) Лингвоспецифичное слово «*грубо*»

В НКРЯ было 98 вхождений по этому слову. Ниже я представлю таблицу с вариантами его перевода на английский язык (и их количеством):

Модель перевода	Количество вхождений
roughly	22
rudely	13
coarse	16
crudely	10
awkward	2
harsh	3
cruelly	3
grim	3
abusive	2
seize	3
broadly	3
curtly	1
gross	2

В это количество не вошли только те слова, которые были опущены в переводе на английский язык в НКРЯ. Так как опущенные слова нельзя назвать «вхождениями», то $98-15=83$ – число вхождений.

Теперь я хотела бы рассчитать среднюю частоту вхождений на одну модель по формуле: $F(O)/NumM$, где $F(O)$ – общее количество вхождений, а $NumM$ – количество разных моделей.

В моём случае $F(O) = 83$, а $NumM = 13$

⇓

$$83/13 = 6,4$$

Чтобы найти отношение абсолютной частоты самой частотной модели перевода ($F(M_{max})$) к количеству различных моделей ($NumM$)

$F(M_{max}) = 22$, $NumM = 13$

⇓

$$22/13 = 1,7$$

По результату последнего вычисления я могу точно сказать, что слов «грубо» лингвоспецифично, потому что самая частотная из всех моделей (roughly) занимает небольшой процент от общего числа возможных соответствий и её частота не сильно отличается от частот остальных.

Я подтвердила свою догадку с помощью НКРЯ. Количество моделей перевода у слова «грубо» достаточно большое и формула (как я говорила выше) это тоже подтверждает.

2) Неспецифичное слово «ложка»

В НКРЯ было 64 вхождения по этому слову. Ниже я представлю таблицу с вариантами его перевода на английский язык (и их количеством):

Модели перевода	Количество вхождений
ladle	8

spoon	46
-------	----

В это количество не вошли только те слова, которые были опущены в переводе на английский язык в НКРЯ. Так как опущенные слова нельзя назвать «вхождениями», то $64-10=54$ – число вхождений.

Воспользуюсь такими же формулами:

$$1) F(O) = 54, \text{ а } NumM = 46$$

⇓

$$54/46 = 1,2$$

$$2) F(M_{\max}) = 46, NumM = 2$$

⇓

$$46/2 = 23$$

Тут мы видим, что отношение абсолютной частоты самой частотной модели перевода к количеству различных моделей равна половине, а это значит, что слово «ложка» неспецифичное (так как частота одной из возможных моделей очень сильно отличается от частот остальных и составляет большой процент от общего числа возможных соответствий).

Таким образом я подтвердила свою догадку с помощью НКРЯ.

Моделей перевода у слова «ложка» только две и формула это так же подтверждает.

Таким образом, я могу сделать вывод, что при помощи НКРЯ можно определить лингвоспецифично или неспецифично слово. Например, я определила, что слово «грубо» лингвоспецифично, а слово «ложка», наоборот, неспецифично.