

Домашнее задание №7

Я выбрала два слова: разгульный и селедка. Мои предположения были следующими: разгульный – это лингвоспецифичное слово, а селедка – неспецифичное. Используя параллельный корпус НКРЯ английского языка, я получила следующие данные:

| | число вхождений | кол-во моделей перевода | самая частотная модель перевода | вторая по частотности модель | третья по частотности модель |
|------------|--------------------|----------------------------|------------------------------------|---------------------------------|---------------------------------|
| разгульный | 21 | 13 | 5 | 3 | 2 |
| селедка | 31 | 4 | 24 | 3 | 2 |

На основании приведенных данных в таблице видно, что у слова «разгульный» количество моделей перевода на 21 вхождение намного больше, чем у слова «селедка» на 31 вхождение. Используя одну из мер разброса моделей перевода, оценим степень специфичности этих слов и подтвердим гипотезу, что «разгульный» - лингвоспецифичное слово, а «селедка» - нет. Посчитаем отношение абсолютной частоты самой частотной модели перевода к общему количеству вхождений.

| | $F(M_{\max})/F(O)$ |
|------------|--------------------|
| разгульный | 0,238095238 |
| селедка | 0,774193548 |

Согласно данным в таблице самая частотная модель перевода слова «разгульный» встретила 5 раз на 21 вхождение. Получается, что отношение частоты самой частотной модели перевода к общему количеству вхождений равно 0.23, что составляет только около 20% от общего числа вхождений. Тогда как со словом «селедка» ситуация совершенно другая – самая частотная модель перевода встретила 24 раза на 31 вхождение. Отношение этих данных получается 0.77, а значит, эта модель перевода занимает почти 80% от всех вхождений. Мы знаем, что лингвоспецифичным является то слово, у которого наблюдается 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, и ее частота не будет сильно отличаться от частот остальных.

Таким образом, у слова «разгульный» и правда самая частотная из всех моделей занимает небольшой процент от общего числа возможных соответствий, из чего следует, что оно лингвоспецифичное. В то время как слово «селедка» является неспецифичным, потому что самая частотная из всех моделей этого слова занимает большой процент от общего числа вхождений.