

HW 7: Использование параллельного корпуса для количественного изучения лингвоспецифичной лексики

Дано: Лингвоспецифичное слово "тварь"

Результаты поиска в параллельном корпусе

Объем всего корпуса: 1 953 документа, 5 844 884 предложения, 76 759 952 слова.

Поиск ведётся по пользовательскому подкорпусу объемом 1 608 376 предложений, 24 681 277 слов.

тварь

Найдено 56 документов, 113 вхождений.

Лингвоспецифичное слово "тварь"	Количество слов
1. abomination	1
2. animal/animals	2
3. beast	5
4. being	1
5. creation/creature/devil's creature	40
6. critter	1
7. cur	2
8. hound	1
9. hussy	1
10. little devil	1
11. living thing	1
12. monster	1
13. night-women	1
14. people and animals	1
15. ruffian	1
16. scoundrel	1
17. slime	1

18. sly dog/common dog	2
19. thing/devil's creature	8
20. vermin	1
21. vile creature	1
22. warmint	1
23. wretch/wretched	3

- Отношение абсолютной частоты самой частотной модели перевода, ($F(M_{\max})$) к количеству различных моделей ($NumM$): $40/23=1,739$
- Средняя частота вхождений на одну модель, ($F(O)/NumM$, где $F(O)$ — общее количество вхождений) : $113/23= 4,913$
- Отношение абсолютной частоты самой частотной модели перевода к частоте второй, ($F(M_{\max})/F(M_{\sec})$): $40/8=5$
- Отношение абсолютной частоты самой частотной модели перевода к общему количеству вхождений, ($F(M_{\max})/F(O)$): $40/113=0,354$

Вывод: Слово «тварь» - лингвоспецифично, поэтому трудно подобрать ему точный вариант перевода на английский. Главная проблема заключается в семантике, как слово интерпретируется автором: либо речь идет о живом существе (creature, beast, living thing etc.), либо это оскорбление (wretch, sly/common dog, scoundrel etc.).

Дано: Неспецифичное слово "деликатный"

Неспецифичное слово "деликатный"	Количество слов
delicate	8

Результаты поиска в параллельном корпусе

Объем всего корпуса: 1 953 документа, 5 844 884 предложения, 76 759 952 слова.

Поиск ведётся по пользовательскому подкорпусу объемом 173 документа, 1 198 469 предложений, 18 089 924 слова.

деликатный

Найдено 8 документов, 9 вхождений.

- Отношение абсолютной частоты самой частотной модели перевода, ($F(M_{\max})$) к количеству различных моделей ($NumM$): 0
- Средняя частота вхождений на одну модель, ($F(O)/NumM$, где $F(O)$ — общее количество вхождений): $9/1= 9$
- Отношение абсолютной частоты самой частотной модели перевода к частоте второй, ($F(M_{\max})/F(M_{\sec})$): 1

- Отношение абсолютной частоты самой частотной модели перевода к общему количеству вхождений, $(F(M_{\max})/F(O))$: **8/9=0,888**

Вывод: В данном случае есть только одна модель перевода. Слово «деликатный» пришло в русский и английский языки из латыни (*dēlicatus* - "изнеженный"), поэтому позаимствованное слово в переводе не имеет разнообразия вариантов: есть единственный точный перевод. Следовательно, «деликатный» является неспецифичным словом.