

Использование параллельного корпуса для изучения лингвоспецифичной лексики

Для исследования были взяты два слова: *habitat* и *comprehensive*; первое было сочтено нелингвоспецифичным, второе, соответственно, лингвоспецифичным.

Рассмотрим сперва переводные эквиваленты первого слова, *habitat*.

переводной эквивалент	количество вхождений
местообитание	8
обиталище	3
естественная среда	2
жильё	2
место обитания	2
ареал	1
жилище	1
место распространения	1
родина	1
среда обитания	1
экологическая ниша	1

Даже без объединения однокоренных слов в одну модель перевода заметно, что один из вариантов встречается значительно чаще всех прочих (из двадцати трёх вхождений восемь приходится на «местообитание»).

Если же сгруппировать однокоренные и близкие слова и словосочетания, картина становится ещё выразительней. Местообитание, обиталище, место обитания, место распространения, среда обитания и естественная среда несколько насильственно, но могут быть объединены в одну группу, и тогда получается, что на эту модель перевода приходится семнадцать вхождений из двадцати трёх. Варианты «жильё» и «жилище» поддаются объединению без всяких натяжек; на их долю приходится три вхождения.

модель перевода	количество вхождений
местообитание, обиталище, etc	17
жильё, жилище	3
ареал	1
родина	1
экологическая ниша	1

Теперь обратимся к слову *comprehensive*, где картина будет совершенно иной.

переводной эквивалент	количество вхождений
всеобъемлющий	4
общий	3
комплексный	2
обширный	2
разумный	2
вместительный	1
выразительный	1
естественный	1
замысловатый	1
изучающий	1
интегральный	1
исчерпывающий	1
обхватывающий	1
обычный	1
ослепительный	1
совокупность	1
совокупный	1

Если даже объединить однокоренные «совокупность» и «совокупный», картину это принципиально не изменит; наиболее частотным останется вариант «всеобъемлющий» — четыре вхождения из двадцати пяти.

Из этих таблиц уже видно, что первое слово, *habitat*, поддаётся переводу вполне предсказуемо; можно даже смело заявить, что все его эквиваленты в целом означают одно и то же, тогда как с моделями перевода *comprehensive* такого обобщения сделать нельзя.

Обратимся теперь к вычислениям, чтобы окончательно подтвердить гипотезу о нелингвоспецифичности одного слова и лингвоспецифичности другого.

Для слова *habitat* результаты будут следующими:

формулы мер разброса	вычисления	результат
$F(M_{max}) / NumM$	17 / 5	3,4
$F(O) / NumM$	23 / 5	4,6
$F(M_{max}) / F(M_{sec})$	17 / 3	5,7
$F(M_{max}) / F(O)$	17 / 23	0,7

Тогда как для *comprehensive* они будут выглядеть вот так:

формулы мер разброса	вычисления	результат
$F(M_{max}) / NumM$	4 / 17	0,24
$F(O) / NumM$	25 / 17	1,47
$F(M_{max}) / F(M_{sec})$	4 / 3	1,33
$F(M_{max}) / F(O)$	4 / 25	0,16

В связи со всем вышесказанным можно прийти к следующему выводу: гипотеза оправдалась, и слова *habitat* и *comprehensive* в самом деле — нелингвоспецифичное и лингвоспецифичное соответственно; это заметно уже при поверхностном анализе вхождений и только подтверждается в дальнейшем. Для *habitat* переводных эквивалентов немного относительно числа вхождений, и все они семантически близки, тогда как *comprehensive* переводится множеством разных способов в зависимости от контекста, и догадаться, что на другом языке всё это — одно слово, практически невозможно.