1)Объем SpecC(arts),источник: “e-flux corpora” на sketchengine:

exhibition - 31,122/5,036,119 (words)

collaboration - 2,905/5,036,119

\*people 2,119/5,036,119

\*food 249/5,036,119

2)RefC(web), источник: “English Web 2013” на sketchengine:

exhibition 681,431/19,685,733,337

collaboration 693,275/19,685,733,337

\*people 31,193,582/19,685,733,337

\*food 7,415,937/19,685,733,337

4)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| #W | Тип | CountSpecC | CountRefC | Log Likelihood | Ранг |  |  |
| exhibition | спец. | 0.62% | <0.01% | +259438.84 | 1 |  |  |
| collaboration | спец. | 0.06% | <0.01% | +10778.91 | 2 |  |  |
| people | общ. | 0.04% | 0.16% | -6100.98 | 3 |  |  |
| food | общ. | <0.01% | 0.04% | -2285.97 | 4 |  |  |

5)

а)Везде, где стоит “<0.01%”, калькулятор показывал 0.00%, однако очевидно, что частота там не нулевая (но и не достигает 0.01%).

б)Log Likelihood во всех случаях рассчитывался в отношении CountSpecC к CountRefC

в) Не особо понял, как именно надо было ранжировать по специфичности, но если ориентироваться на то, что частота специфичных слов в специфичных корпусах выше, чем частота специфичных слов в общем корпусе, а общие слова имеют точность выше в общем корпусе (но эта точность относительно меньше, чем точность спец. слов в спец. корпусе), то результат применения весов оказался точен.