

$w_i$	Тип (специф ичное vs. общеуп отребит ель-ное)	$Count_{Spec}$ с (частота в Specific Corpora)	$Count_{RefC}$ (частота в контрас тном корпусе (Referen ce Corpora) )	LogLikel ihood	Ранг	Weirdne ss	Ранг
Христос	специф ичное	283	0	409.91	1	$\approx 301$	1
церковь	специф ичное	238	12	264.35	2	21	2
год	общеуп отребит ельное	121	141	0.55	4	0.9	4
только	общеуп отребит ельное	263	143	43.73	3	2	3

Для работы были выбраны два подкорпуса НКРЯ (оба со снятой грамматической омонимией)

1. Подкорпус нехудожественных текстов с церковно-богословской сферой функционирования (63 242 слова)
2. В качестве контрастного корпуса - нехудожественные тексты с обиходно-бытовой сферой функционирования (67 235 слов)

Для подсчета LL я воспользовалась [этим калькулятором](#).

Для коэффициента Weirdness - формула из презентации<sup>1</sup>.

## Wierdness

$$\text{Wiredness}(w_i) = \frac{\text{fr}_s(w_i)}{\text{fr}_r(w_i)} \\ = (Ws / Ts) / (Wr / Tr)$$

$\text{fr}_s(w_i)$  – относительная частота слова в коллекции текстов определенной тематической области

$\text{fr}_r(w_i)$  – относительная частота слова в контрастной коллекции (reference corpus)

<sup>1</sup> Мне кажется, в презентации ошиблись в названии...

### **Комментарий**

Применение двух различных весов оправдало ожидания. Специфичные для церковно-богословских текстов слова “Христос” и “церковь” получили первые два места в ранжировании и по LL, и по коэффициенту Weirdness. При этом разница между специфичными и общеупотребительными словами в значении коэффициентов очень значительная.