Женя Сударикова, группа 131

Specific Corpora: подкорпус нехудожественных текстов официально-деловой сферы (снятая омонимия, объём подкорпуса 74 216 слов)
Reference Corpora: подкорпус нехудожественных текстов сферы электронной коммуникации (снятая омонимия, объём подкорпуса 59 508 слов)

Калькулятор Log-likelihood: http://ucrel.lancs.ac.uk/cgi-bin/llsimple.pl?f1=16&f2=42&t1=74216&t2=59508

В качестве второй меры я использовала Wiredness:

Wiredness (w_i) = $fr_s(w_i)/fr_r(w_i)$
= (Ws / Ts) / (Wr / Tr)
$fr_s(w_i)$ – относительная частота слова в коллекции текстов определенной тематической области
$fr_r(w_i)$ – относительная частота слова в контрастной коллекции (reference corpus)

w_i	Тип (специфичное vs. общеупотребитель-ное)	$Count_{SpecC}$ (частота в Specific Corpora)	$Count_{RefC}$ (частота в контрастном корпусе (Reference Corpora))	Log-likelihood	Ранг	$fr_s(w_i)$	$fr_r(w_i)$	wiredness	Ранг
язык	общеупотребительное	34	62	15.64	4	0.0005	0.001	0.5	3
сегодня	общеупотребительное	16	42	18.53	3	0.0002	0.0007	0.29	4
договор	специфичное	384	3	421.92	1	0.005	0.00005	100	1
статья	специфичное	234	7	223.55	2	0.003	0.0001	30	2

Результат: как и ожидалось, специфичные для официально-деловой сферы слова получили большой вес и по LL, и по Wiredness, причём более чем в 10 (а иногда и 100) раз по сравнению с общеупотребительными словами.