

1. Корпус

Файл court-V-N.csv содержит 1020 примеров на употребление слова «суд» + глагол + существительное. В каждом примере 5 слов, следовательно, в корпусе 5100 токенов.

2. Золотой стандарт

По корпусу я выделил следующий топ-10 коллокаций:

принять, решение
удовлетворить, иск
удовлетворить, ходатайство
санкция, арест
санкционировать, арест
вынести, решение
избрать, мера
признать, виновная
наложить, арест
отказать, удовлетворение

3. Метрики

Параметр – минимальная частота коллокации 5 (для осмысленных коллокаций в PMI).

PMI

('ОСВОБОДИТЬ', 'ЗАЛОГ')
('ЛИШЕНИЕ', 'СВОБОДА')
('ПРОЙТИ', 'ПРЕНИЕ')
('МЕРА', 'ПРЕСЕЧЕНИЕ')
('ИЗБРАТЬ', 'МЕРА')
('ВЫДАТЬ', 'САНКЦИЯ')
('ОСТАВИТЬ', 'СИЛА')
('АРЕСТОВАТЬ', 'АКЦИЯ')
('СОГЛАСИТЬСЯ', 'ДОВОД')
('НАЧАТЬСЯ', 'РАССМОТРЕНИЕ')

Из ЗС тут только 'ИЗБРАТЬ', 'МЕРА'. В целом, практически все частотные сочетания являются коллокациями, кроме, пожалуй, (арестовывать, акция), (согласиться довод) и (начаться, рассмотрение), которые просто являются частотными сочетаниями для описания судебного процесса

Likelihood ratio

('СУД', 'ПРИЗНАТЬ')
('ВЫДАТЬ', 'САНКЦИЯ')
('СУД', 'УДОВЛЕТВОРИТЬ')
('УДОВЛЕТВОРИТЬ', 'ИСК')
('УДОВЛЕТВОРИТЬ', 'ХОДАТАЙСТВО')
('ПРИНЯТЬ', 'РЕШЕНИЕ')
('САНКЦИЯ', 'АРЕСТ')
('МЕРА', 'ПРЕСЕЧЕНИЕ')
('НАЛОЖИТЬ', 'АРЕСТ')
('СУД', 'ВЫНЕСТИ')

Из ЗС пять коллокаций. Не является коллокацией ('САНКЦИЯ', 'АРЕСТ'). Все коллокации с элементом СУД в действительности вряд ли стоит считать коллокациями. Это вполне композициональные устойчивые сочетания с агенсом. А, насколько я знаю, коллокации с А-участником в целом выделяются редко.

4. Корреляции

Correlation between PMI and Likelihood ratio

SpearmanrResult(correlation=-0.50303030303030294, pvalue=0.13833369839449197)

Средней величины отрицательная корреляция. Значит, что между рангами отрицательная зависимость. PMI в целом выдает другие коллокации, и более частотные, которые выдает Likelihood ratio, там расположены ниже, и наоборот.

Correlation between PMI and Golden Standard

SpearmanrResult(correlation=-0.57575757575757569, pvalue=0.081552814772602358)

Корреляция между PMI и ЗС тоже отрицательная, по той же причине – в ЗС более частотные коллокации, а в PMI в топе менее частотные и более контекстно зависимые коллокации.

Correlation between Likelihood ratio and Golden Standard

SpearmanrResult(correlation=0.73333333333333317, pvalue=0.015800596250571581)

Корреляция между Likelihood ratio и ЗС довольно выраженная и положительная, что говорит о схожести распределения коллокаций по рангам в ЗС и Likelihood ratio.