*Атнагулова Залина*

*Nlp4-HW1*

Тематический корпус: Серия книг о Гарри Поттере

Характеристика тематического корпуса: фэнтези и подростковый роман с элементами приключений, детектива, триллера и любовного романа; источник текстов – сайт hagread.com, объем корпуса 931473 словоупотребления, тексты в корпусе обрабатывались в сыром виде с разбиением на токены.

Примеры текста:

*— Мой отец сейчас покупает мне учебники, а мать смотрит волшебные палочки, — сообщил мальчик. Он говорил как-то очень устало, специально растягивая слова. — А потом потащу их посмотреть гоночные метлы. Не могу понять, почему первокурсникам нельзя их иметь. Думаю, мне удастся убедить отца, чтобы он купил мне такую… а потом как-нибудь тайком пронесу ее в школу.*

*— Так что мы не ошиблись, решив, что речь идет о философском камне, — спустя какое-то время подытожил он. — А теперь Снегг пытается заставить Квиррелла помочь ему завладеть камнем. Он спрашивал Квиррелла, знает ли тот, как пройти мимо Пушка. И еще он сказал Квирреллу что-то про его фокусы. Я думаю, что камень охраняет не только Пушок, но и самые разные заклинания. Возможно, Квиррелл наложил несколько своих заклятий против Темных сил, и Снеггу надо узнать, как развеять эти чары…*

Контрастный корпус: Серия романов "Наследие"

Характеристика контрастного корпуса: фэнтези и подростковый роман, источник текстов – сайт knigivsem.com, объем корпуса 818723 словоупотребления, тексты в корпусе обрабатывались в сыром виде с разбиением на токены.

Примеры текста:

*Он рассказал Эрагону, как нужно произносить заклинание, и тот, правда после нескольких неудачных попыток, все же сумел поставить защиту на свой меч. Он столь самозабвенно размахивал мечом, что Брому пришлось предупредить его…*

*Ещё только близился вечер, и тени лишь стали немного длиннее, когда Бром вдруг вздрогнул и застыл. Эрагон окликнул его по имени, но он не ответил. Муртаг подбежал к нему, но помочь Брому они уже ничем не могли. Но Эрагону показалось, что, когда он в беспомощном молчании в последний раз заглянул в ещё живые глаза Брома, то прочёл в них одобрение и прощальный привет, умиротворённо вздохнув, он затих, успокоившись навеки. Так умер Бром, великий сказитель и Всадник.*

*— Остановитесь! Опустите оружие! Вы окружены, и шесть десятков лучников сразу выпустят свои стрелы, стоит вам шевельнуться. — И, словно по мановению волшебной палочки, на крышах соседних домов возникли силуэты вооружённых людей.*

Специфичные слова для первого корпуса: палочка, метла

Общеупотребительные для обоих корпусов слова: заклинание, помогать

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| wi | Тип (тематическое vs. общеупотребительное) | CountSpecC  (частота в Specific Corpora) | CountRefC (частота в контрастном корпусе  (Reference Corpora)) | LogLikelihood | Ранг | Вес 2  Wierdness | Ранг |
| палочка | тематическое | 52/931473 | 5/818723 | 3.19 | 2 | 9.1411 | 2 |
| метла | тематическое | 50/931473 | 1/818723 | 5.46 | 1 | 43.9478 | 1 |
| заклинание | общеупотребительное | 53/931473 | 30/818723 | 0.63 | 3 | 1.5528 | 3 |
| помогать | общеупотребительное | 69/931473 | 47/818723 | 0.37 | 4 | 1.2904 | 4 |

Как можно заметить по таблице, оба метода одинаково ранжировали четыре выбранных слова, тематические слова для первого корпуса получили две самые высокие оценки, а два общеупотребительных слова – самые низкие. Стоит заметить, что слово *метла* оказалось настолько тематическим, что во втором корпусе не встретилось ни разу, поэтому мы условно поставили ему абсолютную частоту 1, чтобы избежать деления на 0 при расчетах.