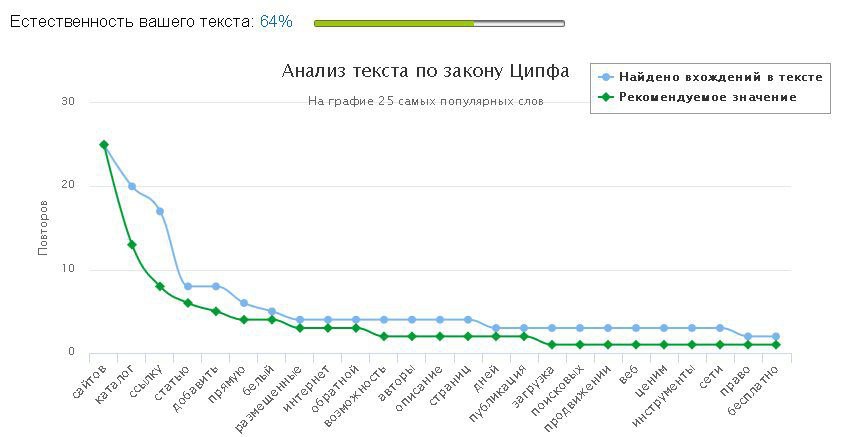
*Лабораторная работа 1. Воропаева, 5.2*

*Уровни 2, 3*

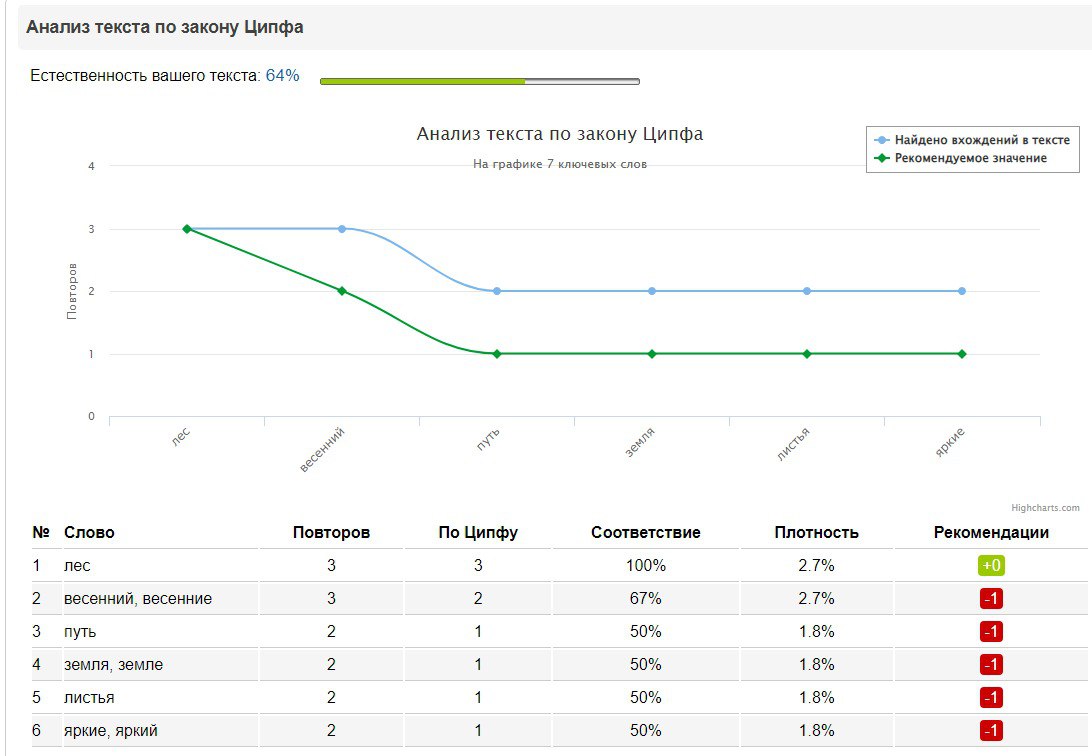
*Первое задание:*

* 1. Распределение по Закону Ципфа:
  2. 1) Текст, написанный человеком: С земли еще не сошел снег, а в душу просится весна. На ясном небе сияет ослепительное солнце. Сосульки на домах начали хныкать, плакать: они не хотят покидать свои уютные места. Зажурчали веселые ручейки. Они серебрятся от солнечных лучей. На улицах началась суета. Маленькие дети ищут удобный момент, когда мама отвернется, – залезть в лужу.  
     Весна пришла в город. На газонах из под снега уже проглянула земля, от которой пахнет сыростью. Мостовая и тротуар покрыты лужами, ребята пускают кораблики и строят плотины.  
     Деревья голы, но уже живут, дышат. На яблонях, на кустах, на деревьях скоро набухнут почки и появятся маленькие листочки.  
     Весна, по моему мнению, – самая живая из всех времен года. Ведь именно с ее приходом просыпаются звери, прилетают птицы, все вокруг оживает, и как будто заново начинается жизнь. Расходятся на небе тучи, и небо становится голубым. Можно лечь на первую травку и, глядя в небо, помечтать. И становится радостно на душе, и ты как будто заново рождаешься. Природа наряжается в яркие краски. Как здорово просыпаться утром под пение птиц! Как приятно выйти на улицу и вдохнуть запах теплого ветерка! Как пахнет весной! Даже люди становятся добрее и чаще улыбаются, потому что пришла весна, долгожданная и всеми нами любимая!
  3. Текст с этого сайта: https://coloring.kz/5/239-sochinenija-o-vesne-dlja-5-go-klassa.html





2) Сгенерированный нейросетью текст: Весенний лес, яркий, живой. Листья, свежесть, ветры нежные, птичьи трели, музыка природы. Тени играют на земле, стволы высокие, солнце пробивается сквозь листья, освещая путь. Окружение цветами, бабочки летят, все вокруг сказывается жизни. Ручей журчит, мелодия тихая, камушки блестят, муравьи работают. Деревья шуршат, словно шепчут тайны. В воздухе запахи, травы, весенние, земля, обновление. За каждым кустом что-то таится, зайцы прыгнут, лисы крадутся. Ветер меняет все, но остается непонятное в этом хаосе. Лес весной – живое искушение, мир, который дышит, но не всегда понятен. Здесь спокойно, но в то же время шумно, удивительность в каждой детали. Краски яркие, но исчезают, смешиваясь с тенями. Не все так просто, но весенний лес, он всегда находит путь к сердцу.



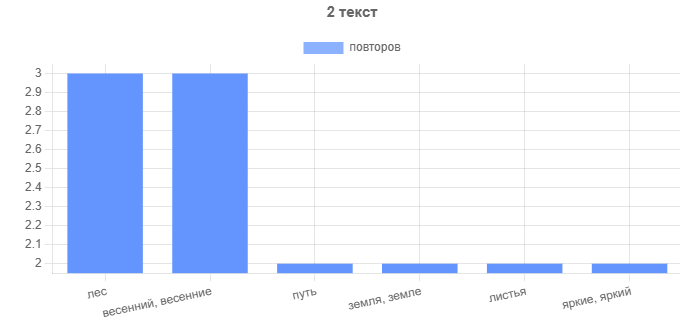
Сайт: <https://p2pi.ru/text-analyze.html?ysclid=m78270rjnu693884248>

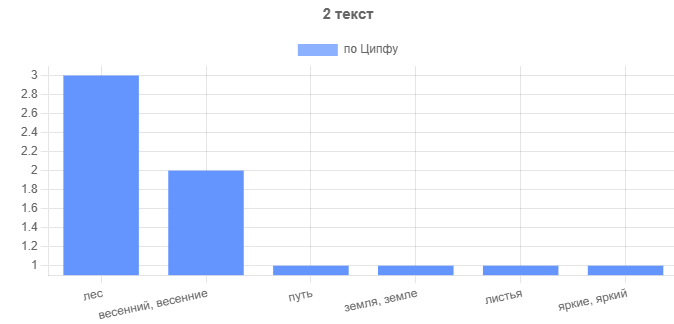
Гистограммы:

1)



2)





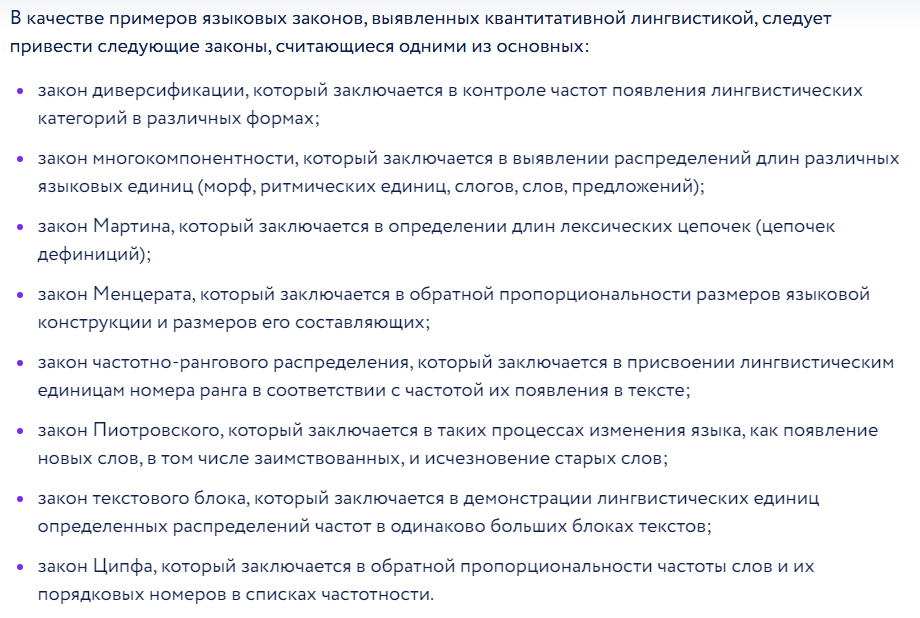
Гистограммы делала с помощью сайта: https://anytools.pro/ru/img/chart/bar

Выводы о естественности текстов:

Первый текст кажется более естественным и органичным, так как он полон эмоций, метафор и четких образов, что создает связь между автором и читателем. Человеческое восприятие природы является основой текста.  
Второй текст, несмотря на красоту образов, менее органичен из-за отсутствия глубоких эмоциональных связей и структуры. Он может показаться менее естественным, чем первый, потому что его стиль напоминает стилизацию, а не органичное наблюдение.

Вывод: Первый текст, скорее всего, написан человеком, поскольку он дышит эмоциями и естественной связью с темой. Второй текст, написанный нейросетью, может быть красивым и выразительным, но ему не хватает того тепла и глубокого понимания, которое присуще рукописи.  
В результате, на основе данного анализа, первый текст можно оценивать как более естественный и эмоционально насыщенный, тогда как второй текст - как менее живой и более механистичный.

Другие законы квантитативной лингвистики:



Источник: <https://spravochnick.ru/yazykoznanie_i_filologiya/kvantitativnaya_lingvistika/>

* Закон диверсификации, который заключается в контроле частот появления лингвистических категорий в различных формах.
* Закон Мартина. Касается лексических цепочек, которые образуются при поиске дефиниций (определений) слов в словаре.
* Закон Менцерата-Альтмана: касается длины языковых конструкций, утверждая, что длина слова обратно пропорциональна длине его составляющих (например, слогов).
* Закон многокомпонентности, который заключается в выявлении распределений длин различных языковых единиц (морф, ритмических единиц, слогов, слов, предложений).
* Закон Пиотровского. Заключается в таких процессах изменения языка, как появление новых слов, в том числе заимствованных, и исчезновение старых слов.
* Закон текстового блока. Заключается в демонстрации лингвистических единиц определённых распределений частот в одинаково больших блоках текстов.
* Закон Ципфа: утверждает, что частота слова обратно пропорциональна его рангу в частотном списке. Это означает, что наиболее частые слова используются значительно чаще, чем менее распространенные.
* Закон частотно-рангового распределения, который заключается в присвоении лингвистическим единицам номера ранга в соответствии с частотой их появления в тексте.
* Модель жизненного цикла слова: показывает тенденцию к увеличению степени абстрактности значений у многозначных слов

Эти законы в файле с условием лабораторной работы

*Второе задание:*

Анализ принципов работы дистрибутивной семантики с использованием ресурса RusVectōrēs.  
Дистрибутивная семантика — это подход к изучению смысловых отношений между словами на основе их контекстного использования в текстах. Сервис RusVectōrēs предоставляет разнообразные инструменты для анализа лексических данных, включая семантические отношения и векторные представления слов. Ознакомление с этим ресурсом позволяет изучить, как дистрибутивная семантика применяется на практике.

Ссылка на сайт: https://rusvectores.org/ru/  
1. Ознакомление с основными принципами работы сервиса RusVectōrēs  
Сервис RusVectōrēs предлагает несколько моделей, разработанных для анализа текста и лексики. Основные принципы его работы включают:

1) Векторизация слов: Каждое слово представляется в виде вектора, что позволяет анализировать семантические связи между ними на основе геометрических свойств векторов.  
2) Контекст: Слова анализируются с точки зрения их использования в различных контекстах, что позволяет выявить семантические связи и отношения.  
3) Семантическое расстояние: Определяются степени близости между словами на основе их векторных представлений.  
2. Анализ инструментов, представленных на сайте

1) Вычисление семантических ассоциатов — это определение набора слов, лексическое значение которых ближе всего к исследуемому слову. Также ассоциативное значение слова отражает его семантику, особенности употребления, а также указывает на его место в ассоциативном поле. Например, материалы ассоциативных словарей основываются на результатах массовых ассоциативных экспериментов и представляют собой сеть понятий, связанных в сознании среднего носителя языка.

2) Визуализация семантических связей между словами — это способ отображения семантической близости слов с помощью графического представления.

Существуют разные подходы к визуализации:

Выдача столбика ближайших к слову «семантических ассоциатов» (то есть слов с наиболее похожими семантическими векторами).

Использование сетей (графов). Для каждого слова строится сеть из его семантических ассоциатов. При этом показатель близости можно отобразить, например, через длину линии: чем короче связь — тем ближе слово в векторной модели.

3) Семантический калькулятор

Здесь можно вычислять отношения: например, «найти слово D, связанное со словом C таким же образом, как слово A связано со словом B». Таким образом можно определять семантические связи между понятиями и решать задачи на аналогии. В форме ввода приведен пример: какое слово относится к слову «Лондон», так же, как «Россия» относится к «Москве»? Ответ — «Великобритания»: Лондон столица Великобритании, а Москва — столица России.

4) Вычисление семантической близости — это процесс определения степени смысловой связанности между двумя концептами (понятиями).

В рамках дистрибутивной семантики, области лингвистики, семантическая близость между лингвистическими единицами вычисляется на основании их распределения в больших массивах лингвистических данных (текстовых корпусах). Каждому слову присваивается свой контекстный вектор, множество векторов формирует словесное векторное пространство. Семантическое расстояние между понятиями, выраженными словами естественного языка, обычно вычисляется как косинусное расстояние между векторами словесного пространства.

Также для расчёта семантической близости пары концептов можно использовать расстояние между соответствующими им статьями в графе ссылок Википедии, при этом для оценки расстояния между вершинами в графе используется длина кратчайшего пути между ними.

Конец формы

3. Выводы о применимости и эффективности инструментов дистрибутивной семантики RusVectōrēs  
На основе проведенного анализа можно сделать следующие выводы:  
Применимость: Инструменты RusVectōrēs являются мощным средством для анализа семантических отношений между словами и могут быть успешно использованы в лингвистических исследованиях, а также в прикладных задачах, таких как обработка естественного языка и создание семантических сетей.  
Эффективность: Большинство инструментов позволяют получать актуальную информацию о лексических связях. Модели, такие как Word2Vec и FastText, предлагают высокую точность и производительность в анализе. Особая полезность заключается в способности обнаруживать как прямые, так и косвенные семантические связи.  
Удобство интерфейса: Пользовательский интерфейс RusVectōrēs является доступным и интуитивным, что позволяет без особых сложностей проводить различные виды лексических исследований.  
Образование и исследования: Сервис может быть полезен не только для ученых и исследователей, но и для студентов, изучающих семантику и синтаксис.  
Заключение  
RusVectōrēs представляет собой полезный ресурс для изучения и анализа дистрибутивной семантики, предлагая широкий спектр инструментов для поиска и анализа семантических отношений между словами. Эффективность его использования позволяет исследователям глубже понять лексическую структуру русского языка и выявить ключевые семантические закономерности.