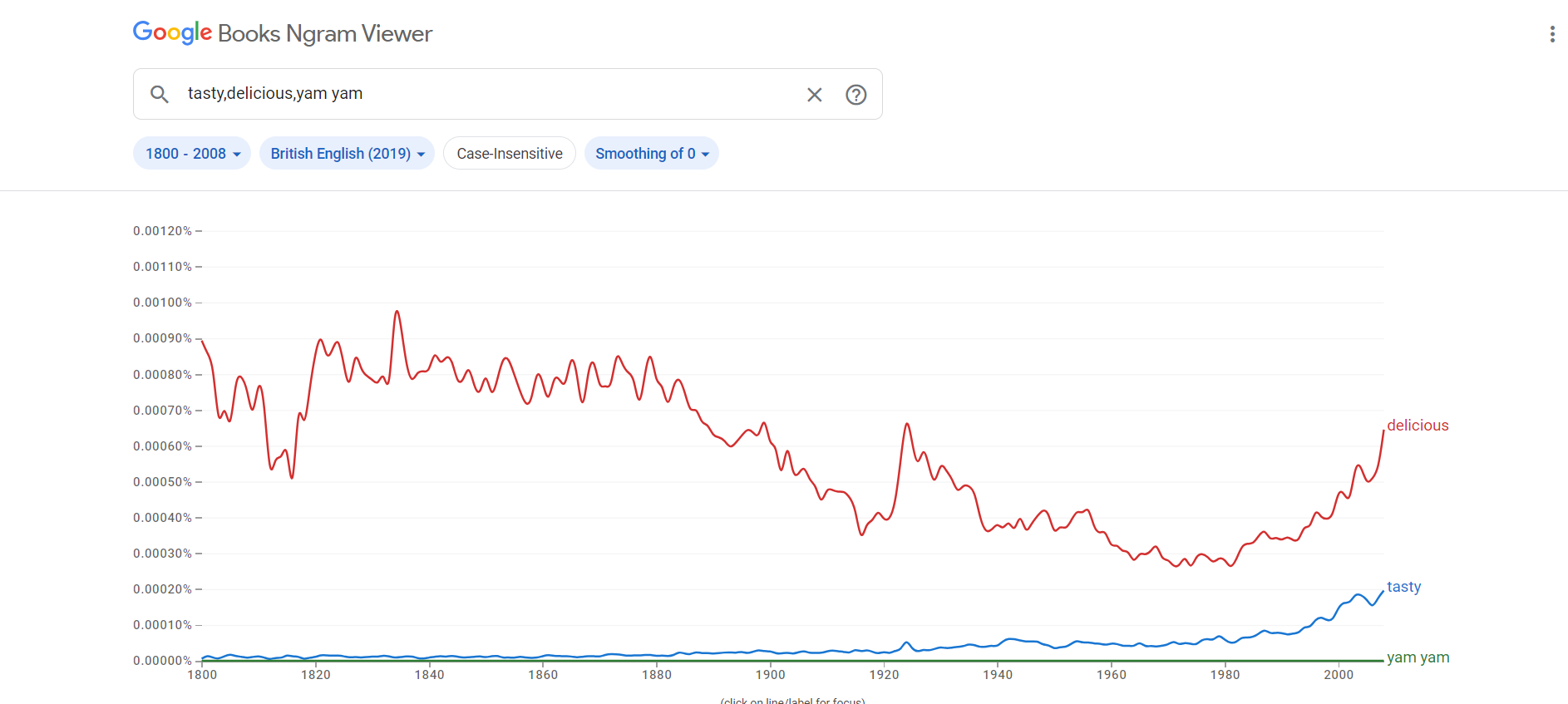
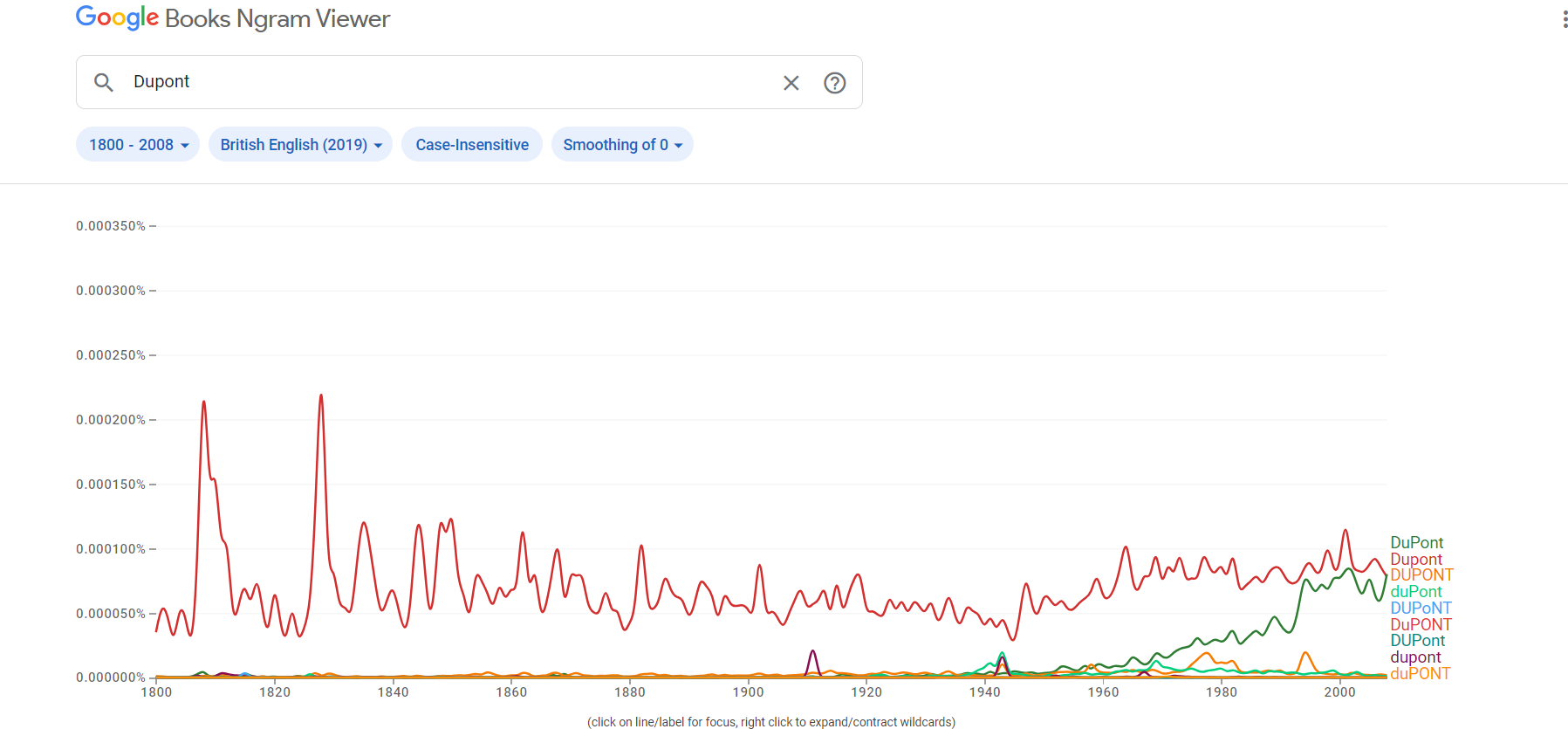
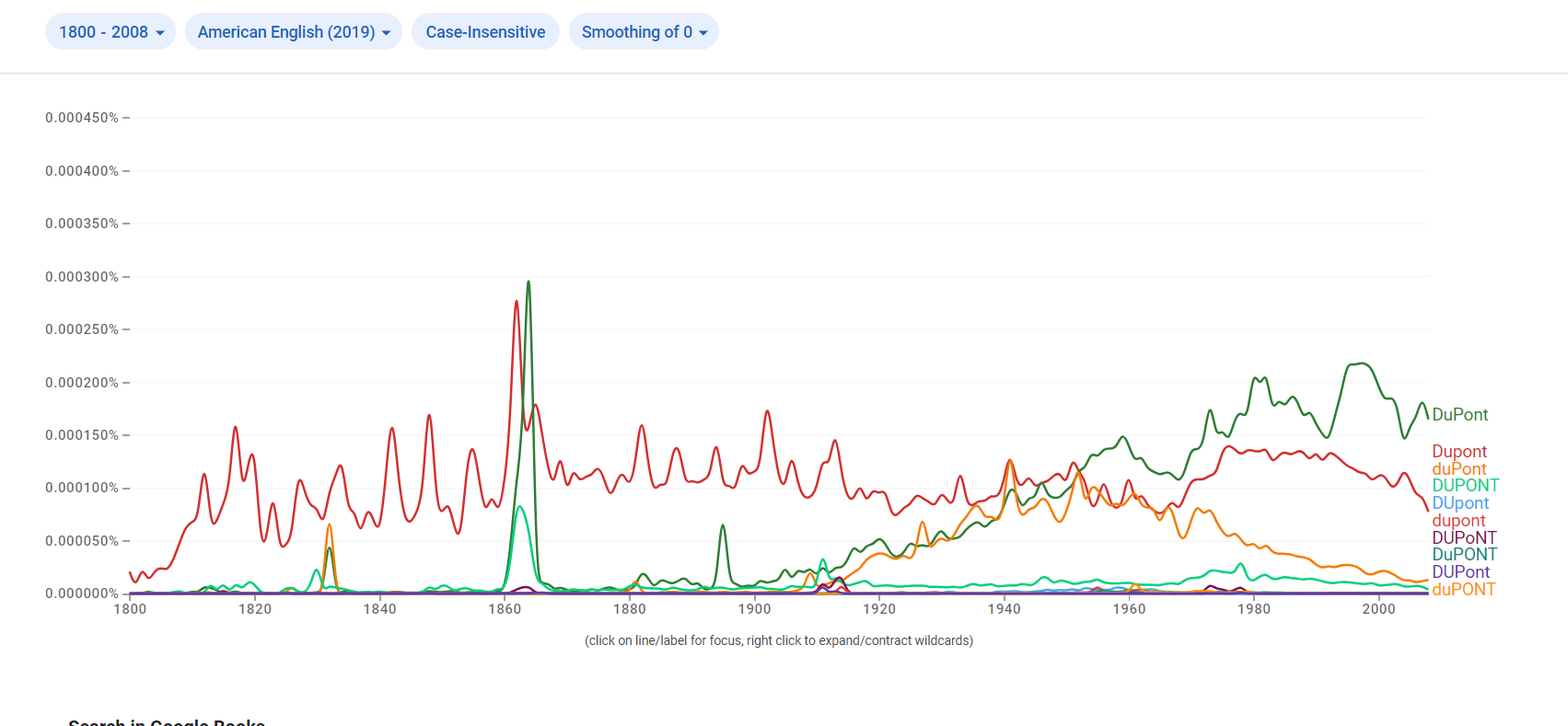
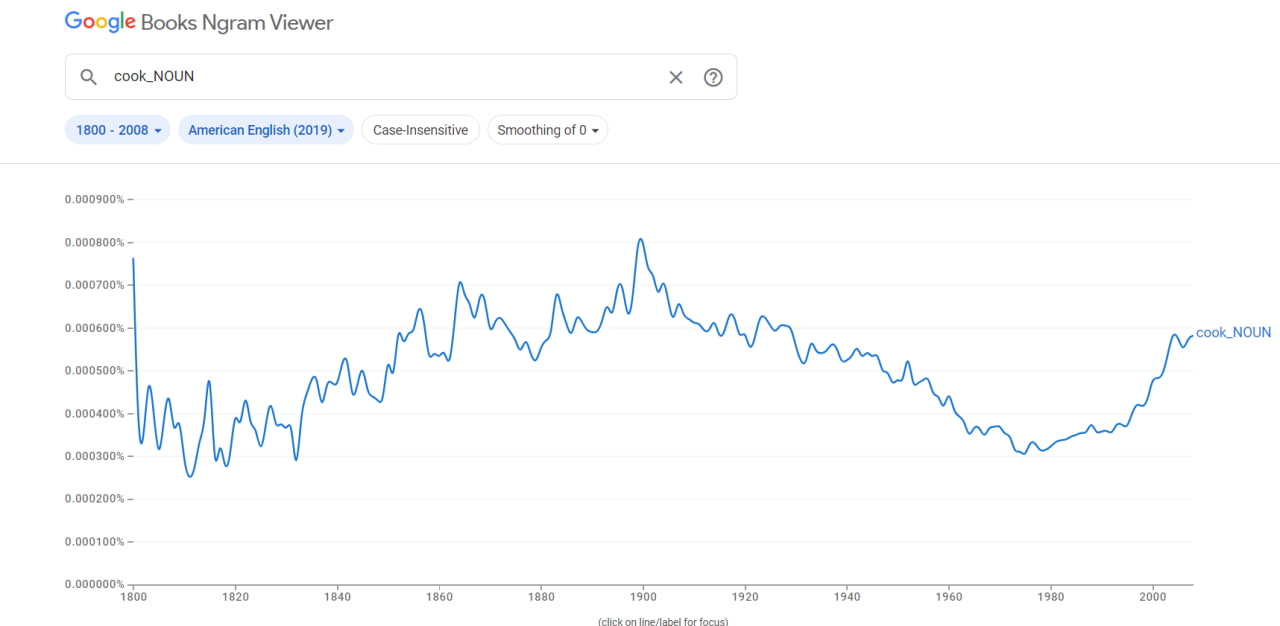
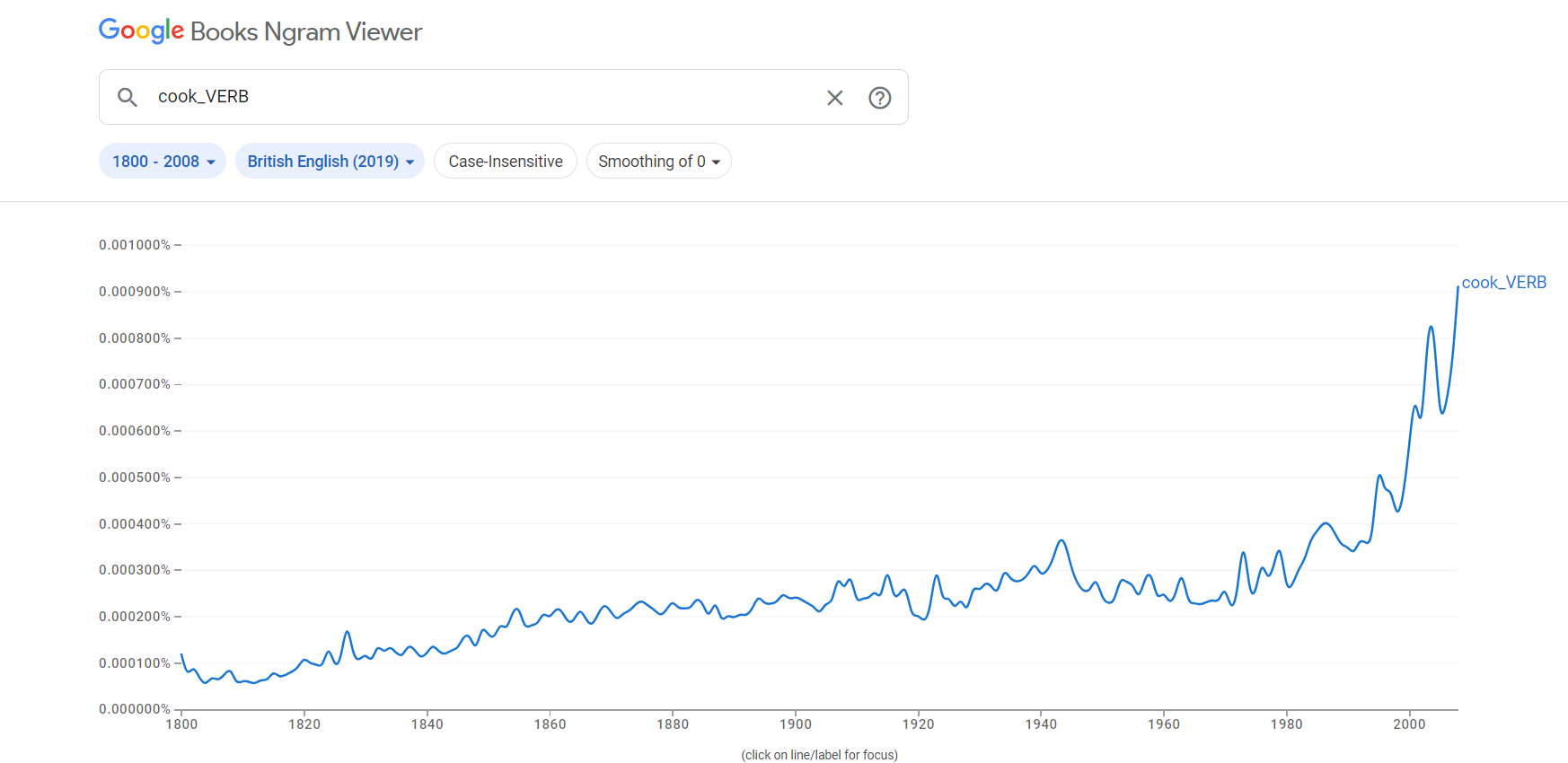
Уровень 1 Задание 1задание 2. 



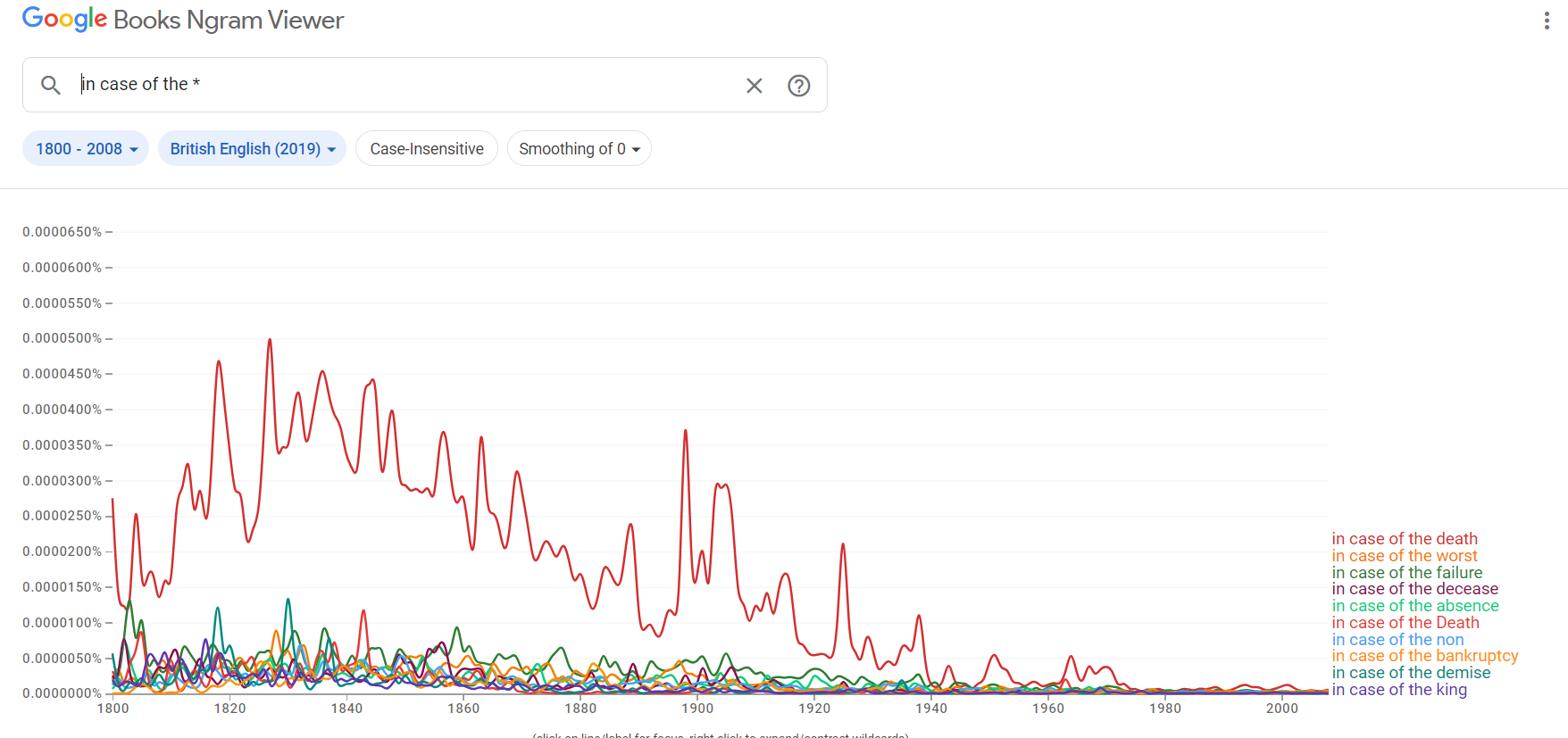
Задание 3.

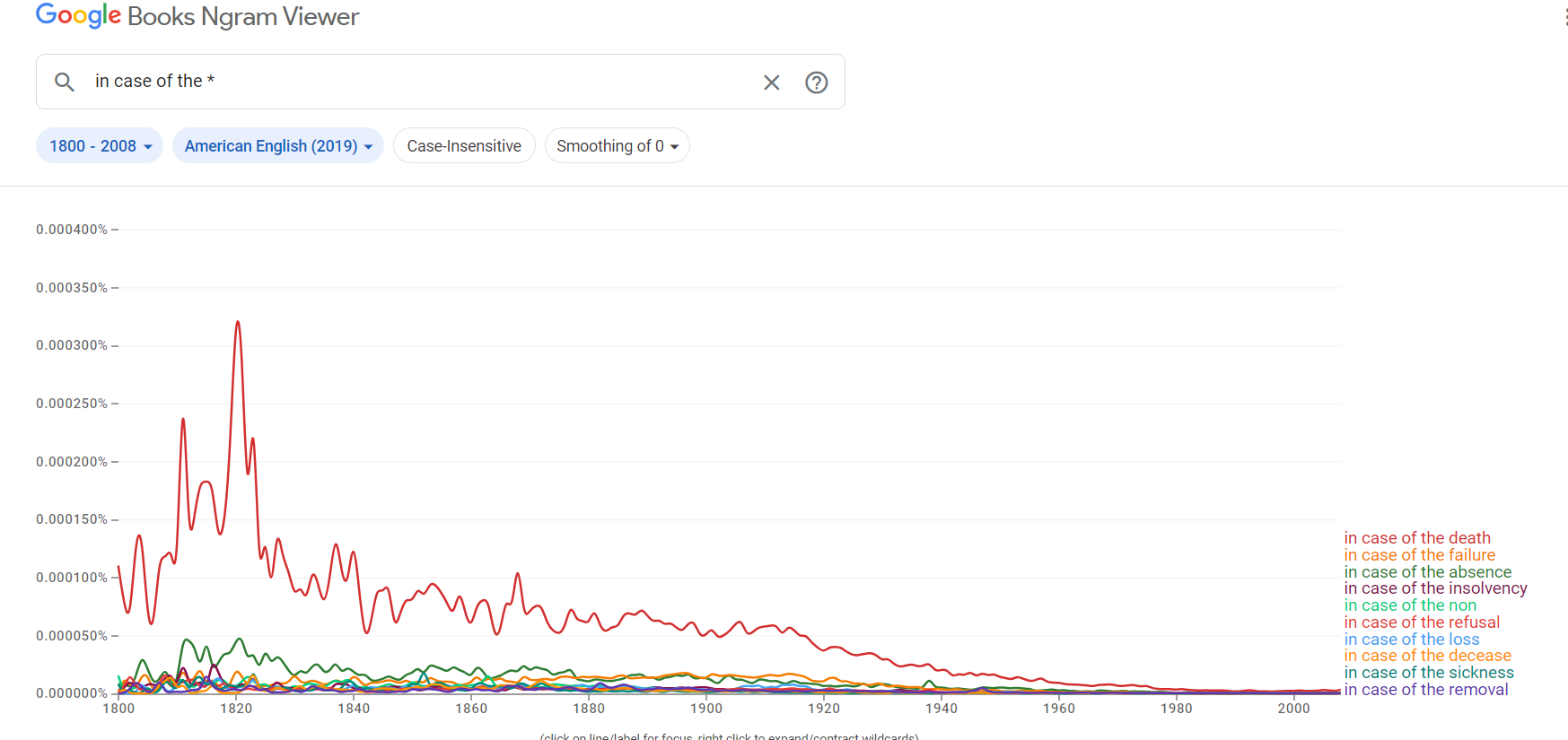




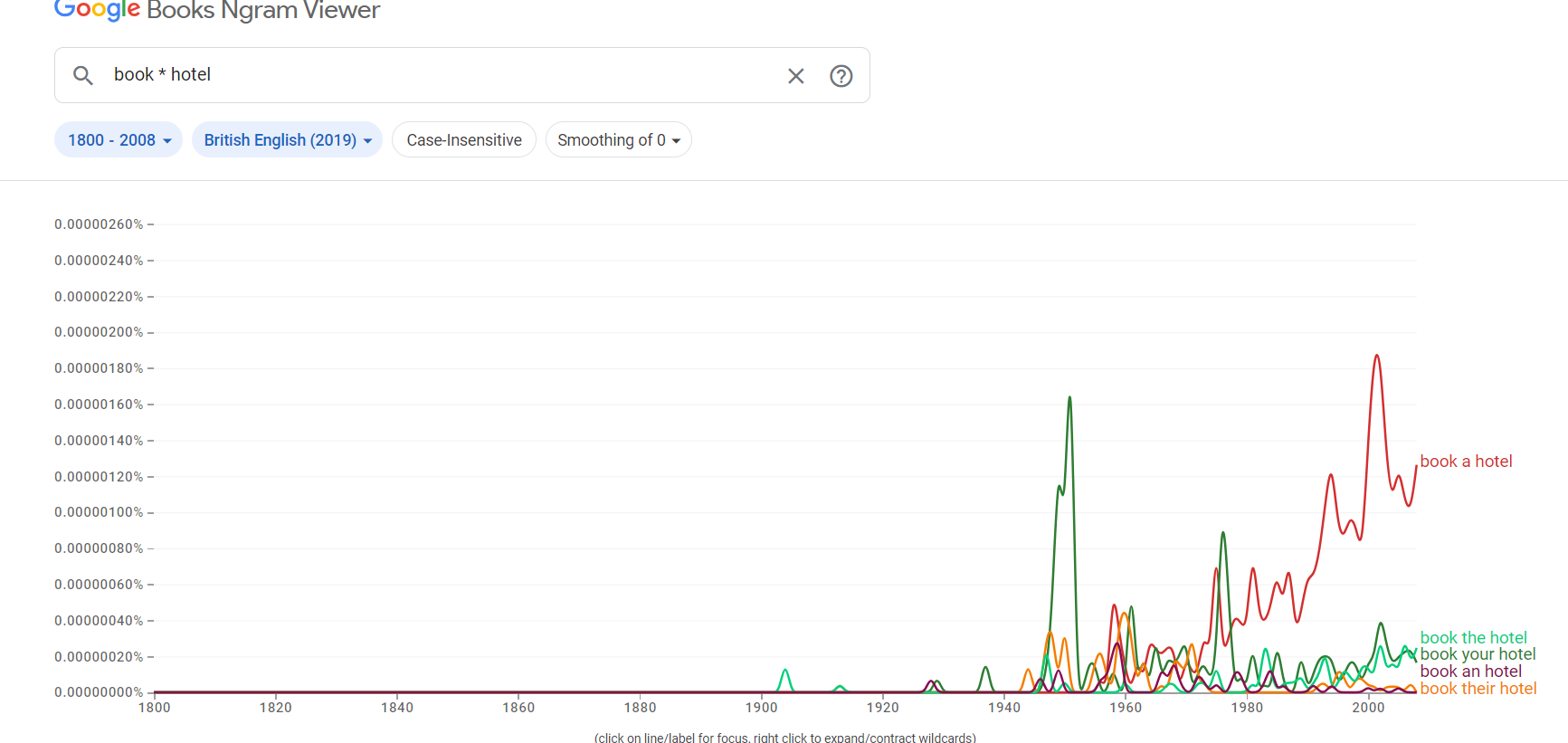


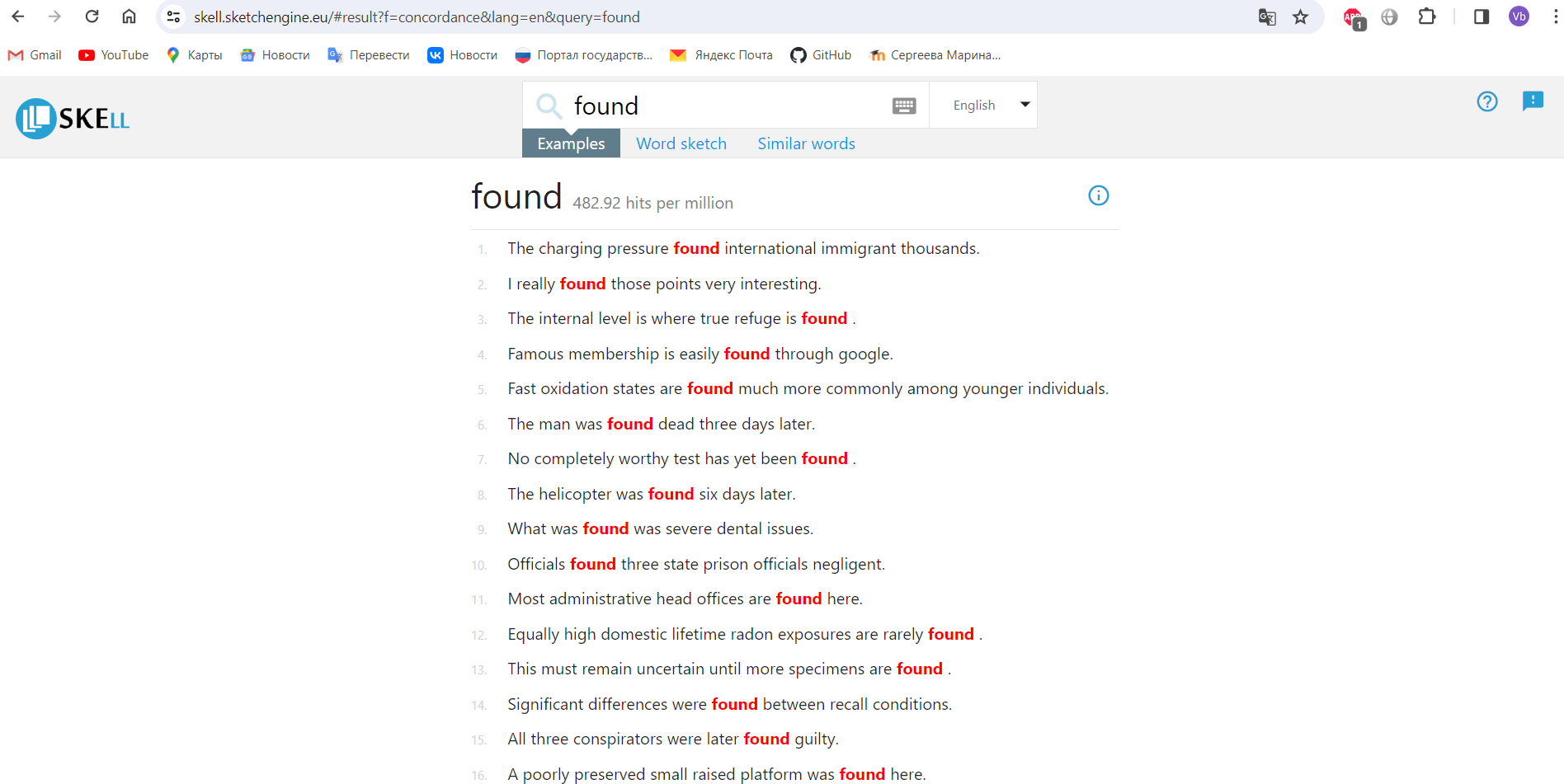
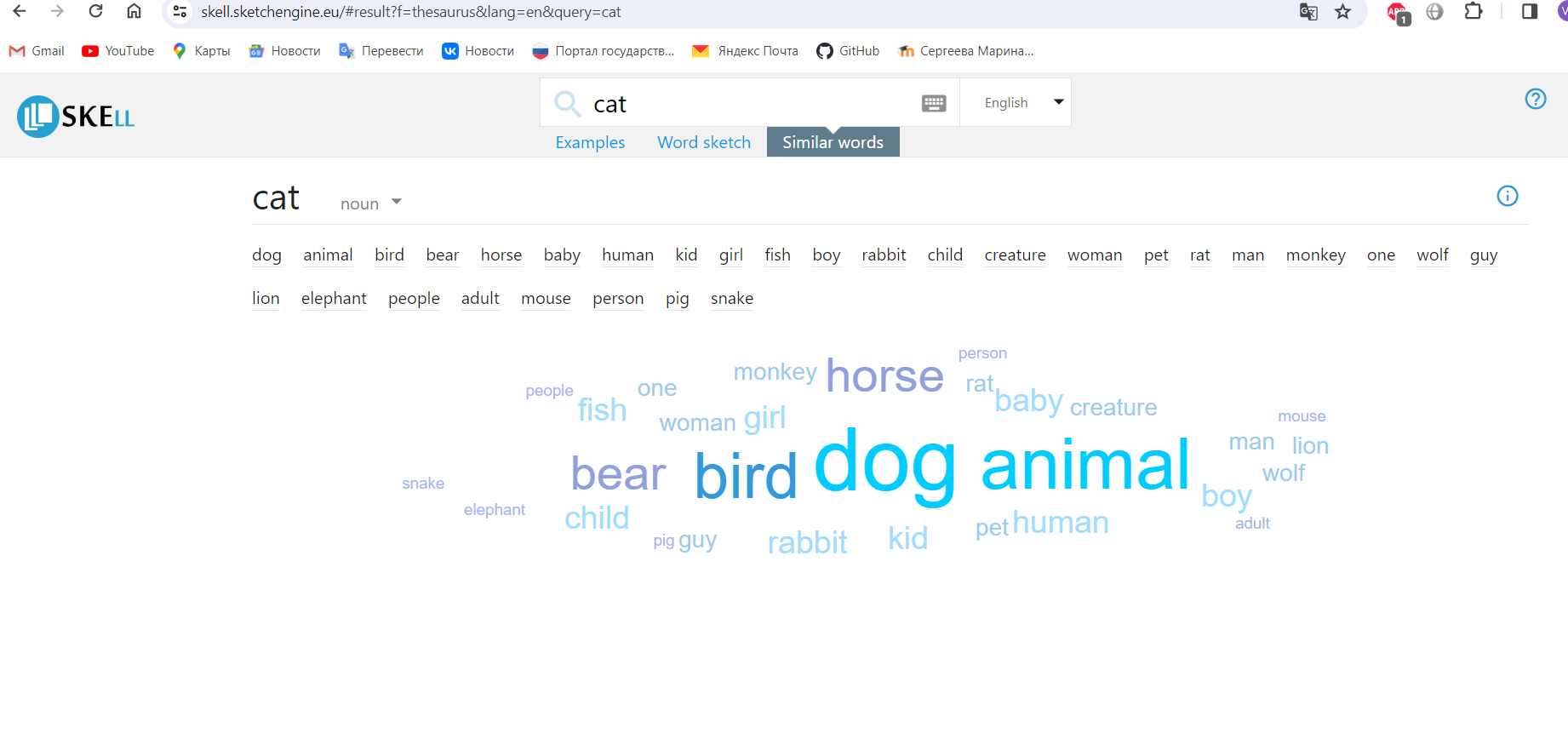
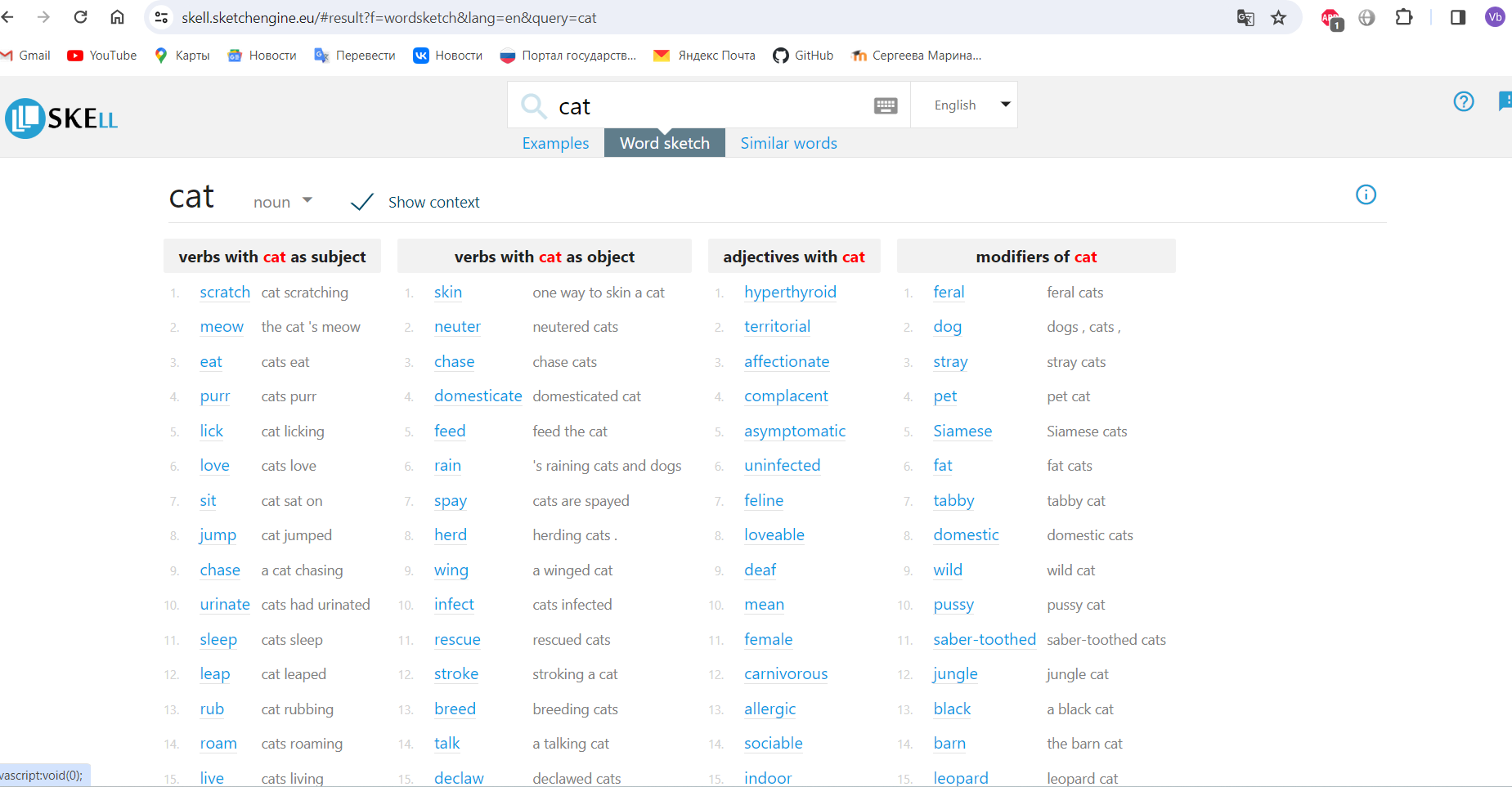
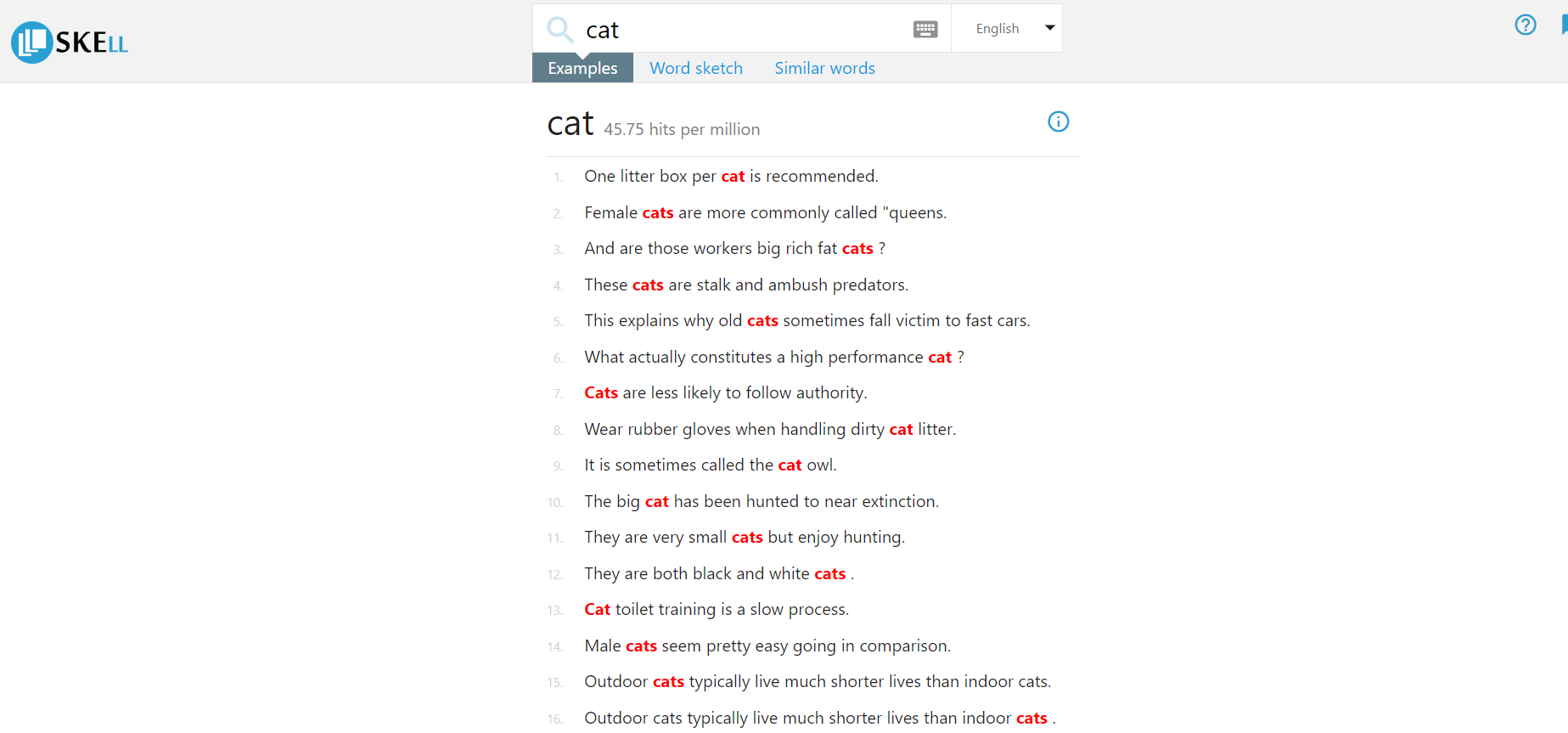


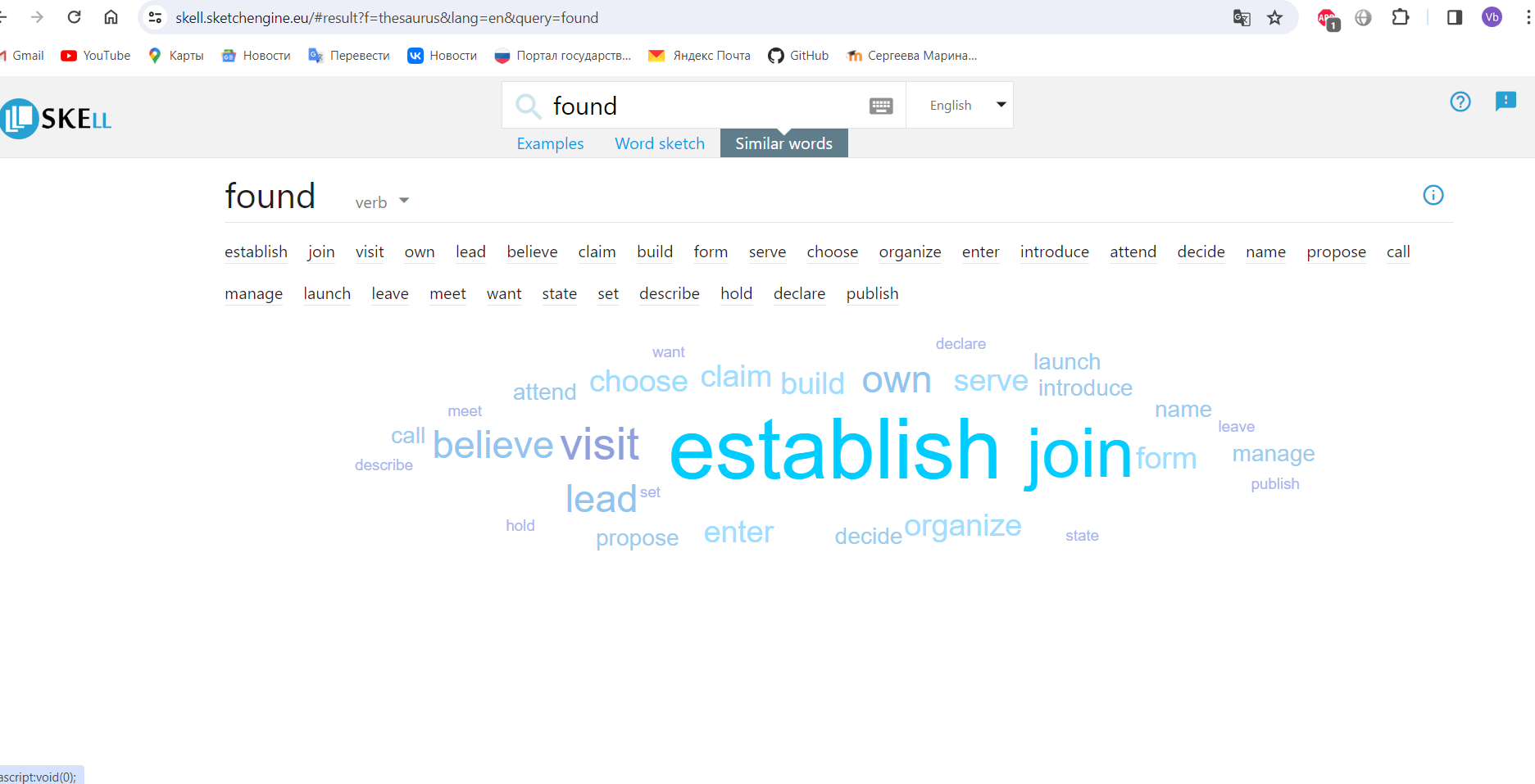
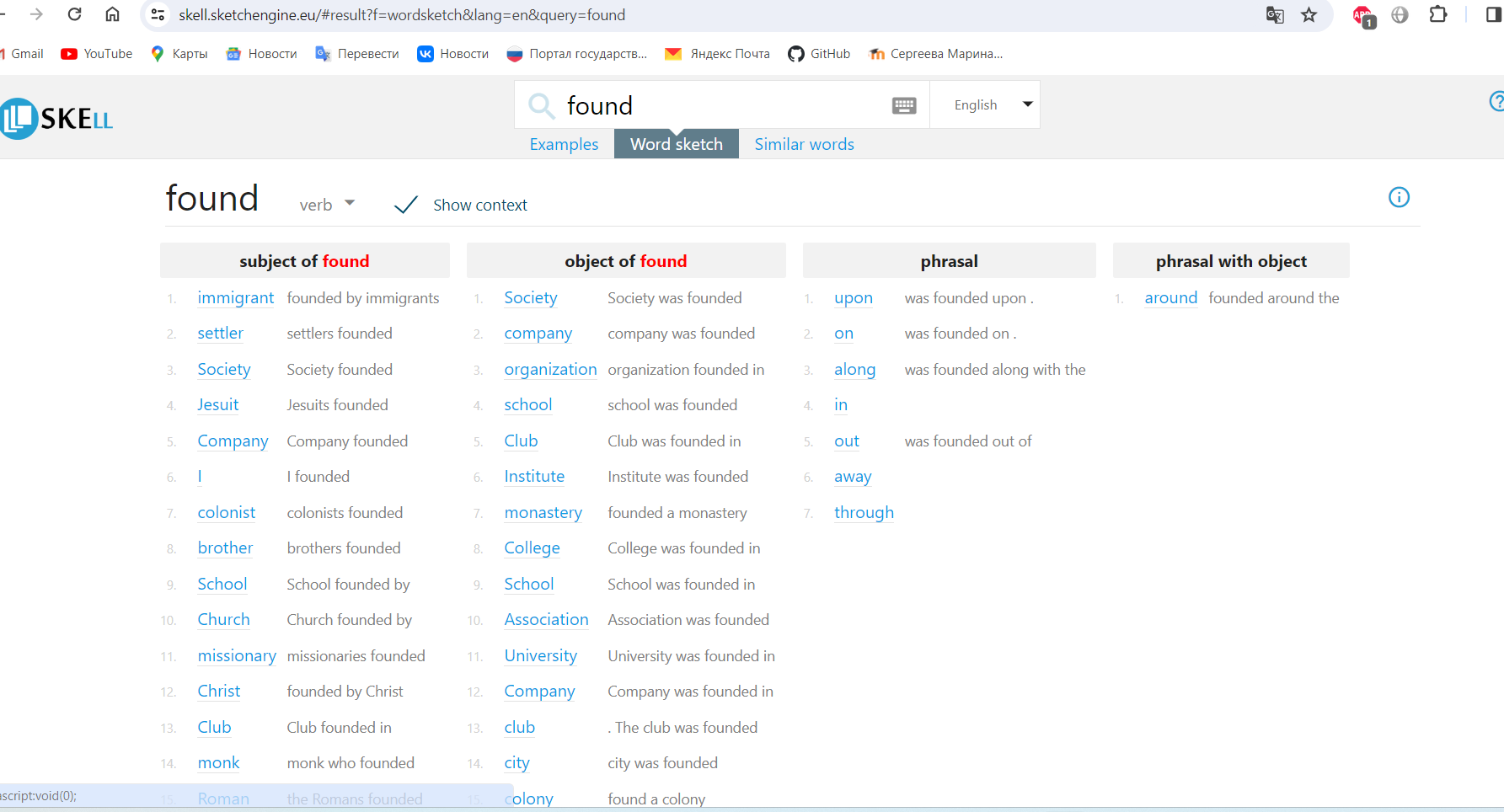
Задание 4. 



Задание 5. 







Задание 3

1 список файлов в корпусе

2 список слов

3 список всех употреблений заданного языкового выражения в контексте

4 инструмент кластеры используется для создания упорядоченного списка кластеров, которые появляются вокруг поиска в целевом файле, перечисленные в левой части главного окна.

1. Concordance (Конкорданс):

Инструмент "Concordance" в AntConc позволяет просмотреть все вхождения определенного слова или фразы в тексте, с контекстом и частотой использования. Это помогает анализировать употребление слова в различных контекстах, выявлять его частотность и использование. Возможные сферы применения: лингвистические исследования, анализ текстов на определенную тему, изучение структуры текстов.

2. Clusters (Кластеры):

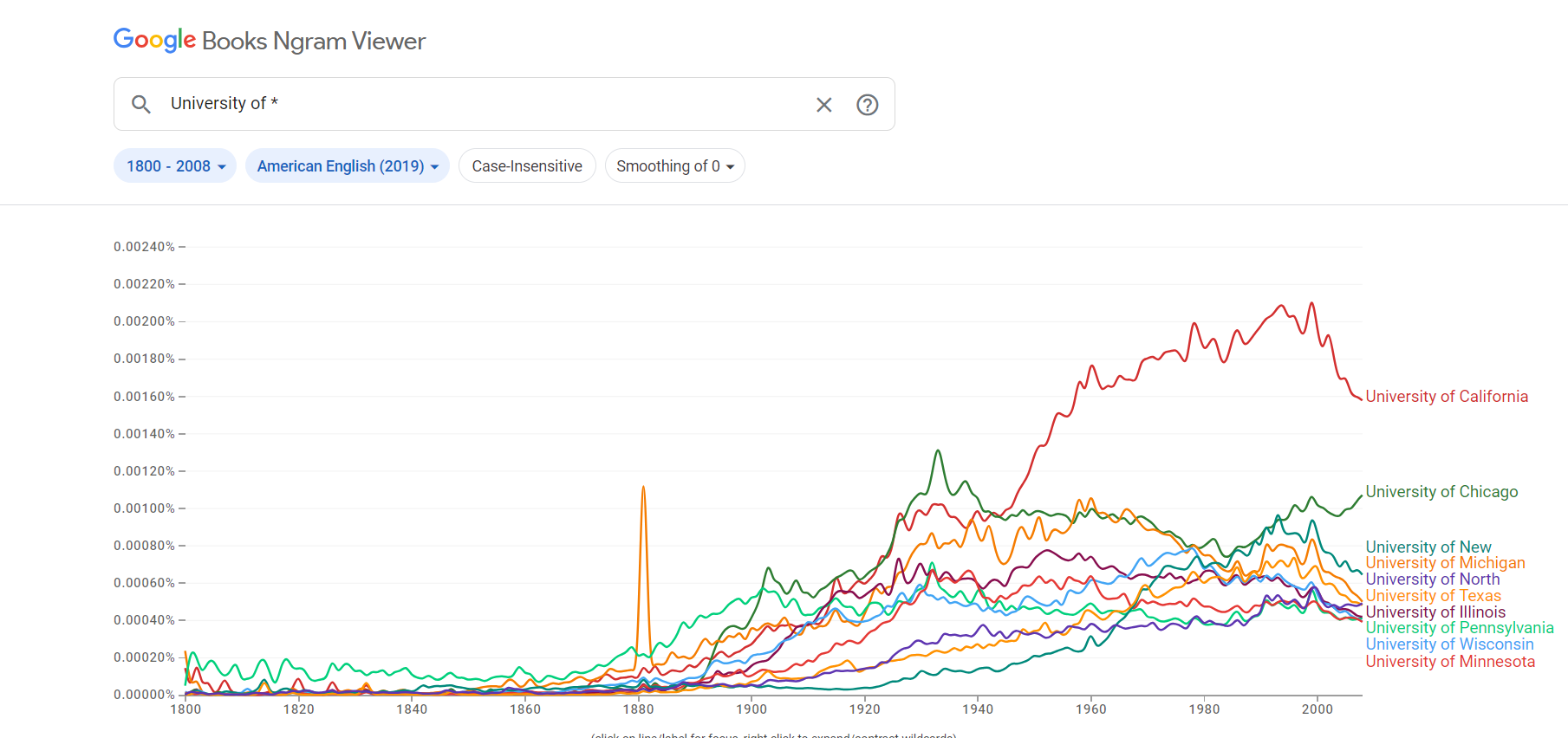
Инструмент "Clusters" в AntConc позволяет выделить группы слов, которые часто встречаются вместе в тексте. Это помогает определить тематику текста, выделить ключевые термины и выявить связи между различными словами. Возможные сферы применения: анализ тематики текстов, изучение ключевых слов и понятий, выявление связей между словами.

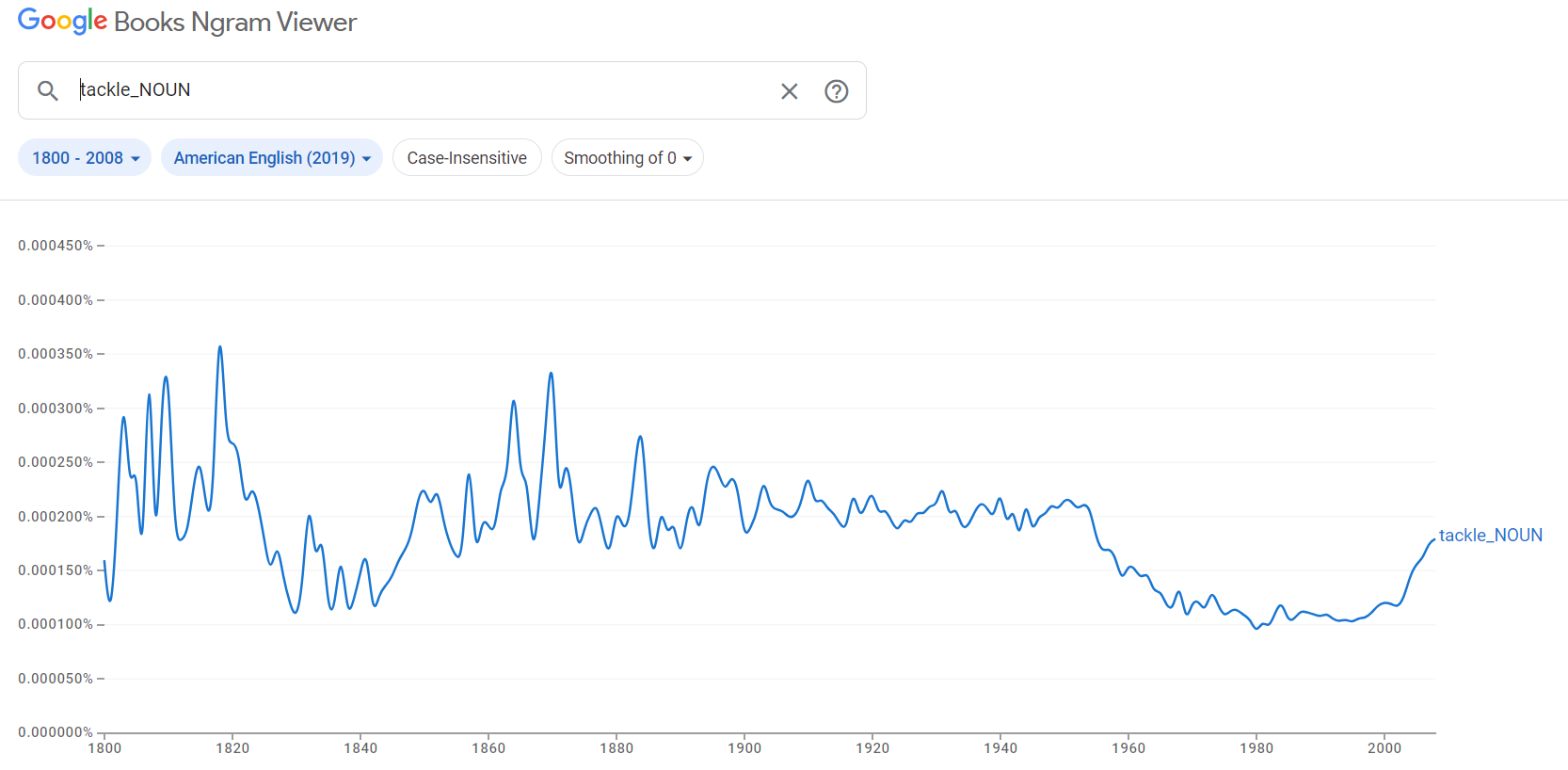
3. Word list (Список слов):

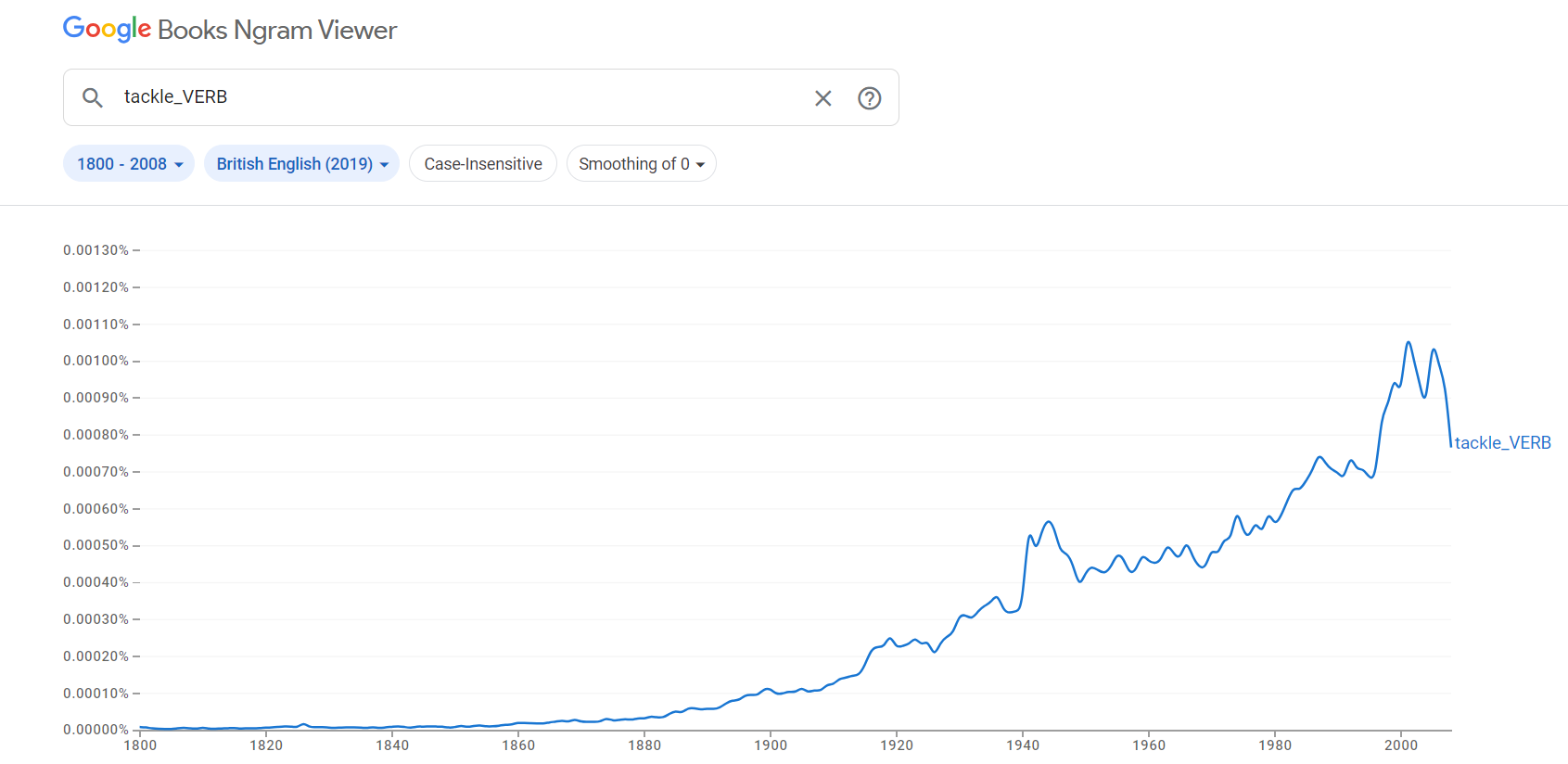
Инструмент "Word list" в AntConc отображает список всех уникальных слов, используемых в тексте, с указанием их частоты встречаемости. Это помогает оценить словарный запас текста, выделить ключевые слова и провести анализ частотности слов. Возможные сферы применения: анализ словарного запаса текста, изучение употребления слов в контексте, подготовка к изучению нового языка.

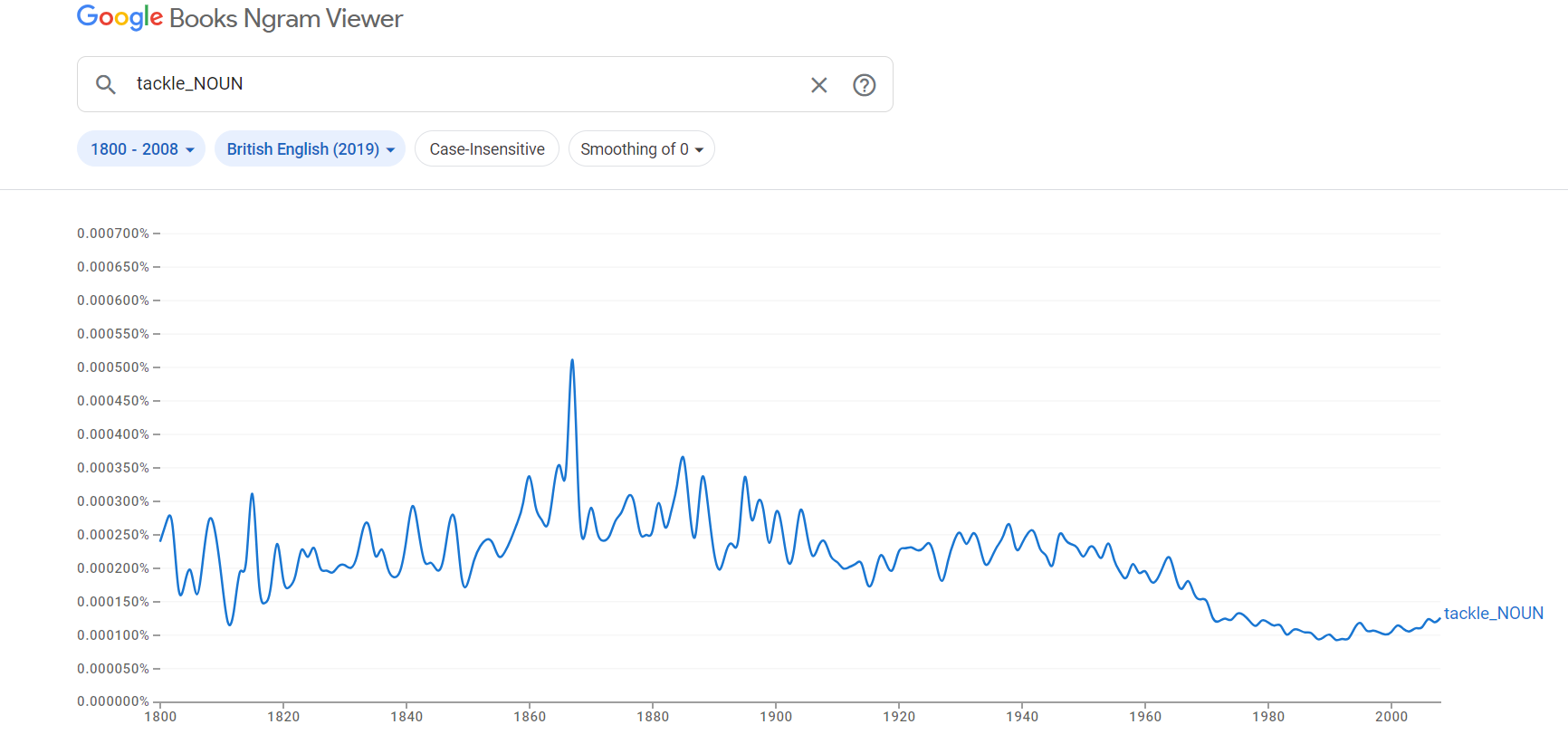
4. Key words (Ключевые слова):

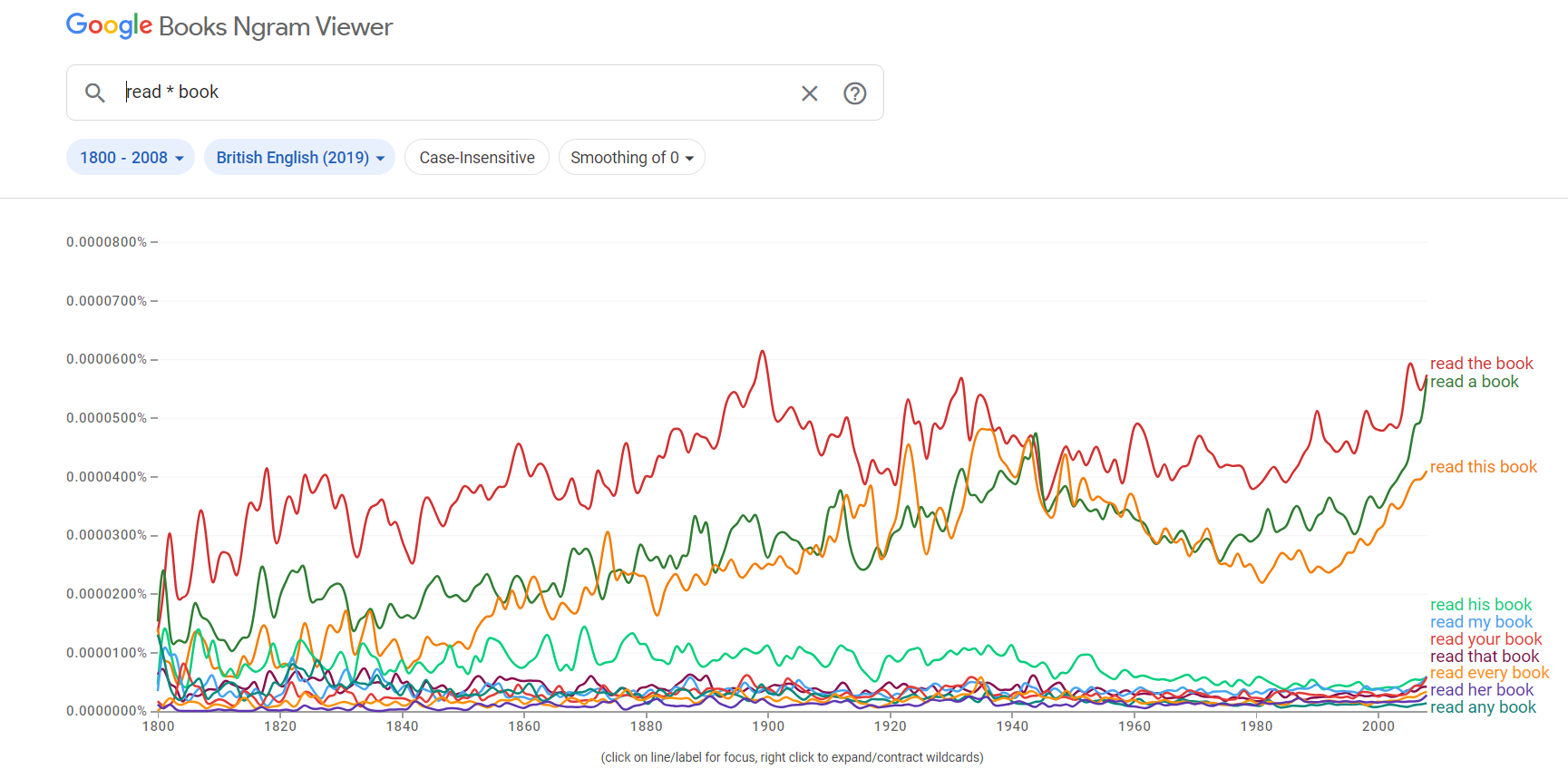
Инструмент "Key words" в AntConc выделяет наиболее значимые и часто употребляемые слова в тексте, помогая определить его основную тему или содержание. Это полезно для быстрой оценки содержания текста и выделения ключевых моментов. Возможные сферы применения: анализ содержания текста, выделение ключевых терминов, определение тематики.

Уровень 2 задание 6. 

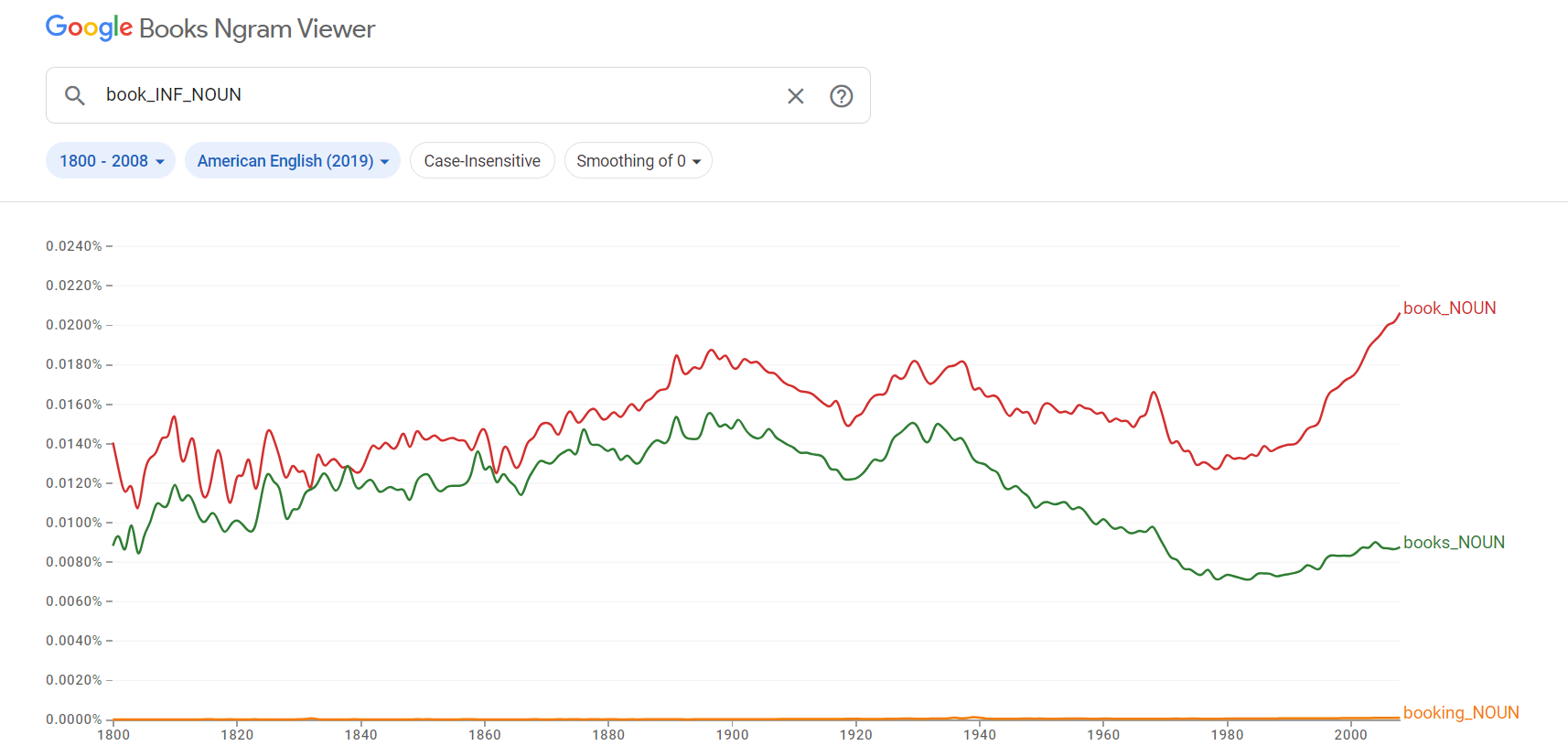
7. 

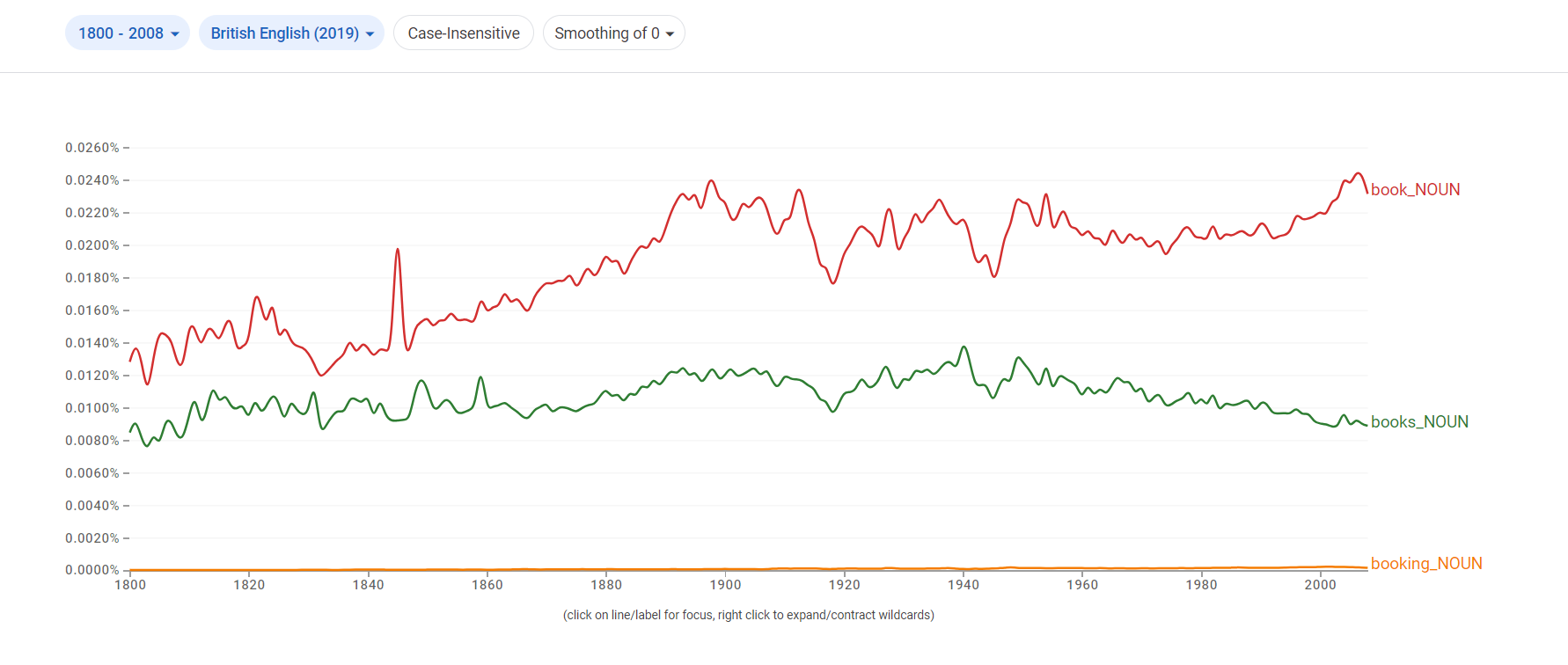




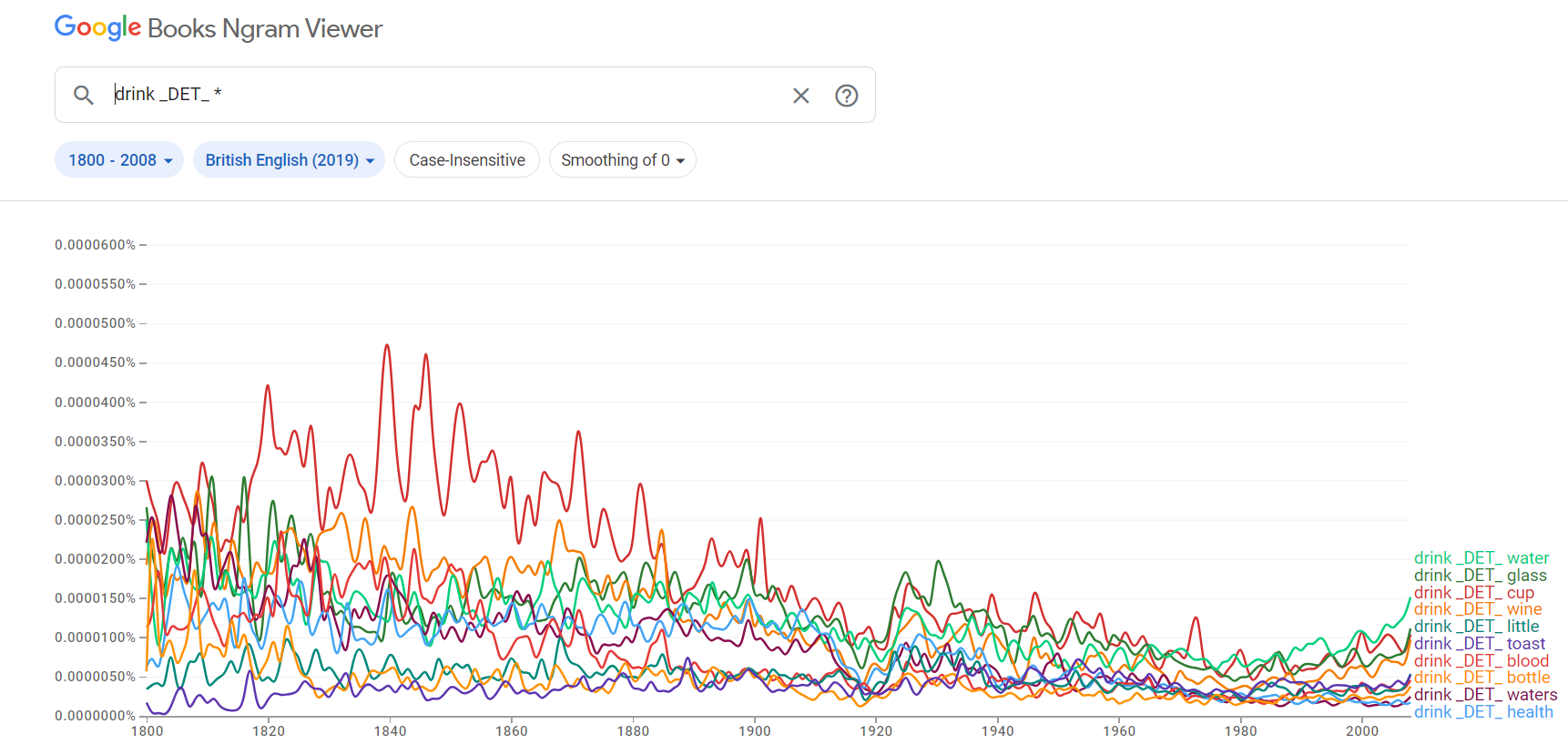
8 

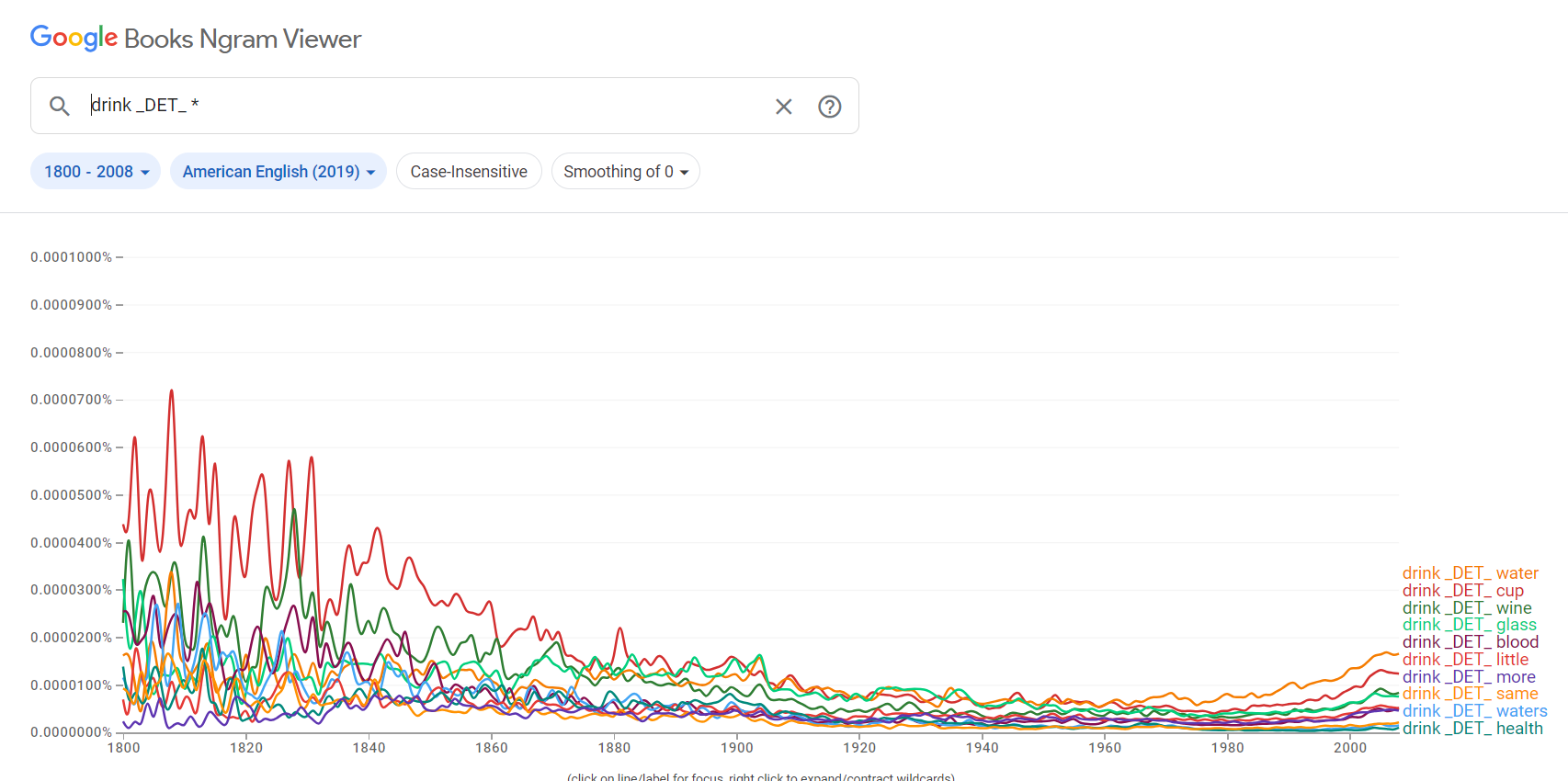


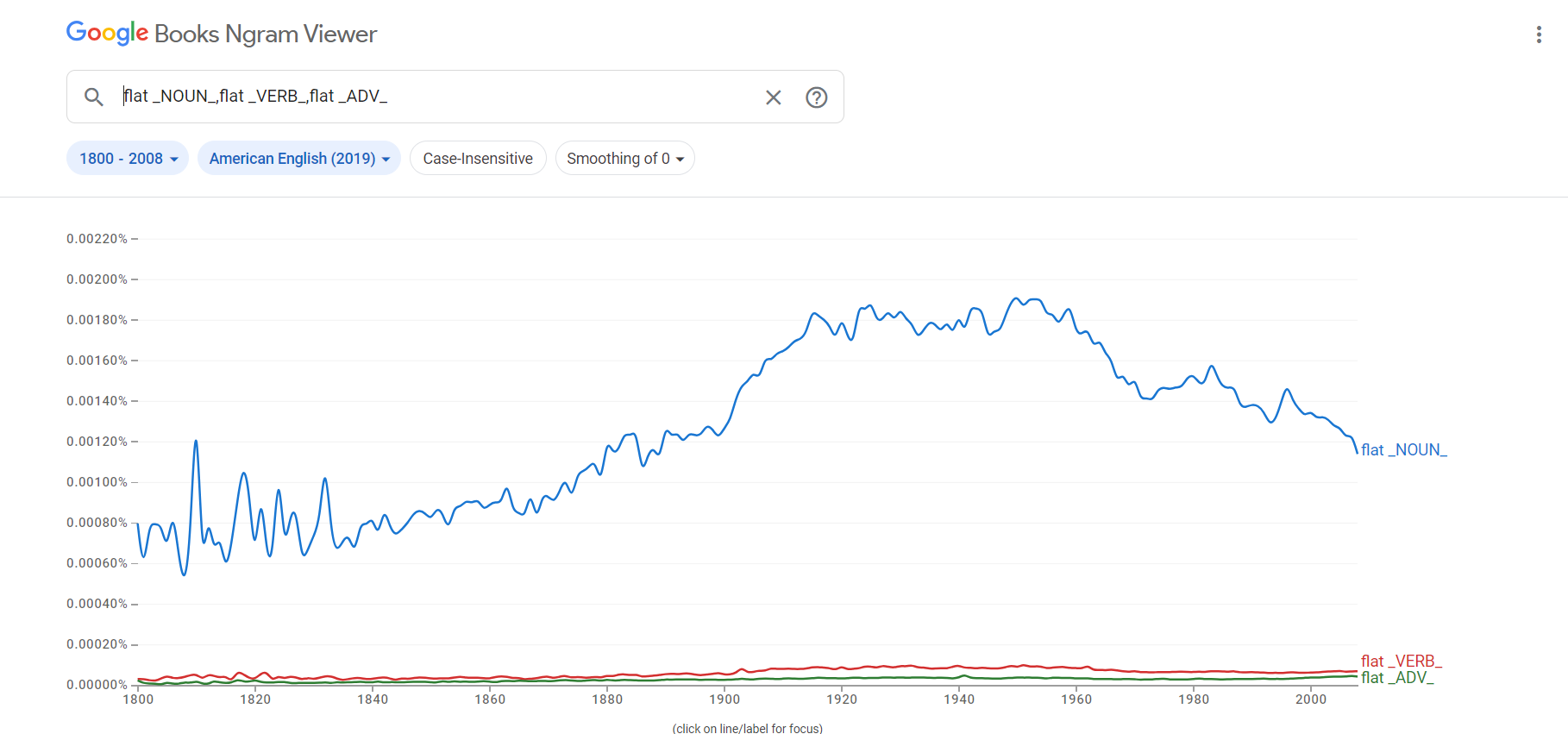
9. 

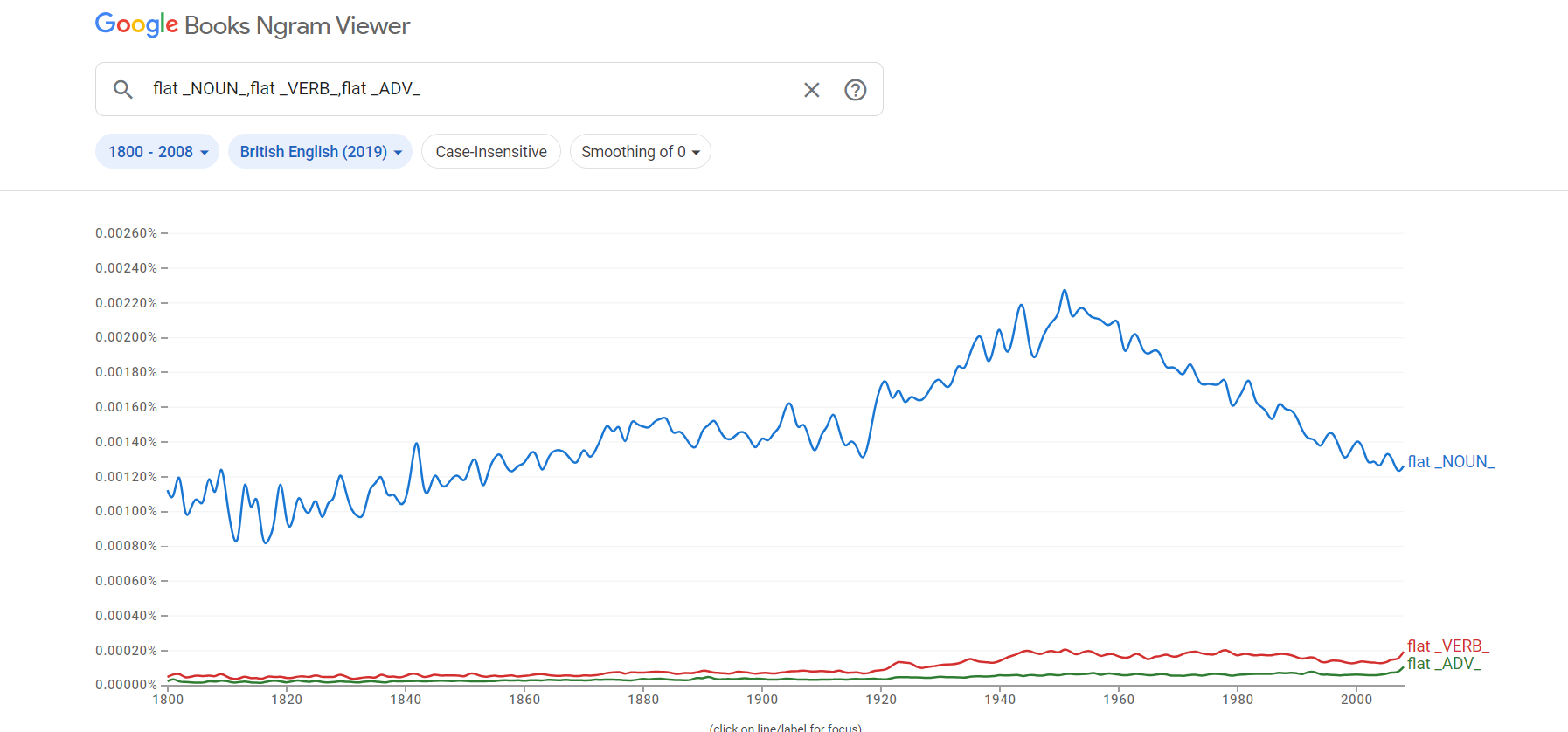


10.

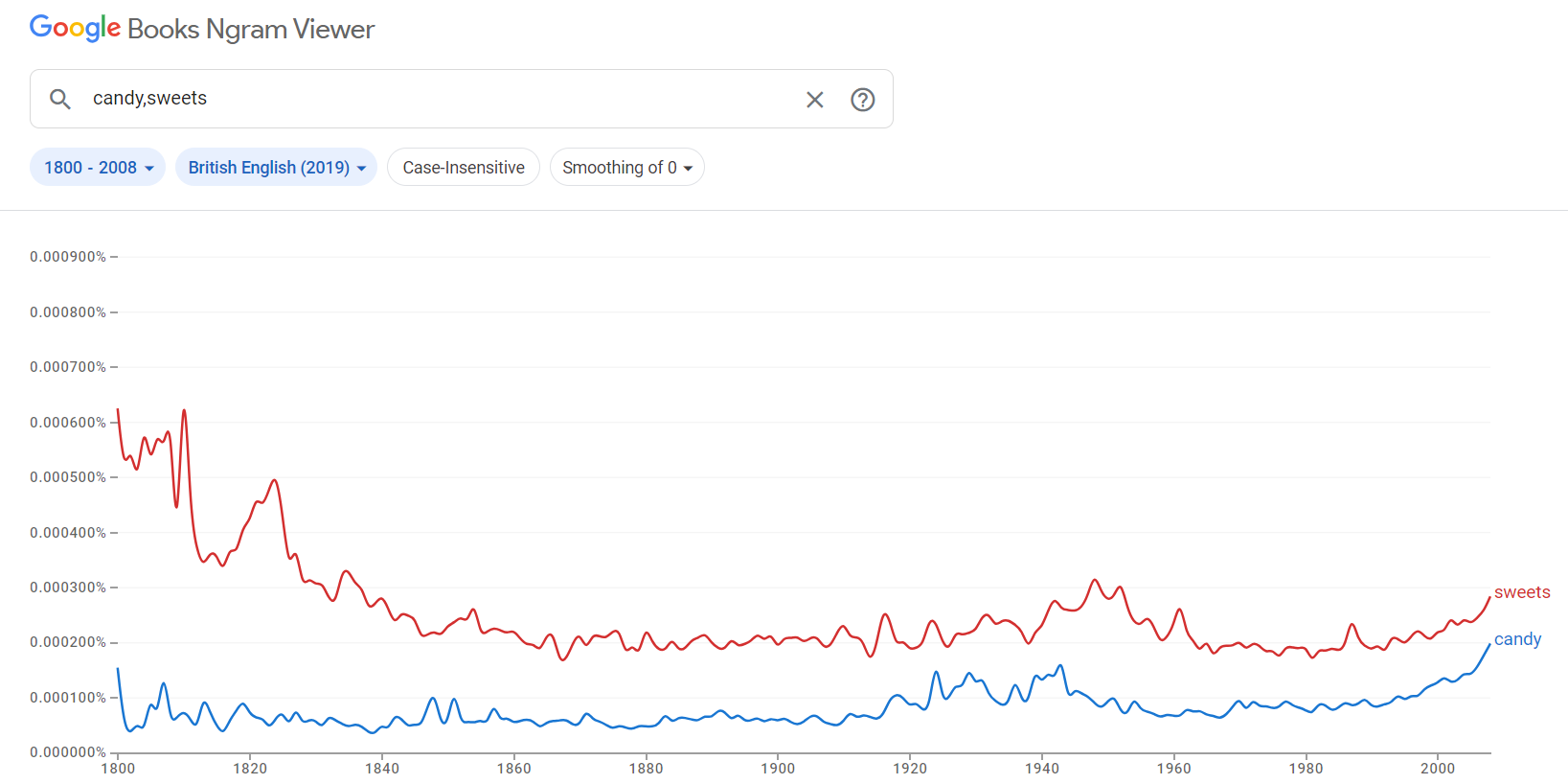


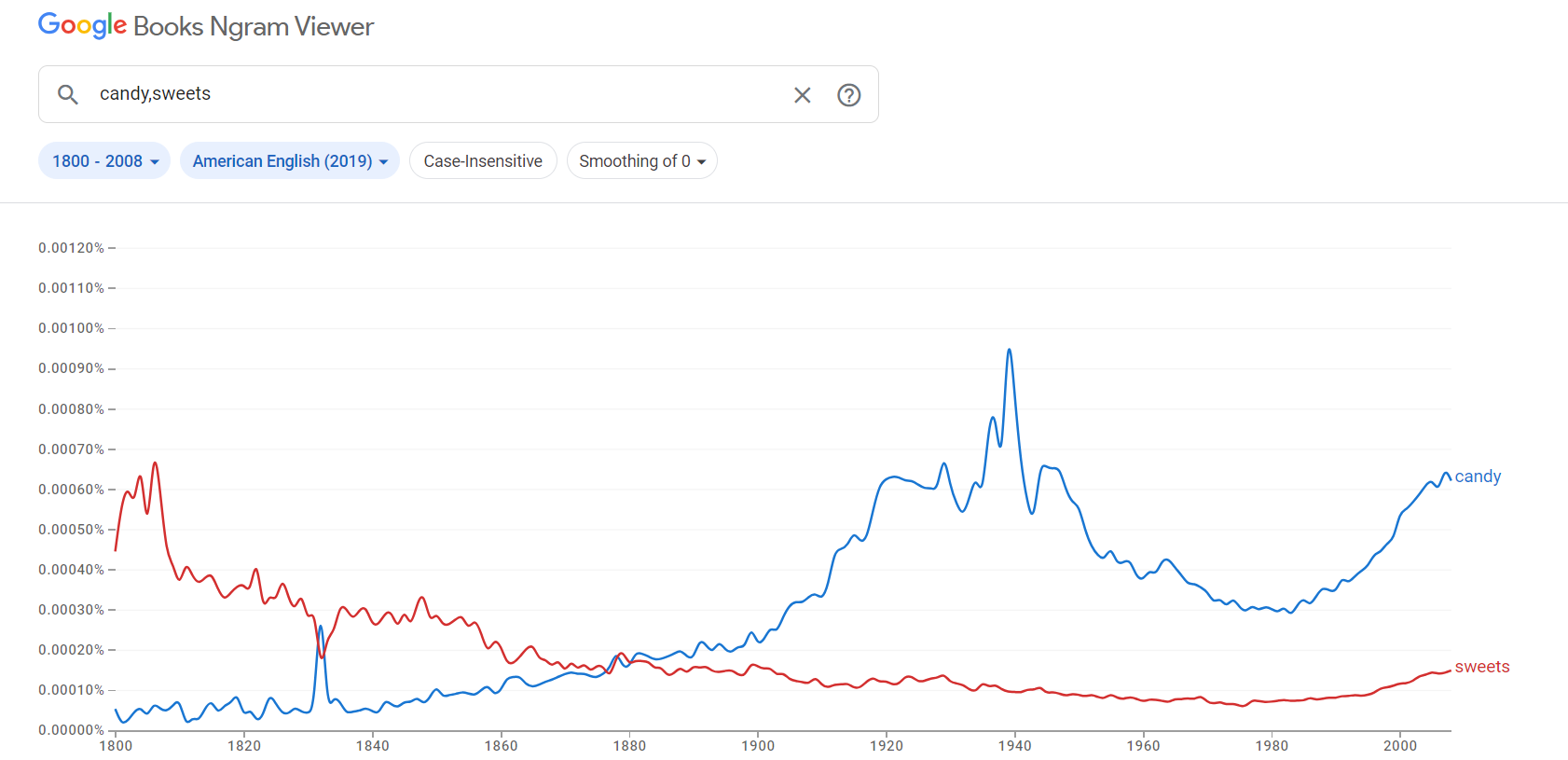


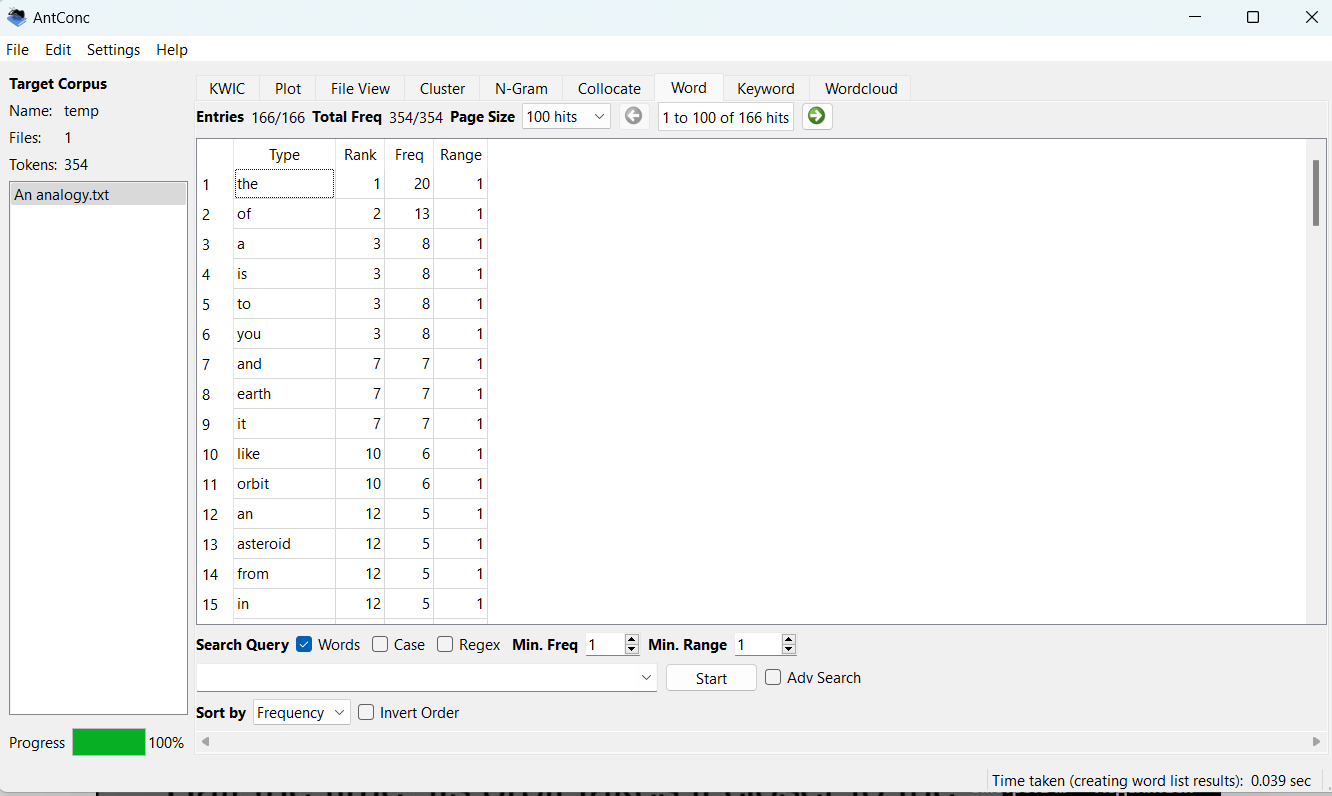
12. 

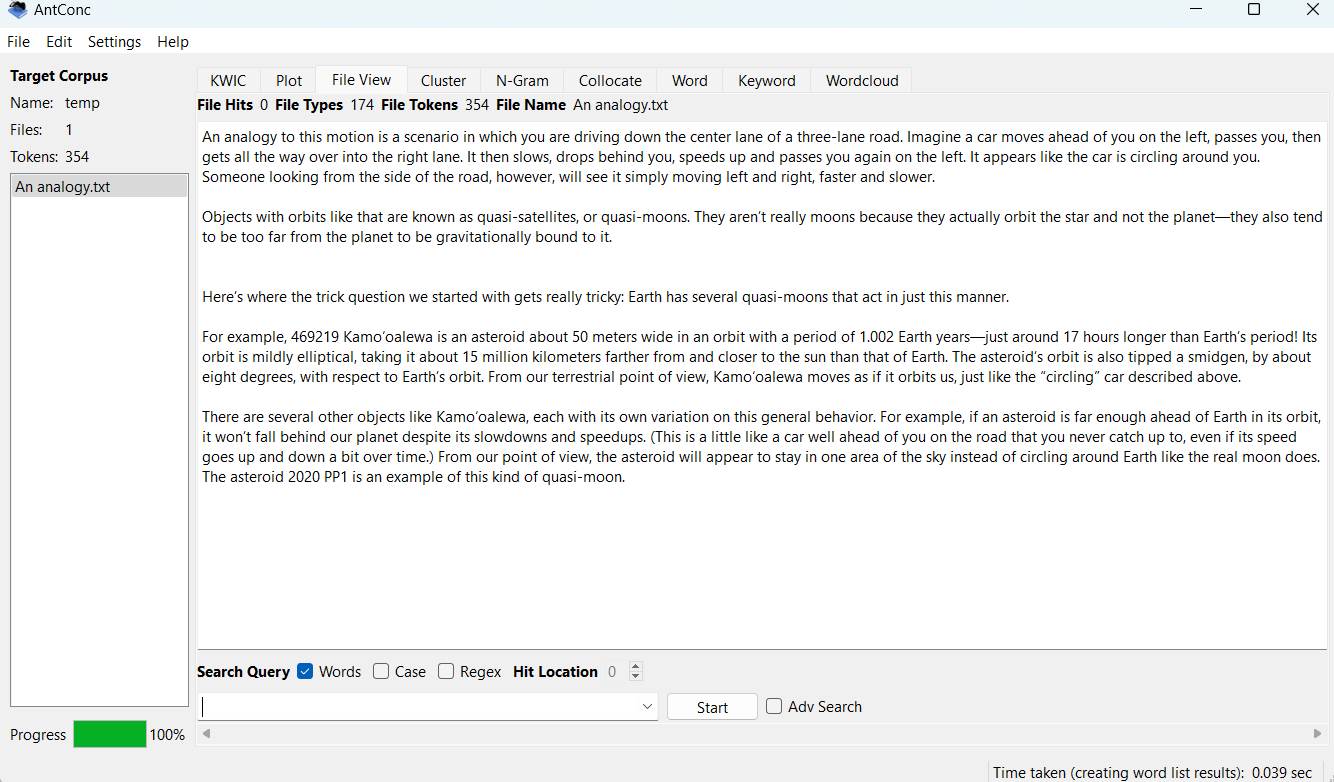


13.









3.Cosine Similarity (косинусная мера) - это метрика, используемая для измерения сходства между двумя векторами в многомерном пространстве. Она определяется как косинус угла между этими векторами и показывает, насколько они близки по направлению.

В лингвистике косинусная мера широко используется для решения различных задач, таких как:

1. \*\*Оценка семантической близости слов\*\*: путем представления слов в виде векторов (например, с использованием word embeddings), можно использовать косинусную меру для определения степени семантической близости между ними. Это помогает в поиске синонимов, антонимов или ассоциаций между словами.

2. \*\*Кластеризация текстов\*\*: косинусная мера может быть использована для кластеризации текстов или документов на основе их содержания. Она помогает определить, насколько похожи тексты друг на друга.

3. \*\*Рекомендательные системы\*\*: косинусная мера используется для ранжирования и рекомендации контента пользователям на основе их предпочтений и сходства с другими пользователями или элементами.

Примеры программ и алгоритмов, использующих косинусную меру:

1. \*\*Python библиотека scikit-learn\*\*: в scikit-learn есть функция cosine\_similarity, которая позволяет вычислять косинусное сходство между векторами.

2. \*\*Gensim\*\*: библиотека Gensim предоставляет инструменты для работы с word embeddings и вычисления семантической близости слов с помощью косинусной меры.

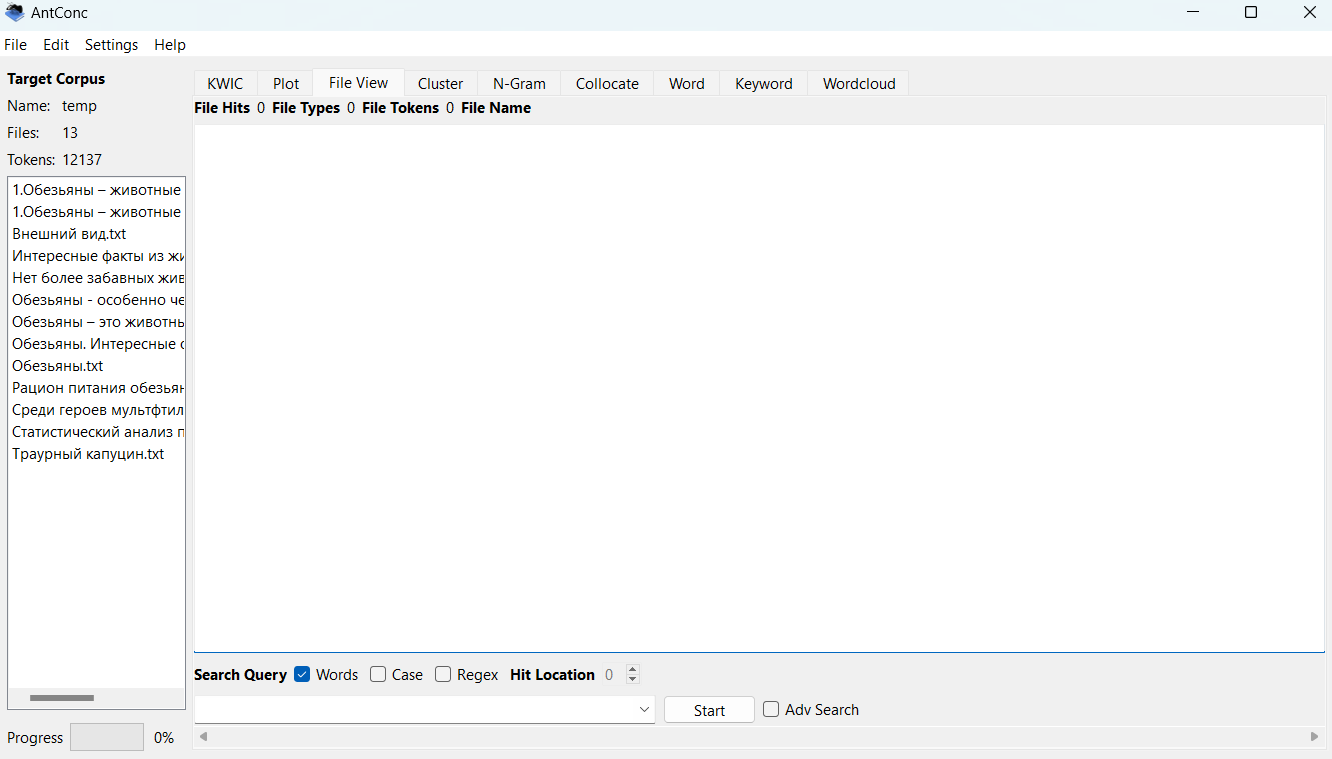
3. \*\*TensorFlow\*\*: TensorFlow также предоставляет возможности для работы с word embeddings и вычисления сходства между векторами с использованием косинусной меры.

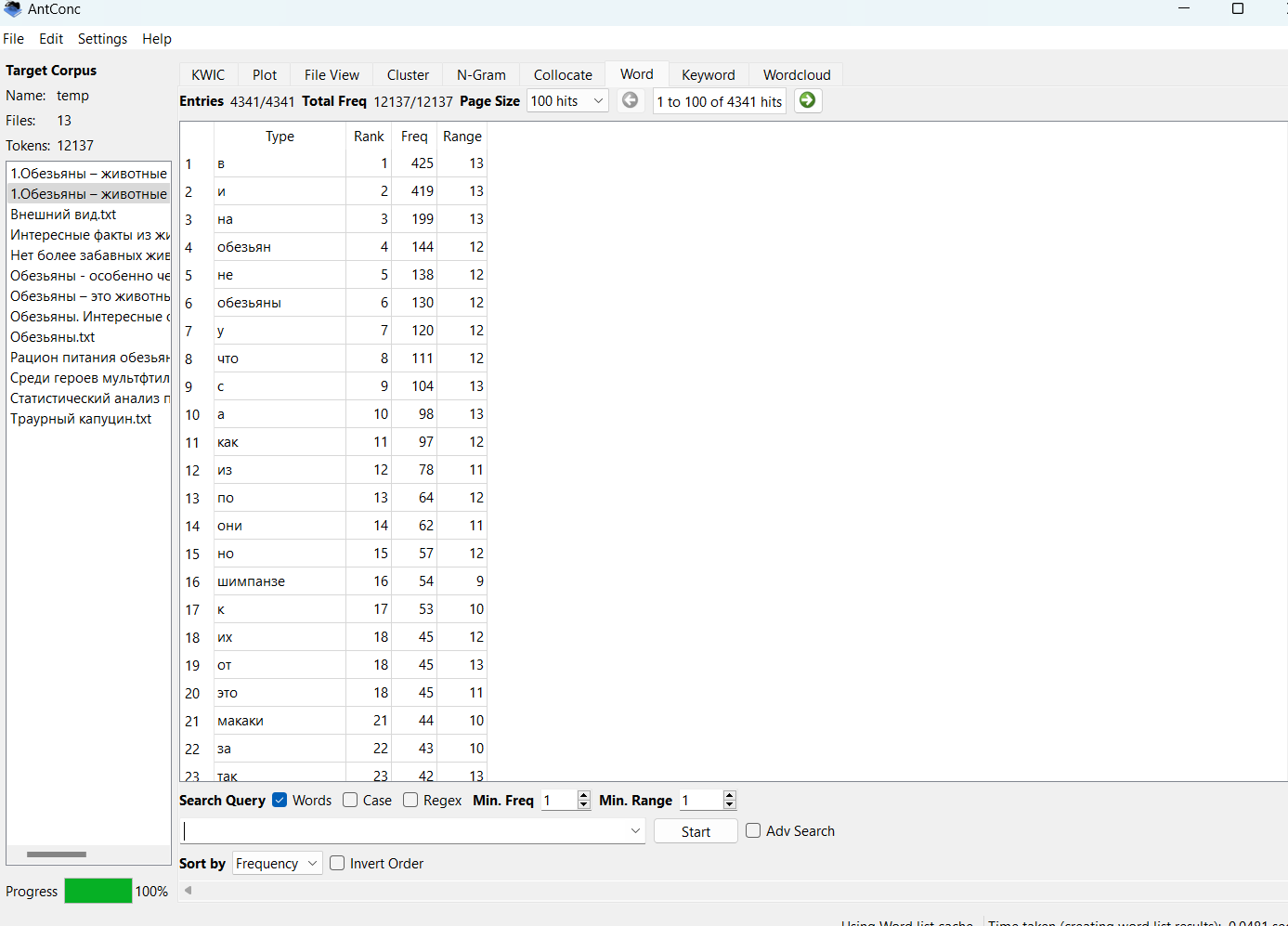
Цель исследования: выявить, какие слова чаще всего встречаются в научных

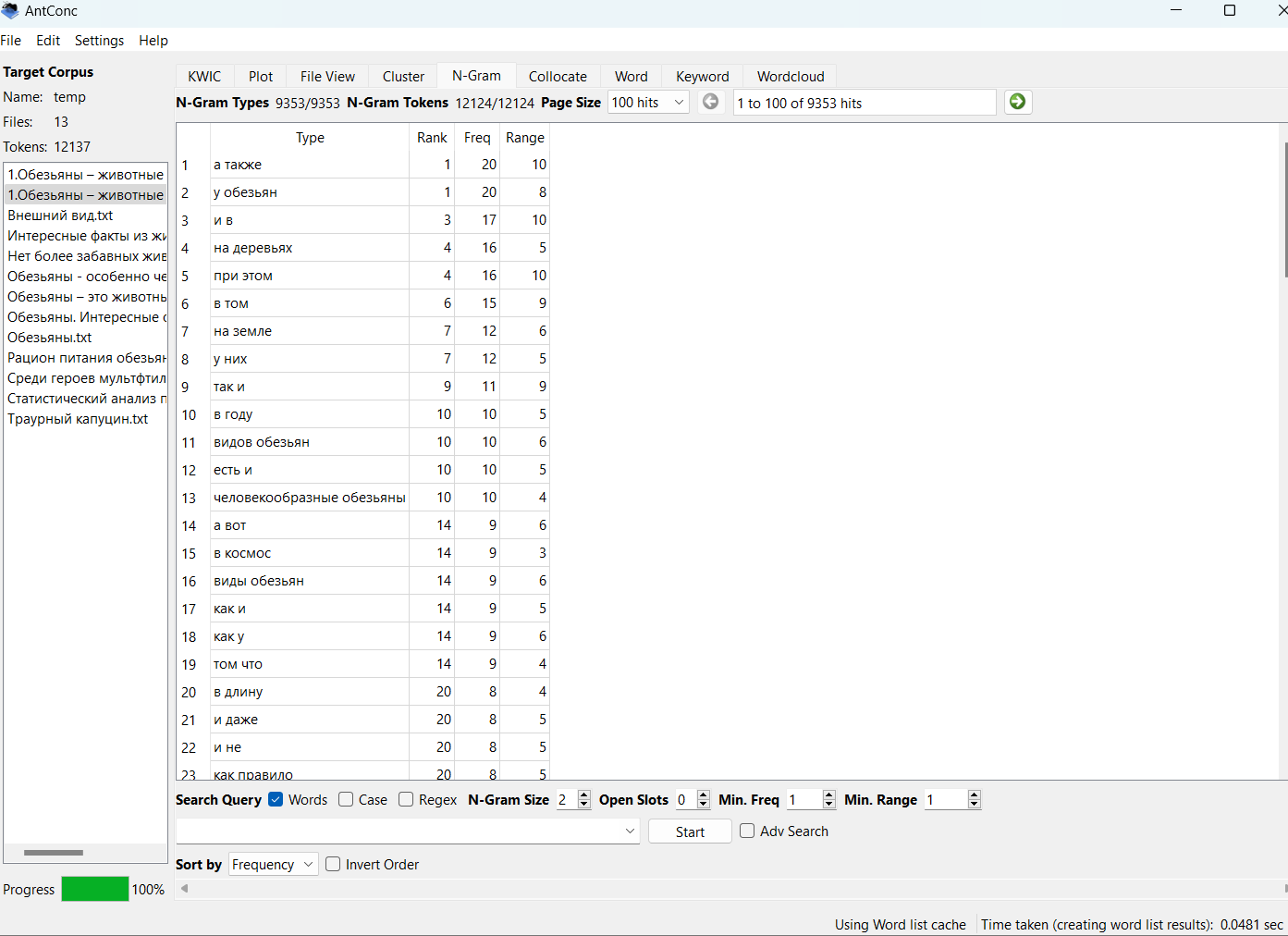
текстах про обезьян, с какими словами обычно употребляется слово обезьяна,

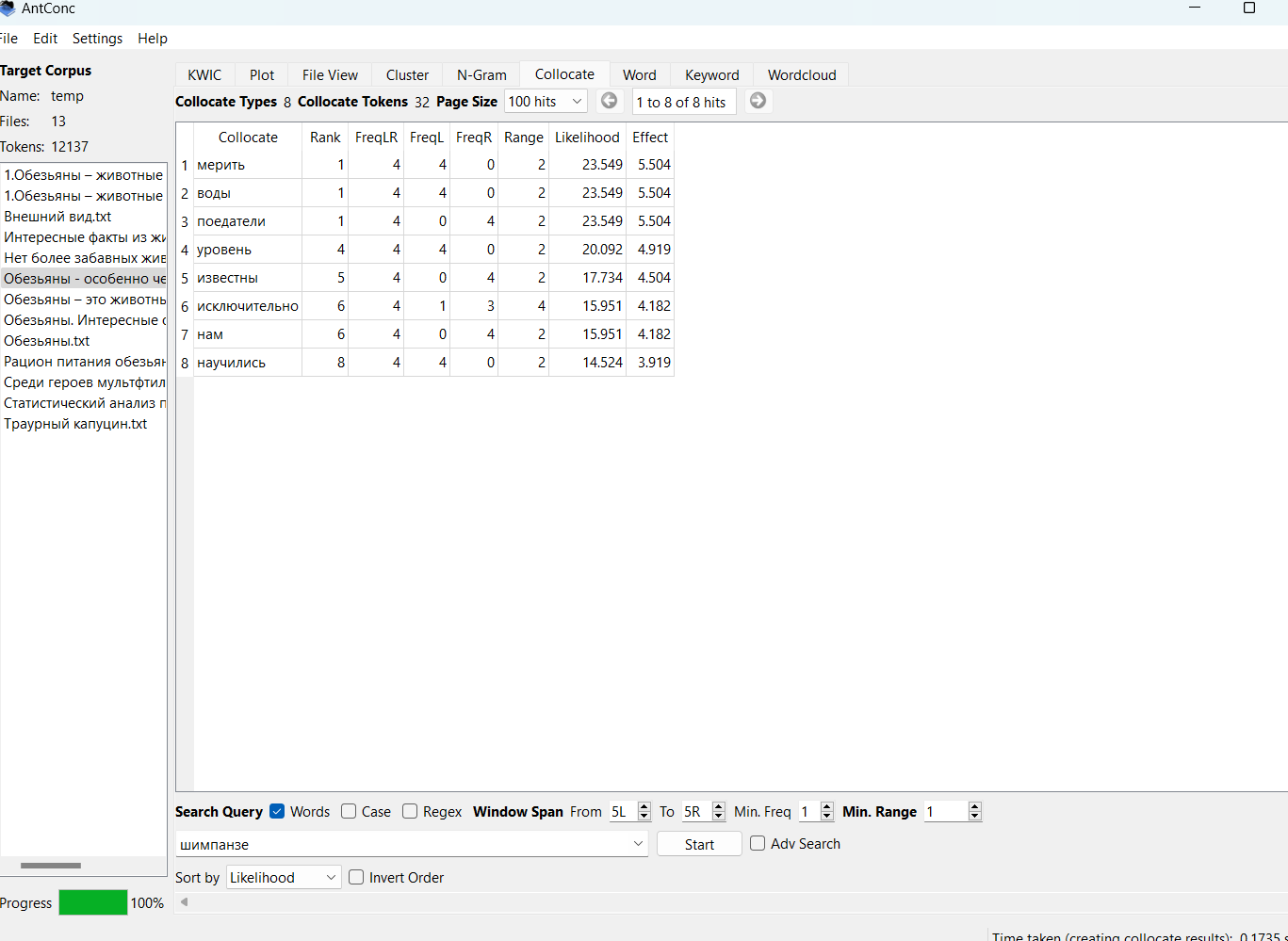
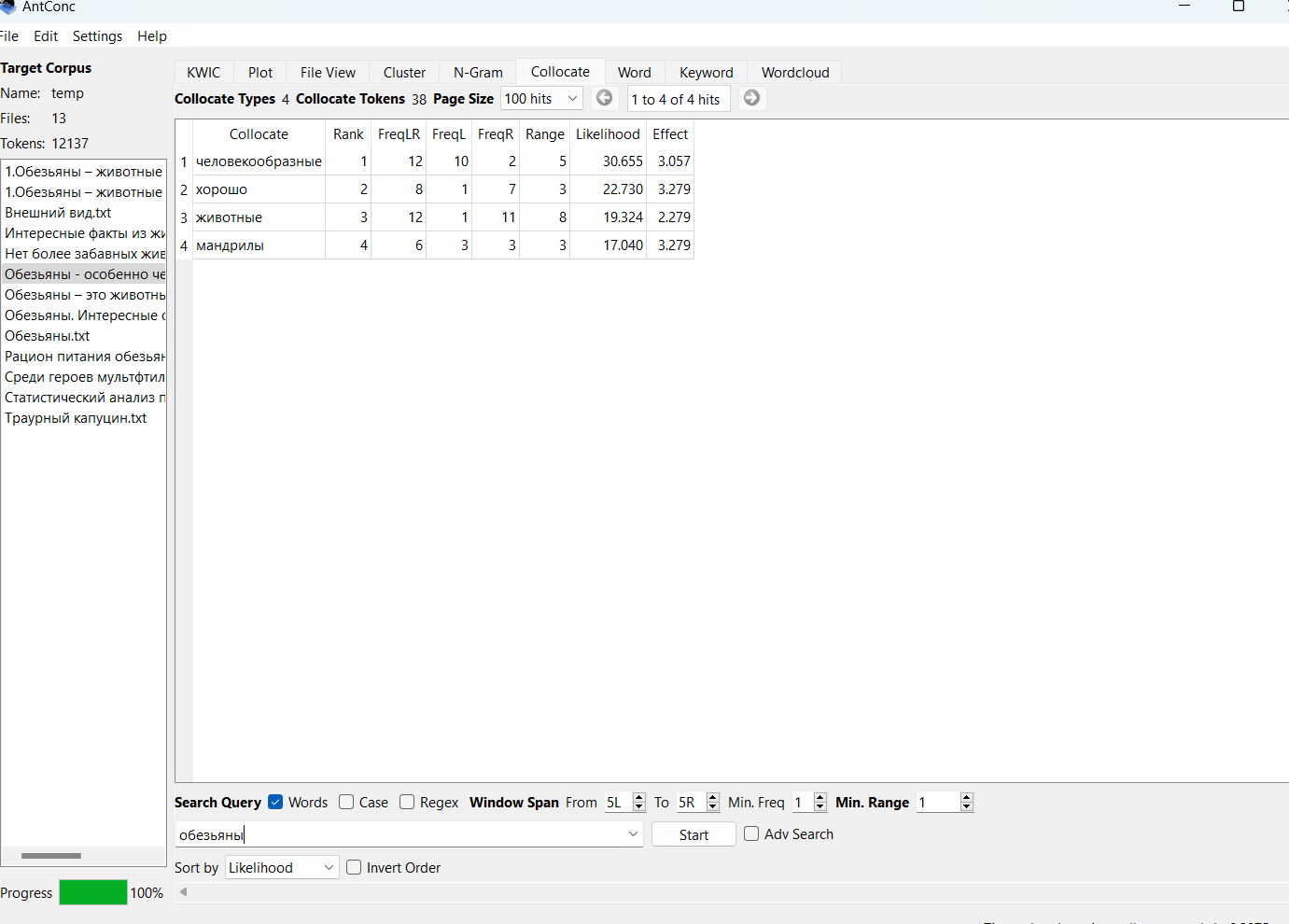
приматы, пищеварительная система и т.д., узнать значение научных терминов из

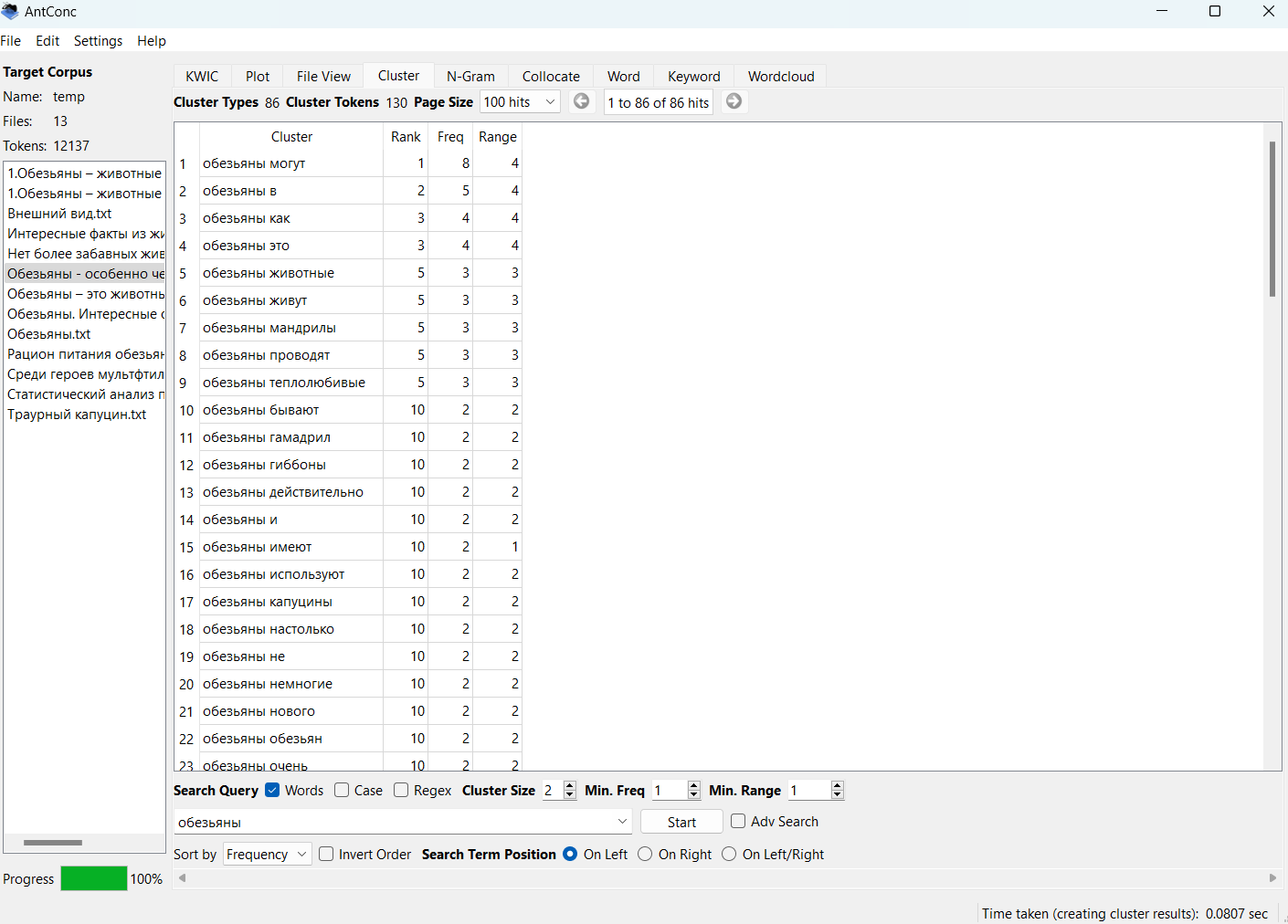
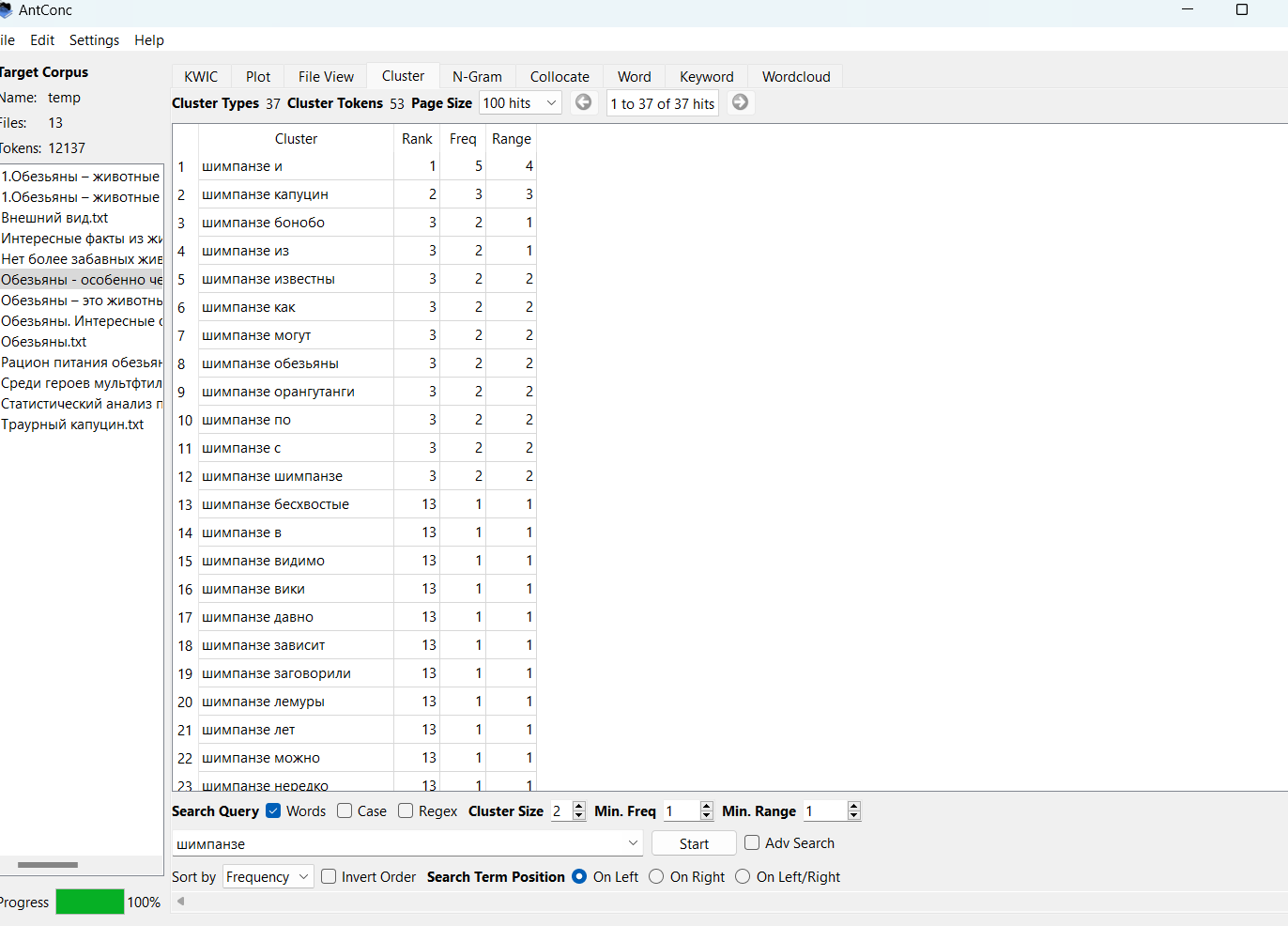
контекста.











Вывод: таким образом, сделаем вывод, что чаще всего в научных статьях

про обезьян употребляются

в 1 425

и 2 419

на 3 199

обезьян 4 144

не 5 138

обезьяны 6 130

у 7 120

что 8 111

с 9 104

а 10 98

как 11 97

из 12 78

по 13 64

они 14 62

но 15 57

шимпанзе 16 54

С помощью программы мы

выявили, с какими словами чаще употребляются выше

перечисленные, на основе этого можно сделать вывод: обезьяна-вид млекопитающих, приматы – отряд млекопитающих, и т.д. Проведенное исследование

может помочь в более глобальных лингвистических исследованиях, в

обучении языкам и других областях, в которых требуется работа с

корпусами.