Sherlocking the Sir Arthur Conan Doyle Corpus
An Exploration of Literary Gems

Khaleeqa Aasiyah Garrett

New York University Abu Dhabi

IM-1511 Introduction to Digital Arts and Humanities Spring 2023

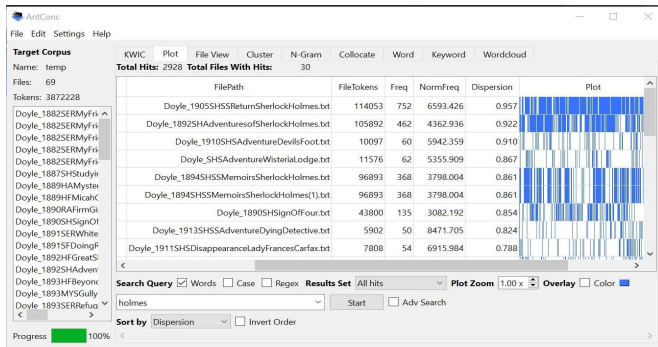جامعة نيويورك أبوظبي
NYU | ABU DHABI

Background

Exploring textual data from "The Adventures of Sherlock Holmes", a corpus comprising works by the renowned author, Sir Arthur Conan Doyle, dating back to the 1930s.

The corpus contains a large amount of textual data that is rich in themes, motifs, and stylistic features unique to the detective fiction genre.

By exploring and analyzing this corpus, we can gain valuable insights into the conventions of the detective fiction genre and how they are exemplified in Doyle's works

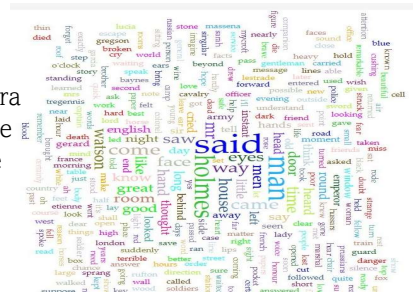Overflow of word "Holmes" in one book, shown via AntConc Plot

Method

Through textual analysis tools, we uncover the unique writing style, themes, and motifs of Doyle's works, and understand how they contribute to the detective fiction genre and its conventions.

Leveraging the power of Voyant tools and AntConc, a comprehensive analysis of the word frequency in all books with "Adventure" in the title within the corpus revealed intriguing observations about the key themes

Before: Word cloud using Voyant tools to showcase the impact of this one book (sup corpora)

After removing this sub corpora with the highest frequency, the word frequency of the entire corpus changed.

Conclusions

The assignment showcased how a small sub corpus can have a significant impact on an entire corpus. Through analyzing the frequency of "Holmes" in one adventure book, we discovered intriguing patterns that extended throughout the entire dataset.

Results after textual analysis and manipulation showing a balanced word frequency graph