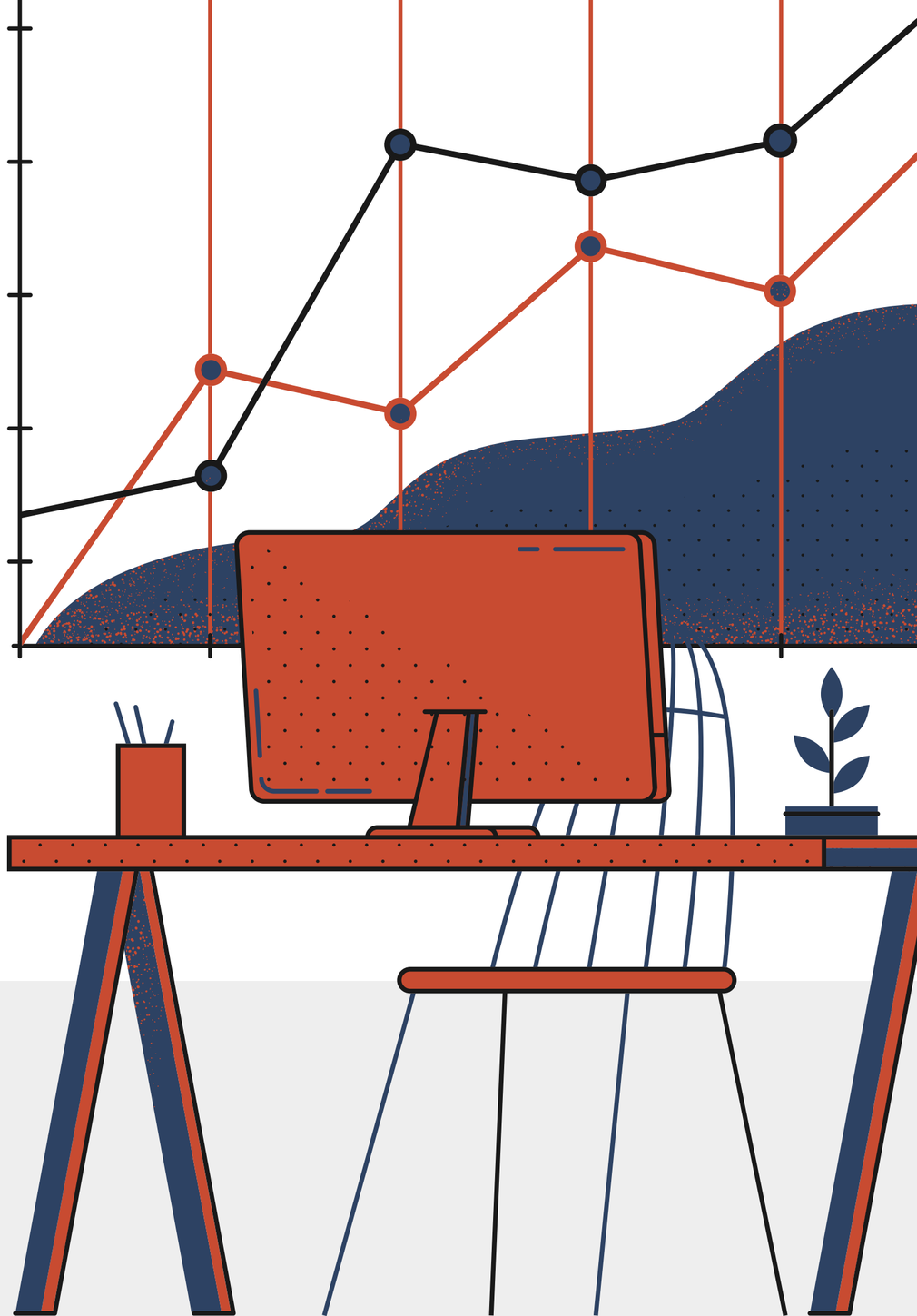
Thursday, 28

Rochester Institute of Technology

Computational Analysis of Gender Perception in the UK Before and After Women's Suffrage



PRESENTED BY

Shaista Syeda

FACULTY ADVISOR

Ashiqur R. KhudaBukhsh

I Progress

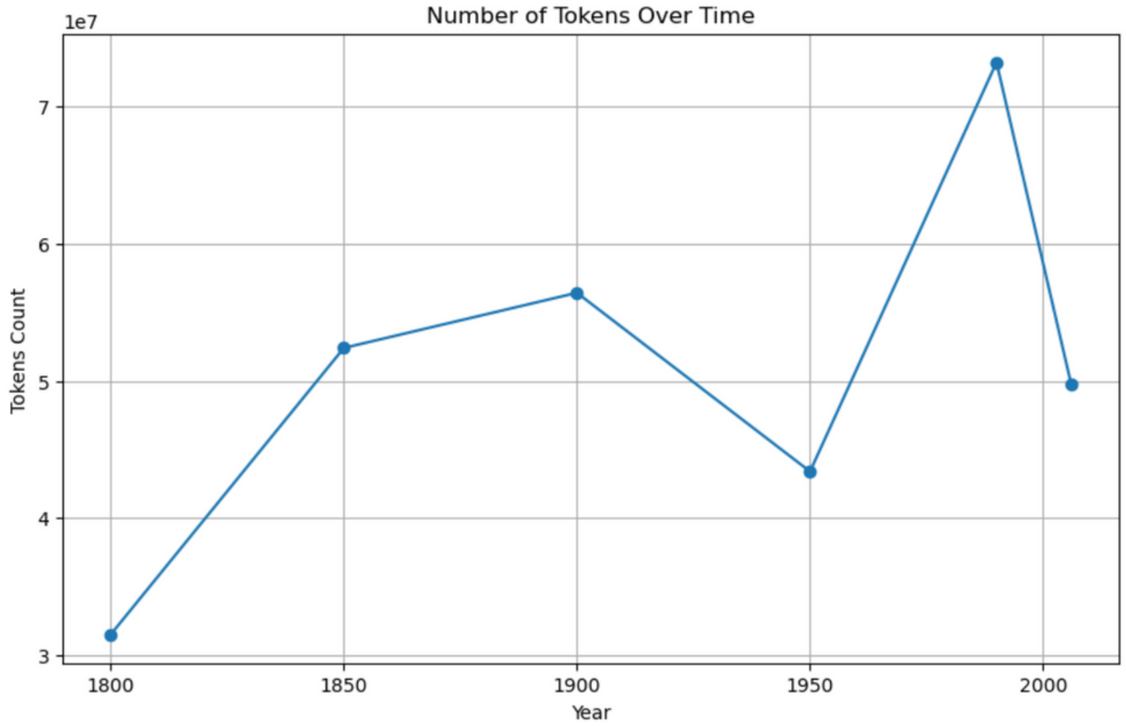
- I have performed Data Segmentation and I have a total of 6 files
- I have tokenized these files
- Performed Basic Statistical Analysis
- Performed Keyword Analysis
- Analysis of Male and Female Pronoun Count

Github Link - https://github.com/xssti/Analysis-of-gender-perception/tree/main/ADS-2

Statistical Analysis

```
Total Documents: 6
Total Sentences: 4360935
Total Words: 306700162
Average Word Count per Document: 51116693.67
Average Sentence Count per Document: 726822.50
```

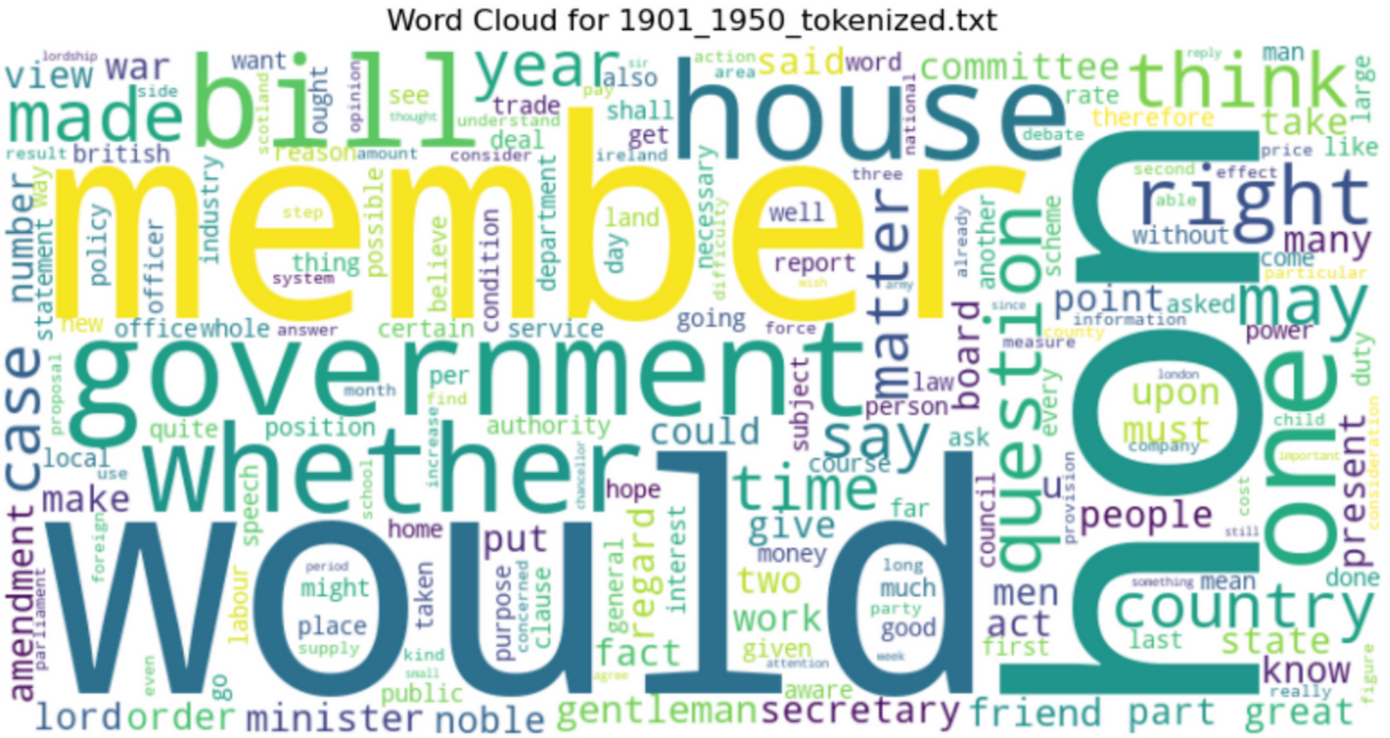
Number of Tokens



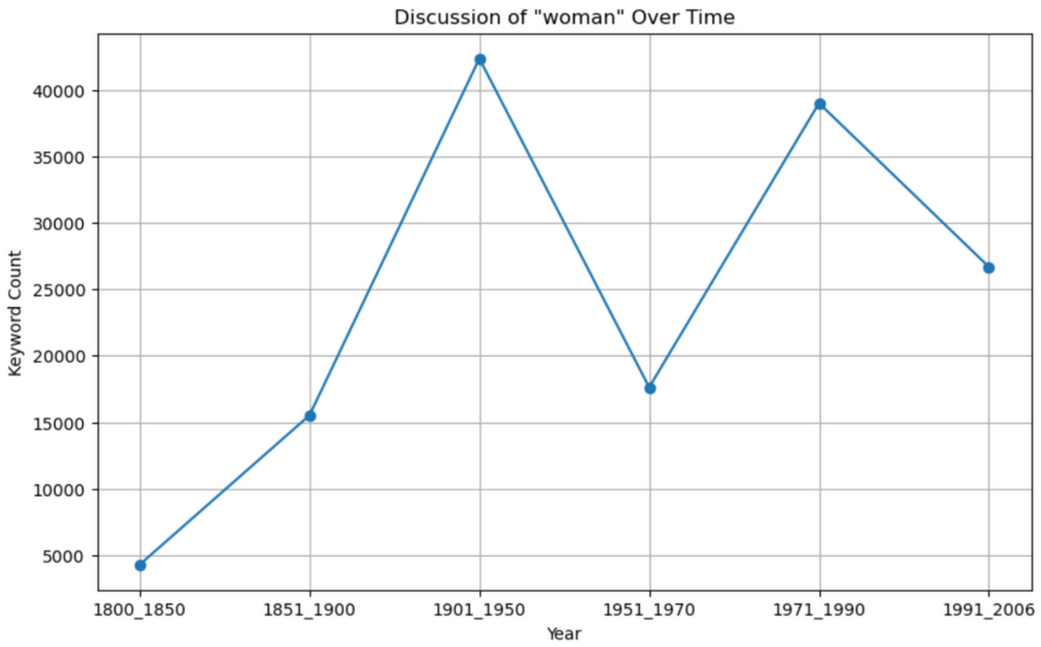
Top 10 frequent words

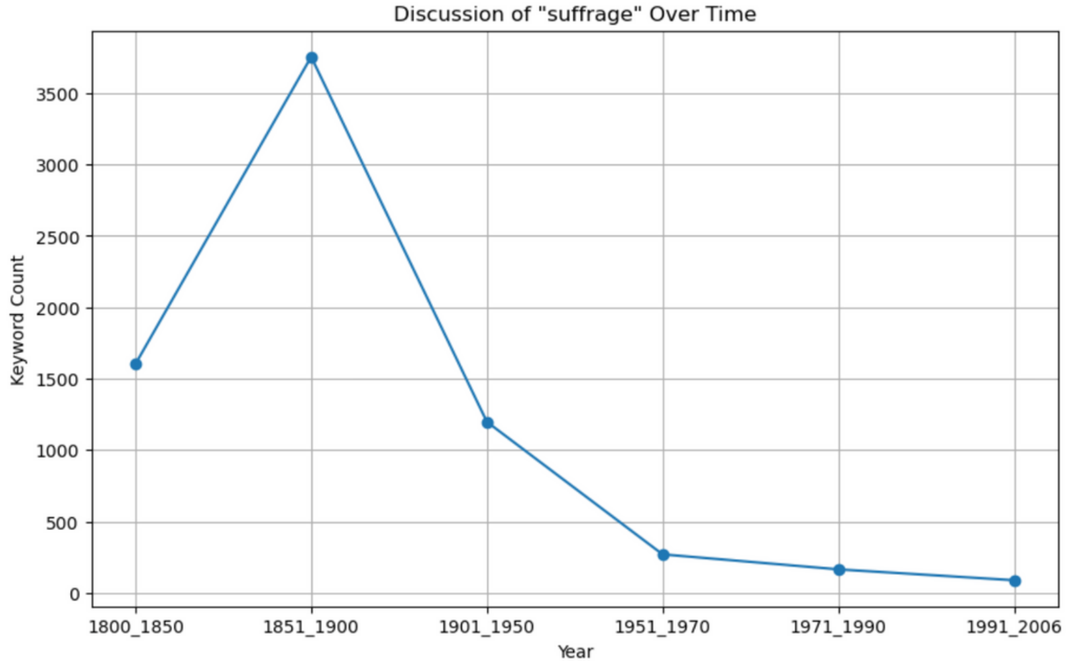
```
[('hon', 3154640), ('would', 3146208), ('government', 2064273), ('member', 2006418), ('house', 1744486), ('state',
1703251), ('right', 1580868), ('one', 1549606), ('year', 1511137), ('lord', 1378509)]
```

Word Cloud of most frequent words for the year 1901-1950

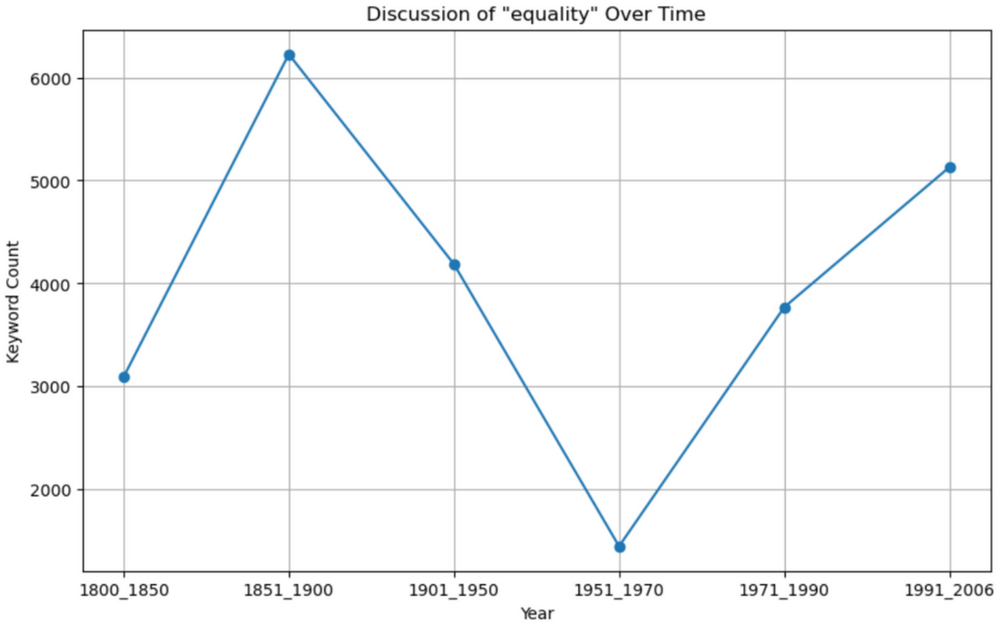


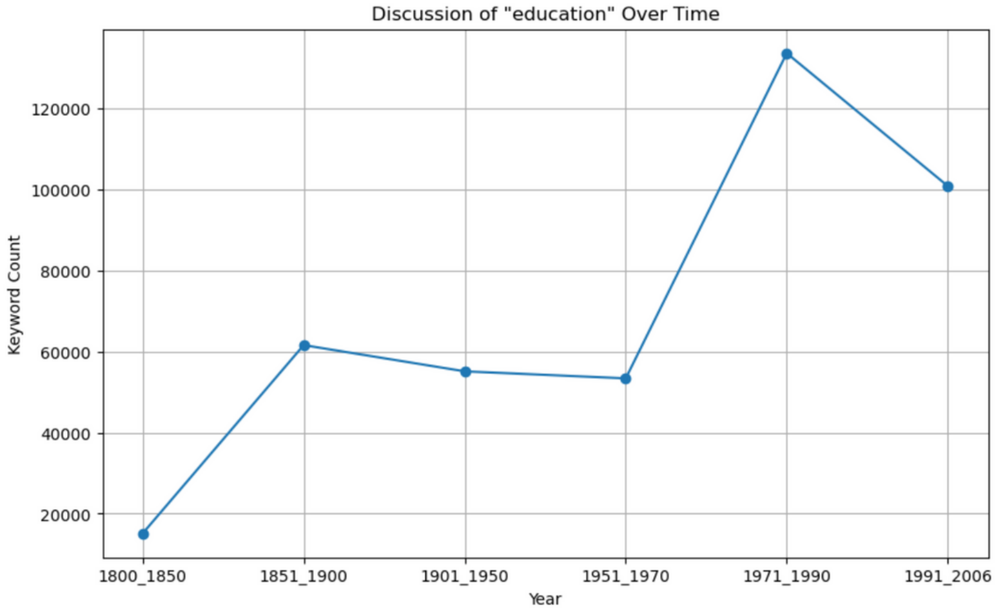
Count of Keywords related to women



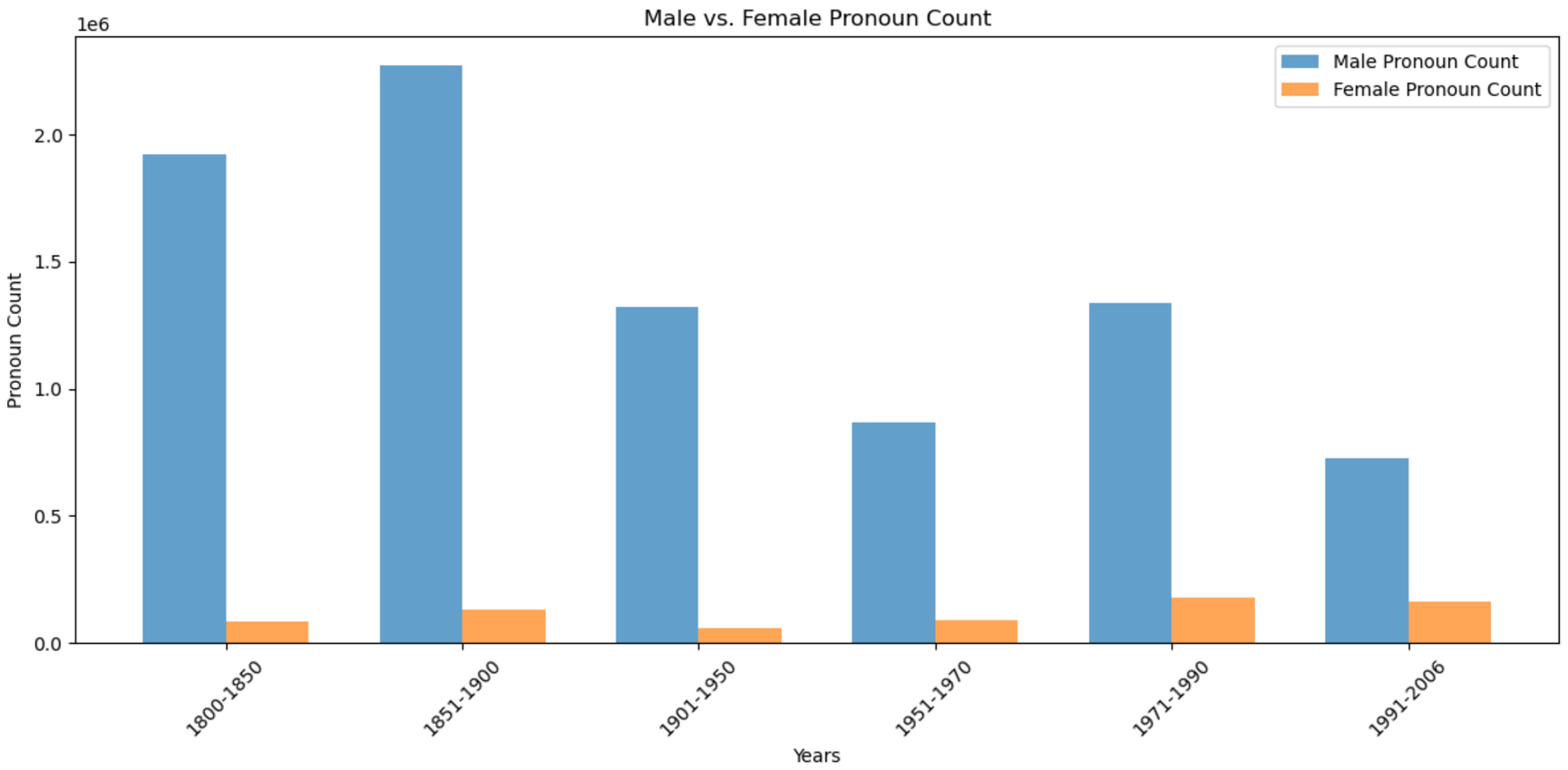


Count of Keywords related to women

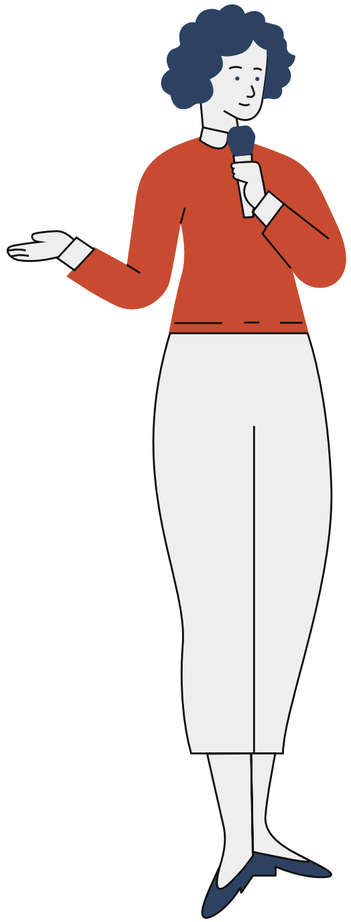




Male vs Female Pronoun Count



VII Work to be done



- Get Access to HPC Servers
- Get the Female and Male pronoun count for each year
- Word-embedding to find the words associated with Men and Women
- A Machine Translation-based analysis between 1850 - 1900 and 1950 - 2000 data
- Fine-tune BERT and GPT2 on the 1850 - 1900 and 1950 - 2000 data and analyze the differences

Thank you