

# PRONOUN WARS: GENDERED LANGUAGE SHOWDOWN IN MID-1800S LITERATURE

A Spark NLP Project

Alaina Holland and Jonathan McGechie

July 25, 2024



## OBJECTIVE/ METHODOLOGY

# 1800s



## Objective

**Aim:** Determine if there's a statistically significant difference in the frequencies of subject and object pronouns for feminine and masculine pronouns in historical texts.

**Explore:** Investigate whether the author's gender influences these results.

**Goal:** Shed light on gender bias and representation in mid-19th century literature.

### Importance

- **Reflects societal norms:** Understanding gendered language in literature reveals and reinforces societal biases.
- **Historical Insight:** Examining mid-1800s texts provides context for gender representation.
- **Contemporary Relevance:** Inform current discussions about gender bias in media and literature.

### Context:

- Understanding gender representation in mid-1800s literature.

## METHODOLOGY

### Books:

- **"Moby-Dick"** by Herman Melville
- **"Little Women"** and **"Little Men"** by Louisa May Alcott

### Technologies:

- **SparkNLP:** For advanced natural language processing.
- **PySpark:** For large-scale data processing.
- **GoogleColab:** For python interface.

### Steps:

- 1.Text Preprocessing:** Clean and tokenize text.
- 2.Part-of-Speech Tagging:** Identify personal pronouns and their grammatical roles.
- 3.Pronoun Frequency Analysis:** Calculate frequencies of subject (e.g., he, she) and object (e.g., him, her) pronouns.
- 4.Statistical Testing:** Perform chi-square tests to determine statistical significance.

### Hypothesis 1: Author's Gender Influences Pronoun Usage

- "Little Women" will have a higher frequency of feminine pronouns.
- "Moby-Dick" will have a higher frequency of masculine pronouns.

**Null Hypothesis (H0):** There is no significant difference in the frequency of masculine and feminine pronouns between the texts.

**Alternative Hypothesis (H1):** There is a significant difference in the frequency of masculine and feminine pronouns between the texts.

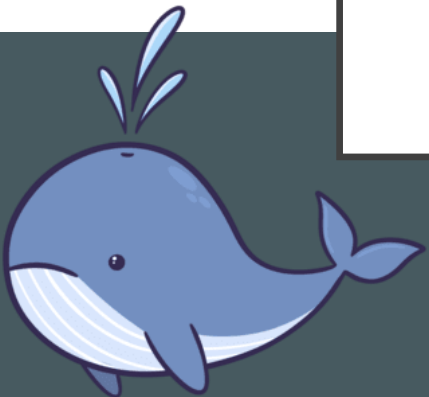
### Hypothesis 2: Gender Narration Perspective Impacts Pronoun Usage

- "Little Women" and "Little Men" will show different feminine pronoun usage patterns due to different narrative perspectives.

**Null Hypothesis (H0):** There is no significant difference in feminine pronoun frequency between "Little Women" and "Little Men".

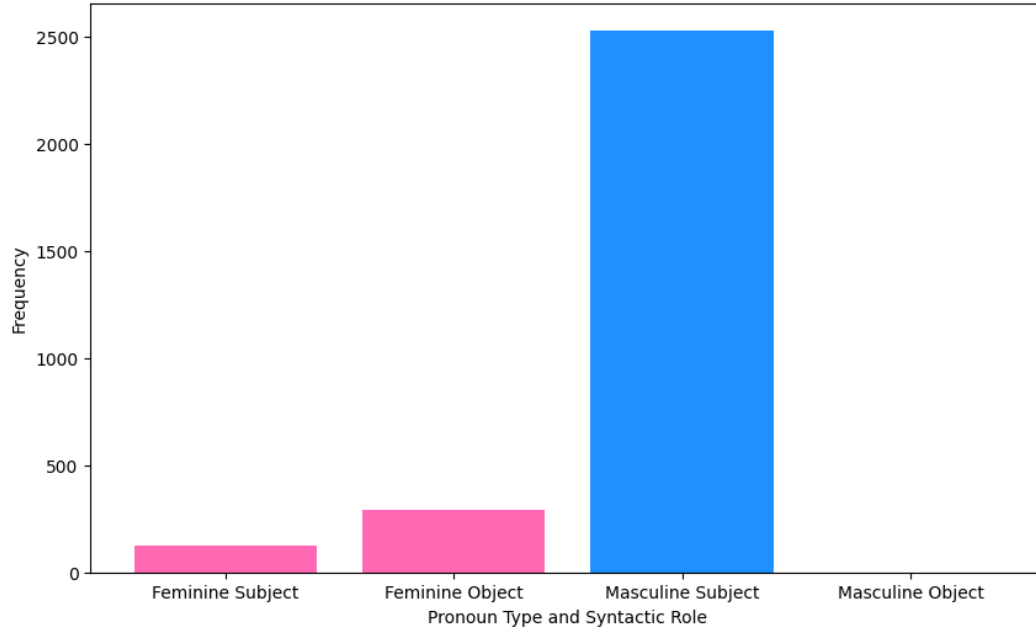
**Alternative Hypothesis (H1):** There is a significant difference in feminine pronoun frequency between "Little Women" and "Little Men".

HYPOTHESIS



# EDA

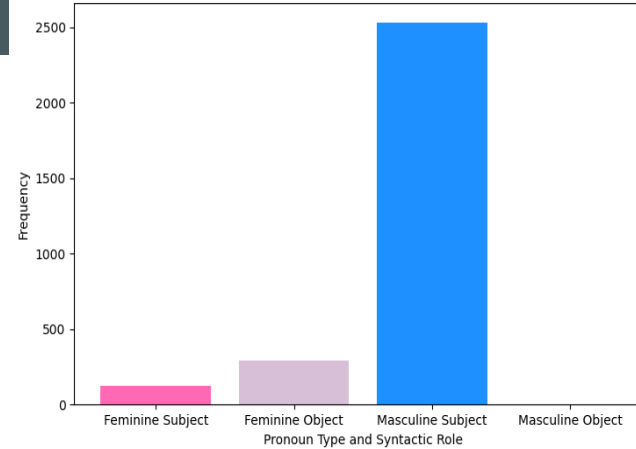
Pronoun Frequencies by Syntactic Role and Text



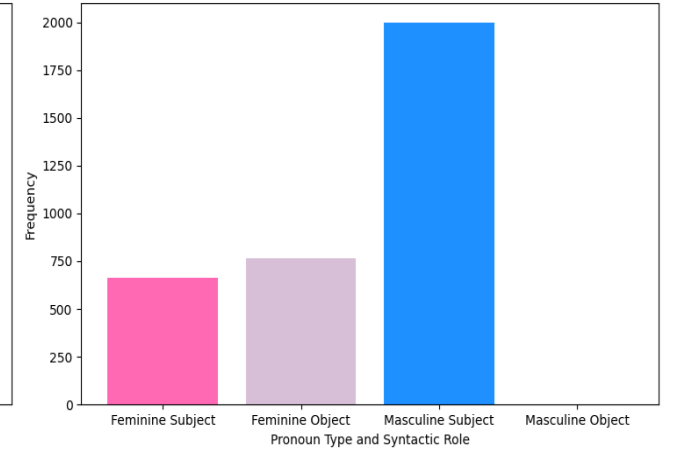
## Overall Pronoun Usage

- Chi2 Statistic: 1947.796335480638
- P-value: 0.0

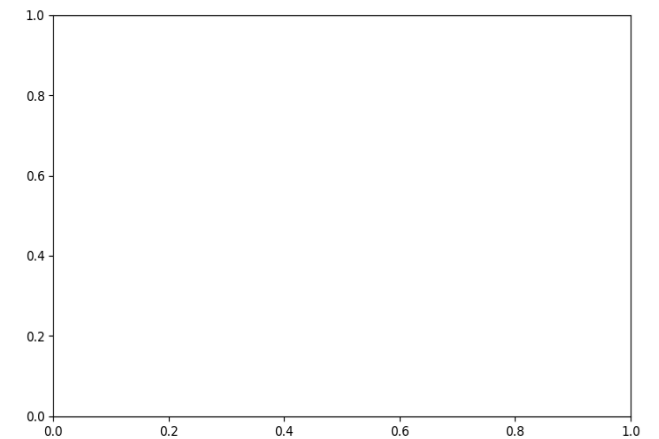
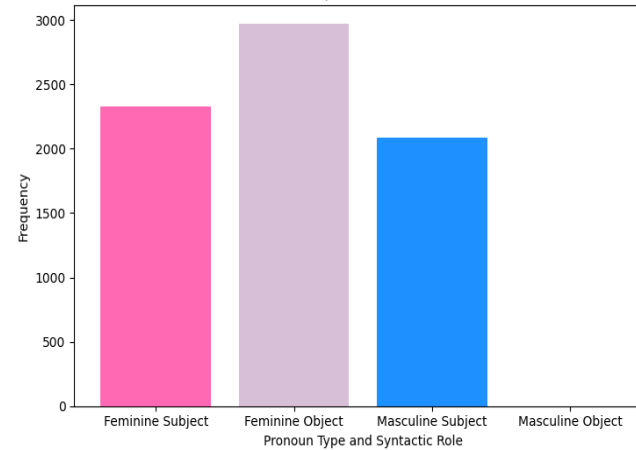
Pronoun Frequencies in "Moby-Dick"



Pronoun Frequencies in "Little Women"



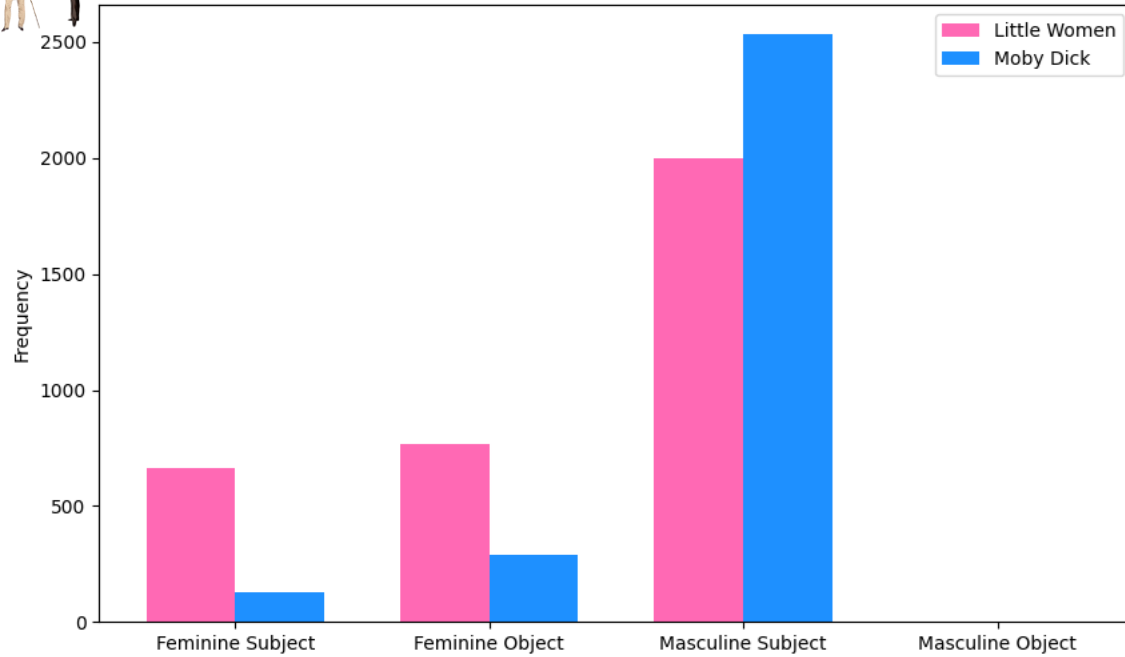
Pronoun Frequencies in "Little Men"



**Pronouns:** Subject (e.g., he, she) and Object (e.g., him, her)



Pronoun Frequencies in Little Women vs. Moby Dick



### Little Women vs. Moby-Dick

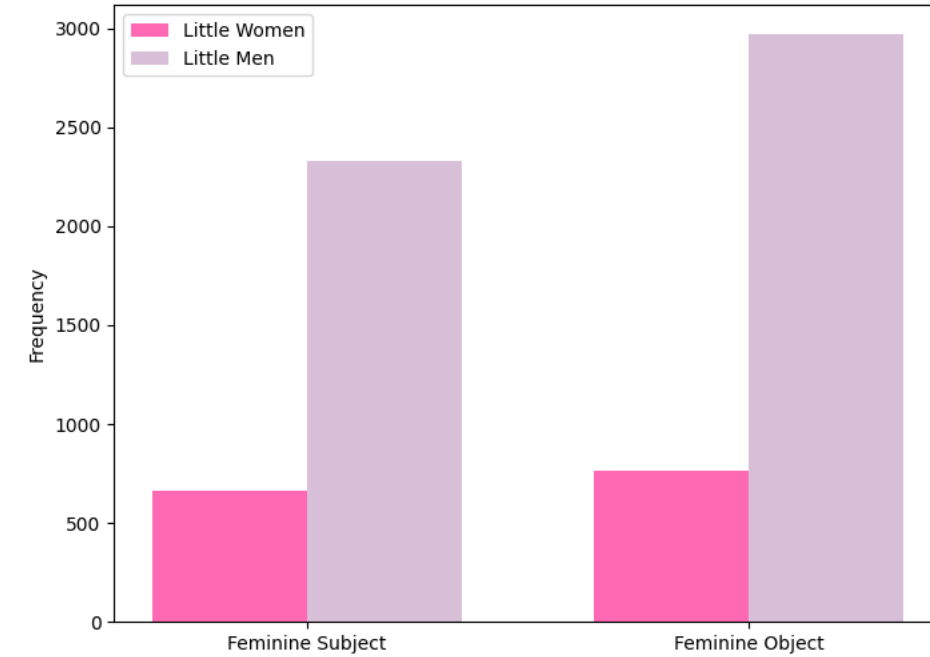
Overall Pronoun Usage:

Chi2 Statistic: 607.79

P-value: 1.0457e-132

**REJECT THE NULL**

Feminine Pronoun Frequencies in Little Women vs. Little Men



### Little Women vs. Little Men

Feminine Pronoun Usage:

Chi2 Statistic: 2.63

P-value: 0.1045

**FAIL TO REJECT THE NULL**

## EVALUATION

### Hypothesis 1: Author's Gender Influences Pronoun Usage

**Null Hypothesis (H0):** There is no significant difference in the frequency of masculine and feminine pronouns between the texts.

**Alternative Hypothesis (H1):** There is a significant difference in the frequency of masculine and feminine pronouns between the texts.

### Hypothesis 2: Gender Narration Perspective Impacts Pronoun Usage

**Null Hypothesis (H0):** There is no significant difference in feminine pronoun frequency between "Little Women" and "Little Men".

**Alternative Hypothesis (H1):** There is a significant difference in feminine pronoun frequency between "Little Women" and "Little Men".





# CONCLUSION

## Findings:

- Significant differences in pronoun usage were found between "Little Women," "Little Men," and "Moby-Dick."
- The absence of masculine object pronouns in all three works potentially reflects the narrative styles and societal norms of the time.
- "Moby-Dick" shows a male-centric narrative, while "Little Women" and "Little Men" highlight female experiences.

## Fallacy:

- Moby Dick:: 214,675 words
- Little Women: 163,750 words
- Little Men 54,750 words

## Implications:

- These differences underline how gender and societal norms influenced literature in the mid-1800s.

