

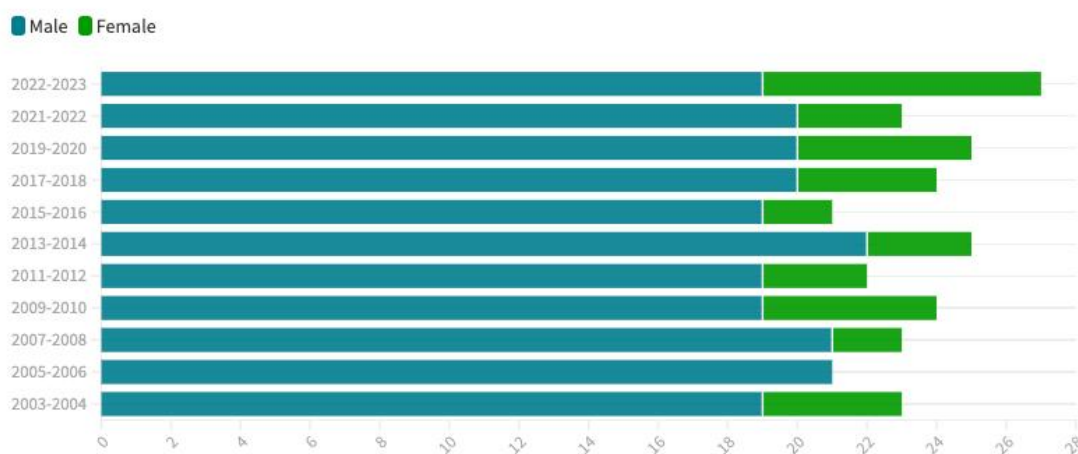
Gendered Narratives in Nobel Prize in Literature

Introduction

A 2015 study analyzing Wikipedia pages across various languages revealed a notable asymmetry: pages related to women frequently included descriptors such as “female,” “woman,” or “lady,” whereas similar terms like “male,” “man,” or “gentleman” were conspicuously absent from pages about men.

A parallel pattern can be observed on the Nobel Prize official website, where dedicated pages exist specifically for female laureates, but no equivalent categorization is made for male laureates. This distinction underscores the persistent framing of women as the “other” or the “second sex” within societal narratives.

Nobel Prize Winners in the Last Two Decades

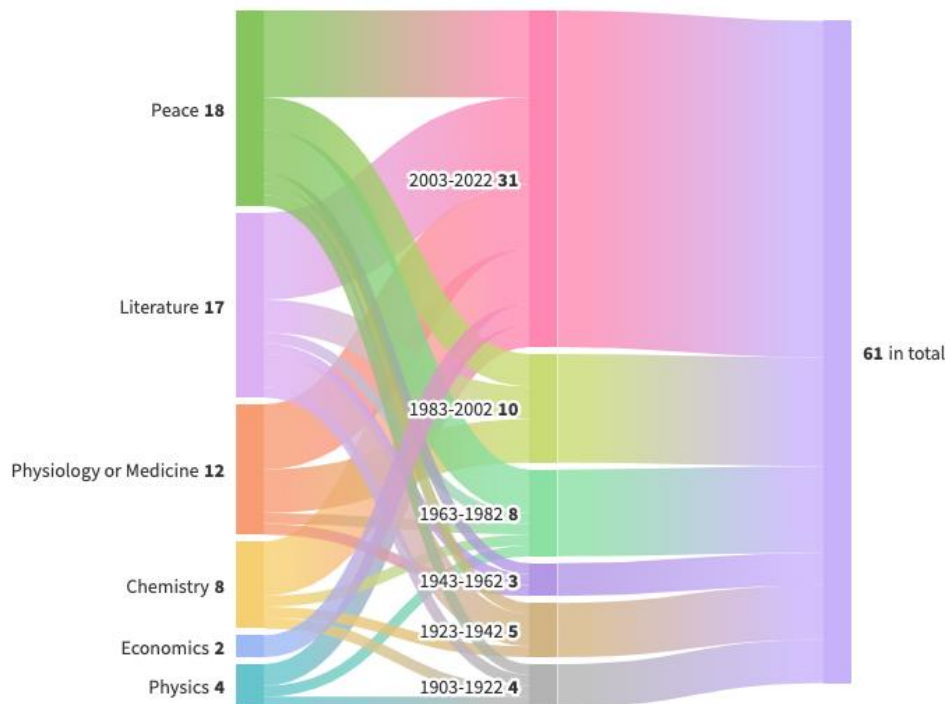


Data source: <https://www.nobelprize.org/prizes/lists/all-nobel-prizes/>

Tool: <https://app.flourish.studio>

Using 20-year intervals as analytical groupings, I employed Sankey diagrams to visualize the number of female Nobel laureates across different categories. The analysis reveals a clear upward trend in the number of women awarded Nobel Prizes over time. Among the categories, the Peace Prize has the highest representation of female laureates, followed by the Literature Prize.

Nobel Prize Awarded Women



Data source: <https://www.nobelprize.org/prizes/lists/nobel-prize-awarded-women/>

Tool: <https://app.flourish.studio>

The Nobel Prize in Literature is one of the most prestigious global recognitions of literary achievement. The prize motivation behind each laureate's award reveals the key factors that made them win. However, little attention has been paid to these motivation's linguistic and rhetorical characteristics, especially from a gendered perspective. This research aims to analyze and compare the prize motivation of female and male Nobel laureates in Literature. By examining word frequency, this research will discuss patterns and differences in how achievements are framed, celebrated, and certificated across genders.

Research Questions

What are the dominant themes expressed in prize motivation across genders?

How does the language used in prize motivation?

Are there significant differences in the prize motivation of female and male laureates? How do these differences reflect broader societal attitudes toward gender and recognition in literary achievements?

Data

This research uses primarily the full text of the Nobel Prize in Literature prize motivation from the official Nobel Prize website. Prize motivation is manually collected and organized into a structured dataset. While organizing my data, I found that 121 people have received the Nobel Prize in Literature from 1901 to 2024, and 117 Nobel Prizes in Literature have been given out since 1901. There were fewer Nobel Prizes given out during World Wars I and II. Seven times, in 1914, 1918, 1935, 1940, 1941, 1942, and 1943, it was not given out. Only four times, in 1904, 1917, 1966, and 1974, have two laureates shared the Nobel Prize in Literature. Within the other Nobel Prize categories, sharing the prize is a more frequent occurrence.

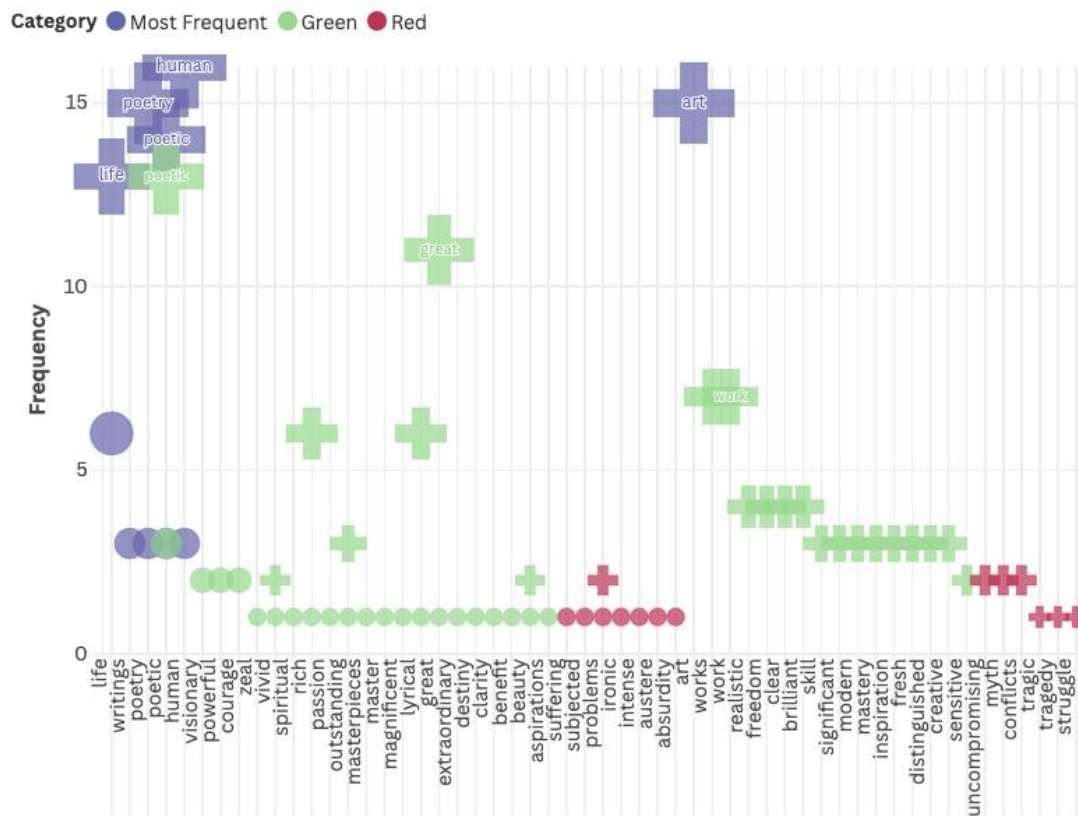
Supplementary information for each prize motivation is also used, including the laureate's name, gender, year of award.

Methods

After I've completed the steps of cleaning up the text and removing HTML tags, punctuation, and extraneous content, I imported the data related to female Nobel Prize in Literature laureates and male laureates into the Voyant Tools (<https://voyant-tools.org/>) separately. Voyant Tools comes with a feature that helped me with the following two steps: tokenization, which breaks the text into individual words or phrases for analysis; and standardization, which normalizes the text to lowercase and eliminates stop words to ensure consistency and focus on meaningful terms. Not only that, it also offers visualization. Cirrus provides a wordcloud view of the most frequently occurring words in the corpus, providing a convenient (though reductive) overview of the content. Collocates graph shows a network graph of higher frequency terms that appear in proximity. Keywords are shown in blue and collocates (words in proximity) are showing in orange. The Cirrus and Links results obtained are as follows:

words)	(1); lyrical (1); great (1); extraordinary (1); destiny (1); clarity (1); benefit (1); beauty (1); aspirations (1).	(3); vivid (2); sensitive (2); profoundly (2); originality (2); monumental (2); love (2); keen (2); ingenuity (2); important (2); impassioned (2); humour (2); ethical (2); contribution (2); consummate (2); compassion (2); beauty (2); achievement (2); won (1); warmth (1); visionary (1); versatility (1); versatile (1); talent (1); stimulating (1); stimulate (1); spiritual (1); scenic (1); richly (1); revival (1); renowned (1); renewed (1); remarkable (1); profound (1); productive (1); powerful (1); perfection (1); noble (1); miracles (1); masterpieces (1); master (1); magnificent (1); leading (1); intelligence (1); innovative (1); ingenious (1); influential (1); imaginative (1); illustrious (1); illuminate (1); honesty (1); happy (1); greatest (1); grace (1); fruitful(1); fearless (1); fantastic (1); fancy (1); famous (1); faithfully (1); exquisite (1); exalted (1); exalt (1); evocative (1); eminent (1); ecstasy (1); earnestness (1); earnest (1); distinctive (1); dignity (1); destiny (1); defeat (1); comprehensive (1); compassionate (1); commitment (1); classic (1); clarity (1); beautiful (1); authentic (1); appeal (1); achievements (1).
Red vocabulary (negative words)	suffering (1); subjected (1); problems (1); ironic (1); intense (1); austere (1); absurdity (1).	uncompromising (2); problems (2); myth (2); conflicts (2); tragic (1); tragedy (1); struggle (1); strict (1); severe (1); revolt (1); resistance (1); predicament (1); prattle (1); outsider (1); mysterious (1); irony (1); illusory (1); illusions (1); fragile (1); disconcerting (1); destitution (1); critical (1); conflict (1); clash (1); challenging (1); bitter (1); barbaric (1); ambiguous (1); abyss (1).

Using the above data, I created a csv file with Word, Frequency, Category (Green, Red), and Gender (Female, Male) columns. Then I uploaded the data to the Flourish studio and created a scatter plot. Each point can represent a word, size for frequency, color for sentiment category, and shape for gender (circle for female, cross for male).



Analysis

Butler (2006) first proposed the idea of gender as performance in her groundbreaking work. Gender is not a natural trait. Instead, it is produced and re-created by connection and activity. Building on Butler's work, Eckert and McConnell-Ginet (2013) put forward that gender is asymmetrical, collaborative, performative, and learned. These frameworks aid in the explanation of how gendered discourses are sustained in the institutional and cultural context. Examples of "practices and structures that are lived out in society from day to day" include the gender inequity associated with specific professions (Baker 2006). Discourses are formed by these structures and practices. Although language and discourse are separate concepts, language analysis can "uncover traces of discourses," according to Baker (2006). Unsurprisingly, when societal problems and relationships of power are

the focus, discourse analysis—more in particular, critical discourse analysis—performs this type of analysis (Mautner 2009). Finding and measuring these power disparities is not an easy task, though. Critical discourse analysis (CDA) and corpus linguistics provide practical tools for uncovering such asymmetries in language. While CDA effectively uncovers gendered discourses, corpus linguistics complements it by analyzing larger datasets to identify systematic patterns, such as the emotional versus neutral descriptors.

Word frequency helped me identify gender-specific thematic priorities, generate word clouds and examine the relative frequency of important terms. The following key points can be summarized:

Women laureates are portrayed in a more emotional and idealized way. Green vocabulary (positive words) for female laureates, such as “poetic,” “visionary,” “courage,” and “beauty” appear frequently, reflecting the often idealized, spiritual, and aesthetic nature of evaluations of their achievements. While red vocabulary (negative words) such as “suffering,” “subjected,” and “absurdity” suggest that the work of female laureates is more likely to deal with personal or social pain and suffering, which may be related to the domestic and emotional expression of women’s literary works.

In contrast to the 174 unique words that appeared in the female laureates’ corpus, 685 unique words appeared in the male laureates’ corpus. Male laureates’ descriptions are doubtlessly broader and more diverse. The green vocabulary of male laureates, such as “great,” “significant,” “brilliant,” and “mastery,” emphasizes the skill, breadth, and authority of their accomplishments, and this lexical diversity suggests that male laureates may be evaluated more comprehensively. Red vocabulary such as “conflicts,” “tragedy,” and “revolt” suggest that male works tend to explore deep philosophical issues and conflict scenarios, indicating a more grandiose narrative features.

Beyond the differences, the commonalities remain remarkable. The high frequency of “life,” “poetic,” and “human” in both groups suggests that regardless of gender, the laureates are generally recognized as showing poetic expression and humanistic concern in literature.

The above inference may lead to further questions, such as whether the female laureates are more concerned with social injustice or personal experiences, while the male laureates are more concerned with universal philosophy or artistic skills; whether descriptions are more genre-oriented (e.g.,

“poetry,” “fiction,” “theater”) or contribution-oriented (e.g., “influence,” “achievements”), and whether these vary by gender; whether women’s contributions were more in terms of personal qualities (e.g., “courage,” “passion”) and men’s contributions were more in terms of the impact or universality of the work. If there is an opportunity at a later date, these questions will probably be explored one by one.

Potential Bias and Limitation

Since 18 women have been awarded the Nobel Prize in Literature, female laureates in Literature constitute a small sample size. I envisioned that cross-temporal trends could be analyzed, such as whether descriptions of female laureates have evolved toward more neutral or professional language over the decades, and whether there has been a shift in the emotional language historically used for women towards a more achievement-oriented portrayal. I also conceived of analyzing regional variation in language use to examine whether laureates are motivated differently by different cultural or geographical backgrounds, which might be intertwined with gender-based bias. And to further explore whether non-Western female laureates are portrayed differently than Western female or male counterparts. But making generalizations from a small sample is challenging, and bias may be more likely to occur. At the same time, it has to be taken into account that as prize motivation is all about articulating the reasons for awarding prizes, it is also difficult to assess each motivation’s emotional tone by assigning sentiment polarity and intensity. Based on the limitations of the number of samples in the corpus, the inferences I have proposed at this stage in the Analysis section may be slightly hasty and arbitrary. After further study, I will continue to compare findings between female and male laureates using statistical methods to determine differences in language use.

As far as the ethical aspect is concerned, I don’t think there are relevant issues involved in this project. After all, all the data used are publicly available on the official Nobel Prize website.

Discussion

In my plan, I had intended to examine the speeches given by the laureates during the month of December, but given that many of them are not native speakers of English, the data collected may be inherently “noisy”. Similarly, the decision to focus on the prize motivation rather than the award ceremony speeches as the object of this project stems from the linguistic complexities inherent in the

latter. Some English speeches are translated from the original Swedish, and this process of code-switching may result in semantic drifts, potentially deviating from the precise subtleties of the initial phrasing. In contrast, the prize motivation is more concise, typically beginning with phrases such as “for her” or “for his” and averaging around 20 words in length, making it a more reliable source for linguistic analysis.

During the development of this project, I also considered age as an alternative entry point for analysis. For instance, are female laureates typically older at the time of recognition compared to their male counterparts? If so, this could reflect a delayed acknowledgment of their contributions. The ages of the youngest laureate, Rudyard Kipling (41), and the oldest laureate, Doris Lessing (87), already suggest notable disparities. This prompted questions about whether younger laureates tend to focus on themes such as innovation or modernity, while older laureates might emphasize reflection or legacy. Additionally, are there significant differences in how age correlates with thematic priorities in the works of female versus male laureates? To explore this, I collected the year of birth of the laureates and the year they won the award, and subtracted the two to get the age at the time of the award. I also created the following boxplot to visualize potential patterns. The boxplot highlights the stark disparity in gender representation. There are far more male laureates than female, especially in the early decades. Female winners are more concentrated in the second half of the 20th and 21st centuries, which may reflect changes in the gender inclusiveness of the Nobel Prize. However, few other significant outliers were identified to warrant further investigation into their implications. The exploration of the factor of age is stuck there.

Reference

- Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Bloomsbury Publishing.
- Butler, J. (2006). *Gender Trouble: Feminism and the Subversion of Identity* (1st ed.). New York: Routledge. <https://doi.org/10.4324/9780203824979>
- Eckert, P., & McConnell-Ginet, S. (2013). *Language and Gender* (2nd ed.). Cambridge: Cambridge University Press.
- Mautner, G. (2009). Corpora and critical discourse analysis. *Contemporary corpus linguistics*, 32-46.