FOEP: Evaluating Female Objectification under the Male Gaze

Bella Zhang
bz2428@nyu.edu

Cindy Luo
kl3108@nyu.edu

Cora Mao
ym1596@nyu.edu

Abstract

Gender bias exists extensively in human-written and model-generated texts. Detecting gender bias in the form of female objectification requires unifying previous efforts of measuring gender bias in semantics and in grammar. We propose the Female Objectification Evaluation Paradigm (FOEP), adopting the concept of the 'male gaze' from gender studies and literature analysis, which quantifies the level of objectification and emphasis on the visual pleasure of female characters in a novel. Applying our method to four novels as examples, the result indicates that restrictive female images are common in English novels both by female and male authors from the 19th and 20th centuries. Our paradigm presents a novel interdisciplinary outlook on the examination of gender bias, showing its efficacy in uncovering gender representations that are imbued with bias.

1 Introduction

Gender bias in texts leads to representational harm when a gender group is stigmatized or stereotyped (Barocas et al., 2017). Although the study of natural language processing (NLP) and gender roles is not new, most of the current research focuses on detecting and mitigating gender bias on the level of gender subspace in word embeddings (Bolukbasi et al., 2016), the association of concepts (Caliskan et al., 2017), or task performance differences across gender (Zhao et al., 2018; Rudinger et al., 2018). However, beyond semantics and grammar, gender bias could extend the patriarchal power dynamics to languages by portraying women as objects and perpetuating harmful gender stereotypes (Mulvey, 1989; Devereaux, 1990).

In this paper, we propose a method to quantify gender bias along two dimensions inspired by the concept of the male gaze (Mulvey, 1989), which are objectification and visual pleasure. The two dimensions are quantified by measuring: 1) the percentage of female characters taking an active role on sentence level as opposed to their male counterparts; 2) how much descriptions of female characters focus on appearance as opposed to those of the male characters. With a novel's whole text as input, our method follows this structure to generate two metrics that can then be used to indicate the existence of a restrictive female image.

We applied the method to four novels and found that various degrees of female objectification existed in novels across the 19 and 20th centuries and both in male and female authors' works. Comparing the quantitative results to our understanding of the novels have also shed light on the shortcoming in our methods and prompted for next steps in refining the method.

2 Related Work

Gender and culture studies have illustrated the problematic phenomenon of female objectification through the male gaze, and NLP researchers have developed various methods to measure gender bias on the semantic or syntactic level which can be combined to quantify female objectification.

Female Objectification through the Male Gaze The term 'male gaze' was first introduced by feminist film theorist Laura Mulvey. It describes how the unequal distribution of social and political power shaped cinematic representations of women and men (Mulvey, 1989). The concept of the male gaze has since evolved beyond its original psychoanalytic context and is now widely used to refer to the depiction of women and the world in arts, literature, and the media through the lens of a heterosexual male perspective. The accompanying concept of female objectification also points out that men's consumption of female imaging reduced women to the status of mere tools for men's purposes and women are objectified through being excessively preoccupied with their appearance (Nussbaum, 1995; Bartky, 2015; Bordo, 1993).

Measuring Gender Bias On the semantics end, Caliskan et al. proposed the Word Embedding Association Test (WEAT) to evaluate the difference in the strength of the connection between concepts and can be used to identify gender bias in the form of unequal man and female portraying. In order to detect gender bias that influences syntactic function, da Cunha and Abeillé (2022) compared the frequencies of men being as a syntactic subject compared to that of women on the sentences level.

3 Methods

3.1 Objectification - Female as passive roles

We first remove personal pronouns and discourse participants. We identify these entities using the coreference resolution model of Gardner et al. (2017) to replace any personal pronouns by the referred persona's name. Notice that the coreference model we used is unable to recognize the first-person singular pronoun. For example, it can't replace 'I' with 'Jane Eyre' in novel *Jane Eyre*. Since 'I' tend to take more active roles as the main narration perspective, to prevent the influence of first-person narration on calculating the ratio of subjective/objective roles, we opted to exclude all occurrences of 'I' and 'me' from our analysis.

Then we apply Named Entity Recognition (NER) tagging in SpaCy (Honnibal et al., 2020) to identify all the character names and manually extract their gender from the novel. There may be various versions of the same name, such as 'Jane', 'Jane Eyre', 'Janet' and 'Miss Eyre', which are synonymous. Thus, the manual merging of these similar entities was required. Additionally, we did not include character last names that could be referring to characters with multiple genders (e.g. 'Mr. and Mrs. Lucas') and only selected names that have appeared more than 3 times. With the list of characters' names, we use dependency parsing (Honnibal et al., 2020) to label the words of characters' names in the whole novel text as subjective or objective. The judgement of *subjective* and *objective* is based on the `.dep_` attributes[1] in dependency parsing model.

[1] https://emorynlp.github.io/nlp4j/components/dependency-parsing.html

We then aggregate the character-level information to the gender-level. For both genders, we calculate the number of times of characters in this gender are labelled as subjective or objective. After aggregation, we compare the ratio ($r = Count_{sub}/Count_{obj}$) between active-role and passive-role occurrences for each gender r_{female}, r_{male}. Meanwhile, we will do two proportions Z test to detect whether there exists difference in proportion of active-role for female and male. The Z test also offers a test statistic z, which can be used as score for quantifying the female objectification in novels.

3.2 Visual pleasure - Focus on female visual appearance

For this metric, we fine-tuned the glove.6B.50d embeddings from (Pennington et al., 2014) using the CBOW objective (Mikolov et al., 2013) on the selected literature corpus.

With the continued trained embedding results, We use WEAT to measure the relationship between the concepts of femininity and visual appearance in the fine-tuned embedding spaces (Caliskan et al., 2017).

WEAT Partially modifying WEAT's method, we have two sets of attribute words (man, male, male character name 1, male character name 2 ...) and (woman, female, female character name 1, female character name 2...), and only one set of target words (appearance, beautiful, ugly, ...). The null hypothesis is that there is no difference between the two sets of attribute words in terms of their similarity to the target words. A two-sided p-value of the permutation test is calculated to measure the likelihood of the null hypothesis. In formal terms, let X be the target words set and F, M the female and male sets of attribute words. Let $cos(m, n)$ denote the cosine similarity between vectors m and n. The effect size is:

$$\underset{f\in F,x\in X}{mean}(cos(f,x)) - \underset{m\in M,x\in X}{mean}(cos(m,x))$$

Thus, a positive value shows that femininity is more strongly connected to the concept of appearance and a negative value shows that masculinity is more strongly connected to the concept of appearance.

For the target word set to construct the concept of visual appearance, we scraped the Oxford Learner's Dictionary website [2] for 1004 suitable

[2] https://www.oxfordlearnersdictionaries.com/us/

words under the main topic ”Appearance”. See Table 4 in Appendix for number words coming from each subtopics.

The gender attribute word sets should construct the concept of femineity and masculinity. Therefore, we aggregated two base sets for the baseline results, and for each novel, we expanded the list to add female and male character names from NER results accordingly. See Table 5 in Appendix for the base set attributes.

4 Results

In order to illustrate our method, we have run our method on four novels: *Jane Eyre* (JE) by Charlotte Brontë published in 1847, *Great Expectations* (GE) by Charles Dickens published in 1861, and *Pride and Prejudice* (PP) by Jane Austen published in 1813, all retrieved from the Gutenberg Project website [3], as well as *Lolita* (LO) by Vladimir Nabokov published in 1955 retreived from Github [4].

4.1 Objectification Measurement Result

For part 1, we first applied coreference resolution to text data then used NER to identify main characters by counting the frequency. To check the accuracy of our NER model, we cross-checked our list of main characters with those from sparknotes website[5].

List Source	Main Characters
NER model	Edward Rochester, Jane Eyre, Sarah Reed, Diana Rivers Alice Fairfax, . . .
Sparknotes	Jane Eyre, Edward Rochester, St.John Rivers, Sarah Reed, Bessie Lee, . . .

Table 1: Top 5 main characters identified by the NER model and Sparknotes.

Take *Jane Eyre* as an example, the list we generated consists of 31 characters. Table 1 shows the top 5 characters in both lists. The overlap rate is high. Also, we calculated the Jaccard similarity[6] for the character list for four novels, and the similarity values are $72.7\%, 76.9\%, 85.2\%, 83.3\%$ respectively, suggesting that our approach was reasonable and accurate in identifying the main characters in the novels.

Then we count the number of times that female and male characters are put as subjective and objective throughout the whole novel. Table 2 shows the result and two-side Z-test scores for four novels. Figure 1 shows the difference between the ratio for female and male characters.

Novels	gender	Number of Times			Z-test
		sub	obj	sub/obj	(P-vale)
JE	female	1277	652	1.959	2.534
	male	1683	728	2.312	(0.01127)
GE	female	770	503	1.531	3.445
	male	2603	1353	1.924	(0.00057)
PP	female	2110	1072	1.968	-4.963
	male	754	536	1.407	(6.94e-07)
LO	female	566	432	1.311	1.698
	male	156	93	1.677	(0.08959)

Table 2: Subjective and objective roles taken by female and male characters.

For *Jane Eyre*, *Great Expectation* and *Lolita*, the ratio for female characters is lower than male, while *Pride and Prejudice* has the opposite trend. The test scores reveal a significant difference: male characters are more frequently the subject of sentences compared to their female counterparts in *Jane Eyre* and *Great Expectations*. Whereas in *Pride and Prejudice*, females take the subjective role more frequently than males.

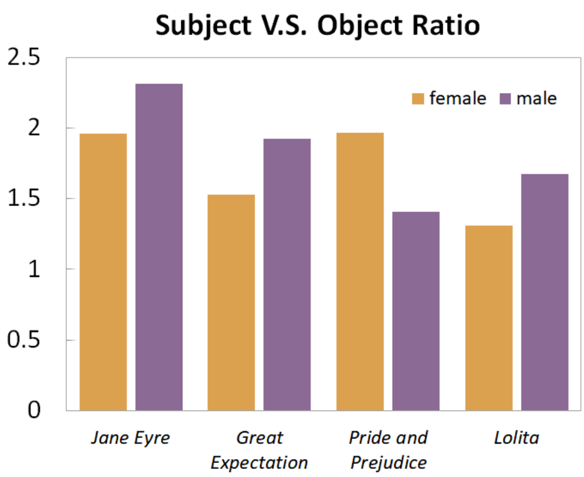

Figure 1: Ratios for the frequency of Subject V.S. Object roles taken by female and male characters in four novels.

[3] https://www.gutenberg.org/

[4] https://github.com/lsb/lolita-versus-twilight/blob/master/lolita.txt

[5] https://www.sparknotes.com/

[6] Jaccard similarity: $J(A, B) = |A \cap B|/|A \cup B|$

4.2 WEAT Result

Using the modified WEAT test, we evaluated if there is any difference between the male and female association to the concept of appearance. For the four selected novel examples, we ran the test for the word embeddings fine-tuned with the novel corpus. Table 3 shows the number of appearance target word in the word embeddings and the WEAT test effective size(P-value).

	Filtered # of target words	Effective Size (P-value)
JE	302	0.032 (0.0002)
GE	267	0.033 (0.0002)
PP	133	-0.004 (0.4178)
LO	432	0.020 (0.0002)

Table 3: WEAT test results.

Reading the results, we can see that *Jane Eyre*, *Great Expectations*, *Lolita* show significant associations between femininity and appearance on comparable levels, while *Pride and Prejudice* shows an insignificant reverse association, in which muscularity has a stronger association to the appearance concept. It is also worth noticing that the fine-tuned embeddings each contain different numbers of appearance-related words from the complete list.

5 Conclusion

Incorporating insights from literature and gender studies, our study proposes the Female Objectification Evaluation Paradigm (FOEP) as a means to assess gender bias and its consequential representational harms. This paradigm offers a comprehensive approach that integrates the concept of the 'male gaze' with current studies on gender bias in NLP. Applying this paradigm to analyze four novels with varying narration perspectives and diverse authorship, including both female and male authors, we measured the level of female objectification present within each novel. Our findings demonstrate the efficacy of this paradigm in capturing and highlighting the pervasive objectification of women across these novels. By shedding light on the implicit reinforcement of the 'male gaze' perspective within literary works from the 19th and 20th centuries, our results prompt a reevaluation of existing methods in NLP for detecting and measuring gender bias, encouraging an interdisciplinary examination of gender representation from multiple perspectives.

Limitations

Although the Female Objectification Evaluation paradigm used in this study yields meaningful results, several limitations should be acknowledged: Firstly, the paradigm faces challenges in accurately labeling instances of 'I' and capturing names from dialogues, particularly in cases of first-person narration. Secondly, the reliance on existing toolkits for coreference resolution revealed less-than-ideal accuracy in matching pronouns with names. Replacing text before applying the dependency parser may also impact the accuracy of correct labeling. Additionally, our current method does not differentiate between subjective and objective roles in sentences with or without an action verb. Refinements are necessary to consider sentence structure for a more accurate analysis. Moreover, the paradigm's effectiveness may be influenced by the relative emphasis on each gender role within a novel. If the narrative predominantly focuses on male characters, appearance word embeddings may skew towards male attributes due to the sparse representation of female attributes. While the paradigm examines appearance-related words, it may only emphasize the 'visual' aspect but less on 'pleasure'. Finally, it is important to note the limited sample size in this study. A larger and more diverse set of novels would provide a more comprehensive understanding of gender representation.

In conclusion, while the Female Objectification Evaluation paradigm has proven valuable, these limitations underscore the need for future refinements to improve accuracy and broaden the scope of analysis.

Contribution Statement

For the analysis work, Bella worked on implementing coreference resolution, NER tagging, and Dependency parsing; Cindy worked on text preprocessing and training of word embeddings; Cora worked on adjusting and applying the WEAT Test and aggregating the word lists.

For the writing of this paper, Bella wrote about related works; Cindy wrote the introduction and the conclusion section; Cora wrote the methods section; all three of us wrote about our own parts of the analysis results and involved in revising and editing of the paper.

Three authors contributed equally to this work.

Ethical Considerations

In our exploration of the 'male gaze' concept, we recognize the limitation of defining gender solely within binary categories. Critiques of the 'male gaze' concept rightly highlight its disregard for the presence of diverse gender identities beyond the binary labels of 'female' and 'male'. To address this limitation, future research should aim to expand upon our current work by adopting a more inclusive definition of gender that accounts for its diverse nature. Furthermore, it is important to emphasize that the primary objective of our study is to propose a new framework for evaluating gender bias, with insights drawn from social studies. However, it is essential to note that the framework's basic structure may not fully capture the nuanced perspectives of authors of the novels and should not be used as a means to discriminate against or attack any individuals or groups. Our intention is to foster a deeper understanding of gender representation and reduce representational harm in NLP, free from any discriminatory implications.

Acknowledgments

We would like to express our gratitude to our course instructor, Prof. Sophie Hao, for her invaluable comments on the paper and the project, as well as for her dedication to the course and its students. We also extend our appreciation to Prof. Laura Harris from Cinema Studies at NYU and Prof. Ann Pellegrini from Performance Studies at NYU for providing valuable feedback on the concept of the male gaze and its relevance to gender studies. Their insights greatly enhanced our understanding and enriched our research.

References

Solon Barocas, Kate Crawford, Aaron Shapiro, and Hanna Wallach. 2017. The problem with bias: From allocative to representational harms in machine learning'. In *SIGCIS conference paper*.

Sandra Lee Bartky. 2015. *Femininity and domination: Studies in the phenomenology of oppression*. Routledge.

Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems* 29.

Susan Bordo. 1993. Unbearable weight: femisisn, western culture, and the body .

Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science* 356(6334):183–186.

Yanis da Cunha and Anne Abeillé. 2022. Objectifying women? a syntactic bias in French and English corpora. In *Proceedings of the First Workshop on Language Technology and Resources for a Fair, Inclusive, and Safe Society within the 13th Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, pages 8–16. https://aclanthology.org/2022.lateraisse-1.2.

Mary Devereaux. 1990. Oppressive texts, resisting readers and the gendered spectator: The new aesthetics. *The Journal of Aesthetics and Art Criticism* 48(4):337–347.

Matt Gardner, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson F. Liu, Matthew Peters, Michael Schmitz, and Luke S. Zettlemoyer. 2017. Allennlp: A deep semantic natural language processing platform.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python https://doi.org/10.5281/zenodo.1212303.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* .

Laura Mulvey. 1989. Visual pleasure and narrative cinema. In *Visual and other pleasures*, Springer, pages 14–26.

Martha C Nussbaum. 1995. Objectification. *Philosophy & Public Affairs* 24(4):249–291.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. pages 1532–1543.

Rachel Rudinger, Jason Naradowsky, Brian Leonard, and Benjamin Van Durme. 2018. Gender bias in coreference resolution. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*.

Association for Computational Linguistics, New Orleans, Louisiana, pages 8–14. https://doi.org/10.18653/v1/N18-2002.

Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang. 2018. Gender bias in coreference resolution: Evaluation and debiasing methods. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. Association for Computational Linguistics, New Orleans, Louisiana, pages 15–20. https://doi.org/10.18653/v1/N18-2003.

A Appendix

Subtopic	Number of words
attractiveness	67
body shape	53
describing hair	134
skin	42
styling hair	77
face	41
hands and feet	29
mouth and teeth	30
parts of the body	63
accessories	32
describing clothes	134
footwear	63
hats	56
items of clothing	149
jewellery	54
parts of clothing	30

Table 4: Number of target words in each subtopics of Appearance

	Base Attribute Words
Female	woman, female, girl, she, her, herself, wife, mother, miss, lady
Male	man, male, boy, he, his, himself, husband, father, mr, sir

Table 5: Base gender attribute words