Analyzing Gender Bias within Narrative Tropes

Dhruvil Gala* Mohammad Omar Khursheed* Hannah Lerner
Brendan O’Connor Mohit Iyyer
University of Massachusetts Amherst
{dgala, mkhursheed, hmlerner, brenocon, miyyer}@umass.edu

Abstract

Popular media reflects and reinforces societal biases through the use of tropes, which are narrative elements, such as archetypal characters and plot arcs, that occur frequently across media. In this paper, we specifically investigate gender bias within a large collection of tropes. To enable our study, we crawl tvtropes.org, an online user-created repository that contains 30K tropes associated with 1.9M examples of their occurrences across film, television, and literature. We automatically score the “genderedness” of each trope in our TVTROPES dataset, which enables an analysis of (1) highly-gendered topics within tropes, (2) the relationship between gender bias and popular reception, and (3) how the gender of a work’s creator correlates with the types of tropes that they use.

1 Introduction

Tropes are commonly-occurring narrative patterns within popular media. For example, the evil genius trope occurs widely across literature (Lord Voldemort in Harry Potter), film (Hannibal Lecter in The Silence of the Lambs), and television (Tywin Lannister in Game of Thrones). Unfortunately, many tropes exhibit gender bias, either explicitly through stereotypical generalizations in their definitions, or implicitly through biased representation in their usage that exhibits such stereotypes. Movies, TV shows, and books with stereotypically gendered tropes and biased representation reify and reinforce gender stereotypes in society (Rowe, 2011; Gupta, 2008; Leonard, 2006). While evil genius is not an explicitly gendered trope (as opposed to, for example, women are wiser), the online tvtropes.org repository contains 108 male and only 15 female instances of evil genius across film, TV, and literature.

To quantitatively analyze gender bias within tropes, we collect TVTROPES, a large-scale dataset that contains 1.9M examples of 30K tropes in various forms of media. We augment our dataset with metadata from IMDb (year created, genre, rating of the film/show) and Goodreads (author, characters, gender of the author), which enable the exploration of how trope usage differs across contexts.

Using our dataset, we develop a simple method based on counting pronouns and gendered terms to compute a genderedness score for each trope. Our computational analysis of tropes and their genderedness reveals the following:

• Genre impacts genderedness: Media related to sports, war, and science fiction rely heavily on male-dominated tropes, while romance, horror, and musicals lean female.

• Male-leaning tropes exhibit more topical diversity: Using LDA, we show that male-leaning tropes exhibit higher topic diversity (e.g., science, religion, money) than female tropes, which contain fewer distinct topics (often related to sexuality and maternalism).

• Low-rated movies contain more gendered tropes: Examining the most informative features of a classifier trained to predict IMDb ratings for a given movie reveals that gendered tropes are strong predictors of low ratings.

• Female authors use more diverse gendered tropes than male authors: Using author gender metadata from Goodreads, we show that female authors incorporate a more diverse set of female-leaning tropes into their works.

---

*Authors contributed equally.

1Our work explores gender bias across two identities: cisgender male and female. The lack of reliable lexicons limits our ability to explore bias across other gender identities, which should be a priority for future work.
We attempt to match titles from film, TV, and literature, excluding other forms of media, such as web comics and video games. We focus on the former forms of media. Descriptions normally consist of multiple paragraphs (277 tokens on average), as well as a set of examples (86), female (23), or N/A (41). N/A represents examples that do not capture any aspect of gender. We then use the lexicon to classify each example as male (precision = 0.85, recall = 0.86, and F1 score = 0.86) or female (precision = 0.72, recall = 0.78, and F1 score = 0.75).

To measure genderedness, for each trope \( i \), we concatenate the trope’s description with all of the trope’s examples to form a document \( X_i \). Next, we tokenize, preprocess, and lemmatize \( X_i \) using NLTK (Loper and Bird, 2002). We then compute the number of tokens in \( X_i \) that match the male lexicon, \( m(X_i) \), and the female lexicon, \( f(X_i) \). We also compute \( m(\text{TVTROPES}) \) and \( f(\text{TVTROPES}) \), the total number of matches for each gender across

---

2 http://github.com/dhruvilgala/tvtropes

3 We match by both the work’s title and its year of release to avoid duplicates.

2.3 Who contributes to TVTROPES?

One limitation of any analysis of social bias on TVTROPES is that the website may not be representative of the true distribution of tropes within media. There is a confounding selection bias—the media in TVTROPES is selected by the users who maintain the tvtropes.org resource. To better understand the demographics of contributing users, we scrape the pages of the 15K contributors, many of which contain unstructured biography sections. We search for biographies that contain tokens related to gender and age, and then we manually extract the reported gender and age for a sample of 256 contributors. The median age of these contributors is 20, while 64% of them are male, 33% female and 3% bi-gender, genderfluid, non-binary, trans, or agender. We leave exploration of whether user-reported gender correlates with properties of contributed tropes to future work.

3 Measuring trope genderedness

We limit our analysis to male and female genders, though we are keenly interested in examining the correlations of other genders with trope use. We devise a simple score for trope genderedness that relies on matching tokens to male and female lexicons used in prior work (Bolukbasi et al., 2016; Zhao et al., 2018) and include gendered pronouns, possessives (his, her), occupations (actor, actress), and other gendered terms. We validate the effectiveness of the lexicon in capturing genderedness by annotating 150 random examples of trope occurrences as male (86), female (23), or N/A (41). N/A represents examples that do not capture any aspect of gender. We then use the lexicon to classify each example as male (precision = 0.85, recall = 0.86, and F1 score = 0.86) or female (precision = 0.72, recall = 0.78, and F1 score = 0.75).

To measure genderedness, for each trope \( i \), we concatenate the trope’s description with all of the trope’s examples to form a document \( X_i \). Next, we tokenize, preprocess, and lemmatize \( X_i \) using NLTK (Loper and Bird, 2002). We then compute the number of tokens in \( X_i \) that match the male lexicon, \( m(X_i) \), and the female lexicon, \( f(X_i) \). We also compute \( m(\text{TVTROPES}) \) and \( f(\text{TVTROPES}) \), the total number of matches for each gender across

---

2 We note that some demographics may be more inclined to report age and gender information than others.

3 The gender-balanced lexicon is obtained from Zhao et al. (2018) and comprises 222 male-female word pairs.
Table 2: Instances of highly-gendered tropes.

| Male Tropes       | \( g \) | Female Tropes | \( g \) |
|-------------------|---------|---------------|---------|
| Motivated by Fear | -1.8    | Ms. Fanservice | 3.4     |
| Robot War         | -1.6    | Socialite     | 3.1     |
| Cure for Cancer   | -1.5    | Damsel in Distress | 2.7     |
| Evil Genius       | -1.3    | Hot Scientist  | 2.2     |
| Grand Finale      | -1.2    | Ditzy Secretary | 2.0     |

We can examine how genderedness varies by genre. Given the set of all movies and TV shows in TVTROPES that belong to a particular genre, we extract the set of all tropes used in these works. Next, we compute the average genderedness score of all of these tropes. Figure 1 shows that media about sports, war, and science fiction contain more male-dominated tropes, while musicals, horror, and romance shows are heavily oriented towards female tropes, which is corroborated by social science literature (Lauzen, 2019).

4.2 Topics in highly-gendered tropes

To find common topics in highly-gendered male or female tropes, we run latent Dirichlet analysis (Blei et al., 2003) on a subset of highly-gendered trope descriptions and examples7 with 75 topics. We filter out tropes whose combined descriptions and examples (i.e., \( X_t \)) have fewer than 1K tokens, and then we further limit our training data to a balanced subset of the 3,000 most male and female-leaning tropes using our genderedness score. After training, we compute a gender ratio for every topic: given a topic \( t \), we identify the set of all tropes for which \( t \) is the most probable topic, and then we compute the ratio of female-leaning to male-leaning tropes within this set.

We observe that 45 of the topics are skewed towards male tropes, while 30 of them favor female tropes, suggesting that male-leaning tropes cover a larger, more diverse set of topics than female-leaning tropes. Table 3 contains specific examples of the most gendered male and female topics. This experiment, in addition to a qualitative inspection of the topics, reveals that female topics

---

7We use Gensim’s LDA library (Rehurek and Sojka, 2010).
We identify implicitly gendered tropes (Glick and Fiske, 1996) — female topics about motherhood and pregnancy display gender differentiation, topics about appearance and nudity can be attributed to heterosexuality, while male topics about money and strength capture paternalism. The bias captured by these topics, although unsurprising given previous work (Bolukbasi et al., 2016), serves as a sanity check for our metric and provides further evidence of the limited diversity in female roles (Lauzen, 2019).

### 4.3 Identifying implicitly gendered tropes

We identify implicitly gendered tropes (Glick and Fiske, 1996)—tropes that are not defined by gender but nevertheless have high genderedness scores—by identifying a subset of 3500 highly-gendered tropes whose titles do not contain gendered tokens. A qualitative analysis reveals that tropes containing the word “genius” (impossible genius, gibbering genius, evil genius) and “boss” (beleaguered boss, stupid boss) lean heavily male. There are interesting gender divergences within a high-level topic: within “evil” tropes, male-leaning tropes are diverse (necessarily evil, evil corporation, evil army), while female tropes focus on sex (sex is evil, evil eyeshadow, evil is sexy).

### 4.4 Using tropes to predict ratings

Are gendered tropes predictive of media popularity? We consider three roughly equal-sized bins of IMDb ratings (Low, Medium, and High). For each IMDb-matched title in TVTROPES, we construct a binary vector $z_i \in \{0,1\}^T$, where $T$ is the number of unique tropes in our dataset. We set $z_i$ to 1 if trope $i$ occurs in the movie, and 0 otherwise. Tropes are predictive of ratings: a logistic regression classifier achieves 55% test accuracy with this method, well over the majority class baseline of 36%. Table 4 contains the most predictive gendered tropes for each class; interestingly, low-rated titles have a much higher average absolute genderedness score (0.73) than high-rated ones (0.49), providing interesting contrast to the opposing conclusions drawn by Boyle (2014). While IMDb ratings offer a great place to start in correlating public perception with genderedness in tropes, we may be double-dipping into the same pool of internet movie reviewers as TVTROPES. We leave further exploration of correlating gendered tropes with box office results, budgets, awards, etc. for future work.

### 4.5 Predicting author gender from tropes

We predict the author gender by training a classifier for 2521 Goodreads authors based on a binary feature vector encoding the presence or absence of tropes in their books. We achieve an accuracy of 71% on our test set (majority baseline is 64%). Interestingly, the top 50 tropes most predictive of male authors have an average genderedness of 0.04, while those most correlated with female authors have an average of 0.89, indicating that books by female authors contain more female-leaning tropes. Eighteen female-leaning tropes ($g_i > 1$), varying in scope from the non-traditional feminist fantasy to the more stereotypical hair of gold heart of gold, are predictive of female authors. In contrast, only.

---

8This process contains noise due to our small lexicon: a few explicitly gendered, often problematic tropes such as absolute cleavage are not filtered out.

9Low: (0-6.7], Medium: (6.7-7.7], High: (7.7-10]

---

### Table 4: Gendered tropes predictive of IMDb rating.

| Topic                        | High Rated | Low Rated |
|------------------------------|------------|-----------|
| Edutainment Show             | 1.2        | Alpha Bitch 2.7 |
| Cooking Stories              | 1.0        | Sexy Backless Outfit 2.6 |
| British Brevity              | -0.9       | Fanservice 1.9 |
| Wedding Smashers             | 0.8        | Shower Scene 1.4 |
| Just Following Orders        | -0.7       | Sword and Sandal -1.0 |
| Male Author | Female Author |
|-------------|---------------|
| Trope       | g              | Trope       | g              |
| Undressing the Unconscious | 1.3 | Cool Old Lady | 2.7 |
| First Girl Wins | 1.3 | Plucky Girl | 2.5 |
| Did Not Get the Girl | 0.9 | Feminist Fantasy | 2.2 |
| God is Evil | -0.8 | Young Adult Literature | 1.7 |
| Retired Badass | -0.8 | Extremely Protective Child | 1.2 |

Table 5: Gendered tropes predictive of author gender.

two such character-driven female-dominated tropes are predictive of male authors; the stereotypical undressing the unconscious and first girl wins; see Table 5 for more. Furthermore, out of 115K examples of tropes in female-authored books, 17K are highly female, while just 2.2K are male-dominated. Since many of these gendered tropes are character-driven, this implies wider female representation in such gendered instances, previously shown in Scottish crime fiction (Hill, 2017). Overall, female authors frequently use both stereotypical and non-stereotypical female-oriented tropes, while male authors limit themselves to more stereotypical kinds. However, it is important to note the double selection bias at play in both selecting which books are actually published, as well as which published books are reviewed on Goodreads. While there are valid ethical concerns with a task that attempts to predict gender, this task only analyzes the tropes most predictive of author gender, and the classifier is not used to do inference on unlabelled data or as a way to identify an individual’s gender.

5 Related Work

Our work builds on computational research analyzing gender bias. Methods to measure gender bias include using contextual cues to develop probabilistic estimates (Ananya et al., 2019), and using gender directions in word embedding spaces (Bolukbasi et al., 2016). Other work engages directly with tvtropes.org: Kiesel and Grimnes (2010) build a wrapper for the website, but perform no analysis of its content. García-Ortega et al. (2018) create PicTropes: a limited dataset of 5,925 films from the website. Bamman et al. (2013) collect a set of 72 character-based tropes, which they then use to evaluate induced character personas, and Lee et al. (2019) use data crawled from the website to explore different sexism dimensions within TV and film.

Analyzing bias through tropes is a popular area of research within social science. Hansen (2018) focus in on the titular princess character in the video game The Legend of Zelda as an example of the Damsel in Distress trope. Lacroix (2011) study the development and representation in popular culture of the Casino Indian and Ignoble Savage tropes.

The usage of biased tropes is often attributed to the lack of equal representation both on and off the screen. The Geena Davis Inclusion Quotient (Google, 2017) quantifies the speaking time and importance of characters in films, and finds that male characters have nearly twice the presence of female characters in award-winning films. In contrast, our analysis looks specifically at tropes, which may not correlate directly with speaking time. Lauzen (2019) provides valuable insight into representation among film writers, directors, crew members, etc. Perkins and Schreiber (2019) study an ongoing increase in the representation of women in independent productions on television, many of which focus on feminist content.

6 Future Work

We believe that the TVTROPES dataset can be used to further research in a variety of areas. We envision setting up a task involving trope detection from raw movie scripts or books; the resulting classifier, beyond being useful for analysis, could also be used by media creators to foster better representation during the writing process. There is also the possibility of using the large number of examples we collect in order to generate augmented training data or adversarial data for tasks such as coreference resolution in a gendered context (Rudinger et al., 2018). The expansion of our genderedness metric to include non-binary gender identities, which in turn would involve creating similar lexicons as we use, is an important area for further exploration.

It would also be useful to gain further understanding of the multiple online communities that contribute information about popular culture; for example, an analysis of possible overlap in contributors to TVTROPES and IMDb could better account for sampling bias when analyzing these datasets.

7 Acknowledgements

We would like to thank Jesse Thomason for his valuable advice.

References

Ananya, Nitya Parthasarthi, and Sameer Singh. 2019. GenderQuant: Quantifying mention-level genderedness. In Proceedings of the 2019 Conference of the North American Chapter of the Association for
Yasmeen Hitti, Eunbee Jang, Ines Moreno, and Car- lorna Hill. 2017. Bloody women: How female authors Jared Capener Hansen. 2018. Why can’t zelda save her- Google. 2017. Using technology to address gender Peter Glick and Susan Fiske. 1996. The ambivalent Rubén H. García-Ortega, Juan J. Merelo-Guervós, David García, Ingmar Weber, and Venkata Rama Ki- Karen Boyle. 2014. Gender, comedy and reviewing Tolga Bolukbasi, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam Kalai. 2016. Man is is to computer programmer as woman is to home- maker? debiasing word embeddings. CoRR, abs/1607.06520. David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. Journal of machine Learning research, 3.

Tolga Bolukbasi, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam Kalai. 2016. Man is is to computer programmer as woman is to homemaker? debiasing word embeddings. CoRR, abs/1607.06520.

Karen Boyle. 2014. Gender, comedy and reviewing culture on the internet movie database. Participations, 11(1):31–49.

David García, Ingmar Weber, and Venkata Rama Kiran Garimella. 2014. Gender asymmetries in reality and fiction: The bechdel test of social media. CoRR, abs/1404.0163.

Rubén H. García-Ortega, Juan J. Merelo-Guervós, Pablo García Sánchez, and Gad Pitaru. 2018. Overview of pictropes, a film trope dataset.

Peter Glick and Susan Fiske. 1996. The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. Journal of Personality and Social Psychology, 70:491–512.

Google. 2017. Using technology to address gender bias in film.

Charu Gupta. 2008. (mis) representing the dalit woman: Reification of caste and gender stereotypes in the hindi didactic literature of colonial india. Indian Historical Review, 35(2).

Jared Capener Hansen. 2018. Why can’t zelda save herself? how the damsel in distress trope affects video game players.

Lorna Hill. 2017. Bloody women: How female authors have transformed the scottish contemporary crime fiction genre. American, British and Canadian Studies, 28(1):52 – 71.

Yasmeen Hitti, Eunbee Jang, Ines Moreno, and Carolyn Pelletier. 2019. Proposed taxonomy for gender bias in text: a filtering methodology for the gender generalization subtype. In Proceedings of the First Workshop on Gender Bias in Natural Language Processing, pages 8–17, Florence, Italy. Association for Computational Linguistics.

Malte Kiesel and Gunnar Aastrand Grimes. 2010. Db- tropes - a linked data wrapper approach incorporating community feedback. In European Knowledge Acquisition Workshop (EKAW).

Celeste C. Lacroix. 2011. High stakes stereotypes: The emergence of the “casino indian” trope in television depictions of contemporary native americans. Howard Journal of Communications, 22(1):1–23.

Martha M Lauzen. 2019. It’s a man’s (celluloid) world: Portrayals of female characters in the top grossing films of 2018’. Center for the Study of Women in Television and Film, San Diego State University.

Nayeon Lee, Yejin Bang, Jamin Shin, and Pascale Fung. 2019. Understanding the shades of sexism in popular TV series. In Proceedings of the 2019 Workshop on Widening NLP, pages 122–125, Florence, Italy. Association for Computational Linguistics.

David J Leonard. 2006. Not a hater, just keepin’it real: The importance of race-and gender-based game studies. Games and culture, 1(1).

Edward Loper and Steven Bird. 2002. Nltk: The natural language toolkit. In In Proceedings of the ACL Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics. Philadelphia: Association for Computational Linguistics.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830.

Claire Perkins and Michele Schreiber. 2019. Independent women: from film to television. Feminist Media Studies, 19(7):919–927.

Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, pages 45–50, Valletta, Malta. ELRA. http://is.muni.cz/ publication/884893/en.

K. Rowe. 2011. The Unruly Woman: Gender and the Genres of Laughter. University of Texas Press.

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), pages 8–14.

Jieyu Zhao, Yichao Zhou, Zeyu Li, Wei Wang, and Kai-Wei Chang. 2018. Learning gender-neutral word embeddings. CoRR, abs/1809.01496.

Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, pages 45–50, Valletta, Malta. ELRA. http://is.muni.cz/publication/884893/en.

K. Rowe. 2011. The Unruly Woman: Gender and the Genres of Laughter. University of Texas Press.

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), pages 8–14.

Jieyu Zhao, Yichao Zhou, Zeyu Li, Wei Wang, and Kai-Wei Chang. 2018. Learning gender-neutral word embeddings. CoRR, abs/1809.01496.