Shirtless and Dangerous: Quantifying Linguistic Signals of Gender Bias in an Online Fiction Writing Community

Ethan Fast, Tina Vachovsky, Michael S. Bernstein
Stanford University
\{ethan.fast, tvachov, msb\}@cs.stanford.edu

Abstract

Imagine a princess asleep in a castle, waiting for her prince to slay the dragon and rescue her. Tales like the famous Sleeping Beauty clearly divide up gender roles. But what about more modern stories, borne of a generation increasingly aware of social constructs like sexism and racism? Do these stories tend to reinforce gender stereotypes, or counter them? In this paper, we present a technique that combines natural language processing with a crowdsourced lexicon of stereotypes to capture gender biases in fiction. We apply this technique across 1.8 billion words of fiction from the Wattpad online writing community, investigating gender representation in stories, how male and female characters behave and are described, and how authors’ use of gender stereotypes is associated with the community’s ratings. We find that male over-representation and traditional gender stereotypes (e.g., dominant men and submissive women) are common throughout nearly every genre in our corpus. However, only some of these stereotypes, like sexual or violent men, are associated with highly rated stories. Finally, despite women often being the target of negative stereotypes, female authors are equally likely to write such stereotypes as men.

Introduction

"Ooh! Ooh! Look at this top! I have to have it." Nina said holding up a gold, glittery top with a boat neck but low cut back. It was gorgeous. "And those shoes!"

– Anonymous (a Wattpad story)

"I highly doubt that," Fanny growled, clenching a hammer in her hands that she pulled from her work belt.

– Anonymous (a Wattpad story)

These two excerpts from stories on an online writing community lay out competing concepts of gender roles in fiction. The first plays on common female stereotypes, while the second rejects them. Fictional worlds offer writers open-ended laboratories where they might explore new forms of social norms. Do these worlds tend to reinforce stereotypes, or push for more balanced gender roles? This paper is about the role of gender to character actions and description. To more easily interpret these signals, we use paid crowdsourcing to group common verbs and adjectives into sixteen high level categories like “sexual” and “submissive” that capture dimensions of gender stereotypes. Over billions of words, clear patterns emerge that separate male and female characters within these categories. For example, we find that male char-
characters are more physically active and violent, while female characters are described by terms that have submissive or childish connotations.

The size and genre variety of our corpus allow us to probe further into more nuanced questions about these stereotypes. Is there any difference in the way male and female authors write gender stereotypes? We might imagine, for example, that female authors would have a different conception of gender roles than male authors, and so write their characters differently. And what about the reception of these stereotypes? Are they associated with higher or lower rated stories? Ultimately, we find that while both genders perpetuate the same set of stereotypes on Wattpad, the relationship of these stereotypes to story rating is mixed.

Related Work

Gender Stereotypes in Modern Storytelling: Studies of gender roles in commercial media offer a useful starting point for investigating stereotypes in amateur fiction. Work analyzing movies, children’s books, and music lyrics has found that women are likely to be portrayed as younger, more emotionally motivated, less rational, and valued for beauty over intellect (Towbin et al. 2008; Gooden and Gooden 2001), whereas men are stronger, more violent, less in control of their sexuality, and expected to express emotion physically (Emons, Wester, and Scheepers 2010; Brethhauer, Zimmerman, and Banning 2007). Women also tend to be portrayed as domestic, with activities centered around the home or family, while men are shown as physically active and economically successful, with a much wider range of employment and often higher ranking and higher paying jobs (Lauzen, Dozier, and Horan 2008; Soulière 2006). Where past studies have focused on gender bias in relatively small datasets of professionally-crafted commercial media, our work addresses the biases of amateurs across orders of magnitude more user-generated data.

Gender Stereotypes on the Web: A second body of work has found that similar gender stereotypes translate to the web. For example, males are over-represented in the reporting of web-based news articles (Jia, Lansdall-Welfare, and Cristianini 2015), an effect which also holds for other forms of social media, like twitter conversations (Garcia, Weber, and Garimella 2014). Gender roles can also shape the substance of web content itself. For example, biographical articles about women on Wikipedia disproportionately discuss romantic relationships or family-related issues (Wagner et al. 2015). Sometimes gender can drive user interaction: for example, on Pinterest, women earn more repins, but fewer followers (Gilbert et al. 2013); on IMDB, reviews written by women are perceived as less useful (Otterbacher 2013); or on social media sites and dating profiles, women are more often judged on physical characteristics than men (Rose et al. 2012; Otterbacher 2015; Fiore et al. 2008). Similarly, in web searches for images of archetypal occupations (e.g., nurse, investment banker), the minority gender is often portrayed and perceived as relatively more unprofessional (Kay, Matuszek, and Munson 2015). Our work extends these analyses to the fictional worlds created by a community of amateur writers.

Gender and User Behavior: The gender of a user or artist can influence how they behave or what kinds of stereotypes occur in creative content. For example, on Pinterest, women more often reciprocate social links, while men specialize in curating specific categories of content (Ottoni et al. 2013). Similarly, in roleplaying games, both genders are likely to choose warriors as novices, but expert female players are more likely to default to the stereotypical healer; men, on the other hand, mainly play healers only if they design the avatar as a woman (Bergstrom, Jenson, and de Castell 2012). In television, the gender of writers can affect plot and narrative: mixed gendered groups of writers are more likely to deviate from stereotypes, writing both men and women in stereotypical feminine roles (Lauzen, Dozier, and Horan 2008). In our work, we examine how the gender of users on Wattpad affects the kinds of characters that appear in their stories.

Analysis of Text Corpora: Our approach to text analysis is informed by a large body of work in the computational social sciences. Researchers often use lexicons like LIWC to analyze texts across a broad range of signals (Pennebaker, Francis, and Booth 2001). Other research has crowd-sourced new lexical categories when they do not yet exist (Wagner et al. 2015; Fast, Chen, and Bernstein 2016), an approach we adopt to build categories that capture common gender stereotypes. To extract relevant signals from text, researchers have leveraged the coefficients of regression models (Gilbert 2012) and the strength and significance of correlations (Pennebaker and Stone 2003), as well as performed significance tests between the means of lexical categories (Kramer, Guillory, and Hancock 2014). In our work, we extend these analyses to target sentence-level relationships that capture behavioral and descriptive attributes of male and female characters.

Data and Methods

We take a quantitative approach to answer our research questions. Here we describe our data and statistical methods.

Data: 1.8 billion words of amateur fiction

We conducted our analyses over amateur fiction from the Wattpad writing community. In aggregate, our dataset contains more than 1.8 billion words selected from a random sample of 600,000 stories, written by more than 500,000 writers across twenty genres. Wattpad provided us with all data, as well as gender annotations for 475,000 authors, 46% of which are by women and 54% of which are by men. Table 1 presents a summary of the stories and genres in our dataset, as well as the gender representation in each genre, captured by counts of male and female pronouns.

How balanced are Wattpad stories between male and female characters? Before addressing our research questions, it is important to note (and control for) the fact that men are commonly over-represented in many forms of media (Ross and Carter 2011; Garcia, Weber, and Garimella 2014). This representational bias also exists on Wattpad (Table 1). Consistent with prior work, male pronouns are more common across most genres: Wattpad stories use 1.24 times more
male pronouns in their writing than female pronouns. Further, both genders write more about men. Male authors use 1.23 times more male than female pronouns, and female authors use 1.25 times more male pronouns (a small but significant difference, \( p < 0.001 \)). Our methodology adjusts for this male over-representation by normalizing our statistics between genders.

### Character Statistics

Social scientists often treat documents as “bags of words,” for example analyzing tweets, books, or status updates with the lexical categories in tools like LIWC (Kramer, Guillory, and Hancock 2014; Pennebaker and Stone 2003). Here we present a more targeted kind of analysis that captures the actions and descriptions of characters by taking advantage of the dependency structure of sentences.

**Character behavior.** Stereotypes are defined in part by how people behave. Do our stories read like Indiana Jones movies, with men initiating violence and activity, and women screaming and needing rescue? To address these kinds of questions, we mined gendered character actions from the dataset.

We extracted actions through subject-verb relationships where the subject is “he” or “she”. For example, in the sentence “she ruined him utterly” we would extract the subject-verb relationship “she ruins”. We lemmatized words in these extractions (e.g., “ruined” to “ruins”) to collapse verbs onto a canonical form.

This process left us with 42 million character actions, of which 25 million are male, and 17 million are female.

**Character description.** Stereotypes also exist through how people are described. Are men expected to be tall and broad-shouldered, while women are small and graceful, as cultural stereotypes would suggest? To find out, we mined gendered adjective descriptors from the dataset.

We extracted two kinds of adjective relationships. First, we targeted adjectives attached to male or female subjects by the verb “to be,” for example “he was clinically deranged.” Second, we targeted the modifiers of words that represent male or female characters like “brother” or “mother,” for example “the upset mother.” Again, we lemmatized all words.

This process left us with 4.3 million adjectives, of which 2 million are applied to female characters, and 2.3 million are applied to male characters.

#### Crowdsourcing a Stereotype Lexicon

Character statistics present a low level view of the differences between genders. To understand stereotypes at a higher level – for example, whether male characters more often act violently or female characters are more often described as weak – we need lexical categories that map individual words onto broader categories. Such categories would tell us that the adjective “fragile” is associated with weak, or the verb “kill” is associated with violence.
To choose these categories of stereotype, we draw from the results of prior work across television, movies, and social media (Towbin et al. 2008; Lauzen, Dozier, and Horan 2008). The literature identifies the following set of categories predominantly associated with men: violent, dominant, strong, arrogant, sexual, angry, and (physically) active. For women, the predominant categories are: domestic, hysterical, childish, afraid, dependent, emotional, beautiful, and submissive. We utilize these categories in our analysis.

We built lexical categories that capture this list of stereotypes by mapping the 2000 most commonly occurring verbs and adjectives in our dataset onto the set of categories through a series of crowdsourcing tasks. Labeling the most common words in a corpus is an efficient way to build a useful lexicon while limiting cost (Mohammad and Turney 2013). For each task, we asked workers on Amazon Mechanical Turk (AMT):

For each word, tell us how strongly it relates to the topic. For example, given the topic VIOLENT, the words “kill” and “hurt” would be strongly related, the word “scratch” would be related, the word “resist” would be weakly related, and the words “laugh” and “patient” would be unrelated.

We created tasks for each category of stereotype. Each task gathers annotations for 20 words, for which we pay workers $0.14 in line with guidelines for ethical research (Salehi, Irani, and Bernstein 2015). We collect labels from three independent workers, enough to generate accurate results given high quality workers, such as the Masters workers we recruit on AMT (Sheng, Provost, and Ipeirotis 2008). To reduce the impact of false positives on our analysis, we added a word to a category of stereotype only if the majority of workers have decided it is related or strongly related to the category in question. Similarly defined tasks have preformed well when constructing lexical categories (Mohammad and Turney 2013).

We generated this lexicon at a total cost of $630. Crowd workers agreed among themselves (voting unanimously for labels of related or strongly related) at a rate of 85%.

Statistical Methods
Given a lexicon that maps low level character statistics onto high level stereotypes, we now present the statistical techniques that translate these data into answers to our research questions. We used two primary methods to investigate our four research questions:

- **RQ1-Action and RQ2-Description.** To discover the presence of gender stereotypes in how characters act and are described, we tested for a difference in mean frequency counts between genders for our sixteen stereotype categories, taking these means over stories and normalizing them on character gender occurrence rates. We used Welch’s t-test, a two-sided test that does not assume equal variance among populations, and we present these results as odds ratios between means. Because we investigated many hypotheses at once, we applied a Bonferroni correction ($\alpha = 0.05/16 = 0.0031$) to the results of each analysis. All the results we present reach this level of significance.

- **RQ3-Ratings and RQ4-Authors.** To discover how stereotypes differ between male and female authors or across high and low rated stories, we trained logistic regression models that use gendered counts from our sixteen stereotype categories to predict either author gender or story rating. This type of model has provided insight into other kinds of linguistic signals (Gilbert 2012). For example, if men and women write different kinds of stereotypes (e.g., if women write more hysterical male characters, or less of them, or follow any other consistent pattern), we would expect a classifier trained on the frequency counts of stereotypes captured by our categories to have significant predictive value on author gender. Similarly, a classifier trained on these features to predict story rating will reveal what stereotypes (e.g., violent men or dominant women) are associated with higher and lower rated stories.

Concretely, we trained these logistic regression models using scikit-learn with L2 regularization (C=1). For both models, the input features were frequency counts of words captured by our stereotype categories for male and female characters, normalized on the baseline character occurrence rates for each gender. This made $16 \times 2 = 32$ features in total. To predict author gender, we trained on two classes (male and female) over a dataset of 429,582 stories. For story rating, we split the dataset into two equal parts above and below the median rating, and trained the model to predict these classes. When reporting on classifier accuracy, we ran both models under 10-fold cross-validation. We applied a Bonferroni correction ($\alpha = 0.05/32 = 0.0016$) when testing the significance of features in these models.

Finally, we computed gendered odds ratios on the 1000 most common verbs and adjectives for both male and female authors. These ratios tell us, for example, how much more likely a male character written by a man will take an action like punch, in comparison to a male character written by a woman. We plotted a log-transform of these ratios for visual clarity, and computed a Pearson correlation to quantify the similarity between authors of different genders.

Results
In this section we present the results of our analyses. We describe our raw statistical results here, then elaborate on them with examples in the Discussion section.
RQ1 and RQ2: Behavior and Representation

All of our stereotype categories are statistically associated with either male or female characters, and these effects are significant after correction (Table 2). We find men are more often strong (2.02 odds), arrogant (1.30 odds), sexual (1.22 odds), active (1.17 odds), dominant (1.15 odds), violent (1.10 odds), beautiful (1.06 odds), and angry (1.05 odds). Women are more often weak (1.73 odds), submissive (1.66 odds), childish (1.54 odds), afraid (1.46 odds), dependent (1.43 odds), hysterical (1.25 odds), domestic (1.16 odds), and emotional (1.03 odds).

When we analyze the same categories within individual genres, we mostly find consistent agreement with the overall effects (Table 3). All genres agree on male associations for strong, active, beauty, and dominant, and female associations for weak, submissive, childish, afraid, dependent, and hysterical. For the other stereotypes, we see a handful of genres where the gender association runs contrary to the overall effect. For example, men are more associated with the emotional stereotype in Action and Science Fiction, and women are more associated with the sexual stereotype in Adventure and Historical Fiction.

Investigating individual words within our stereotype categories, we find 409 significant associations, of which 222 are associated with female characters, and 187 with male characters (Figure 1). While the significant words for male characters are more or less evenly divided between verbs (53%) and adjectives (47%), the significant words for female characters consist mostly of adjectives (72%).

RQ3: How do stereotypes affect story ratings?

A logistic regression model that we trained on normalized stereotype counts for each character gender predicts story rating at an accuracy of 88%. As this model makes predictions over two classes (stories above and below the median rating), chance accuracy is 50%. In other words, gender stereotypes are strongly associated with story ratings. We present the categories of stereotype that are statistically significant for this model in Figure 4.

Positive associations with rating. Sexual, violent, hysterical, and angry stereotypes are positively associated with story rating for both male and female characters.
active, childish, and emotional stereotypes are also positively associated, but only for male characters. Submissiveness is positively associated with rating, but only for female characters.

Negative associations with rating. Domestic, strong, and weak stereotypes are negatively associated with story rating for both male and female characters. Afraid, dominant, emotional, beautiful, dependent, and active stereotypes are also negatively associated, but only for female characters.

RQ4: Do men and women write genders differently?

A logistic regression model that we trained on normalized stereotype counts for each gender predicts author gender at an accuracy of 53% (chance accuracy is 50%). If male and female authors wrote different kinds of gender stereotypes in their stories, this model would perform much higher than chance. So, it appears that both men and women write indistinguishably stereotypical genders. While this model provides little practical predictive value, one feature of the model did achieve significance after correction: the active stereotype for female characters is weakly associated with male authors. This means that men are slightly more likely to write female characters who take physical actions (verbs like “chase” or “lift”) in their stories.

In our second analysis, we find a Pearson correlation of 0.98 over odds ratios that describe the likelihoods of character actions and adjective descriptors written by male and female authors. 94% of these odds ratios agree in their direction (Figure 2). For example, both male and female authors associate punching with male characters, with 2.13 and 2.0 odds respectively. Alternatively, a small number (6%) of verbs and adjectives shift direction between male and female authors. For example, male authors are more likely to apply the adjective “immortal” to female characters, whereas female authors are more likely to apply it to male characters.

The data subsets for male and female authors uniformly support the high level relationships in Table 2, with slight deviations in the magnitudes of these trends. For example, female authors are slightly less likely to apply hysterical adjectives to female characters than male authors, but overall still quite likely to do so.

Discussion

We asked whether amateur writers, limited only by their imaginations, might contradict conventional gender stereotypes. To this question, it seems the answer is largely no. We also investigated whether (forward-thinking) genres or author genders are more likely to buck the trend. The answer, again, seems to be no.

RQ1-Behavior: How do characters act?

The statistically strongest male-oriented actions confirm gender stereotypes like arrogant, active, violent, angry, and sexual. Male characters more often smirk (2.7 odds) and mock (1.6 odds), dodge (1.5 odds) and chase (1.6 odds), abuse (4.4) and kill (1.5), roar (4.3 odds) and curse (1.6 odds), or kiss (2.4 odds) and lick (2.0 odds). For example, “he narrowed his eyes and mocked her laughter,” “he dodged attack after attack,” “he mentally cursed,” or “he tried to lick her ear.” Similarly, the statistically strongest female-oriented actions reinforce stereotypes through categories like childish, hysterical, submissive, afraid or domestic. Female characters more often giggle (6.7 odds) and

<table>
<thead>
<tr>
<th>Positive with rating</th>
<th>Coef.</th>
<th>Negative with rating</th>
<th>Coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>sexual (male)</td>
<td>2.03</td>
<td>strong (female)</td>
<td>-0.96</td>
</tr>
<tr>
<td>arrogant (male)</td>
<td>1.45</td>
<td>domestic (male)</td>
<td>-0.66</td>
</tr>
<tr>
<td>sexual (female)</td>
<td>1.24</td>
<td>afraid (female)</td>
<td>-0.66</td>
</tr>
<tr>
<td>violence (male)</td>
<td>0.92</td>
<td>weak (male)</td>
<td>-0.63</td>
</tr>
<tr>
<td>active (male)</td>
<td>0.90</td>
<td>domestic (female)</td>
<td>-0.57</td>
</tr>
<tr>
<td>hysterical (male)</td>
<td>0.59</td>
<td>strong (male)</td>
<td>-0.51</td>
</tr>
<tr>
<td>hysterical (female)</td>
<td>0.57</td>
<td>dominant (female)</td>
<td>-0.44</td>
</tr>
<tr>
<td>anger (male)</td>
<td>0.56</td>
<td>emotional (female)</td>
<td>-0.44</td>
</tr>
<tr>
<td>violence (female)</td>
<td>0.46</td>
<td>beautiful (female)</td>
<td>-0.39</td>
</tr>
<tr>
<td>childish (male)</td>
<td>0.42</td>
<td>weak (female)</td>
<td>-0.34</td>
</tr>
<tr>
<td>angry (female)</td>
<td>0.20</td>
<td>dependent (female)</td>
<td>-0.25</td>
</tr>
<tr>
<td>emotional (male)</td>
<td>0.12</td>
<td>childish (female)</td>
<td>-0.21</td>
</tr>
<tr>
<td>submissive (female)</td>
<td>0.02</td>
<td>active (female)</td>
<td>-0.03</td>
</tr>
</tbody>
</table>
bounced (2.7 odds), squeal (10.3 odds) and stomp (2.0 odds),
cook (2.3 odds) and wash (1.8 odds), cling (4.0 odds) and
whimper (2.3 odds), or shiver (3.2 odds) and scream (2.2
odds). For example, “she bounced up and down in excite-
ment,” “she would help clean the compound with all other
women in the clan,” or “she clung to him for a while.”

While none of our high-level stereotype categories run
directly against behavioral stereotypes, we do observe a
surprisingly weak male association with angry (1.05 odds).
Despite the “angry man” stereotype, many common behavioral
terms in this category are associated with female characters,
for example, stomp (1.7 odds), huff (1.4 odds), force (1.3
odds), and storm (1.3 odds). In contrast, the male associated
behaviors in this category tend to have a darker connotation,
like beat (2.0 odds), rip (1.7 odds), or snarl (1.3 odds).

RQ2-Description: How are characters described?
The most statistically male-oriented adjectives confirm gen-
der stereotypes through associations like strong, arrogant,
and dominant. Male characters are more often described as
intense (3.1 odds) and intimidating (2.1 odds), cocky (7.1
odds) and rude (1.4 odds), or rich (2.8 odds) and famous
(2.5 odds). For example, “this guy was so intimidating,” “he
was a typical cocky Californian guy,” or “his rich and famous
dad.” Likewise, the most statistically female-oriented adjectives
reinforce stereotypes like submissive, dependent, hyster-
tical. Female characters are more often described as help-
less (2.4 odds) and fragile (5.9 odds), clingy (4.4 odds) and
vulnerable (2.1 odds), or bitchy (12.2 odds) and dramatic
(2.1 odds). For example, “she went weak in the knees,” or
“like all overly dramatic teenage girls, Vanny got over her
breakup,” or “she was too clingy, so he dumped her.”

Against traditional gender stereotypes, we find that beau-
tiful is significantly male associated. This effect size is weak,
however, and both genders are often physically objectified
in our corpus. Women are more often described as beautiful
(3.3 odds), flawless (1.6 odds), curvy (46.0 odds), or slim
(3.5 odds). For example, “her pale skin was nearly flaw-
less.” But men are described with their own comparable
set of terms, including gorgeous (1.3 odds), hot (2.2 odds),
dreamy (3.0 odds), and toned (19.3 odds). For example, “she
watched him walk, shirtless and dangerously gorgeous.”

RQ1 and RQ2: Stereotypes among genres
Gender stereotypes are mostly consistent across genres, sug-
gest they are not driven by any one particularly popular
sub-community on Wattpad. For example, when we first
discovered the male association with beauty, we wondered
whether slash fiction (a popular sub-genre of Fanfiction fo-
cused on gay relationships) might have biased our results.
The genre analysis provides strong evidence to the contrary,
as the beauty association held across every genre in our cor-
pus. Similarly, we find that ten of the fifteen stereotype cat-
egories we analyze are completely consistent across genres.

Beyond providing evidence that stereotypes generalize
across Wattpad, the genre analysis opens up a series of new
questions. For example, why are women more associated
with sexual stereotypes in General Fiction, Science Fiction,
likely to describe male characters as “smarter,” “magical,” or “sparkling,” while men more often apply these adjectives to female characters.

Limitations

Our approach has limitations, which in turn provide opportunities for future work.

Sentence level analysis. In this paper, we analyze language at the level of sentences, examining verbs and adjectives for bias based on gender. While the concreteness of this approach allows us to efficiently draw out quantitative insight, it also excludes the greater context of the story. An independent female character dying after leading a rebellion, for example, tells us something different about gender bias than the death of a submissive and beautiful female character. Looking beyond the sentence level to story structure could allow us to consider how stereotypes interact with each other, or relate to earlier plot events in a story.

External Validity. Our work asks whether modern writers, free from market incentives, would reinforce or contradict gender stereotypes in their stories. There are several ways in which our results may not generalize to fully answer this question. First, while the Wattpad writing community is worth studying in its own right, it may not accurately represent the broader community of writers, amateur or otherwise. Second, writers on Wattpad may not be entirely free of commercial incentives; some writers have used the community as a launching point for their careers in commercial fiction, and many writers desire as wide an audience as possible. However, given the immense scale and diversity of the Wattpad dataset (more than 1.8 billion words, split almost evenly across male and female authors), we suggest that it provides the best available means towards answering our questions.

Conclusion

We ask whether modern amateur writers, limited only by their imaginations, reinforce or contradict gender stereotypes. To find out, we analyze more than 1.8 billion words of fiction on Wattpad, an online writing community. We find that modern amateur writers do reinforce traditional stereotypes, and that male and female authors are just as likely to perpetuate them.

Acknowledgments

First, we would like to thank Jordan Christensen and Mohammad Islam at Wattpad – this work wouldn’t be possible without their help and collaboration. We would also like to thank our reviewers and colleagues at Stanford for their helpful feedback. This work is supported by a NSF Graduate Fellowship.

References


Emons, P.; Wester, F.; and Scheepers, P. 2010. He works outside the home; she drinks coffee and does the dishes: Gender roles in fiction programs on dutch television.


