When the Hero Becomes a Girl: Presenting Characters' Gender with Stereotypes in AO3 Gender-Bending Fanfiction

Yixi Chen[1], and Jianwei Yan[1]

[1] Department of Linguistics, Zhejiang University, Hangzhou, China

Abstract

Gender-bending fanfiction is a subcategory of fanfiction where characters are transformed into a different gender. This study examines gender presentation in gender-bending fanfiction on AO3 (Archive of Our Own). The findings reveal that gender-bent characters are marked by biased words. These biased words predominantly reflect gender stereotypes and even negative experiences of femininity, such as body objectification and vulnerability. It indicates that gender-bending involves not merely a biological sex change, but a modification in characters' personality and behavior through language. This study may provide valuable insights into how fan communities perceive the gender hierarchy in the real world and how their experiences shape their fanfiction creation.

Keywords: gender stereotype, bias modeling, gender-bending, fanfiction, AO3

1 Introduction

Fanfiction refers to fan-created stories based on an existing work of literature. As a non-profitable behavior, fanfiction is primarily motivated by authors' interests and pleasures [4], as well as the recognition from readers. This genre stands out with its freedom in creation and transmission [19]. Although fanfiction is derivative stories from the canon work, authors are allowed to reshape characters and their relationships in an unconventional way [4]. By participating in the re-creation of the canon, authors explore and express their understanding of characters and original works.

Unlike mainstream media, fanfiction normalizes the queering presentation of gender and relationship [12; 15]. A prominent example is slash fanfiction, a major subgenre of fanfiction that typically involves homosexual relationships between (or among) characters and gender-bending [1; 25]. Such queering presentation "rethink[s] and rewrite[s] traditional masculinity" [12, p. 71], challenging the assumption "the [cishetero-]male as neutral, and default" [1, p. 26]. In this sense, fanfiction provides a valuable lens for researchers to examine the perception of queer identity and sexuality in fictional narratives.

While male protagonists are prevalent in real-world works, their presentation in fanfiction is often reshaped. The canon work may present the male protagonist as a hero who embarks on a journey full of obstacles with his allies to save the world [7]. When it comes to fanfiction, fan authors frequently focus on the hero's entanglements with other characters rather than the hero's journey. Some authors even rewrite the hero as a heroine, exploring the imaginative alternative of a male protagonist in a female body [4; 19]. This practice, known as gender-bending, is especially popular in the form of male-to-female transformations. While fewer studies provide discussion on

Yixi Chen, and Jianwei Yan. "When the Hero Becomes a Girl: Presenting Characters' Gender with Stereotypes in AO3 Gender-Bending Fanfiction." In: *Computational Humanities Research 2025*, ed. by Taylor Arnold, Margherita Fantoli, and Ruben Ros. Vol. 3. Anthology of Computers and the Humanities. 2025, 721–735. https://doi.org/10.63744/d0T9FeacxMZ5.

its bias for male-to-female, we can observe this in the subtags of "genderbend" on AO3 (Archive of Our Own), a renowned nonprofit, open source repository of fanfiction).[1]

The prevalence of gender-bending fanfiction naturally follows the question of the motivations behind it. One key motivation is heterosexual validation. Much gender-bending fiction revolves one or more slash relationships (i.e., same sex non-canonical relationships). Some readers comment that gender-bending reveals the author's attempt at "heterosexualizing a homosexual pairing" [1, p. 37] [19]. In male-to-female fanfiction, one of the homosexual couple becomes the female, making the relationship safe for the social norm [6, p. 70]. Some may even interpret regendering as a validation that "(the partners) aren't Gay, they're heterosexual men who just happen to fall in love with each other" [12, p. 75]. Whether the heterosexualization is self-motivated or not, the existence of a female party allows more intimacy in the relationship and speaks for authors' and readers' appreciation of romance [6].

A more important factor lies in the identities of fan authors themselves. The non-profitable and anonymous transmission of fanfiction empowers fan authors to "choose to foreground one identity marker and ignore another" [6, p. 33]. Given that female and queer authors constitute the overwhelming majority of fanfiction creators, their works frequently "talk back" to the male-dominant canon [4; 10; 16], reconsidering "an oversaturated space [by cishetero-male protagonists]" [1, p. 38]. Accordingly, gender-bending serves for the pleasure of non-male writers and readers, as a more gender-inclusive alternative to the (probably) male-centric canon.

Beyond personal enjoyment, the creation of gender-bent heroes is embedded within broader fan community dialogues. The practice of gender-bending raises a reflexive question "as feminist theorists from Simone de Beauvoir to Judith Butler have done, of what it means to be a woman" [6, p. 61] for both authors and readers. Thus, gender-bending fanfiction becomes a meaningful space for marginalized and non-default groups to explore and negotiate the conception of gender in the real world.

To explore the conception of gender in gender-bending fanfiction, the central question is whether gender-bending induces bias in the protagonists' behavior based on gender. In this study, we focus on male-to-female gender-bending fanfiction in English. Therein, *bias* refers to distributional differences in word usage associated with characters. If such biases exist, they may offer deeper insights into how gender influences the portrayal of identical characters' behaviors. These gender-influenced behaviors are referred to as *stereotypes* accordingly, reflecting the social constructs underlying femininity and masculinity.

Unlike previous qualitative studies, this research operationalizes gender presentation by extracting feature words related to characters. We examine the biases and stereotypes in their usage. The goal is to better understand how gender is conceived and presented in the context of gender-bending, particularly how femininity is projected onto typically male characters within a male-default cultural framework. Specific research questions are as follows:

1. Does gender-bending lead to gender-based bias in the linguistic presentation of fictional characters?

2. If so, in what ways are gender stereotypes manifested in the presentation of gender-bent characters in fanfiction?

[1] `https://archiveofourown.org/tags/Changes%20to%20Gender%20or%20Sex`

2 Data and methods

2.1 Data

2.1.1 Scraping and pre-processing

We focused on the top five characters with the most uses under the tag `fem!` on AO3. The exclamation mark `!` is a special marker on AO3. It is preceded by the character's attribute and followed by the character's name. For example, the tag `fem!Harry Potter` stands for Harry Potter as a female character. Other synonymous tags include `genderbent`, `trans!`, and `female!`. Notably, synonymous tags on AO3 are linked together and return the same search results, which greatly facilitated our searching. We thereby searched results with `fem!` as the tag name and "character" as the type. Results are then sorted by uses in descending order. Accordingly, we decided on five characters as shown in Table 1.

Character	Search rank	Raw tag(s)	Fandom
Harry Potter	3, 5	Fem Harry Potter, Fem!Harry Potter	Harry Potter
Stiles Stilinski	4	Fem!Stiles	Teen Wolf
Tony Stark	7	Fem!Tony Stark	Iron Man
Bilbo Baggins	10	Fem!Bilbo - Character	The Hobbit
Crowley	12	fem!Crowley - Character	Good Omens

Table 1: Top five characters in the search results (retrieved on 28 May, 2025).

We then searched AO3 for these characters' works with `fem!` tag (**FEM** works) and without `fem!` tag (**NON** works). To ensure relevance to the original canon, crossover works were excluded. We also controlled for text length by selecting works with word counts between 1,000 and 20,000. Results are sorted by kudos—a metric of readers' recognition of works—in descending order. Our dataset includes the top 100 FEM works and top 80 NON works of each character from the sorted results. This aims to balance the total word count between the two categories because FEM works tend to have lower word counts than the other. General information is presented in Table 2. Notably, the search results of `fem!Stiles`[2] and `fem!Crowley - Character`[3] yielded less than 100 qualifying works. We thus included all available works in our data. All the scraping was conducted on 15 June 2025 without logging in.

To facilitate our extraction of characters' features, we performed coreference resolution in Python 3.7.16 with the `coreferee` library. This tool automatically replaced reference pronouns with the corresponding proper names according to the context. The used library `coreferee` used a mixture of neural networks and programmed rules, demonstrating 81% of accuracy in its training corpora [13], making it a reliable choice for this study.

2.1.2 Extracting feature words

After resolving the coreference, we extracted each character's features with dependency parsing. Dependency parsing was conducted by `SpaCy` 3.7.4 [18] in Python 3.12.0. Based on the prior study [2; 3], we focused on the following four categories of words and adapted their definition to better align with the annotation scheme of `SpaCy` dependency parser. Table 3 summarizes the distribution of word lemmas across the four types.

[2] This tag returned 98 English works in total without any filtering, of which 32 were outside the controlled word count range.

[3] This tag returned 175 English works in total without filtering; 39 were outside the word count range, and 101 involved crossovers.

Category	Character	WC sum	WC median	WC min	WC max	N
FEM	Harry Potter	732,644	5,769.0	1,053	19,632	100
(N = 415)	Stiles Stilinski	354,138	3,652.0	1,012	17,967	62
	Tony Stark	626,759	4,577.0	1,041	18,441	100
	Bilbo Baggins	625,598	4,285.5	1,040	19,916	100
	Crowley	284,044	4,041.0	1,008	18,840	53
	(Total)	2,623,183	4,391.0			415
NON	Harry Potter	753,515	9,867.5	1,200	19,675	80
(N = 400)	Stiles Stilinski	660,658	6,325.5	1,423	19,773	80
	Tony Stark	654,997	6,647.5	1,499	19,587	80
	Bilbo Baggins	575,321	5,253.0	1,028	19,751	80
	Crowley	542,985	5,373.0	1,100	19,856	80
	(Total)	3,187,476	6,295.0			400

Table 2: General information on the dataset (WC, word count).

1. Agent. Actions conducted by a character as the agent (i.e., verbs in a `nsubj` relation with the character),
2. Patient. Actions conducted on a character as a patient (i.e., verbs in a `nsubjpass` or `dobj` relation with the character),
3. Possessive. Objects that belong to a character (i.e., nouns in a `poss` or `compound` relation with the character), and
4. Predicative. Attributes related to a character (i.e., nouns, adjectives, and adverbs in an `advmod`, `acomp`, or `attr` relation with the character's agent verbs).

Category	Role	LC sum	LC median	LC mean	LC std
FEM	Agent	25,929	31.0	62.48	80.86
(N = 415)	Patient	5,993	8.0	14.44	17.94
	Possessive	8,958	10.0	21.59	30.85
	Predicative	9,319	11.0	22.46	30.83
NON	Agent	57,506	102.5	143.77	130.92
(N = 400)	Patient	11,851	22.0	29.63	27.41
	Possessive	21,082	32.5	52.71	57.15
	Predicative	24,097	41.0	60.24	56.22

Table 3: Distribution of agent, patient, possessive, and predicative word lemmas (LC, lemma count).

2.2 Methods

2.2.1 Detecting bias in feature words

The first research question investigates the bias in word usage by character gender. Bias was measured as the strength of association between characters' features (word lemmas) and their gender

(context). For this goal, pointwise mutual information (PMI) is a widely applied metric of word-context association in prior studies [8].

Differently, the current study utilized normalized PMI (NPMI) to quantify the association strength [5]. Compared to PMI, NPMI reduces the sensitivity for low-frequency cases and provides a comparable result in a unified scale despite the varied data size [5]. This is particularly important here, as low-frequency words constitute a considerable portion of our data, and there is an imbalance in the number of feature words by gender. To ensure representativeness, we set a frequency threshold of ten occurrences. Any tuple (r, w) occurring less than ten times across all works was excluded from the following analysis. [4]

We calculated NPMI for all the word lemmas and sorted them in descending order by role and gender as Formula 1:

$$\text{NPMI}\left((r,w);g\right) = \frac{\text{PMI}\left((r,w);g\right)}{h\left((r,w),g\right)} = \frac{\log \frac{p\left((r,w),g\right)}{p(r,w)p(g)}}{-\log\left(p\left((r,w),g\right)\right)} \tag{1}$$

Therein, (r, w) is the tuple of role (r, one of agent, patient, possessive, predicative) and word lemma (w). $p(r, w)$ equals the frequency of the tuple (r, w) divided by the number of all word lemmas specified by the role r. g is the category of works (FEM or NON). For example, the agent verb *love* is represented as $(\texttt{agent}, \texttt{love})$. Its probability $p(r, w)$ thus equals the frequency of $(\texttt{agent}, \texttt{love})$ divided by the total number of agent verbs.

NPMI ranges between -1 (negative association) and +1 (positive association). A greater value indicates a stronger association between the (r, w) tuple and the given gender g. Eventually, the top 100 tuples (r, w) were respectively selected from FEM and NON works to represent the biased featured words of characters.

2.2.2 Word embedding and clustering

After extracting the biased feature words, we examined the presentation of characters' gender by exploring their topic clusters, responding to the second research question. We assumed that characters' gender is reflected in their behavior, represented by the latent topics of extracted feature words ($N_{\text{word}} = 100 \times 4 \times 2 = 800$). These latent topics were operated as word clusters based on their embeddings [14].

We first obtained the 300-dimensional embeddings of the selected feature words from the word2vec model [17; 21] trained on all the scraped works ($N_{\text{work}} = 815$, see Table 2). This model generates word embeddings based on their co-occurrences in the text, on the assumption that words that occur in similar contexts are similar in meaning [11]. We present the most similar words to the sample word *love* in our word2vec model in Table 4.

Rank	Word lemma	Similarity	Rank	Word lemma	Similarity
1	hate	0.7324	6	honest	0.5381
2	trust	0.5762	7	miss	0.5373
3	forgive	0.5659	8	owe	0.5353
4	mate	0.5651	9	marry	0.5238
5	deserve	0.5611	10	wish	0.5223

Table 4: Top ten most similar words to *love* (similarity is defined as Levenshtein similarity in gensim [22]).

[4] Distribution of filtered word lemmas by role can be found in Table 6 in Appendix A.

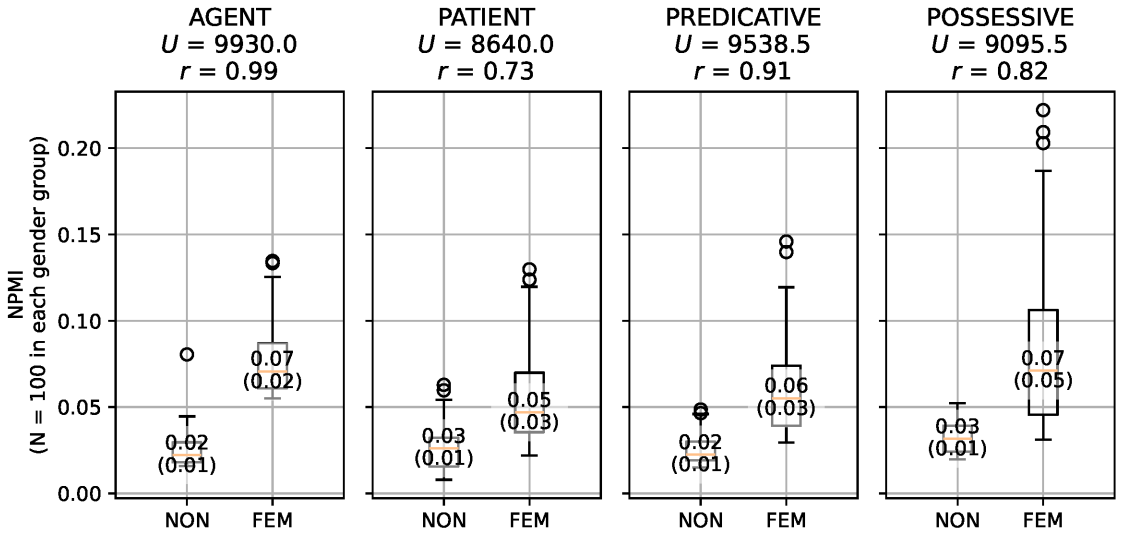

Figure 1: Comparison of NPMI by role and gender. Mann-Whitney U test is used with the rank-biserial correlation r as the effect size. Median and standard deviation (bracketed) are also annotated.

Word embeddings were then reduced to two dimensions with t-SNE, which is capable of capturing the local structure of high-dimensional data while reducing dimensions [23]. In other words, this method can retain the semantic similarity among 300-dimensional embeddings while projecting them in a two-dimensional space. We applied t-SNE to eight groups of biased feature words by role (i.e., $r \in \{\texttt{agent}, \texttt{patient}, \texttt{possessive}, \texttt{predicative}\}$) and gender (i.e., $g \in \{\texttt{FEM}, \texttt{NON}\}$). The perplexity parameter for each t-SNE model was optimized to maximize the trustworthiness score.

After the dimensionality reduction with t-SNE, we applied k-means clustering to obtain the k latent topic clusters within each group. Similarly, the number of clusters k was optimized by maximizing the Silhouette score. Clustering the word embeddings enabled us to discover the latent topics shared by the word lemmas in the same cluster. Both t-SNE and k-means clustering were conducted with the help of `scikit-learn` 1.6.1 [20] in Python 3.12.0.

3 Results and discussion

3.1 Biased feature words for gender-bent characters

NPMI indicates the strength of association between feature words and the given gender. As aforementioned, we sorted feature words by NPMI and defined the top 100 feature words in each group by role and gender as the biased feature words. Notably, no overlapped biased words of the same role were found in FEM and NON works. The top ten biased words are shown as samples in Table 7 in Appendix A.

Figure 1 presents the distribution of NPMI by role and gender. Therein, we conducted the Mann-Whitney U test (two independent samples) to compare the group-wise difference within each role.[5] All biased words in FEM works have a significantly higher NPMI than their counterparts in NON works given the same role, demonstrated by their effect sizes r greater than 0.5 [9]. It indicates that biased words in FEM works reveal a stronger association with their gender compared

[5] We applied the Mann-Whitney U test instead of t-test because the NPMI of biased words is not normally distributed (Shapiro-Wilk test: $p < .0001$).

to those in NON works.[6] The stronger association aligns with the concept of "markedness" [24], or non-defaultness. In this context, gender-bent characters are marked by their behaviors, possessions, and attributes—represented by the feature words analyzed in this study—whereas their cis-gendered counterparts in NON works are comparatively less marked.

While the bias is proven significant, NPMI remains overall low in biased agent, patient, and predicative words (all below 0.15). In contrast, possessive words (median = 0.07) demonstrate a higher NPMI than the other three roles in both FEM (U = 17,562, p = .01, r = .17) and NON works (U = 21,660.5, p < .001, r = .44). This suggests that characters' possessions are more salient indicators of gender, although behaviors and attributes also contribute to gender representation.

Our results confirm the biased usage of feature words in FEM and NON works, where possession words stand out with their higher NPMI therein. While *bias* in the current study is defined as the statistical difference in distribution, it indicates that femininity is marked in the context of gender-bending. However, the non-defaultness of femininity only partially accounts for this markedness. It also arises from the authors' performance of gender-bending [19]. In other words, gender-bending in fanfiction is not merely a shift in biological sex but also a modification of the character's personality as conveyed through the text. To depict a protagonist as female, the narrative must elaborate on what being female means to the character, or how the gender-bent character differs from their cis-gendered counterpart. Thus, biased word usage in gender-bending fanfiction provides a window into how fan authors present gender in their writing. We further the discussion by examining their latent topics in the next section.

3.2 Stereotypes in biased feature words

Figure 2 illustrates the clusters of biased feature words by role and gender. Therein, feature words of each role are classified into four to eight clusters, optimized based on the trustworthiness score. Generally, the results indicate that feature words belong to several topics. Agent and possession words can be segmented into four major topics (k = 4) while patient (k = 8) and predicative words (k = 6) may be more diverse. Each topic has its biased words for both femininity (represented by FEM ◯) and masculinity (represented by NON ♢). We summarized the topics for all word clusters in Table 5.

Agent Biased agent words involve four topics: AFFECTIVE, COMMUNICATIVE, INTERACTIVE, and ABRUPT. Therein, FEM-biased words are more densely distributed in the AFFECTIVE, COMMUNICATIVE, and ABRUPT clusters. This distribution suggests that FEM-biased words are strongly associated with communication and affection, conveying tenderness (*cuddle*, *bend*, *clap*) and lightheartedness (*giggle*, *chuckle*, *tease*).

By contrast, NON-biased words are predominantly found within the INTERACTIVE category. While some of them are related to communication (*invite*, *confirm*, *suggest*), their meanings tend to be more neutral and equal in terms of social power. Interestingly, FEM-biased INTERACTIVE words often appear at the boundary with the AFFECTIVE and COMMUNICATIVE clusters, implying an overlap among the three topics.

Patient Patient words cover the widest range of topics among the four categories. In the ASSAULTIVE, INTIMATE, and INTERACTIVE clusters, FEM-biased words reflect more intensity and concreteness in expressions of both intimacy (*hug*, *embrace*, *kiss*, *cradle*) and violence (*wipe*, *slap*, *slam*, *beat*, *attack*). These two topics tend to be more general in their NON-biased counterparts (INTIMATE: *touch*, *ease*; ASSAULTIVE: *hit*, *punch*, *kick*).

FEM-biased words also dominate the MOVEMENT and COMMUNICATIVE topics, highlighting gender-bent characters' active engagement in these areas. FEM-biased COMMUNICA-

[6] As aforementioned, NPMI enables us to avoid the effect of low-frequency data as well as the unbalanced data size on the strength of association, see [5].

TIVE cluster is adjoined to a considerable part of AFFECTIVE words (*attract, resent, betray, like*) and AFFE-INTERACTIVE words (*involve, miss, remember*). Whereas, this association is confined to the overlap between AFFECTIVE and AFFE-INTERACTION for NON-biased words, with a clear-cut separation between AFFECTIVE and COMMUNICATIVE words. It thus suggests that communication is distinctly defined in FEM and NON works.

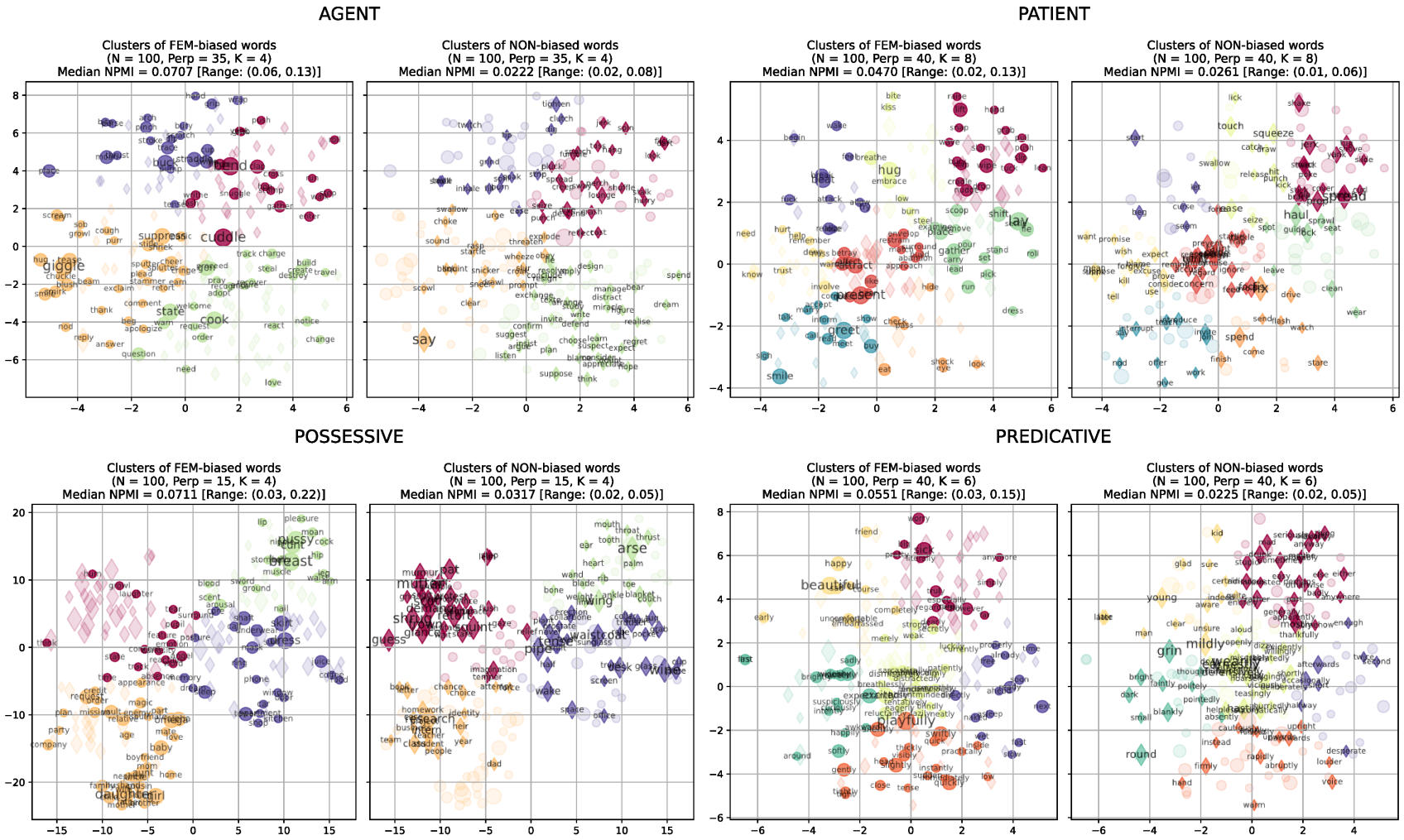

Figure 2: Clusters of biased words by role and gender. Each cluster is represented by different colors, while we marked two genders with ◯ (FEM, left panel for each subplot) and ♢ (NON, right panel for each subplot) respectively. Each dot represents the annotated feature word in the embedding space, whose size and alpha positively correlate with their NPMI.

Possession Possessive words involve four topics: AFFE-COMMUNICATIVE, RELATIONAL, PHYSICAL, and TANGIBLE. Among these, AFFE-COMMUNICATIVE, referring to possessive words that convey affection or emotion within communication, is the only topic cluster where NON-biased words dominate (see Figure 2, lower-left subplot). Specifically, NON-biased AFFE-COMMUNICATIVE words are more mild and neutral in emotion (*shrug*, *mutter*, *squint*) than some FEM-biased ones (*emotion, laughter*, *tear*, *growl*).

The other three clusters are relatively balanced in distribution. FEM-biased TANGIBLE words contain more words on feminine clothing items (*skirt, dress*), while NON-biased ones are generally less gender-specified (*wine*, *trousers*, *pocket*; except for *waistcoat* and *pipe*). Similarly, FEM-biased RELATIONAL words are also feminine-specific (*daughter*, *girl*), in contrast to the less marked NON-biased ones. While some PHYSICAL words are sexually-related due to fanfiction's focus on relationship, these parts of words yield a higher NPMI in FEM works than NON ones, demonstrated by their larger dot sizes.

Predicative Of six topics of predicative words, MANNER is FEM-dominant while NEG-AFFECTIVE is NON-dominant in distribution. FEM-biased MANNER words typically convey playfulness (*playfully*, *swiftly*, *slightly*), whereas NON-biased MANNER words tend to express more intensity or firmness (*rapidly*, *abruptly*, *firmly*). In the NEG-AFFECTIVE cluster, FEM-biased words are related to vulnerability (*sick*) or serve as intensifiers of degree (*pretty*, *dearly*, *truly*). Similarly, FEM-biased AESTHETIC words also implies feminine vulnerability (*beautiful*, *sad*, *embarrassed*), compared to NON-biased words being more neutral (*mildly*, *young*).

While slash fanfiction is defined by its focus on homosexual relationships, our analysis of biased words suggests that its gender-bending subcategory often portrays characters in ways that align with real-world gender stereotypes. These stereotypes typically include biological femininity (PHYSICAL in possession words), tenderness (AFFECTIVE and COMMUNICATIVE in agent words, MANNER in predicative words), and emotional subtlety (AFFE-COMMUNICATIVE in patient and possession words, NEG-AFFECTIVE and AESTHETIC in predicative words), even though their subjects are originally male.

On the one hand, the conformation to feminine stereotypes is commented to "turn one half of a popular slash ship into a 'straight' ship" [1, p. 37] [12]. With one of the characters transformed into a female, the homosexual relationship shifts to a heterosexual one, naturally adopting the norms and expectations of heterosexuality. This shift in sexuality induces changes in how the gender-bent characters behave and are treated in the relationship. Accordingly, it raises the question "how much of their feelings are generated by the person they love and how much are merely a reaction to their biological sex" [6, p. 52] to those in the text and the question "how much of their personality are left when their sex is subverted" to those out of the text. Nevertheless, the results in the current study indicate that the presentation of gender, rather than personality, may play a more decisive role in presenting gender-bent characters.

On the other hand, as aforementioned, the presentation of gender reflects fan authors' perception of gender in the real world, whether consciously intended or not. Some biased words further depict the negative experiences of being a female: sexual objectification (PHYSICAL in possession words) and intimacy-mixed violence (ASSAULTIVE and INTIMATE in patient words). When these experiences are transferred to a male protagonist—who used to be the hero of stories—the misplacement provides access to non-default authors (as well as readers) to challenge the cisheteronormative conventions of mainstream media and express their desire for intimacy and sexuality [6; 10]. Last but not least, gender-bending is a part of the ongoing negotiation among fan authors and readers, where they rewrite the character and reconsider the legitimacy of their canonical personality [1]. Whether the presentation is canon or not, characters therein are nevertheless agents for non-default gendered readers' perception, experience, and even critiques of gender hierarchy in the real world.

Topic	Sample words (FEM)	Sample words (NON)
Agent (k = 4)		
AFFECTIVE	cuddle, bend, clap	dive, lounge, crash
COMMUNICATIVE	giggle, suppress, tease	say, obey, slur
INTERACTIVE	cook, state, question	write, think, suppose
ABRUPT	buck, straddle, moan	tighten, dig, twitch
Patient (k = 8)		
ASSAULTIVE	wipe, lift, nudge	spread, prop, jerk
INTIMATE	hug, breathe, embrace	squeeze, ease, touch
INTERACTIVE	beat, break, realize	start, beg, let
MOVEMENT	lay, place, gather	haul, lock, clean
AFFECTIVE	present, attract, restrain	focus, resign, chide
AFFE-INTERACTIVE	involve, miss, remember	think, prove, expect
NEUTRAL	eat, check, pass	fix, spend, finish
COMMUNICATIVE	greet, smile, buy	invite, teach, join
Possession (k = 4)		
AFFE-COMMUNICATIVE	angel, emotion, state	shrug, mutter, squint
RELATIONAL	daughter, girl, omega	research, intern, class
PHYSICAL	breast, pussy, cunt	arse, wing, blanket
TANGIBLE	dress, skirt, tower	wine, pipe, tense
Predicative (k = 6)		
NEG-AFFECTIVE	sick, however, dearly	mostly, rather, drunk
AESTHETIC	beautiful, happy, sad	mildly, young, aware
TEMPORAL	already, ahead, soon	afterwards, occasionally, twice
AFFECTIVE	excitedly, dismissively, tenderly	wearily, earnestly, defensively
POS-AFFECTIVE	expectantly, shyly, softly	grin, round, blankly
MANNER	playfully, swiftly, silently	firmly, backwards, rapidly

Table 5: Topics and sample words (top three words with the highest NPMI) of word clusters.

4 Conclusion

This study focuses on the gender presentation in gender-bending fanfiction on AO3. We operationalized gender presentation with feature words extracted with dependency parsing and investigated their biased usage and topic distribution within fanfiction.

The results indicate that biased feature words mark gender-bent characters in slash fanfiction. Possession words, in particular, show the strongest gender associations. In addition, these biased words strongly align with conventional feminine stereotypes, such as biological femininity, tenderness, and emotional subtlety. These findings suggest that gender-bending involves not just a change in biological sex but also a performed modification of personality and behavior through language.

Our study highlights the complex dynamics of gender presentation in fan communities in a descriptive manner. Despite the prevalence of stereotypes therein, gender-bending serves as a "ironic playground" [6, p. 62] for non-default gendered writers and readers to mock the cisheteronormative norms by repositioning male protagonists within feminine experiences. It offers insights into how fan communities perceive and experience the gender hierarchy they live with and how these experiences shape their reading and writing. Methodologically, our study contributes a robust framework for modeling gender stereotypes in character presentation by integrating feature

extraction with topic modeling techniques. This approach provides nuanced insights into gender presentation in fanfiction and practical tools for further research on character representation and language bias.

The study's main limitations lie in its quantitative focus, which may overlook individual motivations and contextual nuances best captured through qualitative inquiry. Additionally, focusing solely on English-language fanfiction may limit the generalizability of the results. Future research that integrates qualitative methods and examines broader cultural contexts could offer a more comprehensive understanding of gender performance in fanfiction.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities [grant number S20240030]. We also thank the anonymous reviewers for their helpful comments.

References

[1] Baker, Lucy. *Media and Gender Adaptation: Regendering, Critical Creation and the Fans*. 1st ed. London: Bloomsbury Academic, 2023. DOI: 10.5040/9781501370076.

[2] Bamman, David, O'Connor, Brendan, and Smith, Noah A. "Learning Latent Personas of Film Characters". In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, ed. by Hinrich Schuetze, Pascale Fung, and Massimo Poesio. Sofia, Bulgaria: Association for Computational Linguistics, Aug. 2013, pp. 352–361.

[3] Bamman, David, Underwood, Ted, and Smith, Noah A. "A Bayesian Mixed Effects Model of Literary Character". In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, ed. by Kristina Toutanova and Hua Wu. Baltimore, Maryland: Association for Computational Linguistics, June 2014, pp. 370–379. DOI: 10.3115/v1/P14-1035.

[4] Barnes, Jennifer L. "Fanfiction as Imaginary Play: What Fan-Written Stories Can Tell Us about the Cognitive Science of Fiction". In: *Poetics 48* (Feb. 2015), pp. 69–82. DOI: 10.1016/j.poetic.2014.12.004.

[5] Bouma, Gerlof. "Normalized (pointwise) mutual information in collocation extraction". In: *Proceedings of GSCL 30* (2009), pp. 31–40.

[6] Busse, Kristina. *Framing Fan Fiction: Literary and Social Practices in Fan Fiction Communities*. University of Iowa Press, 2017. DOI: 10.2307/j.ctt20q22s2. JSTOR: j.ctt20q22s2.

[7] Campbell, Joseph. *The Hero with a Thousand Faces*. 3rd ed. Bollingen Series XVII. Novato, Calif: New World Library, 1949.

[8] Church, Kenneth Ward and Hanks, Patrick. "Word Association Norms, Mutual Information, and Lexicography". In: *Computational Linguistics 16*, no. 1 (1990), pp. 22–29.

[9] Cureton, Edward E. "Rank-Biserial Correlation". In: *Psychometrika 21*, no. 3 (Sept. 1956), pp. 287–290. DOI: 10.1007/BF02289138.

[10] Duggan, Jennifer. ""Worlds. . .[of] Contingent Possibilities": Genderqueer and Trans Adolescents Reading Fan Fiction". In: *Television & New Media 23*, no. 7 (Nov. 2022), pp. 703–720. DOI: 10.1177/15274764211016305.

[11] Firth, J. R. "Applications of General Linguistics". In: *Transactions of the Philological Society 56*, no. 1 (Nov. 1957), pp. 1–14. DOI: 10.1111/j.1467-968X.1957.tb00568.x.

[12] Green, Shoshanna and Jenkins, Cynthia. “3 “Normal Female Interest in Men Bonkbaring”: Selections from the Terra Nostra Underground and Strange Bedfellows”. In: *Fans, Bloggers, and Gamers: Exploring Participatory Culture*. New York University Press, Sept. 2006, pp. 61–88.

[13] Hudson, Richard. “Richardpaulhudson/Coreferee”. Oct. 2025.

[14] Lewis, Molly, Cooper Borkenhagen, Matt, Converse, Ellen, Lupyan, Gary, and Seidenberg, Mark S. “What Might Books Be Teaching Young Children About Gender?” In: *Psychological Science 33*, no. 1 (Jan. 2022), pp. 33–47. DOI: 10.1177/09567976211024643.

[15] Llewellyn, Anna. ““A Space Where Queer Is Normalized”: The Online World and Fanfictions as Heterotopias for WLW”. In: *Journal of Homosexuality 69*, no. 13 (Nov. 2022), pp. 2348–2369. DOI: 10.1080/00918369.2021.1940012.

[16] McInroy, Lauren B. and and Craig, Shelley L. “Online Fandom, Identity Milestones, and Self-Identification of Sexual/Gender Minority Youth”. In: *Journal of LGBT Youth 15*, no. 3 (July 2018), pp. 179–196. DOI: 10.1080/19361653.2018.1459220.

[17] Mikolov, Tomas, Chen, Kai, Corrado, Greg, and Dean, Jeffrey. “Efficient Estimation of Word Representations in Vector Space”. Sept. 2013. DOI: 10.48550/arXiv.1301.3781. arXiv: 1301.3781 [cs].

[18] Montani, Ines, Honnibal, Matthew, Boyd, Adriane, Landeghem, Sofie Van, and Peters, Henning. “spaCy: Industrial-strength Natural Language Processing in Python”. Zenodo. Oct. 2023. DOI: 10.5281/ZENODO.1212303.

[19] Oulton, Harry. “Harry Potter and the Social Construct. Does Gender-Swap Fanfiction Show Us That We Need to Re-consider Gender Within Children’s Literature?” In: *Childrens Literature in Education 55*, no. 3 (Sept. 2024), pp. 465–482. DOI: 10.1007/s10583-022-09518-4.

[20] Pedregosa, F. et al. “Scikit-Learn: Machine Learning in Python”. In: *Journal of Machine Learning Research 12* (2011), pp. 2825–2830.

[21] Rehurek, Radim and Sojka, Petr. “Gensim–python framework for vector space modelling”. In: *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic 3*, no. 2 (2011).

[22] Rehurek, Radim and Sojka, Petr. “Software Framework for Topic Modelling with Large Corpora”. English. In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Valletta, Malta: ELRA, May 2010, pp. 45–50.

[23] Van der Maaten, Laurens and Hinton, Geoffrey. “Visualizing Data Using T-SNE.” In: *Journal of machine learning research 9*, no. 11 (2008).

[24] Waugh, Linda R. “Marked and Unmarked: A Choice between Unequals in Semiotic Structure”. In: *Semiotica 38*, no. 3-4 (Jan. 1982), pp. 299–318. DOI: 10.1515/semi.1982.38.3-4.299.

[25] Yang, Xiaoyan and Pianzola, Federico. “Exploring the Evolution of Gender Power Difference through the Omegaverse Trope on AO3 Fanfiction⋆”. In: *Proceedings of the Computational Humanities Research Conference 2024*. Vol. 3834. Denmark: CEUR Workshop Proceedings, Dec. 2024, pp. 906–916.

A Descriptive statistics and sample of biased feature words by role in FEM and NON works

See Tabl 6 and Table 7.

	FEM		NON		(Total)	
Role	LC sum	Unique LC sum	LC sum	Unique LC sum	LC sum	Unique LC sum
Agent	24,771	1,463	54,396	1,934	79,167	2,219
Patient	5,851	874	11,545	1,229	17,396	1,430
Possessive	8,913	1,255	20,946	1,919	29,859	2,297
Predicative	7,330	1,179	18,478	1,961	25,808	2,273

Table 6: Distribution of agent, patient, possessive, and predicative word lemmas (LC, lemma count) after filtering out low-frequent words (threshold = 10).

		FEM		NON	
Role	Rank	Word	NPMI	Word	NPMI
Agent	1	giggle	0.1346	say	0.0805
	2	cuddle	0.1334	dive	0.0446
	3	bend	0.1334	lounge	0.0436
	4	cook	0.1253	tighten	0.0423
	5	state	0.1202	obey	0.0387
	6	suppress	0.1168	write	0.0382
	7	buck	0.1168	think	0.0377
	8	gon	0.1104	suppose	0.0366
	9	straddle	0.1066	fle	0.0361
	10	tease	0.1061	crash	0.0351
Patient	1	lay	0.1298	fix	0.0629
	2	present	0.1238	spread	0.0597
	3	hug	0.1197	haul	0.0543
	4	greet	0.1164	squeeze	0.0469
	5	beat	0.1039	ease	0.0464
	6	smile	0.1	prop	0.0464
	7	attract	0.1	focus	0.0464
	8	place	0.0983	spend	0.0462
	9	gather	0.096	touch	0.0446
	10	shift	0.0948	jerk	0.0446
Predicative	1	playfully	0.1459	wearily	0.0486
	2	beautiful	0.1398	grin	0.0465
	3	sick	0.1194	mildly	0.046
	4	excitedly	0.116	earnestly	0.045
	5	swiftly	0.1104	round	0.0411
	6	expectantly	0.1052	defensively	0.0411
	7	silently	0.1043	hastily	0.0396
	8	happy	0.1029	young	0.0388
	9	quickly	0.1024	mostly	0.0385

Role	Rank	FEM		NON	
		Word	NPMI	Word	NPMI
	10	dismissively	0.102	rather	0.0351
Possessive	1	breast	0.222	shrug	0.0523
	2	daughter	0.2093	mutter	0.0519
	3	pussy	0.2029	arse	0.0516
	4	dress	0.1869	wine	0.051
	5	girl	0.1851	pipe	0.0495
	6	cunt	0.178	squint	0.049
	7	skirt	0.1675	tense	0.049
	8	omega	0.164	scoff	0.0485
	9	aunt	0.1606	frown	0.0484
	10	baby	0.1572	retort	0.048

Table 7: Top ten biased feature words by role r and gender g. NPMI may repeat in the sample group by role and gender because the frequency (and thereby the probability) of these word lemmas are the same.