Exploring Trends in High Fantasy Literature:

A Dive Into Readers' Opinions Towards Women Protagonists in the Contemporary Literary World

Ranjani Yegneshwara Iyer

MPhil. Digital Humanities and Culture

Trinity College Dublin

2024

Abstract

This study computationally analyzes gender trends in post-2000s high fantasy literature by examining 1,082,691 Goodreads reviews across 132 novels using Python-based text mining and VADER sentiment analysis. After preprocessing data (removing non-English reviews, converting emojis, and stripping HTML tags), results reveal women authored 74.7% of analyzed works and were significantly more likely to feature women protagonists (35% of their novels vs. 2% by male authors). The analysis shows 85.1% of sampled books were series starters, and women protagonists received overwhelmingly positive reception (69.5% positive sentiment overall). These findings demonstrate how computational methods can uncover meaningful patterns in literary reception while highlighting persistent gender disparities in protagonist representation.

Methodology

Data Collection

Data collected between August 19th to August 24th 2024, inclusive.

Created a researcher account (Trinity email) to access Goodreads' tagged "high fantasy" books (35,402 total).

Manual extraction was required due to:
1. Platform limitations (only first 25 pages/1,250 books loaded)
2. API restrictions preventing automated scraping

Inclusion Criteria for Initial Screening (Google Sheets):
- Published ≥2000
- Standalone or series-starter (no cross-universe titles)
- Prose only (excluded graphic novels/anthologies)
- >1,000 reviews

Challenges Encountered:
- Some non-qualifying books slipped into List 1 during manual entry
- Automation attempts failed due to:
- Goodreads' API key requirements
- IP blocking risks for unauthorized scraping

```python
#MODULE IMPORTS - Web requests and data handling
import requests #HTTP library for API calls
import json #JSON data parsing
import time #for adding delays between requests


def get_all_reviews(api_url, initial_payload, headers):
        #Initialize storage for all reviews and page counter
        all_reviews = []
        page_number = 1

        while True: #Continuous loop until break conditions met
```

```python
        print(f"Fetching page {page_number}")
        try:
                #API Request with timeout safety
            response = requests.post(
            api_url, headers=headers, data=json.dumps(initial_payload), timeout=10 #timeout
            )
        except requests.exceptions.Timeout:
            print(f"Request timed out for page {page_number}.")
            break
        except requests.exceptions.RequestException as e:
            print(f"Request failed for page {page_number} with error: {e}")
            break

        #HTTP STATUS CHECK
        if response.status_code != 200:
            print(
            f"Failed to fetch page {page_number}, status code: {response.status_code}")
            break

        #JSON DATA PARSING
        data = response.json()

        #ERROR HANDLING
        if 'errors' in data:
            print(f"API returned errors: {data['errors']}")
            break

        #DATA EXTRACTION LOGIC
    if 'data' in data and 'getReviews' in data['data']:
        reviews = data['data']['getReviews']['edges']
        all_reviews.extend(reviews)
                #PAGINATION HANDLING
        total_count = data['data']['getReviews']['totalCount']
        page_info = data['data']['getReviews']['pageInfo']
        next_page_token = page_info.get('nextPageToken')

        if next_page_token: #more pages available
        initial_payload['variables']['pagination'] = {
                "after": next_page_token, "limit": 30
        }
        page_number += 1
```

```
time.sleep(1) #rate limiting
print(f"Got {len(all_reviews)} reviews")
Else: #no more pages
break
else:
print(f"No more data available or error in data structure.")
break


return all_reviews #return complete dataset
```

Code Block 1: Snippet of data retrieval code

Preprocessing

Non-English reviews were removed, and emojis were mapped to words (e.g., 😊 → 'happy')
using regex substitutions (code snippet B).

```
# Emoji-to-text mapping for sentiment analysis preprocessing
# Converts emojis to standardized lexical representations to preserve emotional tone
# in Goodreads reviews while maintaining text-analytic compatibility

emoji_dict = {
        # Positive sentiment emojis (frequently used in favorable reviews)
        "😊": "[happy]",          # Common in positive character/appreciation comments
        "😂": "[laughing]",       # Indicates humorous engagement with text
        "❤️": "[love]",           # Strong emotional attachment to book/characters
        "😍": "[adoring]",        # Extreme character/story appreciation
        "👍": "[approval]",       # General endorsement of content

        # Negative sentiment emojis (critical or disappointed tones)
        "😔": "[displeased]",     # Mild dissatisfaction with plot/characterization
        "🥶": "[chills]",         # Could indicate either thrill or discomfort
        "🤦‍♀️": "[frustration]",  # Common in critical reviews of plot decisions

        # Neutral/context-dependent emojis
        "🙏": "[pleading]",       # Often appears in hopeful/hyped reviews
        "💕": "[affection]",      # Romantic subplot engagement marker
        "😘": "[flirtatious]",    # Character shipping-related usage
```

```
    "😎": "[cool]",            # Stylistic appreciation
    "⭐": "[star_rating]",     # Frequently substitutes for star ratings

    # Additional emojis observed during preliminary review analysis
    "😭": "[crying]",          # Strong emotional reaction to plot events
    "🤔": "[pondering]",       # Indicates critical analysis of themes
    "👎": "[disapproval]"      # Explicit negative rating
}
```

Code Block 2: Snippet of preprocessing code

Analysis

For this exploratory study, I focused on descriptive statistics (proportions, averages) to identify trends in:

- Author gender distribution
- Protagonist gender representation
- Sentiment polarity (positive/neutral/negative)

While more advanced statistical testing could reveal deeper correlations, these metrics provide a foundation for understanding gendered patterns in high fantasy reception.

This study's reliance on **descriptive statistics** (proportions, averages) highlights trends but cannot establish causality or statistical significance. For example, the **7:1 ratio of women to men authors** in the screened sample may reflect either genuine authorship trends or Goodreads' algorithmic biases in recommending books. Similarly, the **manual screening process** (prioritizing books with >1,000 reviews) likely skewed results toward mainstream, commercially successful works. These limitations underscore the need for future research with balanced sampling and inferential statistics. Nevertheless, the findings offer valuable groundwork for SSH discussions about **gender, authorship, and reader expectations** in speculative fiction.

Results

The data reveals a striking gender disparity in contemporary high fantasy authorship: **93.9% of the screened books** (124/132) **were written by women**, compared to only 4.5% by men and 1.5% by non-binary authors. This aligns with the broader trend observed in the initial dataset (74.7% women-authored among 419 books), suggesting that women dominate the genre—especially when the protagonist is female. Notably, **women authors were significantly more likely to write stories centered on a single woman protagonist (35% of their works) compared to men (2%)**, reinforcing a gendered pattern in character representation. This could reflect either women's preference for writing women-led narratives or market-driven incentives favoring such stories.

Across **1,142,962 analyzed sentiments**, reader reviews skewed overwhelmingly positive (**69.5%**), with only **14.8% negative** and **15.6% neutral** responses. While sentiment distribution was consistent across author genders, the near-absence of men-authored books with women protagonists (4.5% of the sample) makes gendered comparisons difficult. However, the high positivity toward women protagonists (regardless of author gender) challenges stereotypes about reader bias against female-led fantasy. The **prevalence of series starters (84.98% women-authored, 88.88% men-authored)** further suggests that high fantasy's commercial success hinges on multi-book arcs, though women authors displayed slightly more variety in publishing standalone works.