SI630 W19: <EXPLICIT - Sex Stories + NLP>.

Zi Huang, ziihuang@umich.edu, https://medium.com/@ziihuang

Abstract

In this study, I examined linguistic characteristics of erotic stories written by men and women searching statistically significant differences between the two.

I found that the distribution story perspective (first and third person), story topics (sex and not sex), and the proportion of VAD words from the NRC Valence, Arousal, Dominance Lexicon is significantly different across author gender. I also find that stories written be men tend to contain more male entities than stories written by women.

1 Introduction

This must be written by a man. That's a phrase that every woman (and potentially men too!) has said at some point when reading about a female fictional character.

Whether it's the irrational obsession with female anatomy, the absurd sexualization of female characters in inappropriate locations, or other strange-minded prose, there's clearly a pattern with male authors writing female characters in a... bizarre light. Since oversexualization was already prominent in these stories, why not go straight to the sex itself and analyze how men and women write sex stories? Thus, this study seeks to examine the differences between the male and female-authored sex storeis using NLP methods.

Today, there is not much NLP research in the realm of sex stories, because people have better things to do. However, scholars have analyzed the differences between male and female-authored fiction through the use of applications such as BookNLP. I explore this approach to fiction analysis and offer a few new methods of my own.

I found that it was difficult to build an accurate classifier distinguishing between male and female authored stories. However, there are statistically significant differences in the choices that men and women made in their stories. Some of these differences matched expectations and previous research, while others were rather astounding.

2 Problem Definition and Data

2.1 Problem and Motivation

Success in this project was defined as finding significant differences between stories written by men and those written by women.

As previously mentioned, lamenting that a work must be written by a man is commonplace complaint for many people. I'd like to believe that we're not just grappling the air, but that there are legitimate characteristics that distinguish work written by women from work written by men. Thus, this project is motivated by the core motivation of all projects: legitimizing complaint.

2.2 Data (downloading 8k sex stories)

This data was obtained by scraping popular online sex story and social networking site lushstories.com. The benefit of this site is that story authors are more inclined to include personal information about themselves, such as their gender and age.

Overall, the corpus consists of 8228 stories with the following characteristics:

- 2423 female-authored stories
- 5805 male-authored stories
- 1510 author biographies
- 529 female authors, 981 male authors

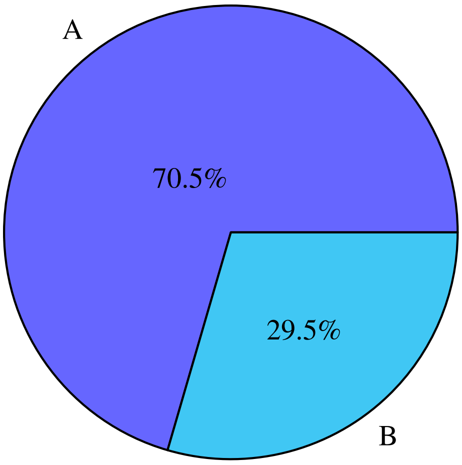


Figure 1: Pie chart of the percent of stories authored by each gender. A = male-authored stories, B = female-authored stories

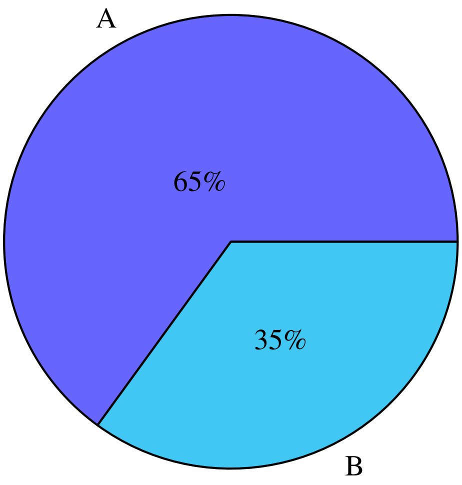


Figure 2: Pie chart of the percent authors by each gender. A = male authors, B = female authors

** I was able to download 40k stories, but I used a subset for this study. I intend to recreate this study with the full corpus when I am able to.

3 Related Work

3.1 Sex Stories Research

Though there is little NLP work on sex stories, a plethora of work on male and female responses to written erotica is helpful to understanding the types of stories that they might be inclined to write.

First, stories depicting sex are equally arousing to men and women. However, women experience more experiences of emotional avoidance and instability than men do after reading them (Schmidt et al., 1973). Though this study is almost 50 years old, the womens' increased reaction of "disgust, shock, quarrelsomeness" and internal avoidance may account the lower percentage of few female authors compared to male ones.

Second, both men and women prefer sex stories with emotions to those of plain physical nature (Schmidt et al., 1973). As a result of this one might expect the proportion of emotional words to be distributed equally across author gender.

Third, upon reading sex stories, men "reported more sexual arousal and less negative affect when the story described the male character as dominant" while women "reported more sexual arousal and less negative affect when the female character was described as dominant" (Garcia et al., 1984). This implies that men and women prefer to read about characters that they can empathize with (Garcia et al., 1984).

3.2 NLP and fiction

Elsewhere, methods of performing NLP on fiction also influence this study. Most NLP methods in this arena utilize BookNLP. Unfortunately, BookNLP (and other NLP applications), are not particularly accurate when inferring the gender of entities in first-person stories, or at resolving coreference to characters in first-person stories(Underwood et al., 2018).

Meanwhile, most sex stories are written in first-person, complicating this analysis. This project is thus an uncharted exploration into the linguistic characteristics of sex stories.

4 Methodology

As this project was mostly data-exploration, the methods were improvised based on previous analyses' results.

4.1 Preprocessing

Preprocessing for the data varied by step, but most of the time it involved lowercasing all words and removing stopwords where they were not semantically significant. Preprocessing specifics are described for each step below.

4.2 Unigram, Bi/Trigram Classification

The natural place to begin when trying to understand the differences between two texts is classification. Classification of the data was first performed with both unigrams using sklean CountVectorizer and TfidfVectorizer and bi/trigrams using sklearn CountVectorizer and TfidfVectorizer.

Naive-bayes and logistic regression classification were attempted. Naive-Bayes was used because it is a generative classifier. Logistic Regression was used because the two-class system is perfect for delineating the classes of male and female authorship, and the coefficients showed which unigrams, bigrams and trigrams were most likely to be written by men and women.

4.3 Topic Modelling

Topic modelling was performed on the stories to understand bigger themes at play within the stories, and to see if different genders preferred to write about different topics.

Topic models were trained for topic numbers ranging from 2 to 38. Two topic models (n=2 and n=14) were more deeply explored due to their high coherence scores.

I performed a chi-squared test to determine if there is any difference between men and women's writings across topics (chi).

4.4 Emotional Word Scoring

Emotional word scoring was performed to see if the writings confirmed Schmidt's findings that both genders reacted equally strongly to stories with emotional words in them.

I scored the presence of emotional words in a text by first counting the occurrence of emotional words in the text according to the NRC Word-Emotion Association Lexicon (Mohammad and Turney, 2013). This lexicon is "a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust)" (Mohammad and Turney, 2013). After counting the word frequency, I multiplied them by the number of emotions that each word represented. I normalized this score by dividing it by the length of each document, creating an emotion-score-per-word measurement.

Finally, I performed the two-sample Kolmogorov Smirnov (ks) on the distribution of male and female-authored emotional scores to see if the scores were differently distributed.

4.5 VAD Word Scoring

A similar procedure was applied using words from the NRC Valence, Arousal, Dominance Lexicon to score the stories. "The NRC Valence, Arousal, and Dominance (VAD) Lexicon includes a list of more than 20,000 English words and their valence, arousal, and dominance scores" (Mohammad, 2018). "Valence is the positive–negative or pleasure–displeasure dimension; arousal is the excited–calm or active–passive dimension; and dominance is the powerful–weak or 'have full control'–'have no control' dimension" (Mohammad, 2018).

After scoring and normalizing the stories, the two-sample Kolmogorov Smirnov test was performed to determine if their was a different distribution in score across genders.

4.6 POV Inference Using Coreference Resolution

Coreference clusters generated by AllenNLP's coreference resolution algorithm were used to infer the point-of-view from which the stories were written in (Gardner et al., 2017). This had to be performed on an even smaller subset of the data (718 stories) due to computer limitations.

If the clusters contained more first-person pronouns - "I", "we", etc - the story was judged to be first-person. If the clusters contained more second or third-person pronouns - "he", "you", "they" - the story was judged to be third-person.

Following this POV inferencing, a chi-squared test was performed to see if the writing's point-of-

	uni nb	uni logreg	bi nb	bi logreg
avg acc	0.7	0.81	0.72	0.83
baseline	0.7	0.7	0.72	0.72

Table 1: Table of average performance of non tfidf (count vector) classifiers

view (first or third person) was contingent on the gender of the author.

4.7 Named Entities Recognition

SpaCy's Named Entity Recognition was used to label all the person entities in the stories by gender using the PyPI gender package. This entity count was normalized by dividing the total number of entities by the number of 1000-words in the story.

The number of male and female named entities in the stories was analyzed using the two-sample Kolmogorov Smirnov to identify if the distribution of male and female entities per thousand words was equal or unequal across genders.

5 Evaluation and Results

5.1 Unigram, Bi/Trigram Classification

Count vectorizer classification methods were always more successful than TFIDF vectors on both unigram and bi/trigram classifiers. The average performance for all count vector classifiers tested is in Table 1. Logistic regression performed far better than naive-bayes on all counts.

To ensure that the most accurate classifiers were actually successful, five-fold cross validation was performed on them. The logistic regression classifiers maintained their accuracy against the baseline of 70%. This baseline is acquired by simply guessing the most common class (male-authored) for all texts.

Finally, the most common unigram and bi/trigram vectors associated with each gender are shown in Table 2 and Table 3.

Findings from the unigrams indicate that certain names were effective at distinguishing female-authored stories from male-authored ones ("stevie", "javon", "brianna", "sonya"). Additionally, the presence of genitals seemed to be an accurate indicator of male-authored stories ("dick", "shaft", "crotch", "anus").

Findings from the bi/trigrams indicate that *both*

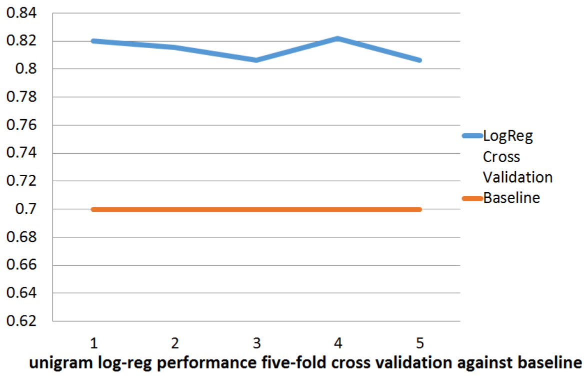


Figure 3: Cross validation accuracy of Unigram Logistic Regression classifier

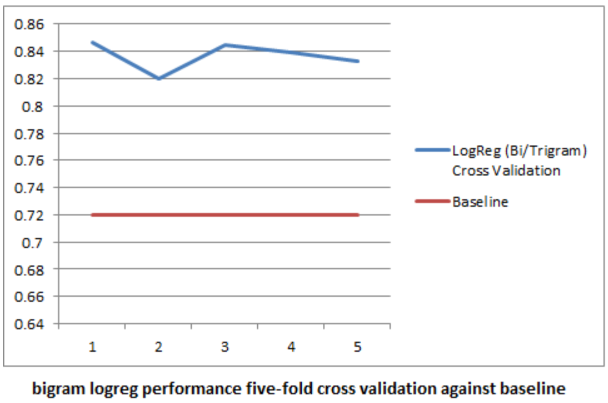


Figure 4: Cross validation accuracy of Bi/Trigram Logistic Regression classifier

female		male	
stevie	-0.49	partner	0.47
fun	-0.46	scene	0.41
donald	-0.45	lover	0.40
audrey	-0.43	belly	0.40
lip	-0.43	dick	0.37
cried	-0.43	shake	0.37
javon	-0.40	ten	0.36
definitely	-0.40	crotch	0.36
steven	-0.39	goes	0.35
usual	-0.39	lady	0.33
teeth	-0.38	shaft	0.33
gripping	-0.38	cora	0.33
sonya	-0.37	sissy	0.33
vagina	-0.36	christy	0.33
fucked	-0.36	embrace	0.33
brianna	-0.36	pushes	0.33
sira	-0.36	bottoms	0.33
weeks	-0.35	anus	0.32
wanting	-0.35	beauty	0.32
ramble	-0.35	whilst	0.32

Table 2: 20 most heavily-weighted unigrams associated with female and male authors in logistic regression classifier

men and women like to write from the first-person perspective about their anatomy (”my clit”, ”my pussy”, ”my breasts”; ”my cock”, ”my dick”, ”my balls”). Once again, more names are present in the female bi/trigrams than the male one.

5.2 Topic Modelling

Coherence was high for the 2-topic model and then peaked again for the 14-topic model. Based on these scores, I decided to further examine the 2 and 14 topic models.

The null hypothesis for each of these examination is that topics would be equally distributed across men and women-authored stories. The alternative hypothesis is that these topics are not equally distributed across author gender.

5.2.1 2-topic model

The *p-value=3.75e-11* from the chi-squared test indicates that the two topics are also distributed

female		male	
my clit	-0.26	my cock	0.35
when i	-0.19	my dick	0.23
my skirt	-0.18	my wife	0.23
my wet	-0.18	her belly	0.18
amy s	-0.17	she said	0.18
my pussy	-0.17	my balls	0.16
beth ann	-0.16	i ll	0.15
was so	-0.16	said she	0.15
against her	-0.15	with her	0.15
wants you	-0.14	the head	0.15
my breasts	-0.14	his dick	0.14
had been	-0.14	of cum	0.13
his hand	-0.14	her husband	0.13
grabbed her	-0.14	year old	0.13
down my	-0.14	your pussy	0.13
you i	-0.14	as his	0.13
to make	-0.14	a very	0.12
on and	-0.14	to see	0.12
my dress	-0.14	my pants	0.12
miss jones	-0.13	then i	0.12

Table 3: 20 most heavily-weighted unigrams associated with female and male authors in logistic regression classifier

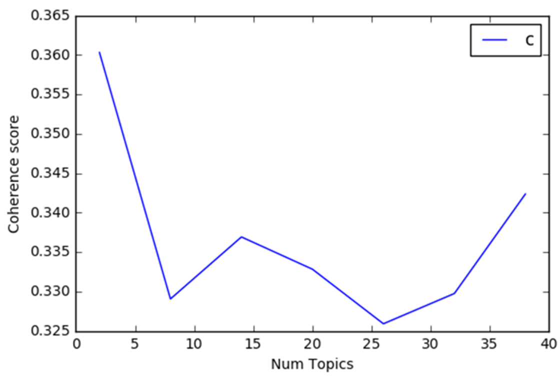


Figure 5: Coherence scores of topic models trained. There is a peak at 14 topics.

topic_0 (sex)		**topic_1 (not sex)**	
cock	0.017	time	0.011
hand	0.015	make	0.008
feel	0.012	good	0.007
back	0.011	back	0.006
pussy	0.009	man	0.006
mouth	0.008	give	0.006
head	0.008	love	0.005
fuck	0.008	thing	0.005
finger	0.008	feel	0.005
pull	0.007	day	0.005

Table 4: Table of top ten words for 2-topic model.

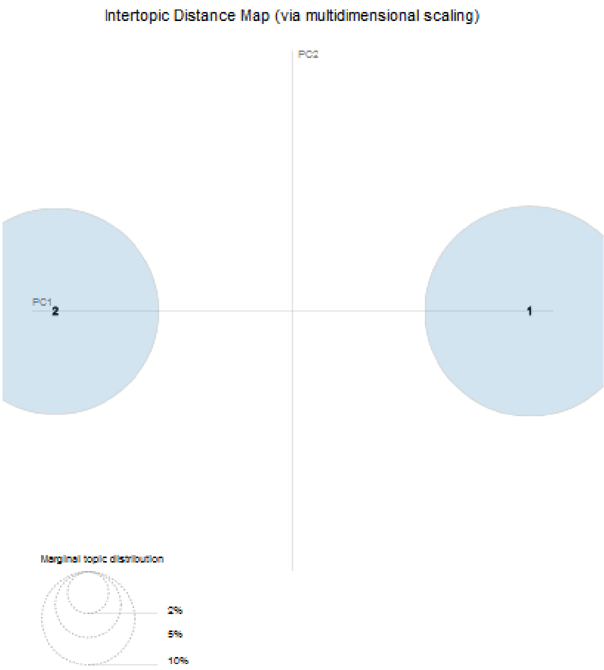


Figure 6: Distance between 2 topics. Here, 1 is topic_0 (sex) and 2 is topic_1 (not sex)

Dominant_Topic	0.0	1.0
sex		
F	1539	884
M	3227	2578

```
chi2_contingency(contingency)
```

```
(43.741936611601666,
 3.746569531209265e-11,
 1,
 array([[1403.50243072, 1019.49756928],
        [3362.49756928, 2442.50243072]]))
```

Figure 7: Contingency table and chi-squared test of topic contingency across gender. The results prove that at the 5 percent significane level, the topics are different distributed across gender.

differently across author gender (Figure 7). Furthermore, there is a greater proportion of female writers writing about topic_0 (sex) than there are male writers (Figure 7).

I also attempted to train Naive-Bayes, Logistic Regression, Random Forest, and LinearSVC classifiers using topic_0 scores as features. However, none of these classifiers performed better than baseline (70%) on the test set. This means that while topics are distributed different between men and women, they are not the sole differentiating feature of their writings.

5.2.2 14-topic model

The *p-value=1.15e-43* from the chi-squared test indicates that the fourteen topics are also distributed differently across author gender (Figure 9). I have included top ten words for each topic in the appendix as it was too long to fit in this report (Table 6).

Once again, I trained Naive-Bayes, Logistic Regression, Random Forest, and LinearSVC classifiers using topic scores as features, dropping topic_7 to prevent collinearity from occuring. Once again, these topics were not successful as classification features, failing to beat the baseline (70%). This continues to prove that while topics are distributed differently across gender, knowing the

Figure 8: Distance between 14 topics. Once again, all the topic numbers are shifted forward by 1. Thus topic_0 is 1, topic_1 is 2, and so on.

Dominant_Topic	0.0	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	11.0	12.0	13.0
sex														
F	126	240	45	72	182	86	71	746	35	276	78	155	145	166
M	327	806	215	212	579	182	254	1066	228	513	160	384	464	415

```
chi2_contingency(contingency)
```

```
(239.1265871449666,
 1.1538832811149942e-43,
 13,
 array([[ 133.40046184,  308.02843947,   76.56538649,   83.63296062,
          224.1009966 ,   78.92124453,   95.70673311,  533.60184735,
           77.44883325,  232.34649976,   70.08677686,  158.72593583,
          179.33969373,  171.09419057],
        [ 319.59953816,  737.97156053,  183.43461351,  200.36703938,
          536.8990034 ,  189.07875547,  229.29326689, 1278.39815265,
          185.55116675,  556.65350024,  167.91322314,  380.27406417,
          429.66030627,  409.90580943]]))
```

Figure 9: Contingency table and chi-squared test of topic contingency across gender. The results prove that at the 5 percent significane level, the 14 topics are different distributed across gender.

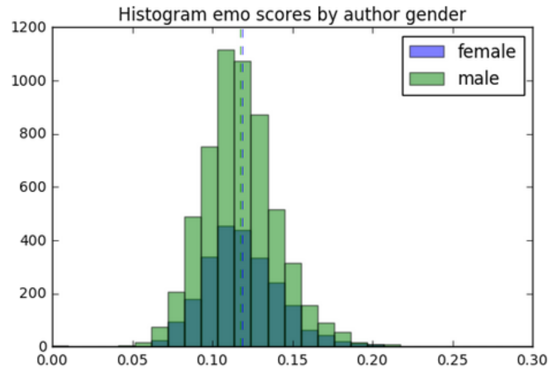


Figure 10: Distribution of emotional scores across male (green) and female-authored (blue) stories.

topic makeup of a document alone is not enough to know which gender wrote the story.

5.3 Emotional Word Scoring

After compiling the emotional scores for the stories, and separating them by author gender, the distribution of scores is shown in Figure 10, which indicates that they look very similar.

I examined the emotion scores' histogram, skewness statistic and performance on the k-s test for normality, concluding that the scores were not normally distributed.

Following this, I used the k-s 2-sample test to determine whether the emotion scores were distributed differently between men and women. The null hypothesis for each of these examination is that emotion scores are equally distributed across men and women-authored stories. The alternative hypothesis is that these scores are not equally distributed across author gender.

The test result of *pvalue=0.073* was indicates that at the 5% significance level, emotion score distributions are not significantly different from between male and female authors.

5.4 VAD Word Scoring

I followed a similar procedure for the VAD scores - evaluate normalcy (all were not normal on the basis of histogram (Figure 11, Figure 12, Figure 13), skewness, and ks test for normality) and then perform the ks 2-sample test. Once again, the null hypothesis for each of these examination is that the scores are equally distributed across men and women-authored stories. The alternative hypothe-

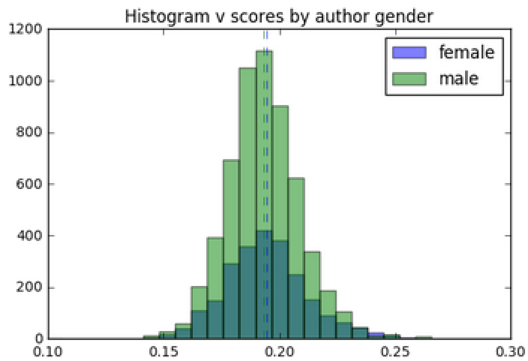


Figure 11: Distribution of valence scores across male (green) and female-authored (blue) stories.

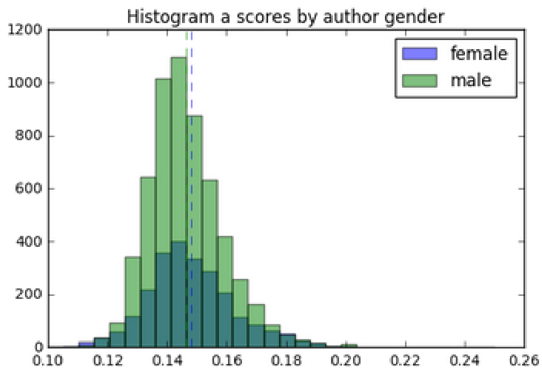


Figure 12: Distribution of arousal scores across male (green) and female-authored (blue) stories.

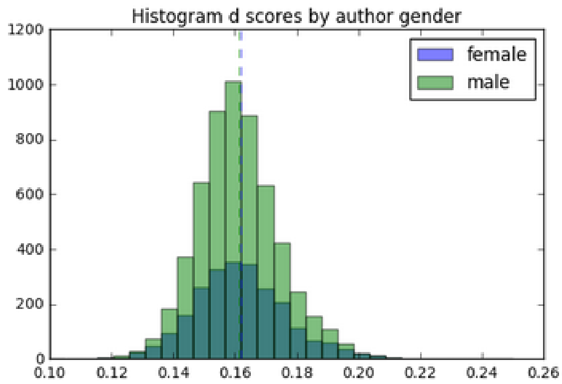
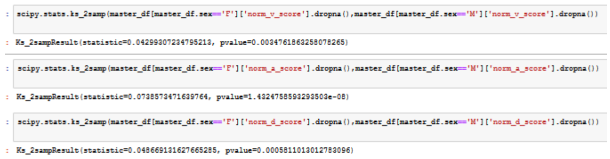


Figure 13: Distribution of dominance scores across male (green) and female-authored (blue) stories.



```
: scipy.stats.ks_2samp(master_df[master_df.sex=='F']['norm_v_score'].dropna(),master_df[master_df.sex=='M']['norm_v_score'].dropna())
: Ks_2sampResult(statistic=0.04299307234795213, pvalue=0.003476186325807826S)
: scipy.stats.ks_2samp(master_df[master_df.sex=='F']['norm_a_score'].dropna(),master_df[master_df.sex=='M']['norm_a_score'].dropna())
: Ks_2sampResult(statistic=0.0738573471639764, pvalue=1.4324758593293503e-08)
: scipy.stats.ks_2samp(master_df[master_df.sex=='F']['norm_d_score'].dropna(),master_df[master_df.sex=='M']['norm_d_score'].dropna())
: Ks_2sampResult(statistic=0.04866913162766S285, pvalue=0.0005811013012783096)
```

Figure 14: VAD scores are differently distributed across stories authored by men and women

sis is that these scores are not equally distributed across author gender.

Contrary to the emotion scores, I found that the distribution of VAD scores were indeed different across author gender at the 5% significance level (Figure 14). For VAD, the *pvalue=0.0034, pvalue=1.4324e-08, and pvalue=0.0005* respectively. Women's stories all averaged slightly higher than men's stories on all scores, indicating that there are more words from the VAD lexicon in women's stories than men's (Table 5).

	avg_v	avg_a	avg_d
M	0.193	0.146	0.161
F	0.194	0.148	0.162

Table 5: Average VAD scores for men and women. All are slightly higher for women.

5.5 POV Inference Using Coreference Resolution

After inferring the perspective from cluster pronouns, I examined the distribution of first and third-person stories across gender. It appeared that women authored proportionately more third-person stories than men (Figure 15). Thus, the null hypothesis for this test was that men and women authored the same number of stories across perspective, while the alternative hypothesis was that they authored different numbers of stories across perspective.

This alternative hypothesis was confirmed by the chi-squared result of *p-value=0.0377*. Thus, at the five-percent significance level, the writing perspectives (first and third person) are different between women and men.

5.6 NER Recognition

The distribution of male and female entities per 1000 words of text were both found to not be normal (Figure 16, Figure 17). Thus, the ks two-sample test was used again.

The null hypothesis for this analysis is that men and women include the same number of male and

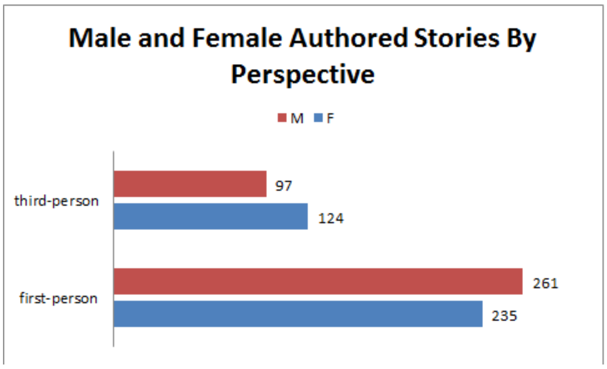


Figure 15: Bar graph of perspectives of male and female authored stories. As can be seen, women author proportionately more third-person stories than men.

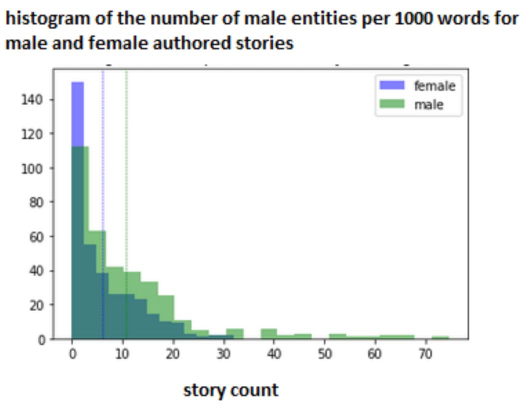


Figure 16: Histogram of the number of male entities per thousand words in male (green) and female (blue) authored stories. Entity count is on the y-axis.

female of entities in their stories.

At the 5% significance level, men and women include different numbers of male entities in their stories. The *p-value=3.977e-06* confirms this.

At the 5% significance level, men and women include the same number of female entities in their stories. The *p-value=0.970* confirms this.

6 Discussion

6.1 Results in Context

Things that are statically the same between men and women:

- Emotion scores
- Number of female named entities

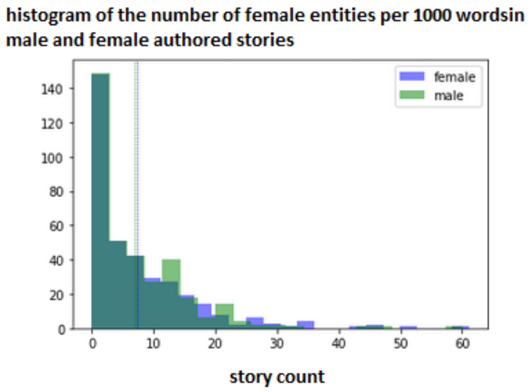


Figure 17: Histogram of the number of female entities per thousand words in male (green) and female (blue) authored stories. Entity count is on the y-axis.

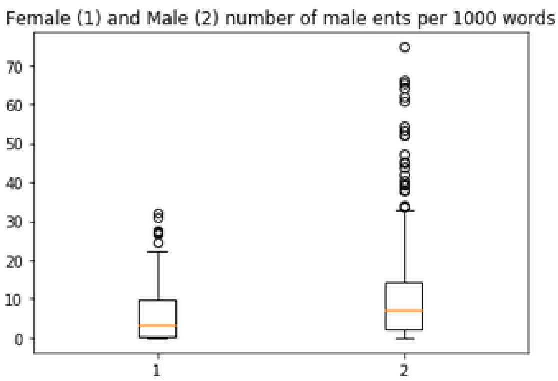


Figure 18: Box plot of the number of male entities per thousand words in male (2) and female (1) authored stories. Entity count is on the y-axis. Men write significantly more male entities than women on average.

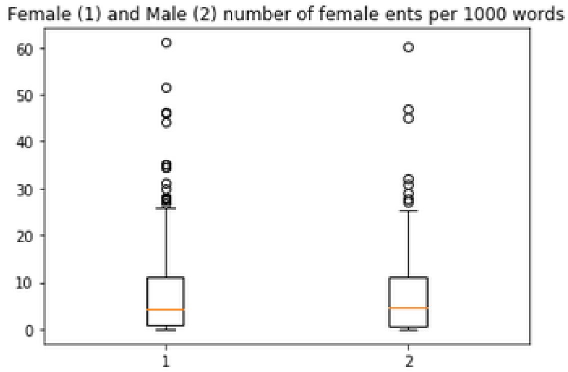


Figure 19: Box plot of the number of female entities per thousand words in male (2) and female (1) authored stories. Entity count is on the y-axis. Men and women write approximately the same number of female entities on average.

Given Schmidt's findings that both men and women prefer sex stories with more emotional wording, the sameness of the emotion scores is not surprising. However, to someone unacquainted with sex stories, it may come as a shock that both men and women use an equal distribution of emotional words in their work.

The finding that men and women write approximately the same number of female entities supports those found by Underwood et al. that women tend to write the same number of male and female entities (Underwood et al., 2018).

Things that are statistically different between men and women (at the 5 percent level):

- Unigram and bigram frequency
- Topic distribution
- VAD Scores
- Perspective (first vs third person)
- Number of male named entities

These were surprising and further research is required to understand why they happened. For a few of the findings, I have some thoughts.

As mentioned in the Garcia study, both men and women were more comfortable when a protagonist of their own gender was dominant in the story. Using first-person is a natural method of making oneself dominant (assuming that the gender of the first-person character matches the gender of the author). This corroborated by the bigrams, which indicated that both sexes tended to write in first-person about their own anatomy.

The finding that men write more male entities than women also supports those in Underwood et al.'s analysis in a large corpus of fiction, which also found that men wrote more male entities than women (Underwood et al., 2018). Taken together, this suggests that the gender of entities in sex stories follows the patterns found in fiction at large. However, more research is needed for this to be conclusive.

6.2 Limitations

6.2.1 Limitations of Internet

The biggest caveat to this study is its reliance on authors to self-report their gender: who says they have to be honest? The social pressure of the networking site might encourage authors to be more authentic about their gender, but there is always some uncertainty.

Thus, findings should be taken with a grain of salt.

6.2.2 Limitations of Technology

There are also limitations on the technology I used to perform this study.

Most notably, spaCy's NER cannot detect personal pronouns, so we have no idea of the gender of any of the pronoun-entities ("I", "me", etc) in these stories. Considering that the majority of stories seem to be written in first-person, this is a substantial limitation that could potentially overhaul the NER results if refined. I chose to use spaCy because the coreference resolution could not accurately cluster or gender many of the first-person stories. For example:

Cluster 1:

- 'Hand', 'in'
- 'your', 'strong'
- 'Your', 'thumb'
- 'Your', 'fingers'
- 'your', 'face'

Cluster 2:

- 'your','manhood'

- ’you’, ’while’
- ’your’, ’zipper’
- ’you’, ’to’

These clusters are meant to belong to the same character. However, there is no way to infer the gender of Cluster 1 and no way of linking Cluster 1 to Cluster 2. I’m determined to find a work-around for this solution: it could be another method of counting characters, or an assumption that a first-person story is more likely to contain two characters; a solution is out there and I will find it!

Another limitation is that this study’s definition of gender is not inclusive of non-gender-binary people. Because the gender package used to assume entity gender only outputs ”m” for male and ”f” for female, it may not truly reflect the gender of entities in the story. Furthermore, as queer sex stories and non-binary people are generally underrepresented on the site, this study’s findings may not reflect these populations well.

7 Conclusion

This study has managed to find some statistically differences in sex stories authored by men and women. Differences at the 5% level were found for the following: topic distribution, VAD Scores, perspective (first vs third person), and number of male named entities.

However, it’s no closer to legitimizing the complaint that I put forth at the start of the study. There are differences between work written by men and women, but do these differences indicate anything about the quality of work written by men?

To do this, I’ll need to do more research into the quality of the stories, using the ratings that they have earned on the site. If stories written by one gender perform worse than stories written by another gender, I may have a real complaint on my hands.

Furthermore, I’d also like to analyze the adjectives used to describe male and female entities across author gender to see if patterns of adjectives match those found in general fiction. These are just a few of the avenues I intend to explore; if you’d like to see more, feel free to follow the Medium account below.

Updates: https://medium.com/@ziihuang
GitHub: https://github.com/aureliuszi/EXPLICIT

8 Other Things I Tried

SRL with practnlptools and coreference resolution with AllenNLP

I spent several hours using the outdated package practnlptools to perform semantic role labeling. I selected this package because it was old, like my computer, but it was very difficult to work with. I first tried to install the python 3.x forks of the project, but those were buggy and refused to work properly. I ended up creating an environment running python 2.7 to get the package to work, only to discover that the output was too confusing for me to use: I couldn’t discern the gender of the agents (A0) and patients (A1) without coreference resolution.

When I attempted coreference resolution using AllenNLP, it was still inaccurate for many stories. The inaccuracy stemmed from the fact that most NLP systems are not good at performing coreference resolution on first-person narration, as the plethora of ”you” and ”your” words can refer to any number of parties. As a result of this, I was unable to get accurate information on agent gender and could not analyze agents and patients.

Acknowledgments

I thank my professor David Jurgens and the GSI Heeryung Choi for putting up with my shenanigans and awkward questions.

References

???? Data analysis. http://learntech.uwe.ac.uk/da/Default.aspx?pageid=1440.

Luis T. Garcia, Kathleen Brennan, Monica DeCarlo, Rachel McGlennon, and Sandra Tait. 1984. Sex differences in sexual arousal to different erotic stories. *The Journal of Sex Research* 20(4):391–402. https://doi.org/10.1080/00224498409551235.

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

Saif M. Mohammad. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*. Melbourne, Australia.

Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a word-emotion association lexicon 29(3):436–465.

Gunter Schmidt, Volkmar Sigusch, and Siegrid Schfer M.D. 1973. Responses to reading erotic stories: Male-female differences. *Archives of sexual behavior* 2:181–99. https://doi.org/10.1007/BF01541755.

William E Underwood, David Bamman, and Sabrina Lee. 2018. The transformation of gender in english-language fiction. https://doi.org/10.22148/16.019.

topic_0		**topic_1**		**topic_2**	
back	0.018	cock	0.075	man	0.017
water	0.010	fuck	0.038	cock	0.012
time	0.009	pussy	0.030	mike	0.010
hand	0.009	cum	0.028	time	0.010
head	0.008	ass	0.022	feel	0.010
turn	0.008	mouth	0.020	sam	0.009
walk	0.008	hard	0.020	david	0.008
sit	0.007	suck	0.017	big	0.008
start	0.007	start	0.013	wife	0.007
stand	0.007	feel	0.011	smile	0.007
topic_3		**topic_4**		**topic_5**	
make	0.018	back	0.016	bottom	0.014
mom	0.018	guy	0.013	mum	0.012
feel	0.017	girl	0.011	girl	0.011
love	0.017	hand	0.011	sarah	0.011
hand	0.017	turn	0.009	spank	0.011
back	0.013	door	0.009	time	0.009
time	0.012	room	0.008	give	0.008
good	0.011	walk	0.008	hand	0.008
kiss	0.011	pull	0.008	back	0.008
lip	0.011	kelly	0.008	dave	0.008
topic_6		**topic_7**		**topic_8**	
man	0.015	hand	0.029	jack	0.014
time	0.011	feel	0.022	kiss	0.013
woman	0.009	finger	0.020	smile	0.012
sex	0.009	lip	0.016	wife	0.011
feel	0.009	body	0.014	night	0.010
cock	0.007	back	0.013	bed	0.010
give	0.007	kiss	0.013	love	0.010
enjoy	0.007	tongue	0.012	time	0.009
mary	0.007	eye	0.011	hand	0.009
good	0.007	move	0.010	amy	0.008
topic_9		**topic_10**		**topic_11**	
eye	0.012	emma	0.013	time	0.019
make	0.006	begin	0.012	feel	0.017
face	0.005	hand	0.009	make	0.015
feel	0.005	head	0.008	love	0.014
voice	0.005	time	0.007	sue	0.011
moment	0.005	lisa	0.007	good	0.010
body	0.005	move	0.006	night	0.010
word	0.004	fuck	0.006	wife	0.010
head	0.004	body	0.006	thing	0.009
sound	0.003	feel	0.006	kiss	0.009
topic_12		**topic_13**			
dress	0.011	good	0.012		
hand	0.009	time	0.011		
girl	0.008	work	0.011		
leg	0.008	man	0.009		
wear	0.007	make	0.009		
pantie	0.007	day	0.009		
skirt	0.007	thing	0.008		
stand	0.007	call	0.007		
smile	0.007	woman	0.007		
foot	0.006	year	0.006		

Table 6: Table of top ten words for 14-topic model.