

FanFiction Pronoun Analysis

Jake Warmflash

Dataset

```
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/horror_1.txt" horror_1.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/horror_2.txt" horror_2.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/horror_3.txt" horror_3.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/horror_4.txt" horror_4.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/horror_5.txt" horror_5.txt
```

```
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/romance_1.txt" romance_1.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/romance_2.txt" romance_2.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/romance_3.txt" romance_3.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/romance_4.txt" romance_4.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/romance_5.txt" romance_5.txt
```

```
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/warewolf_1.txt" warewolf_1.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/warewolf_2.txt" warewolf_2.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/warewolf_3.txt" warewolf_3.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/warewolf_4.txt" warewolf_4.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/warewolf_5.txt" warewolf_5.txt
```

```
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/harrypotter_1.txt" harrypotter_1.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/harrypotter_2.txt" harrypotter_2.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/harrypotter_3.txt" harrypotter_3.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/harrypotter_4.txt" harrypotter_4.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/harrypotter_5.txt" harrypotter_5.txt
```

```
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/twilight_1.txt" twilight_1.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/twilight_2.txt" twilight_2.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/twilight_3.txt" twilight_3.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/twilight_4.txt" twilight_4.txt
!curl "https://raw.githubusercontent.com/Jwarmflash/PronounTextAnalysis/main/twilight_5.txt" twilight_5.txt
```

Text Cleaning

Remove:

- Extra Spaces, Tabs, Newlines
- Unique Characters
- Comment Numbers

afraid to show off in her clothes, although that might need to change now she was moving to the damp Forks.

60

.....(♡).....

"Come on, Bell" Riley gestured for her older sister to move along in the busy airport that they were currently cramped in, Bella quickly

hand as I did so. I smile again when I make contact, I couldn't wait to meet my sister or brother.

4

~ • ~

"It's been a while since I've seen you smile," Mom admits, taking the spot next to me on the sofa.

Split Into sentences:

```
horror_sentences = re.split(r'(?<!\b(?:Mr|Mrs|Ms|Dr)\.)(?<!\w\.\w.)(?<=[.!?"']?) +', cleaned_text)
```

Tag and Count

- Annotate
- Extract Parts of Speech
- Fuse POS to word
- Map Reduce
- Filter to he/she/him/...
- Sum up Subject, Object, total counts
- Calculate Proportion

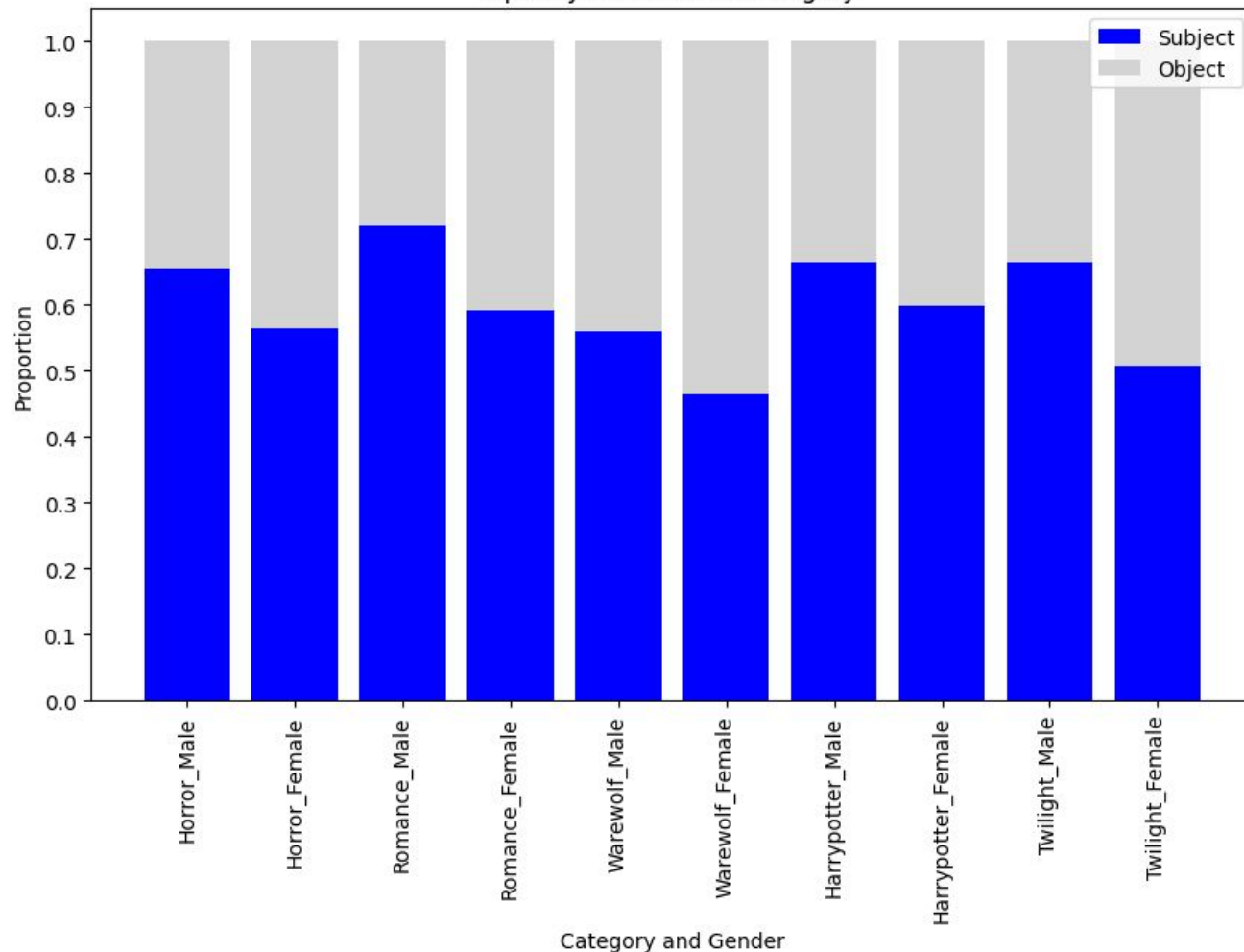
Fun Fact: The longest piece of English literature is a FanFiction called “The Loud House: Revamped” by user Jamesdean5842 at over 35 million words. The longest published writing is “A La Recherche du Temps Perdu” by Marcel Proust at 1.2 Million words.

```
#pronoun proportions
horror_male_pronouns_subject_proportion = summed_male_pronouns_subjects / summed_male_pronouns
horror_female_pronouns_subject_proportion = summed_female_pronouns_subjects / summed_female_pronouns

print("Horror male subject pronoun proportion: " + str(horror_male_pronouns_subject_proportion))
print("Horror female subject pronoun proportion: " + str(horror_female_pronouns_subject_proportion))
```

```
[('shePRP', 2391), ('hePRP', 3809), ('herPRP', 408), ('herPRP$', 2165), ('himPRP', 1100), ('hisPRP$', 2592), ('hersNNS', 2), ('hersPRP', 1)]
Horror male subject pronoun proportion: 0.6544460738568191
Horror female subject pronoun proportion: 0.5635192268975237
```

Proportion of Subject and Object Pronoun Usage in Online FanFiction
Split by Gender and Category



Conclusion

- Proportions statistically significant
- Male higher proportion than female
- Romance and Twilight largest differences between genders
- Harry Potter least difference
- Wattpad has higher proportion of female users, mostly teenagers

Thanks, here's some funny lines

- Can I please bury my face in your neck to smell more of your delicious earthy scent that I pick up with my supernatural nose?
- "Yes, pretty boy. You finally found someone that doesn't want to kiss you. You want an award?"
- "My dick is so long it goes from a all the way to z."
- "You never knew your parents. You were dead before you got into this world. And it's your fault that your parents were taken. If you had stayed alive, they would have been alive." Hadrian replied, taking a different route for dealing with the Boggart. He simply got over the fear.
- I hate Lego. Its like toys made by the devil, especially when you stand on those tiny blocks and you feel like you are going to die.