

Gender in *The Simpsons*



Student: Justin Carlou "Jus" Lim | Instructor: Adam Anderson
Digital Humanities 100: Theory and Methods | Summer 2021

Table of Contents

- | | |
|---------------------------|--------------------|
| 1. Research Question | 2. Hypothesis |
| 3. Dataset | 4. Tools & Methods |
| 5. Gender Dist. Over Time | 6. Word Clouds |
| 7. N-Gram Analysis | 8. Topic Modeling |
| 9. Conclusion | 10. Resources |
| 11. Acknowledgements | |

Research Question

How are gender differences presented in *The Simpsons*?

Specifically:

1. How has the inclusion of female characters changed over time?
2. How does dialogue between male characters differ from dialogue between female characters?

Hypothesis

Because *The Simpsons* is a show which follows pop culture closely-- and discussions of gender have become more commonplace- I do believe that some increased inclusion of non-male characters will be observed. However, as a major network television show, I don't expect there to be a significantly drastic change in gender inclusion and representation.

Specifically, I expect that:

1. The proportion of female characters will change, but only slightly.
2. Conversations between folks of the same sex will be stereotypically gendered.

Dataset

- Prashant Banerjee's "The Simpson's Dataset" (2020) posted on Kaggle.
 - This dataset was originally web-scraped by Tod Schneider.
- I focused my analysis on the character dataset and the script dataset.
 - For simplicity's sake, I merged these two tables into one and dropped the irrelevant columns.

Link to dataset: [Kaggle Dataset](#)

```
list(s_script.columns)
```

```
['id',  
'episode_id',  
'number',  
'raw_text',  
'timestamp_in_ms',  
'speaking_line',  
'character_id',  
'location_id',  
'raw_character_text',  
'raw_location_text',  
'spoken_words',  
'normalized_text',  
'word_count']
```

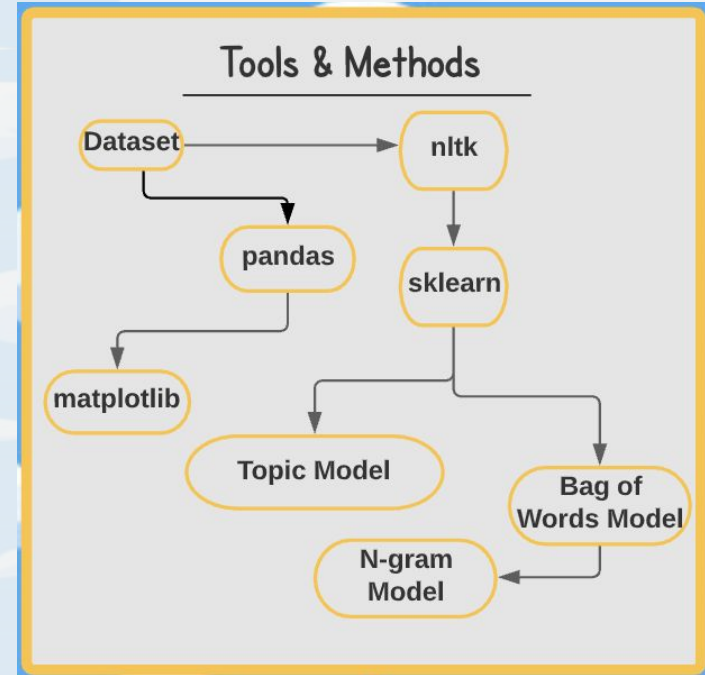
```
list(s_char.columns)
```

```
['id', 'name', 'normalized_name', 'gender']
```

	episode_id	normalized_name	char_id	gender	line	timestamp_in_ms	raw_location_text
0	32	miss hoover	464	f	No, actually, it was a little of both. Sometim...	848000	Springfield Elementary School
1	32	miss hoover	464	f	I don't know. Although I'd sure like to talk t...	856000	Springfield Elementary School
2	49	miss hoover	464	f	Thank you, Ralph. Very graphic. Lisa Simpson? ...	784000	Springfield Elementary School
3	49	miss hoover	464	f	Dear God!	803000	Springfield Elementary School
4	51	miss hoover	464	f	I question the educational value of this assem...	226000	Springfield Elementary School

Tools & Methods

- Organize data into distributions
 - Visualize results
- Word Cloud Model for differences in dialogue
- N-gram analysis for differences in dialogue
 - Visualize Histogram
- Topic Model for dialogue per gender group
 - Visualize Network



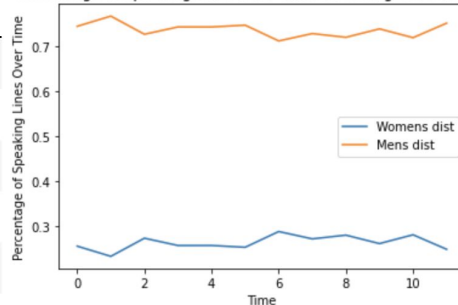
Gender Distribution Over Time

Takeaways:

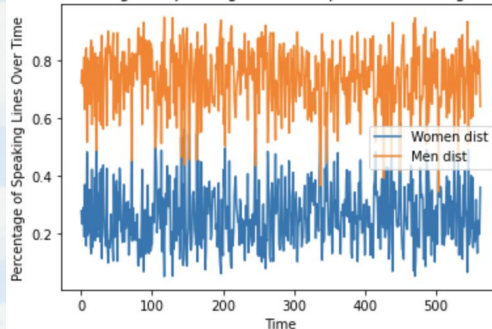
- Women are underrepresented
- There is an increase in women's representation, but it is negligible

index	womens_dist	mens_dist	total
0	0	0.278997	0.721003
1	1	0.233202	0.766798
2	2	0.257143	0.742857
3	3	0.291045	0.708955
4	4	0.188755	0.811245

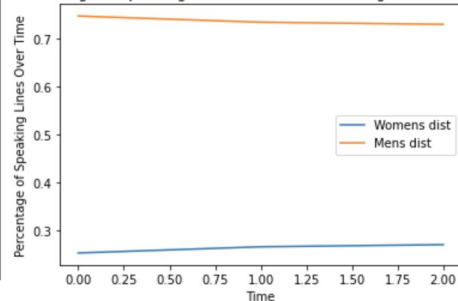
Percentage of Speaking Lines Over Time for each gender (AVG 50)



Percentage of Speaking Lines Per episode for each gender



Percentage of Speaking Lines Over Time for each gender (AVG 200)

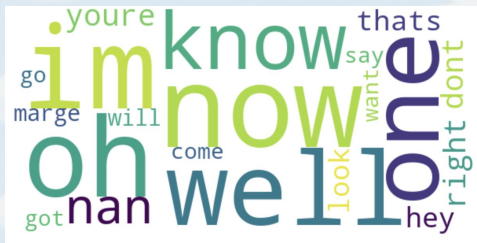


Comparing Dialogue: Word Cloud

Women's



Men's



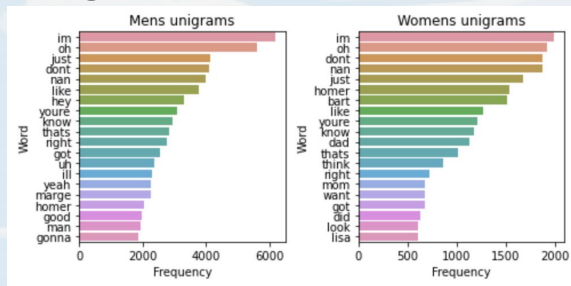
	line
normalized_name	
marge simpson	13193
lisa simpson	10769
edna krabappel-flanders	725
selma bouvier	566
patty bouvier	463
...	...
barbara	20
ms melon	18
evelyn	17
rachel	14
darcy	2

Interpretation:

1. The cast of women is mostly comprised of maternal figures to Bart Simpson, the son. These figures include Bart's mom, Bart's sister, and Bart's teacher.
2. Lisa Simpson comprises a large portion of women's speaking roles.

Comparing Dialogue: N-Gram Analysis

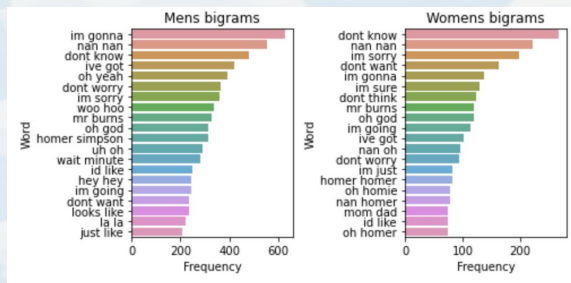
Unigram



Interpretation:

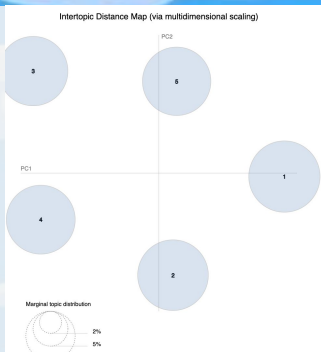
1. The cast of women is mostly comprised of maternal figures to Bart Simpson, the son. These figures include Bart's mom, Bart's sister, and Bart's teacher.

Bigram



2. Lisa Simpson comprises a large portion of women's speaking roles.

Comparing Dialogue: Topic Model



Topic #1:
youre, want, did, look, okay, make, love, theres, kids, say, yes,
home, school, help, great, honey, weve, boy, remember, talk

Topic #2:
oh, like, know, little, really, time, ive, gonna, maybe, simpson,
mean, better, wait, god, isnt, youll, baby, night, things, bad

Topic #3:
im, think, right, mom, lisa, going, ill, hes, come, sure, hey, didnt,
yeah, doing, let, father, marge, said, big, wont

Topic #4:
dont, homer, bart, good, sorry, people, need, whats, tell, day, youve,
theyre, new, house, maggie, thought, nice, lot, shes, ooh

Topic #5:
nan, just, dad, thats, got, homie, stop, mr, way, man, thing, lets,
thank, uh, guess, family, feel, id, doesnt, believe



Topic #1:
oh, dont, like, youre, ill, uh, good, man, little, come, say, simpson,
make, dad, lisa, let, life, help, mean, place

Topic #2:
know, marge, gonna, want, think, boy, day, sorry, old, god, son, work,
guys, money, mom, guy, school, youll, thanks, away

Topic #3:
im, just, nan, thats, got, yeah, homer, look, okay, ive, did, hes,
really, theres, thing, sir, sure, kids, new, ya

Topic #4:
right, time, yes, way, lets, love, going, whats, tell, mr, great, didnt,
id, big, home, theyre, said, stop, youve, hell

Topic #5:
hey, bart, need, people, wait, better, huh, gotta, thank, maybe,
springfield, thought, family, hello, long, aw, whoa, ooh, isnt, em

Interpretation:

1. Womens' dialogue is very centered around the topic of family and mother-daughter bonding
2. Mens' dialogue focuses mostly on Homer and his friend group or Bart and his friend group.

Conclusions

In conclusion, American tv show, The Simpsons, is shown to follow the legacy of commercial gender representation.

Findings:

- The distribution of men and women characters on the show is consistently unequal.
- Marge and Lisa make up a huge part of women's speaking roles. Following them is Edna Krabappel-- all three characters being maternal figures to Bart Simpson.
- Women are written to support men in the show.

Jupyter Notebook

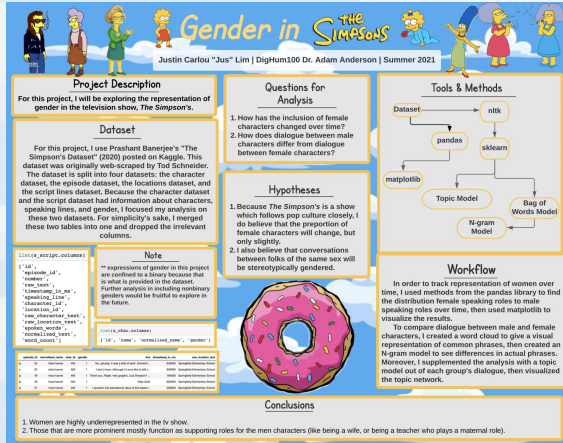
Github Repository

Works cited:

- <https://www.kaggle.com/huchi728/word2vec-analysis-the-simpsons-notebooks/Unigrams-Bigrams-and-Trigrams>
- <https://www.kaggle.com/rousselle111/the-simpsons-dataset>

Initial Storyboard

Resources



Poster



Video

Acknowledgements

- Banerjee, P. (2020). *The Simpsons Dataset*. Kaggle: <https://www.kaggle.com/prashant111/the-simpsons-dataset>
- Cowles, D. L., Austin, M., & Houston, G. T. (1994). Feminist Criticism. In *The critical experience: literary reading, writing, and criticism* (pp. 214–234). essay, Kendall/Hunt.
- Jessup, J. B. (2021). *Topic Modeling* [.ipynb notebook]. Retrieved from University of California, Berkeley Digital Humanities 101: Practicing the Digital Humanities.