The Quiet Ones: How Do Women Speak in Films?

Amal Alabdulkarim
Department of Computer Science
Columbia University
New York, 10027, NY
aa4235@columbia.edu

Jing Qian
Department of Computer Science
Columbia University
New York, 10027, NY
jq2282@columbia.edu

1 Abstract

With the dialogue analysis of 600 movies and over 3000 characters, we explored how women were portrayed in different decades and genres. We observed the gender gap in the role importance which was evaluated by dialogue length and roles in conversations: women had fewer lines, started fewer conversations and women conversations were under-represented in almost all genres. Highly rated movies tended to have lower female presence and the relationship between movie ratings and female presence had genre-dependence. Through text analysis with latent Dirichlet allocation (LDA) and Linguistic Inquiryand Word Count (LIWC) methods, we found that although men references still appeared frequently in women conversations, the topics of women dialogue became more diverse in recent years.

2 Introduction

Women role in movies have changed through the years, and women recently have stronger appearances in more genres(Donald et al., 2015). For example, in 2017 the movie Wonder Woman and 2019's Captain Marvel featured strong female superheros. These two examples in the male dominated genre of superhero movies and many others led us to take a closer look on the changes of female characters roles in films.

The representation of women role in film has been widely studied on multiple dimensions across literature. (Glascock, 2001)(Lauzen, 2008), suggested that women are more represented in recent films and television shows than older ones. (Glascock, 2001) also suggests that the type of characters of women has also changed with time, more females characters now have jobs in comparison to older television shows when they were mostly unemployed. (Agarwal and Dey, 2015) investigates the correlation between the importance of roles of women and the Bechdel test (Bechdel, 1986), through looking at womens topics among other features such as network analysis. (Movies, 2018) used linguistic features to inspected the speech patterns of female and male characters in superhero movies.

This study focuses on understanding the characteristics of female character roles in movies by analyzing movie dialogues. More specifically, their use of male-references, changes in topics, amount of speech, role in conversation and if there is a relation between the movie rating and female presence.

Firstly, We anticipate that female characters have higher importance in specific genres, such as romance and drama. Which we will measure using the number of lines, and the significance of their roles in those genres. Secondly, we expect that movies with higher rating tend to have a stronger male presence. Lastly, female characters dialogue topics became less men focused and more diverse in recent movies compared to older ones.

3 Previous Research

3.1 Understanding Women Roles in Media

Understanding women roles in media, especially in movies and television shows is widely studied. (Neville and Anastasio, 2018) Studied the relation between age and gender of actors and their role in films. Their findings suggested that although female major characters increased significantly in later years, male characters still outnumbers the female characters.

The Center for the Study of Women in Television and Film produced a report (Lauzen, 2008) comparing the change in the role of women in television and film between 1995 and 2008. In this report, they analyzed several features such as,

the change in the frequency of women movie lines through the years. They also inspected gender diversity across multiple dimensions such as television networks, genres, settings, age, marital status, and race. Moreover, they inspected the gender diversity of the other crew members on multiple dimensions. Their findings suggested that there was an increase in female characters in more recent television shows and movies. Also, female characters were more likely to appear in some genres such as dramas and reality programs and less likely to appear on others such as situation comedies.

(Glascock, 2001) Conducted a study on prime-time television shows. Their primary goal is to determine how the roles of female and male characters have changed on prime-time television since the 1970s. Their first research questions focused on the change of the demographics of these movies. To answer this question, they measured several demographic features. Examples of these features are the number of female/male characters, marital status, age, and appearance features. Their second research question focuses on the differences in the behavior of female/male characters and how it changed in recent years. Their general hypothesis is that the female role in movies has improved through time, and they measured this using their demographic and behavioral data. They extracted the demographic and behavioral features from the shows manually. Their findings showed that there was a change in how television shows depict female characters. While the gender gap on television has shrunk, the change has been slower than what the researchers initially expected. One reason could be that some of these discrepancies are merely reflective of the real world. Another reason may be innate in the movie industry, which the researchers attributed to the relations between the off-screen representation the and on-screen demographics as they showed evidence of male-dominance in the industry in their findings.

3.2 Using Movies Dialogues to Study Women Roles

Several studies utilized movie dialogues in studying the gender roles in fiction. (Skowron et al., 2016) Studied movie dialogues to detect character types from the movie dialogue. Other researchers studied the portrayal of professional women in pop culture. (DeTardo-Bora, 2009) Examines how fiction depicted female criminal justice professionals. Moreover, (Painter and Ferrucci, 2017) studied female journalists characters roles the television show House of Cards.

One method for analyzing gender roles in movies is the Bechdel test, which was introduced by Alison Bechdel in her comic Dykes to Watch Out For (Bechdel, 1986). The test aims to measure womens representation in fiction, and it consists of three questions.

1. Does the movie have at least two women in it?
2. Do they talk to each other?
3. Do they talk to each other about something besides a man?

In (Agarwal and Dey, 2015), the researchers proposed a method to study gender in films by automating the Bechdel test. For the third question, the researchers combined several feature vectors, such as topic and linguistic features. The features that best answered this question were network analysis features. They found that word unigrams, topic modeling features, and features that capture mentions of men in conversations are less effective in their experiments. Their results suggested that there is a significant correlation between the importance of roles of women in movies with the Bechdel test. Movies that fail the test tend to portray women as less-important and peripheral characters.

(Movies, 2018) Used movie dialogue to study gender roles in two superhero movies Wonder Woman and Man of Steel. The researchers used five linguistic features, amount of speech, interruptions, questions, minimal responses and hedges. Their findings showed a correlation in three of those features (amount of speech, interruptions, and questions) and no apparent patterns in the other two features. More specifically, the findings showed that men speak and interrupt more than women, while women ask more questions than men.

(Danescu-Niculescu-Mizil and Lee, 2011) Studied linguistic convergence between movie characters by constructing and analyzing the Cornell movie dialogues corpus. Among their findings, they looked at the gender role in linguistic convergence, and their results suggested that women have a more influential role in language

convergence. This unintentional finding suggested that this corpus has more to offer regarding gender roles in movie dialogue.

4 Method

Our first hypothesis has two sub-hypotheses, and both can be tested with similar statistical analysis. We will examine if the length of their dialogues (measured by the number of words and number of lines) differ across genres. We will also measure the appearance of the characters in those movies by the length of their dialogues and their mentions in other characters dialogues. We will specifically look at if the number of times female character started a dialogue versus when they responded to someone else is different across genres.

For the second hypothesis, we will measure how the rating of the movie relate to the female character role (identified using the previous statistical analysis for the second hypothesis), the ratio of female to male characters in the movies and ratio of female to male characters dialogues lengths (measured by number of words and number of lines).

For the third hypothesis, to test the diversity of the female character dialogues, we define diversity in dialogue in two main aspects. The first aspect is when female characters are talking, they will be talking about something other than men. The second aspect is whether the female character conversations topics vary with time. To test the first aspect, we will inspect the use of male references in female characters lines. To evaluate the second aspect, we will look deeper into the dialogues, using the latent Dirichlet allocation (LDA) and Linguistic Inquiryand Word Count (LIWC) for topic modeling and comparing the topics across the years.

In evaluating these hypotheses, we will be testing for statistical significance. And we will also consider the other factors and extraneous variables that may affect the findings of the analysis and make sure we address them correctly. We will use graphs and exploratory visualization for the results and the data to discover the relations and patterns and explain the findings. The Results and Discussion section exhibits the results of these evaluations in more details.

We will use the Cornell movie dialogue dataset. The Data Exploration section inspects this dataset in details.

Table 1: Information of data used from the Cornell Movie Dialogues Corpus.

Years	1930-2009
Female Characters	966
Male Characters	2049
Lines	242023
Movies	600

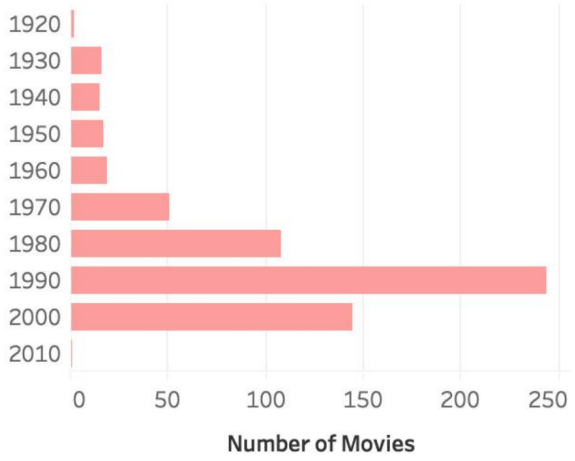

Figure 1: Release year distribution of movies.

5 Data Exploration

In this paper, we used the Cornell Movie Dialogues Corpus(Danescu-Niculescu-Mizil and Lee, 2011). This corpus contains a large metadata-rich collection of fictional conversations extracted from raw movie scripts, including movie metadata (genres, release year, IMDB rating, etc), character metadata (character name, gender, movie, etc) and conversation metadata.

Because our hypotheses are gender-related, we extracted only the dialogues from characters that with gender labeled. The information regarding the data we analyzed in this paper is shown as following: We could see that the number of identified male characters is about twice as that of female characters, which should be kept in mind when doing later difference analysis between genders.

We showed the release year distribution of the 600 movies in Fig. 1. One may observe the time distribution of the movies is skewed towards recent years.

We also showed the genre distribution of the 600 movies we analyzed in Fig. 2. One should know that one movie may be classified to multiple genres, like the movie *2001: a space odyssey* was labeled as 'adventure', 'mystery' and 'sci-fi'.

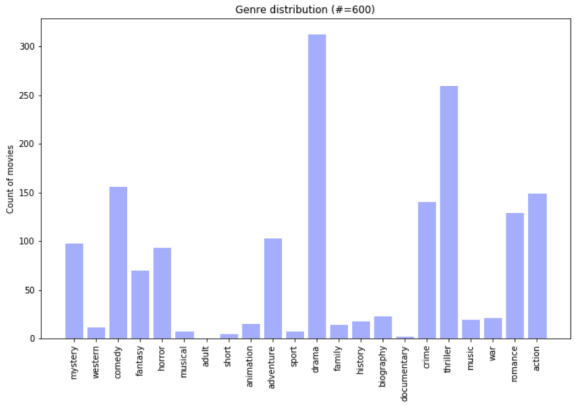

Figure 2: Genre distribution of movies.

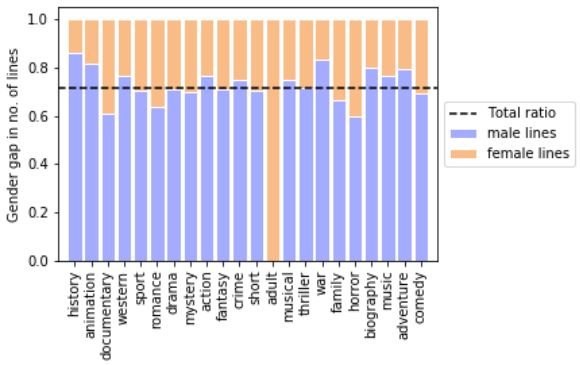

Figure 3: Dialogue length proportion of each gender in different genres. Orange bars represent female proportion while purple ones represent male. The black dashed-line represents the averaged male dialogue length proportion from all movies.

We could see from the plot that the genre distribution of movies was highly uneven. The top three genres with the highest number of movies are drama, thriller and comedy, in which drama has over 300 movies, more than half of the total movies. The least three are adult, documentary and short, each of which has less than 10 movies.

When discussing the female characters in different genre and period in the following sections, the readers should notice that results of some genres and years are based on relatively small number of movies.

6 Results and Discussion

6.1 First hypothesis

Firstly, we examined the importance of female character roles in movie dialogues and how this importance varied from genre to genre. So our first hypothesis is:

Female characters have a higher importance in specific genre.

There are various ways to evaluate whether female characters are important in the movies: like showing gender distribution(Neville and Anastasio, 2018) or comparing speaking time of different genders (Glascock, 2001). Here we discussed importance of female characters from dialogue corpus in two aspects: dialogue length and the role in conversation.

6.1.1 Do Females have more lines in dramatic and romantic movies than male characters?

In this sub-hypothesis, we defined the dialogue length as the number of lines in one movie. We also checked the definition as number of words and got similar result. In Fig. 3, we showed the dialogue lengths from each gender for individual genre. Considering the variant dialogue lengths from movie to movie, genre to genre, here we showed the ratio of dialogue length instead of actual counts. For the same reason, we calculate the ratio in each genre instead of each movie, through which we expect to balance out the fluctuation in the dialogue lengths in different movies. That is to say, the purple bars represent the ratio between male line numbers and total line numbers while the orange bars represent the ratio between female and total. Note here the total line numbers equal to male line numbers plus female line numbers, which don't include the lines spoken by characters without gender labels.

The ratio between male line numbers and total line numbers from all genres, which is shown by the black dashed line, is above 0.7. Hence male characters dominate movie dialogues in general. This ratio is close to the ratio of male character numbers to the total character numbers.

For individual genre, we use orange bars to present female dialogue length ratio while use purple bars for male. One could see that male characters dominate movie dialogues in almost all genres with a minimum about 0.6 and maximum less than 0.9. There is one peculiar genre, adult, which is denoted by the single orange bar and means only female characters are involved in the movie dialogue. However, as we showed in the genre distribution in Fig. 2, there is only one adult movie in our analysis. So this result of all female dialogues may not be general. The top three genres with the highest female dialogue length ra-

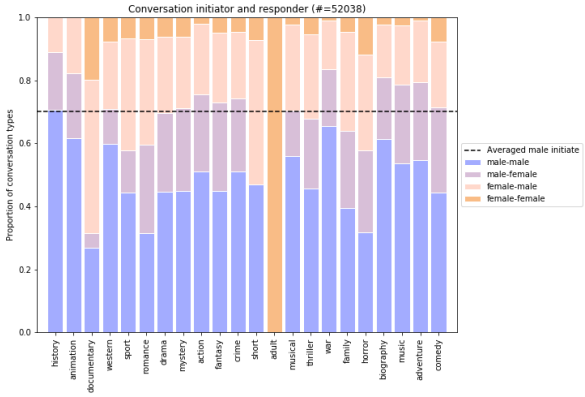

Figure 4: Conversation role proportions in different genres. From top to bottom are female-female, female-male, male-female and male-male conversations. The black dashed-line represents the averaged male initiating proportion in all movies.

tio are horror, documentary and romance (exclude adult) while those with the highest male dialogue length ratio are history, war and animation. So this sub-hypothesis fails: the ratio of female dialogue length vary from genre to genre, but in general, females tend to have fewer lines than males. The female dialogue length ratio of drama is close to the averaged value of all genres while that of romance is higher than average, close to 0.4.

6.1.2 Do Females have more significant conversation roles in dramatic and romantic movies?

In this sub-hypothesis, we examined what conversation roles female characters in movies are. Here we differed the conversation role as initiator and responder:initiator is defined as the first character of a conversation and responder is the second character of a conversation. We inferred that the initiator is the more proactive role in one conversation and hence if there are more female conversation initiators than male in one movie, women have higher importance than men in that movie.

Then there are four types of conversation: female-female, female-male, male-female and male-male, which are denoted by orange, nude, lavender and purple from top to bottom in Fig. 4.

Similar to previous plot, males started most conversations in movie dialogues with an initiate rate around 0.7, represented by the black dashed line. For most genres, the probability that males instead of females start conversations is higher. This may due to the fact that there are twice as many male characters of female characters and nearly half of the conversations are between males.

As mentioned before, the adult genre only has one movie and only includes female dialogues, it is not surprising that all conversations in that movie are classified female to female. In short and documentation, females start more conversations than male, this is mainly due to their high proportion of female-male conversations.

If we only look at the cross gender conversations, males and females have equal opportunities to start a conversation. We used the two sample t-test for the hypothesis that there is no difference between female-male and male-female proportions and got p-value equal to 0.096, which suggested that there is no significant difference between the probability that a female and a male initiates a conversation to the other gender.

We would also like to point out that there are almost no female-female conversations in history and animation. A movie without conversations between two female characters is considered lack of female representation according to Bechdel test mentioned above(Bechdel, 1986). On the opposite, the top three genres with the highest proportion of female-female conversations are documentary, horror and western. Other than the conversation roles, we discussed the conversation content between different types of conversations in detail in Hypothesis 3.

So our second sub-hypothesis of hypothesis 2 fails. Females do have more significant conversation roles (evaluated by conversation initiator roles ratio) in some genres such as short and documentation, but neither in drama or romance.

6.2 Second hypothesis

Following previous discussion of female role importance, we would like to explore whether there is any relationship between the female role presence and the people's opinion of movies. It is well known that how people evaluate a movie is subjective and changing. Here we choose a quantitative value, averaged movie rating from IMDB to evaluate whether people like a movie or not. Then our second hypothesis is:

Do movies with higher ratings tend to have stronger male presence?

A little different from previous analysis, here we define the female role importance of singular movie as the ratio of female dialogue length to the

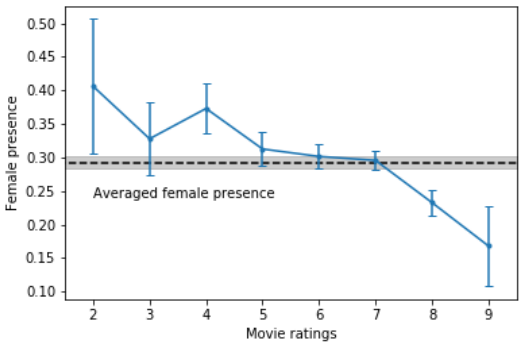

Figure 5: Relationship between female presence and movie ratings. The black dashed-line represents the averaged female presence from all ratings.

total dialogue length and named it as the female presence of that movie. So the male presence = 1 - female presence. Then we have the plot of ratings vs. female presence in Fig. 5. The x-axis is movie ratings, range from 2 to 9[1] while y-axis is the averaged female presence at given movie ratings. The black dashed line represents the averaged female presence of all movies at all ratings, which is 0.3 and in agreement with our previous result in Fig.3.

This plot suggests that the second hypothesis is true. The averaged female presence decreases with the increasing movie ratings, which supports that highly rated movies have stronger male presence. The movies with the highest rating, 9, have around 16% female dialogues in average while those with the lowest rating, 2, have over 40%.

Since this plot is based on average of movies, the numbers of movies at each rating vary and hence affect the error bar of each data point. There are much fewer movies rated the highest or lowest scores than the middle ones, which leads to the relatively large error bars at the highest and lowest scores.

We also looked into the relationship between ratings and female presence in specific genres. Here we depicted the rating-female presence distribution from three genres to illustrate the genre-dependence of such relationship in Fig.6. Based on previous analysis, we chose drama, romance and action, each of which has over 100 movies. Moreover, these three movies have quite different female presence. So we expect them to be representative to show the genre-dependence.

[1]The actual movie ratings are averaged on the votes of each movie and hence represented by float numbers. Here we round down these ratings to integers.

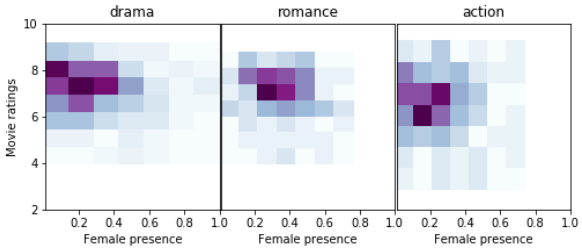

Figure 6: Movie rating vs. female presence distribution in different genres. From left to right, drama, romance and action. Dark purple shows high frequency while light blue shows low frequency.

In Fig.6, we showed the 2-dimensional distribution of movies in three genres, x-axis is the female presence distribution while y-axis is the rating distribution. Dark purple shows high frequency while light blue shows low.

Looking at the dark purple region, which represents the highest distribution density, one could observe that the female presence is generally highest in romance and lowest in action. The action movies have widest distribution and lowest mean of ratings among the three genres. Although we barely see a clear trend of the relationship between movie rating and female presence, it is obvious that such relationship is different from genre to genre.

6.3 Third hypothesis

Female characters dialogue topics became less men focused and more diverse in recent movies compared to older ones. To test this hypothesis, we address the following questions.

6.3.1 Is there an association between the character gender and the use of male references in the character lines?

We used several categories of Linguistic Inquiry and Word Count (LIWC). The primary dimension we are employing for this analysis is the male reference category, which has the percentage of male-related words in every line. These words can either be pronouns such as he and his or nouns such as man and male. The value of a category in a movie line shows how much of this line is from this category. We also chose the personal concerns and biological processes categories to represent some of the topics that could appear in the movies. The personal concerns categories are topic categories according to LIWC manual. Also, we added the biological processes to the comparisons because it

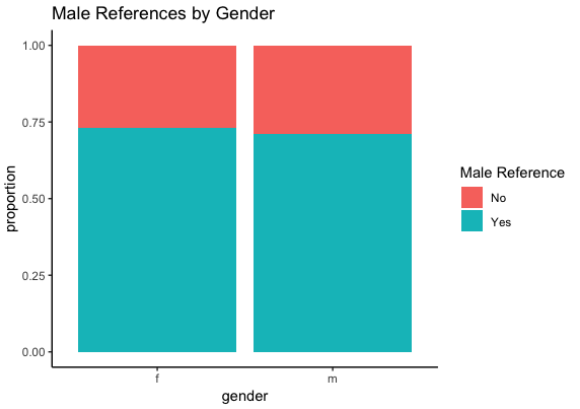

Figure 7: Male References in movie dialogues by gender, this bar chart shows a slight difference in the percentage of female characters use of male references in their dialogue

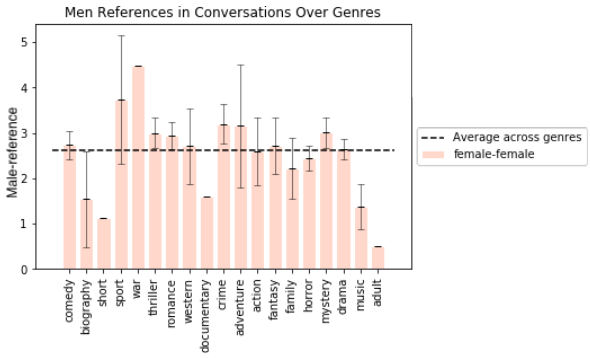

Figure 8: Male References in movie dialogues by genre, this bar chart shows the distribution of the use of male references across genres

also can be a theme of dialogue in movies, such as the line "They are not delivering. I am going to go pick up the pizza." which contains the word pizza from the Ingest category in LIWC but does not match the other topics categories.

We introduced a binary variable that indicates if the male-reference value is higher than any of these topics. Then, we performed a chi-square test of independence to examine the relation between the gender of the character and the use of male-references as shown in figure 7. The association between these variables was significant, $p < .01$. Among female characters, the likelihood of using male references in their lines is 1.02 times higher than male characters using male references in their lines.

We also wanted to check if the genre is a confounding factor in this relation. However, we could not do this because every movie in our dataset has from one to three genres. If we flatten the genres and duplicate the line to its corresponding genres, we would introduce bias in the data.

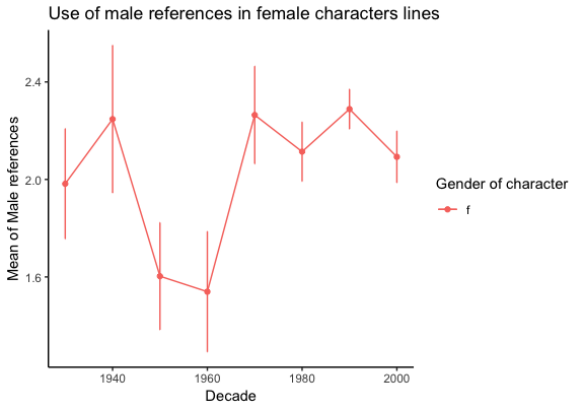

Figure 9: The frequency trends in the use of male-references in female characters lines in the years

We inspected the use of male-references across genres as shown in figure 8. Most genres are within the average of male-references across all genres. Some genres have lower means such as music and biography. On the other hand, war and crime have higher means. Which suggests that genre may not have a significant effect on the relation, but we cannot say that until we account for it as a confounding factor.

6.3.2 How does the frequency of men references change over time in female characters lines?

To address this question, we calculated the means and the confidence interval of the male-reference values in the movie lines grouped by decades as shown in figure 9. From this line plot, we can notice the apparent drop in the use of male-references in 1960 and 1950. Moreover, there is no noticeable difference when we compare the decades with higher male-references such as 1930 with 2000 as their confidence intervals intersect.

We also wanted to measure the association of female characters lines and the years, so we ran a linear regression model and got a p-value less than 0.01. This result indicates that there is a significant association between the years and the use of male-references in female characters lines. The frequency female characters use of male references in their lines is affected by the years.

6.3.3 Do the topics of female characters dialogues change with time?

We approached this question using two methods, topic modeling using latent Dirichlet allocation

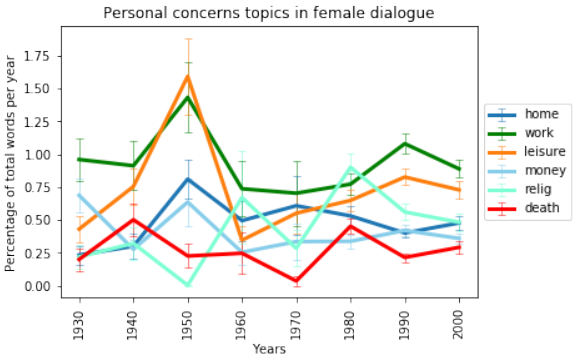

Figure 10: The change in topics distribution of the movies across the years

(LDA) and LIWC topics categories. We trained the LDA model on the full dataset, and then we used it to predict the topics of each movie line. We did cross-validation to choose a suitable number of topics number our measure was the topic coherence. The result of the cross-validation indicated that 18 topics did very well as they had the highest coherence value. Moreover, then we did a logistic regression on the resulted topics and their occurrences in each decade. The result of the logistic regression shows a significant association between the two variables with a $p-value < .01$.

This finding suggests that the topics are associated with the years. Still, a more in-depth inspection of the results of the analysis suggests that some topics are changing with the years. For example, in 1950 work had a strong positive correlation with female characters lines. Also another interesting example, in the 1940 decade there was a strong negative correlation between swear topic in female characters lines, but in the 1990 decade, there is a strong correlation between the swear topic in female characters lines which indicates an increase in their use in the female character lines.

A limitation of this approach is that the resulted topics are somewhat noisy. However, they reflect the most used words in every movie line, which makes them sufficient because the goal is to analyze the change in the topics distribution in each year.

The other approach was to use LIWC topic categories, which in this part were the personal concerns categories and the health dimension from the biological processes. Figure 10 shows the result of this analysis.

LIWC topics analysis confirms the previous finding that the topics are changing with time and are becoming more diverse and equal spread in recent years. Also, it clearly shows how these topics change with time. For example, the work and leisure dimensions had their highest mean value in 1950, and then their values dropped in the following decades.

This hypothesis had three parts in it, decreasing men focus and diversity of topics. The analysis results imply that the focus on men did not decrease as figure 9 shows. However, the topics are changing with time which confirms the second part of this hypothesis. These findings suggest that even though female characters use of male references did not decrease with time, the topics of their conversations are changing with time.

Limitations of this analysis include the increase in the number of movies in the dataset since the 1980s might have biased the topics modeling. Also, genres were not accounted for in the analysis due to the fear of introducing bias into the data.

This analysis revealed an interesting finding from the data. In the 1950s and 1960s, the topic Work arose, and the male-references were at its lowest. Although this can be an effect of the relatively low number of years which might introduce a low diversity in the sample, we decided to do more research to find out if something worth mentioning happened during these years. This era was post World War II and the beginning of the women liberation movement. After the war, popular culture encouraged women to leave their careers and return naturally to domestic life which contradicted with the conditions of these women lives during that time (Newman and Grauerholz, 2002). Also in our dataset, we have movies featuring strong working women during the same period such as the movie ”All About Eve”. All of this could have factored in in the changes during that period, but to confirm this assumption more research needs to be conducted on the associated cultural factors.

7 Conclusion

In this paper, using dialogues with labeled genders from the Cornell Movie Dialogues Corpus, we performed a systematic analysis of the female dialogues in both structure and contents, how they evolve with time and vary from genre to genre.

We found that for almost all genres, male roles are the dominant roles, having more lines and starting more conversations. The gender gap in di-

alogue length is most severe in history, war and animation and less severe in horror, documentary and romance. The conversations mainly consisted of male-male and cross-gender conversations. Some of the genres have almost no female-female conversations, especially in history and animation. For the cross-gender conversations, men and women have equal opportunities to start a conversation.

The female role importance of movies showed by female dialogue ratio is related to the movie ratings. We observed the trend that highly rated movies tend to have a stronger male presence. Moreover, this tendency has genre-dependence.

We also studied the dialogue topics, especially in the female-female conversations. We observed that men are in most women conversations even if they were not the main topic. This dominance of male reference did not show any clear trend through the years. The topics of female dialogue changed from time to time, which may reflect social change. One good sign is that female dialogue topics are more diverse and distributed evenly in recent years.

To sum up, female characters are underrepresented in almost all genres and highly rated movies tend to have a lower female presence. Although male reference dominates women conversations, the topics of female dialogues becoming more diverse with time.

8 Future Work

As we showed in the data exploration, our results are based on skewed release time and genre distributions of movies. The data samples from some years or some genres are very limited, which renders the corresponding results from being fully representative of those years or genres. We expect to have more data with higher diversities in time and genre.

Also, as shown in Table 1, there are many more male characters than female characters in our data. This fact may be explained as that the number of male characters in movies is indeed twice as many as that of female characters. Or maybe due to the limitation of the gender-labeling technique. We may implement different methods to identify the gender of characters (like from text analysis or the gender of the actor or actress of such character) and see whether these different methods would affect the results.

Moreover, we could use other ways to evaluate the importance of roles in a movie, like being mentioned frequently, instead of having more dialogue or start conversations often.

Finally, we may dig more into the relationship between movie ratings and female presence. With more and diverse data, we could study that relationship in different genres in detail. The movie rating is averaged over votes on IMDB, which is affected by release years and the number of votings and hence may affect the relationship we observe.

We expect that with more and diverse data and improved techniques, we may get more systematic results in future studies.

References

Apoorv Agarwal and Shirin Ann Dey. 2015. Automating the Bechdel test. pages 830–840.

A Bechdel. 1986. *Dykes to Watch Out for*. Dykes to Watch Out for Series. Firebrand Books.

Cristian Danescu-Niculescu-Mizil and Lillian Lee. 2011. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs.

Kimberly DeTardo-Bora. 2009. Criminal Justice Hollywood Style: How Women in Criminal Justice Professions Are Depicted in Prime-Time Crime Dramas. *Women & Criminal Justice*, 19(2):153–168.

Alexandra Donald et al. 2015. Girls against the world: Female representation in modern hollywood. *Screen Education*, (76):60.

Jack Glascock. 2001. Gender Roles on Prime-Time Network Television. 8151.

Martha M Lauzen. 2008. Study of Women in Television and Film. Technical report.

Contemporary Superhero Movies. 2018. The Superman Speaks and the Wonder Woman Keeps Quiet :. pages 61–90.

Conor Neville and Phyllis Anastasio. 2018. Fewer, Younger, but Increasingly Powerful: How Portrayals of Women, Age, and Power Have Changed from 2002 to 2016 in the 50 Top-Grossing U.S. Films. *Sex Roles*, pages 503–514.

David M. Newman and Elizabeth Grauerholz. 2002. *Sociology of families*. Pine Forge Press.

Chad Painter and Patrick Ferrucci. 2017. Gender Games: The portrayal of female journalists on House of Cards. *Journalism Practice*, 11(4):493–508.

Marcin Skowron, Martin Trapp, Sabine Payr, and Robert Trappl. 2016. Automatic Identification of Character Types from Film Dialogs. *Applied Artificial Intelligence*, 30(10):942–973.

9 Contributions

In this paper, Amal and Jing worked together as a group, from initial proposal to analysis to the final paper writing. The introduction, previous work, method and hypothesis 3 are mainly written by Amal while the abstract, data exploration, hypothesis 1-2, conclusion and future work are mainly written by Jing.